Solution Description

For a really effective solution to a problem, it is always mandatory to identify the main challenges first. Therefore, in this case as well, to implement a time-series model to correctly predict the wind power output at various wind parks in India, we need to identify the challenges.

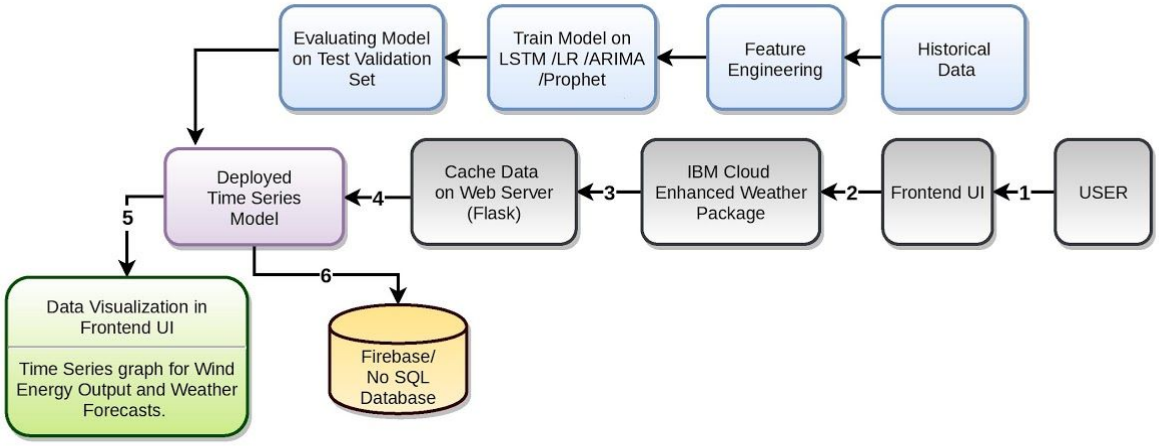
The primary challenge for the accuracy of any machine learning model is with the data that it is trained on and the inputted for prediction. One part of this is finding relevant data to train our model. Second part lies in getting an accurate future weather forecast (upto 72 hours) to input into our model. Another challenge is to finalise which prediction method works best for our data. Wind energy output predictions have been implemented using various methods like linear regression, ARIMA, SVR, Decision Tree Regressor and different methods work best for different scenarios. In this project, we will be testing with Linear regression, ARIMA, LSTM and Prophet (by facebook).
To get the correct weather forecast, we will be using IBM Cloud's Enhanced Weather Data package to get accurate weather forecasts for upto 72 hours.Moreover, we will be training our model on data compiled from various farms in different countries like USA, Brazil on these 3 major attributes for prediction -

- Wind Speed
- Wind Direction
- Temperature

Our aim is to provide an easily deployable and extendable model which predicts the wind power output for all the sites it is deployed on. The end to end process starts with our web UI where the user inputs 3 parameters for prediction -

- Region/pincode
- Duration
- Start time.

Once these details are inputted, our system accesses the IBM Cloud's weather data packages for **Enhanced Forecast** package for its 15 day standard forecast to retrieve the weather details for required duration. Once weather details for the required period are stored in a node/python server, then these details are inputted in the model to predict the respective wind energy output. Once the model processes the data and predicts the output, these results will be displayed on the web app for the perusal of the user. Along with a time-series graph for wind energy output, time-series of weather forecasted data from IBM Cloud (with 3 major attributes) will also be published for the user making it easy for him to verify the corresponding weather details.



Novelty/Uniqueness

This project helps wind park managers to predict the wind power output variation in future upto 72 hours. According to the predicted value, energy suppliers can coordinate energy production with other suppliers from different sources to avoid overproduction of energy.
The web app designed here takes the parameters from the user and retrieves the relevant data for that period and location using IBM Cloud's enhanced forecast package. Specifically, these features of this end to end solution are unique -

- Dataset for Training Model: Due to lack of availability of India's wind farm output data, we will be training our model on data compiled from various wind farm sites like USA, Brazil. Since, Spatial scales and geographical features have significant effect in efficiently predicting power output of wind farms, compiling data from different sites will help in minimizing that bias. Training dataset can be found here.

- Dataset features - Majority of the models implemented so far have only been trained on data containing only wind speed and wind direction. However, this model will be trained on an additional important feature which is temperature. As this research paper suggests in Fig. 6, temperature is one of the main features for the prediction.

- Prediction time - Considering stochastic nature of weather, it has been difficult to predict the correct weather forecast in future and that too for mid-range duration of 72 hours. Many wind power forecast models have only implemented some hours to 36 hours prediction at most. Like this model developed by google.

- Prediction Methods - Most models implemented so far for wind energy prediction have used methods like Linear regression, ARIMA, ARMA, ARMAX, SVR, decision tree regressor etc. However, we will be testing with LSTM, a type of recurrent neural network, as ANN methods have been proving successful for this objective. Also, we will be using the Prophet model by facebook for forecasting. In this study on Tamil Nadu wind park conducted, authors have stated the successful nature of ANN methods for prediction.

Business/Social Impact

- The above model can be easily deployed and extended to work according to the weather conditions at any of the 9 places in India where the wind energy is being produced. Different sites can keep training the model with their own data with the unique IBM Watson Studio feature called **Continuous Learning** and improve personalised forecasts.
- Accurate energy forecasts can help in more informed decisions with regards to electricity transmission sizing, sizing of nearby DERs, distribution of the electricity load, sizing of energy storage systems, and other issues related to renewable energy integration in a smart grid or microgrid.
- Our country is pursuing ambitious wind energy goals. This model's prediction can be used by energy suppliers to make better decisions and collaborate with more stable energy sources producers for efficient production reducing overproduction and costs related to that overproduction.

Technology Stack

- Programming Languages - Python (Flask)
- Database - Firebase
- IBM Cloud - Enhance Forecast Weather Data package, for deployment
- Packages - Scikit, Pandas, Tensorflow, Keras, Matplotlib etc.
- IBM Watson Studio

Scope of Work

This product will consist of these modules with specific functionalities -

- Weather Data Collector - This module will make the API call to the required IBM Cloud weather forecast package and retrieve required data and store it on a web server.

- Data Analyser - This module will do all the preprocessing required (Cleaning,Reduction,Transformation...) on the data stored in the web server to make it suitable for input in our prediction model predicting respective wind energy forecasts.

- Data visualization - This part will be responsible for visualizing the results received from the Data Analyzer Module and the weather data retrieved from the IBM Cloud's weather data package.

- Frontend UI - This module will interact with the end user taking inputs and displaying the results of the data visualization part.

Useful Links

1) IBM Purchased - The Weather Company to integrate with Watson for better weather prediction. https://analyticsindiamag.com/from-weather-forecasting-to-nowcasting-how-google-is-using-ml-to-predict-precipitation/
2) A github repo with different methods of predicting wind energy output (Go to "models" folder for methods like - ar/arima/arma/rnn)- https://github.com/fernandarp/Wind-Energy-Forecasting
3) Using LSTM for various time series prediction tasks - https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/
4) Tensorflow tutorial for using RNN for time series model - https://www.tensorflow.org/tutorials/structured_data/time_series
5) Hack challenge - https://smartinternz.com/ibm-hack-challenge-2020
6) ARIMA vs RNN - https://www.quora.com/When-should-I-use-an-RNN-LSTM-and-when-to-use-ARIMA-for-a-time-series-forecasting-problem-What-is-the-relation-between-them

7) New IBM weather data package API - http://biz.weather.com/WU-Data-API_Data-Package-Demo-Request.html

Repositories with relevant datasets

1) Go to datasets folder - https://github.com/2012vaibhav/Wind-energy-Forecast-Time-series-Analysis
2) Check in data folder - https://github.com/fernandarp/Wind-Energy-Forecasting
3) https://github.com/jon-lo/wind-energy-forecasting
4) By the name of WindPark VI - https://github.com/derevirn/wind-power-forecasting

Common attributes in these: temperature, wind speed, wind direction
Most of them have used average monthly weather data.
https://docs.google.com/spreadsheets/d/1oh-41p0dQ9zDlfmjnCDPMAJrURNozS2r5XbigwcOXds/edit?usp=sharing

2 more research papers

1) Read sara hi and focus on page 20 table - https://cora.ucc.ie/handle/10468/1735
2) The model (Projektor) mentioned in the table - https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwjy1YWymLTpAhWd4zgGHRPMDXEQFjAAegQIBBAB&url=http%3A%2F%2Fhenrikmadsen.org%2Fwp-content%2Fuploads%2F2016%2F02%2FLandberg_et_al-2003-Wind_Energy.pdf&usg=AOvVaw3f_02xHcYt4zPRMknOQMKV

https://www.researchgate.net/publication/224147384_Wind_power_forecasting_prediction_methods

Possible Data Collection Sources:

USA data collection site used in link1: https://www.ncdc.noaa.gov/

Spain wind farm data: http://www.sotaventogalicia.com/en/technical-area/real-time-data/historical/

Research gate ans: https://www.researchgate.net/post/Where_can_I_find_data_on_power_generation_from_wind_and_PV_plants_and_corresponding_weather_data

South africa weather data: https://sauran.ac.za/

Spatial scales and geographical conditions have significant effect in efficiently predicting power output of wind farms. Since there is poor availability of Wind Power Mills Power output data of India, we plan to develop an initial model using data available from other countries like USA, Brazil. We will divide available data into two major groups:on-shore and off-shore wind farms and then use this data separately to develop models, which can then be deployed for predicting wind power in sites of India according to the geographical characteristics of that particular site.