**Abdulrhman Abdullah**

**Vector Wind Forecast System**

**Project proposal**

**4 Dec 2021**

**Question/need:**

In this project I will analyze the data form weather station to forecast hourly wind vector (wind speed and direction), in industrial area that will help to take control of the effects of an environmental accident. An example of this is when a gas leak occurs, based on the predict can decide the best time and way evacuating residents. I chose this project because it’s related to my work under environmental protect and control.

And answer these few questions:

    ⁃    Will predict the wind vector help to protect from effect of environmental accident?

    ⁃    Can forecast wind vector for specific hour?

    ⁃    What features has the strong correlation with forecast wind vector?

**Data Description:**

I will be using Jena Climate dataset recorded by the [Max Planck Institute for Biogeochemistry](https://www.bgc-jena.mpg.de/wetter/). This dataset contains 14 different features such as air temperature, atmospheric pressure, wind speed, wind direction and humidity etc, recorded once per 10 minutes, beginning in 2003 to 2016.

**Location**: Weather Station, Max Planck Institute for Biogeochemistry in Jena, Germany

**Time-frame Considered**: Jan 10, 2009 - December 31, 2016

The target value: hourly wind vector which is indicate wind direction and speed.

I will calculate the wind vector vx and vy by using wind vector formula vx = ws \* sin(θ) , vy = ws \* cos(θ)  
where θ is the wind direction in radian, and ws is the wind speed. And add it to the dataset also I will just deal with hourly predictions only

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Features | Format | Description |
| 1 | Date Time | 01.01.2009 00:10:00 | Date-time reference |
| 2 | p (mbar) | 996.52 | The pascal SI derived unit of pressure used to quantify internal pressure. Meteorological reports typically state atmospheric pressure in millibars. |
| 3 | T (degC) | -8.02 | Temperature in Celsius |
| 4 | Tpot (K) | 265.4 | Temperature in Kelvin |
| 5 | Tdew (degC) | -8.9 | Temperature in Celsius relative to humidity. Dew Point is a measure of the absolute amount of water in the air, the DP is the temperature at which the air cannot hold all the moisture in it and water condenses. |
| 6 | rh (%) | 93.3 | Relative Humidity is a measure of how saturated the air is with water vapor, the %RH determines the amount of water contained within collection objects. |
| 7 | VPmax (mbar) | 3.33 | Saturation vapor pressure |
| 8 | VPact (mbar) | 3.11 | Vapor pressure |
| 9 | VPdef (mbar) | 0.22 | Vapor pressure deficit |
| 10 | sh (g/kg) | 1.94 | Specific humidity |
| 11 | H2OC (mmol/mol) | 3.12 | Water vapor concentration |
| 12 | rho (g/m \*\* 3) | 1307.75 | Airtight |
| 13 | wv (m/s) | 1.03 | Wind speed |
| 14 | max. wv (m/s) | 1.75 | Maximum wind speed |
| 15 | wd (deg | 152.3 | Wind direction in degrees |

**Tools:**

The Plan is to predict one-hour wind vector, in addition answer the previous questions. First, I will just deal with hourly predictions, so start by sub-sampling the data from 10-minute intervals to one-hour intervals, then

I will process the data by finding any null values or erroneous values and doing other cleaning process, and the then I’ll try to do some features engineering to make sure the data appropriately formatted. Second split the data into a training, validation, and test sets. After that I’ll Normalize the data then build the AI model to predict the wind vector. By using python and its libraries.

**Jupyter Notebook**: Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations.

**Python libraries might be used**:

    ⁃    **NumPy**: (Numerical Python) open source library used in data science and engineering.

    ⁃    **Pandas**: open source library for data manipulation and analysis.

⁃    **scikit-learn**: Python module for machine learning built on top of SciPy

    ⁃    **Matplotlib**: for data visualization

    ⁃    **Seaborn**: data visualization library for statistical graphics plotting in Python.

    ⁃    **Tensorflow**: open source library to help you develop and train ML models.

    ⁃    **Keras**: Python deep learning API, runs on top of tensorflow.

**MVP Goal:**

The wind vector is the result of the system that will help to take control of the effects of an environmental accident.  The next step after the proposal is the MVP, which will contain the data after processing and date engineering to forecast the wind vector.

References:

[https://jupyter.org](https://jupyter.org/)

[https://www.tensorflow.org](https://www.tensorflow.org/)

<https://www.analyticsvidhya.com/blog/2020/11/top-13-python-libraries-every-data-science-aspirant-must-know-and-their-resources/>

<https://www.bgc-jena.mpg.de/index.php/Main/HomePage>

<https://www.tensorflow.org/>

<https://numpy.org/>