Izvorni naučni članak / Original Scientific Article
dostupan online / available online: 17. 6. 2022.

DOI: 10.53028/1986-6127.2022.13.1.11
Uprava 2022; 1: str. 11-16

APPLICATION OF MACHINE LEARNING ALGORITHMS IN DECISION MAKING OF STATE ENTERPRISES

PRIMJENA ALGORITAMA MAŠINSKOG UČENJA U DONOŠENJU ODLUKA DRŽAVNIH PREDUZEĆA

Dženan Šašić[1], Zerina Mašetić[2]
International Burch University
[1] dzenan.sasic@stu.ibu.edu.ba, [2] zerina.masetic@stu.ibu.edu.ba

Primljen / Recieved : 7. 3. 2022.
Prihvaćen / Accepted: 18. 3. 2022.

Abstract

One of the conditions that Bosnia and Herzegovina needs to meet in order to become a member of the European Union is to increase the share of electricity production from renewable sources. The aim of this paper is to use a Linear Regression algorithm, Support Vector Machine algorithm and Random Forest algorithm to make a contribution in this area from the aspect of machine learning - more precisely to determine the parameters that most affect the wind speed in Mostar region. The obtained model showed that parameter that most affects the wind speed in Mostar are short-term wind accelerations. The model can be easily generalized and applicable to other data sets.

Key words: European Union, linear regression, random forest, support vector machine, renewable sources, model, machine learning, datasets

Sažetak

Jedan od uslova koje Bosna i Hercegovina treba da ispuni da bi postala članica Evropske unije je povećanje udjela u proizvodnji električne energije iz obnovljivih izvora. Cilj ovog rada je da se algoritmima linearne regresije, pomoćnih vektora, te algoritmom slučajna šuma da doprinos ovoj oblasti sa aspekta mašinskog učenja, tačnije da se odrede parametri koji najviše utiču na brzinu vetra. Dobijeni model je pokazao da parametar koji najviše utječe na brzinu vjetra u Mostaru su kratkotrajna ubrzanja vjetra. Model se može lako generalizirati i primijeniti na druge skupove podataka.

Ključne riječi: Evropska unija, linearna regresija, slučajna šuma, algoritam pomoćnog vektora, obnovljivi izvori, model, mašinsko učenje, skupovi podataka

1. INTRODUCTION

The first weather forecast was published in the British newspaper Daily News in 1848. Since then, the measurement of atmospheric conditions has progressed. Meteorological stations are present everywhere. They are equipped with thermometers, hygrometers, anemometers, rain gauges and barometers. In addition to stations, data is also collected through meteorological balloons, meteorological aircraft, meteorological satellites, and radars.[1]

People have used renewable energy sources since ancient times. Hundreds of years ago, the soil of Europe was covered with Dutch windmills. Before the grid became ubiquitous, farmers used wind turbines to charge batteries, and further used batteries to run equipment in the workplace. Again, the world faces the task of literally returning to the initial forms of energy use due to global climate change.

Many countries around the world are encouraging the production of energy from renewable sources, and the population is increasingly accepting these forms of energy production. Thus, the essence is to reduce carbon dioxide emissions from fossil fuels through renewable energy sources. The United Nations Intergovernmental Panel on Climate Change (IPCC), the International Institute for Applied Systems Analysis (IIASA) and the World Energy Council are working to make plans for how almost all electricity will come from renewable sources in the future. It is claimed that in 2100, 65% of the total electricity produced in the world will be from renewable sources. However, for now, these are just goals. These are very sensitive data and it is uncertain to predict these things. Some countries have greater resources for energy from renewable sources, and some are limited. The price of fossil fuels will greatly affect these things. The big question is how to adapt all physical infrastructure to renewable sources [1]. Machine learning contributes in some way to global world goals. With the help of machine learning, it is possible to obtain certain predictions about the weather in the future, based on the historical weather data.

Therefore, we aim to find the optimal parameters to create the most accurate model for wind power prediction.

This greatly facilitates the work of the management of electric power companies and utility companies. This contributes to the goal that Bosnia and Herzegovina need to meet in order to become a member of the European Union.

[1] J.F. Manwell and J.G. McGowan, Wind energy explained: Theory, Design, Application.

2. LITERATURE REVIEW

There is a lot of talk in the world about electricity. There are many reasons and many different approaches. They can be social, economic and environmental views. At the moment, we cannot imagine a world without electricity. It is the basic driver of all processes of the modern age. Electricity supply strategies are very important. However, countries do not have the same resources to generate electricity.

The researches in the field of renewable energy resources and wind power predictions were done before.

Z. Liu and W. Gao; Y.-H. Wan and E. Muljadi, used genetic algorithms to predict wind strength based on historical wind speed and direction data.[2] The paper is divided into two steps. The collected data is entered into an information system that filters the probability of a neural network. So, this step prepares the data that will be used for prediction. In the second step, the prediction itself is performed using a neural network. It was shown that the model achieves high accuracy in relation to the measured data.

There are works whose only goal is data processing and model optimization in case of wind power prediction like Like the work done by Nicolae-Buturache and Stancu. Wind energy production is soon linked to weather conditions. Nicolae-Buturache and Stancu are for initiating multiple machine learning with short-term, medium-term and long-term predictions. This paper offers theoretical considerations for the development of a technical solution that can be applied to support actual operations in the wind turbine industry. The purpose of this paper is to compare different models in the problem of wind power prediction.[3]

3. MATERIAL AND METHODS

3.1. Data collection and preparation

In this section, along with the description of the data set and the application of the EDA method, the purpose of the data itself and how they can be used will be explained. The data set was purchased on the website www.worldweatheronline.com, which contains weather data in all territories. It is necessary to select the desired territories and we will get all the historical measurements. In this case, the start date is July 1, 2008, and the end date is September 17, 2021. The number of rows is 115849. This dataset contains all measurements per dey, at the Fortica location near

2 Z. Liu and W. Gao; Y.-H. Wan and E.Muljadi , Wind power plant prediction by using Neural Network.

3 Adrian Nicolae-Buturache and Stelian Stancu, Wind energy prediction using machine learning.

Mostar where the wind farm is located. It is extremely important to predict wind strength at this location. The data set contains 32 columns. Thar are: location_id, date, isDayTime, tempC, tempF, windspeedMiles, windspeedKmh, winddirdegree, winddir16point, weatherCode, weatherIconUrl, weatherDesc, precipMM, precipInches, humidity, visibilityKm, visibilityMiles, pressureMb, pressureInches, cloudcover, heatIndexC, heatIndexF, DewPointC, DewPointF, WindChillC, WindChillF, WindGustMilles, WindGustKmph, FeelsLikeC, FeelsLikeF, uvIndex.

At this stage, columns not relevant for the model building are removed. We replaced the text values with numerical values in those columns that concern model creation. Without preprocessing the data, it is unthinkable to create the models themselves. This phase ensures that clean data is obtained for model making. Dataset has 3 different types: float64, int64 and object.

After pre-processing the data, encoding the data and ejecting unnecessary columns, the data set was reduced to 6 columns from which the optimal parameters will be selected. That are: time, isdaytime, tempC, windGustKmph, winddir16point, humidity.

Table 1. Preprocessed data

time	isdaytime	tempC	windspeedKmph	winddir16point	humidity
0	0	13	15	1	57
100	0	15	15	1	58
200	0	16	15	1	59
300	0	18	14	1	59
400	0	20	15	1	59
500	0	22	16	1	59
600	0	24	17	1	60
700	1	26	14	1	52
800	1	28	12	4	45
900	1	30	10	4	38
1000	1	31	8	1	37

3.2. Model training and testing

Model testing provides information about a model’s performance when it comes to its application in real life. 70% of the data set will be used for model training and 30% for model testing.

3.3. Model building

It is necessary to include a linear regression library through the Jupiter environment. The fit method corresponds to the model and is used to estimate the parameters in the model. The .predict method is used to obtain the predicted response using the model and takes the X_train predictor. It is also possible to calculate the average deviation. It is necessary to do the same for X_val. For Support Vector Machine, an SVM plugin is included in the Jupyter Notebook environment. The goal of all three models is to predict wind strength in the future. Algorithms were chosen mostly because of the structure of the data itself and the type of problem. For example, genetic algorithms are unacceptable for solving this problem due to the specifics of the field.

4. RESULTS AND DISCUSSION

Root Mean Square Error (RMSE) has been used as a measure for Model reliability and evaluation and evaluation of the model itself. The ultimate goal is to obtain the smallest possible RMSE by combining different dependent variables. In general, a lower RMSE is better than a higher one. Normalization and standardization of data was applied before the model was built. There is a large range between variables in the data set itself. Correlation matrix was applied that tells exactly what the relationship is between the variables in the model.

Table 2. Results od different models

Model	RMSE	Parameters
Linear regression	1.0044741170101779	Wind gust kmph
Random forest	0.957516237635529	Wind gust kmph
Support Vector machine	1.1074874511186479	Wind gust kmph

The parameter that most affects the wind speed at Fortica is 'Wind gust kmph'. The correlation matrix showed that the best-selling parameter is wind gust km km and algorithms were tested and trained with this parameter.

5. CONCLUSION AND FUTURE WORKS

The aim of this work is to predict the wind speed on Fortica location, above Mostar and find the optimal parameters that affect it. The most optimal parameter for the model itself is 'wind gust kmph'. The advantage of this work is that it is flexible and will be easily applied to different data sets and for different locations.

In the future, the entire research paper could be implemented in one web application. The application would use an API call with pages that have time data. The data would be stored in one of the stores (Google, AWS, Oracle, Microsoft). For machine learning only, python scripts would be called in the background, and the operations themselves could be performed on the application through a simple UI. It would be very easy to create data sets, jobs, selection of parameters, as well as data on accuracy itself through the platform. Of course, it would be good for people who know the problem to work on the platform itself. This is exactly the great benefit of this work, which allows it to train and use the model itself in different locations and data choices.

References

- J.F. Manwell and J.G. McGowan, Wind energy explained: Theory, Design, Application
- Z.Liu and W. Gao; Y.-H. Wan and E.Muljadi , Wind power plant prediction by using Neural Network
- Adrian Nicolae-Buturache and Stelian Stancu, Wind energy prediction using machine learning

Other references

- [1] Renewable energy sources and climate change Mitigation, retrieved from https://archive.ipcc.ch/pdf/special-reports/srren/SRREN_FD_SPM_final.pdf