Engineering and Technology Journal e-ISSN: 2456-3358

Volume 09 Issue 09 September-2024, Page No.-5103-5110

DOI: 10.47191/etj/v9i09.13, I.F. – 8.227

Estimating the Wind Power by Using K-Nearest Neighbors (KNN) and Python

Abumuslim KHUJAEV[1], Osman YILDIRIM[2], Faiz BENZAIBAK[3]

[1]MS Student, Electrical and Electronic Engineering, Istanbul Aydin University, Istanbul, Turkiye

[2]Professor, Electrical and Electronic Engineering, Istanbul Aydin University, Istanbul, Turkiye

[3]MS Student, Electrical and Electronic Engineering, Istanbul Aydin University, Istanbul, Turkiye

ABSTRACT: In recent years the production and distribution of renewable energy has become a popular trend in power generation, and renewable energy sources such as hydro, solar, and wind energy are examples of natural and clean energy sources. Among the renewable energy sources, wind energy source is one of the trending and crucial energy sources that has outstanding electricity power generation. This research investigates an estimation model of wind power to address the real-life issues in renewable energy power generation by estimating an accurate amount of wind power production by applying the machine learning (ML) K-Nearest Neighbors (KNN) technique integrated with Python programming. To achieve the expected outcome for this research, it is expected to collect data from wind energy, wind speed forecasts, and all the relevant information that supports the research in analyzing the technique. There are various meteorological methods to evaluate the performance of wind power generation, but in this research, the methodology is to employ machine learning techniques such as KNN integrated with Python coding to achieve estimated wind energy output. Python is a software program and is diversely used for coding various algorithmic formulas to obtain results. In this research, Python is best suited for investigating the performance of the KNN technique for estimating the most accurate result for wind power generation. Here, the report presents the methods of applying the KNN technique and displaying the outcome on screen by a graph representation. The outcome of the research showed that the KNN is a more reliable technique to make an accurate prediction on wind power generation. A similar approach can be implemented in various renewable energy data to estimate an accurate energy output.

KEYWORDS: Renewable Energy, Wind Energy, Machine Learning, K-nearest Neighbors, Python.

1. INTRODUCTION

The use of renewable energy is becoming a global trend in current world and it has become an integral part of our life [1]. Renewable energy such as hydro, solar and wind are the most popular trending energy production that are committed to generating power in commercial, transportation, industrial and healthcare sectors. Countries, are in stage of economic development and the number of infrastructures, industries, technological advancement, and manufacturing are growing rapidly [2], [3]. Therefore, it is crucial to build more energy production facilities to maintain balance and required power supply for the variety of loads. Wind energy is one of the renewable energy system which as a good performance and a good power supply in renewable energy production [3]. In addition wind energy does not require a specific time to generate electricity. Wind energy system has the potential to provide large amount of electrical energy, it cost less expensive than other renewable energy system, and it can be installed in any location with enough wind speed. The wind energy production is dependent on the performance of the wind speed, internal mechanism of wind turbine, weather condition, and efficiency of the wind turbine [4]. There are various methods that can be used to eliminate the limitations in wind energy production and develop methods to address the limits for better power generation. In addition, machine learning is one of the necessary methods which can give a positive respond such as balance power estimation in power output of the wind energy system. The machine learning require input data from wind energy to apply algorithm to provide estimated power output from wind power to the power grid.

There are several type of machine learning techniques to calculate for estimated power output of wind energy system [5]. In addition, KNN is an algorithmic technique that is used for estimating new data with more accuracy data output. It takes the existing data and estimate new data for wind power and the estimated output power can be used as the power supply to the grid [6]. The word K refers to the number of nearest neighbors taken into consideration for estimating predictions, in which the higher number of K makes better decision on predicting a new data set [7]. The similar method can be used in various research works where engineers can estimate new data set for a performed research work. The machine learning techniques performed by using various programming applications. Furthermore, Python is one of the popular software program which can be used in

processing the KNN technique for estimating new data set related to wind energy output. Despite having a great tool in different domains, KNN has great estimating data estimating technique for especially small to medium dataset, where it can compute better performance and serves for various machine learning models.

Python is an open source, programming language that has a wide range of implementations, and sources that used in different applications [8], [9]. In addition, it support various machine learning techniques to process input data to a new set of data. Here, the research process a wind energy dataset to provide an output dataset for wind energy system. Two datasets were selected from Waterloo Wind Farm and KNN algorithm was applied to select a model with closest neighbor to k value. Then the second code used to find the regression result. Hence, the third KNN code was used to predict accuracy result for the wind energy.

2. LITERATURE REVIEW

In machine learning KNN algorithm is one of the best option to work with, and it is used in many applications for processing, detecting, and sensing obstacles. There are many research conducted by using KNN algorithm to process variety of data input. According to Jian [10] a data received from a fall detector and sent to mobile phone and program running based on K-NN algorithm will process and give a message to other people in charge to come and help the fallen person. This paper proved that 95% of the cases were reported positively and the elderly people who experience such incidents will get immediate help by the healthcare centers. Theresia Diah [11] studied emotion recognition based EEG signals. There were several methods to classify different types of time frequency domain elements. The author also used some methods to select the perfect elements for classification. The results were generated as dataset and implemented in DEAP dataset. Furthermore the DEAP dataset then represented in time-frequency domain, processed by K-NN algorithm, and it showed a good performance. The K-NN classifier obtained 60.68% recognition as compare to other classifying methods which is high rate. Yuni Yamasari [12] studied students' academic achievement based on K-NN technique. The technique enables professors to evaluate students' performance and constantly update their teaching methods. The study categorized students' achievements in different class categories and created a data set for processing in K-NN method. The study achieved a good result and showed 84.52% accuracy to the students' performances data.

The machine learning can be used in many applications such as predictive and estimating conditions. One of the cause of traffic accidents can be reduced by daily collecting data and road movements. The data can be collected and implemented in machine learning processing [13] and predict a solution to minimize the road accidents. The techniques used in machine learning algorithm includes various techniques such as KNN, decision tree, and random forest. The data used to predict a solution with accuracy results which can improve the system's behavior for future accident prevention. The author claimed that the solution can be reported to related authorities and include them in their policies so that the system can help in reducing the fatality rates. The KNN technique is employed in medical cares. The study employed electroencephalography data in machine learning to predict accuracy for whether eyes is open or closed [14]. The data was obtained by manually recording every open and closed eye condition and the data was processed using KNN machine learning technique. The results obtained showed a better result in prediction accuracy that achieved in lesser time comparing with complex machine learning methods. Senthil [15] studied the heart disease and cardiovascular diseases which occur in various range of heart conditions. The author claimed that machine learning suing mathematical models can provide a better and accurate response to the data obtained from the heart condition. The data for processing in methodology obtained by recording heart pulse hear measurements or tomography devices. The KNN machine learning technique helps the cardiologist to give proper treatment for sick person in advance. The results obtained by the author gave an accurate data output for giving an efficient and accurate response to minimize the risk for cardiac arrest.

Yang Gong [16] conducted research on air quality level prediction using KNN machine learning technique. In the first instance, the author generated an input data by collecting data from various sources of air quality, and saved the data in CSV file. The data was shown in figures and the KNN algorithm applied to evaluate the accuracy of the air quality grade. The research showed a better result and the work achieved the expected result. Water quality grade is another important task for governments to provide for the community. A research conducted by Xudong [17] analyzed the water quality using different analysis. At first the data was obtained from kaggle source and the data was computed in Python programming by applying KNN machine learning technique to estimate accuracy for the water quality level. The results were compared with other conventional methods and KNN algorithm showed a better result. The modern economic developments in many countries has increased the complexity of the stock market performance. Qian Yunneng [18] claimed that the accurate analysis to forecast the stock price has become a challenging issue in many companies. Various prediction models are exist to analyze and predict the most accurate result for the market price. The author picked KNN machine learning technique and compare the result with the conventional KNN algorithm. The analysis for new method had shown a better result in prediction performance of stock prices than the conventional algorithm.

3. METHODOLOGY

In this section the research explains the steps and methods to estimate a result for the KNN machine learning technique. A

set of data gathered from Waterloo Wind Farm Sothern Australia [19], [20]. The first turbine model was Vetas V90 2000 and the second model was chosen GE 1.7. The dataset was provided by MERRA-2 (Global) Modern Era Retrospective Analysis for Research and Applications at Renewables ninja. The inputs need to be carefully obtained to predict a model between the existing Vetas V90 2000 and GE 1.7 models. The models show the data in kW from January 2019 to late December 2019. It is necessary to consider all meteorological factors to investigate and provide a positive contribution in wind energy system.

The information collected for the two wind turbine models collected from Waterloo wind farm itself as shown in Figure 1. Waterloo wind farm is located in southern part of Australia with localization information latitude: -33° 59' 47.3" longitude: 138° 54' 54.7" [19], [21].

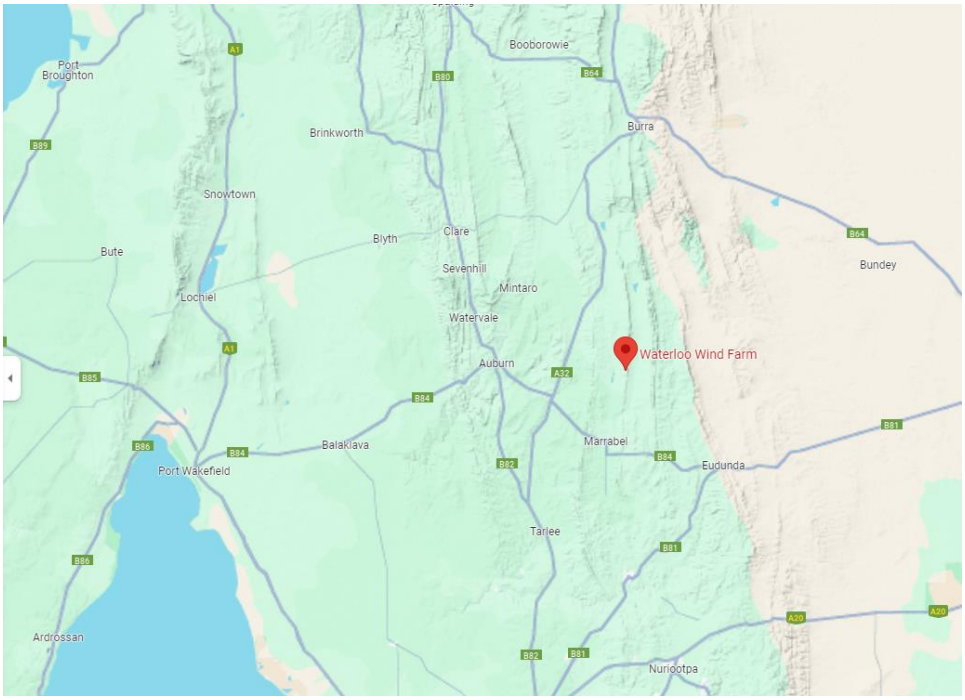


Figure 1. Shows the location of Waterloo Wind Farm location [20].

In this dataset the wind speed and electrical power was collected 10 times in a day and the same recording was continued from 1 January 2019 to 31 December 2019. The measurement was recorded for both Vetas V90 2000 and GE 1.7 models. The data set made up to 8750 samples and the mean values for each month were calculated that was going to be used for machine learning processing. The data set for GE 1.7 model was collected from the same location and similar calculations were made for the second model [20].

Waterloo Wind Farm is one of the largest energy farm which generates electricity and the amount of electricity generated in a year is shown in Table 1. The power generation for model 1 and model 2 which are V90 2000 and GE 1.7 models perform different. The first turbine generator model 1 generated 3.431498 MW and turbine generator model 2 generated 2.836934 MW electrical power. Variation in wind speed performance has impact on electrical output power in each month data collection. The following (Figure 2 and Figure 3) shows the power generated from January 2019 to December 2019 for model 1 and model 2 by considering the mean values of monthly data collection for both wind speed and power. Python was used to scatter the figures. The wind speed below 3.5 m/s and above 25.5 m/s were considered as zero for electrical power output as it is outside the accepted power range performance of the wind turbine. Furthermore, the accepted range for active power can be evaluated by the data obtained from model 1 and model 2. The wind speed characteristics is shown in Figure 4 with data from model 1 and model 2

.

```
import matplotlib.pyplot as plt

x = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
y = [310.272,253.428,242.256,320.548,249.135,
     239.605,253.031,322.421,274.764,315.96,350.067,300.011]
plt.title('Monthly Power Generation Model1')
plt.xlabel("Months")
plt.ylabel("Power kW")
plt.scatter(x, y)
plt.show()
```

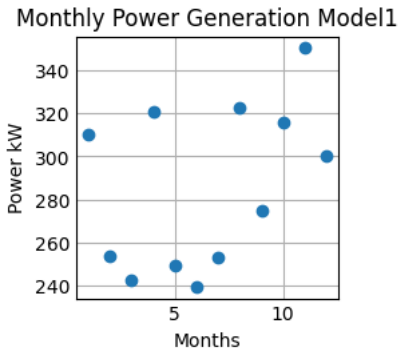


Figure 2. Shows the power generated for Vetas V90 2000 model (Model 1).

```
[48]: import matplotlib.pyplot as plt

x = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
y = [203.174,183.464,205.852,226.02,263.368,207.789,
     271.476,270.271,238.925,257.744,284.15,224.701,]
plt.title('Monthly Power Generation of Model2')
plt.xlabel("Months")
plt.ylabel("Power kW")
plt.scatter(x, y)
plt.show()
```

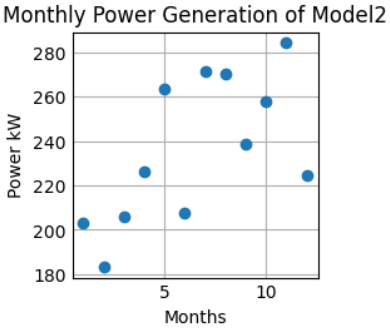


Figure 3. Shows the power generated for GE1.7 model (Model 2).

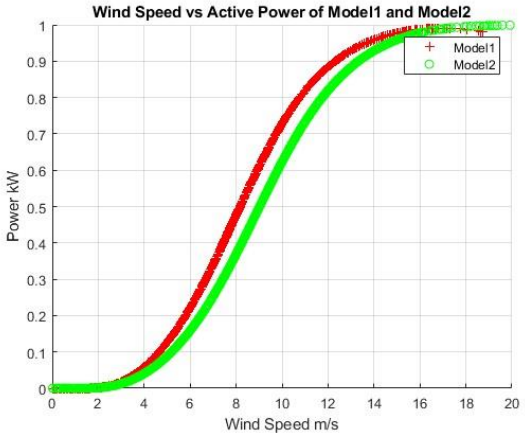


Figure 4. Shows the performance of wind speed and active power.

Table 1. Power generation in 2019 for Vetas V90 2000 and GE 1.7 wind turbine models.

Wind turbine model types	Power Generated in 2019
Vetas V90 2000	3.431498 MW
GE 1.7	2.836934 MW

The prediction data set for the selected models were evaluated using KNN machine learning technique to estimate which model is more accurate to use in Waterloo Wind Farm for electricity generation. Considering the economic aspects, weather condition, and wind speed at wind farm location, KNN algorithm estimates which model is more suitable to install for generating electricity. The regression and accuracy predictions were analyzed for the dataset obtained in this research. The processes conducted by implementing machine learning technique in Python programming.

4. RESULTS AND DISCUSSION

In this section the paper discusses the performance of KNN machine learning technique and results obtained for wind power data. The technique was applied and analyzed the results in Jupyter.org Notebook Python. The computer specification for this analysis had 16GB RAM, RYZEN 9 4000 SERIES, Base speed 3.3 GHz, NVIDIA GTX 1660 Ti, and SSD memory. The hardware characteristics were fine to run and obtain results for wind power dataset.

The wind power dataset was collected from Waterloo Wind Farm from January 2019 to December 2019 on daily basis [20]. In this paper the dataset merged into monthly basis by considering the average wind power and wind speed for each month. The dataset merged from 8750 samples to 12 samples that each samples represent each month accordingly and similar merging applied for wind power. At first a model selection technique was used to select the best suitable wind turbine for the wind farm. The data set shown in figure 2 and figure 3 were plotted and using KNN machine learning technique model 1 which is Vetas V90 2000 was closest neighbor to the predicted wind power point. The KNN regression and KNN accuracy estimation for wind power data for model 1 were proceeded. The result for model selection is shown in Figure 5. The dataset for model 1 is represented by red circles and the model 2 dataset represented in green circles. The blue circle is closer in neighboring with model 1 so the Vetas V90 2000 data used for the next procedure

.

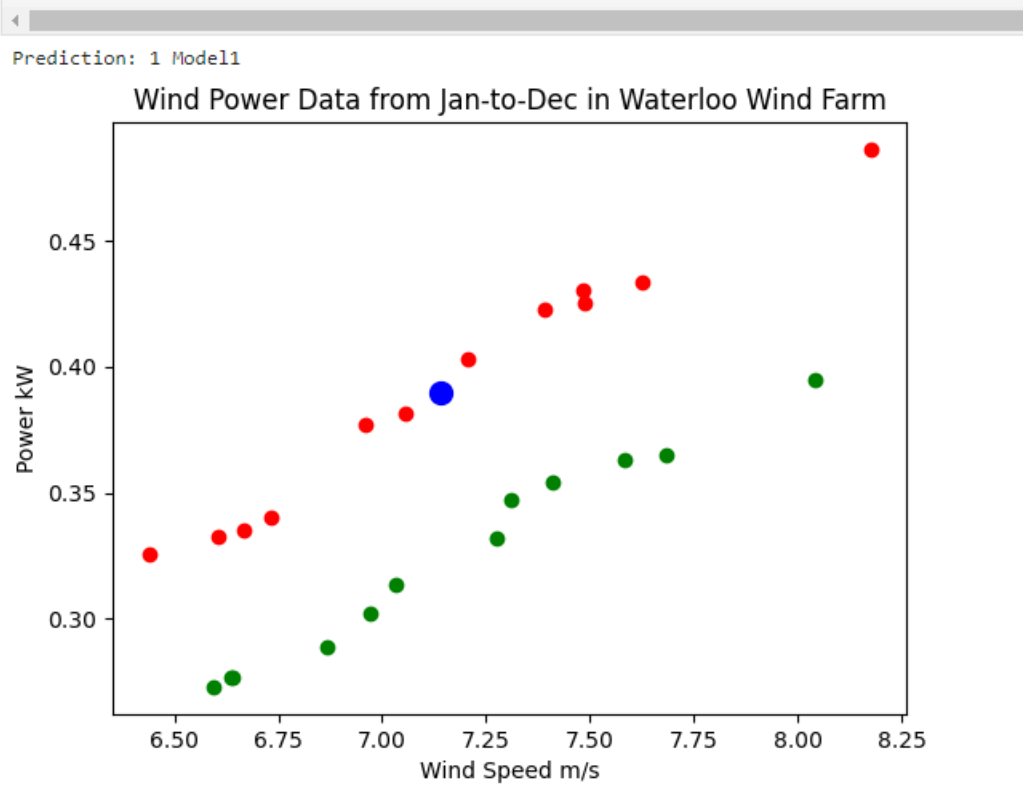
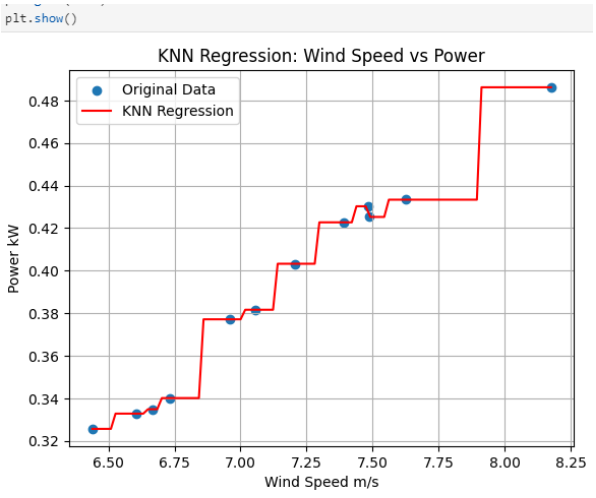
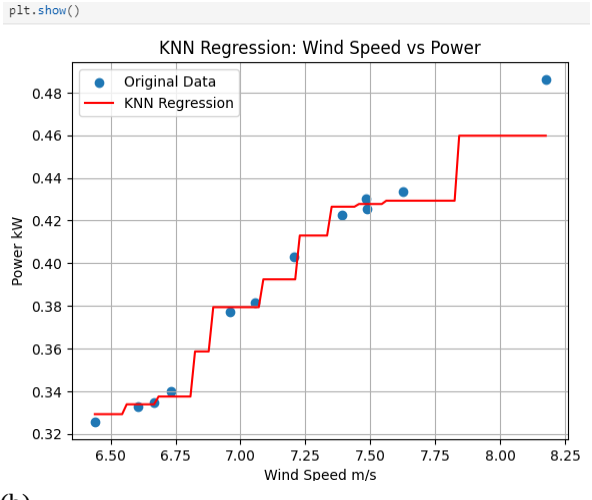


Figure 5. Shows the mean power and mean wind speed in 2019.

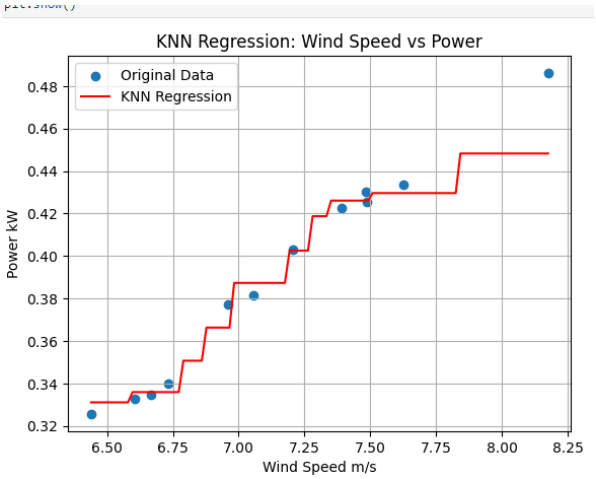
The wind power estimator was able to predict data accuracy by employing KNN algorithm in Python programming. Figure 6 shows the KNN methods in regression plot to shows the accuracy of the data. As the research has 12 samples the k value range applied between k1, k2, k3, and k4. The predicted values are plotted as follow.



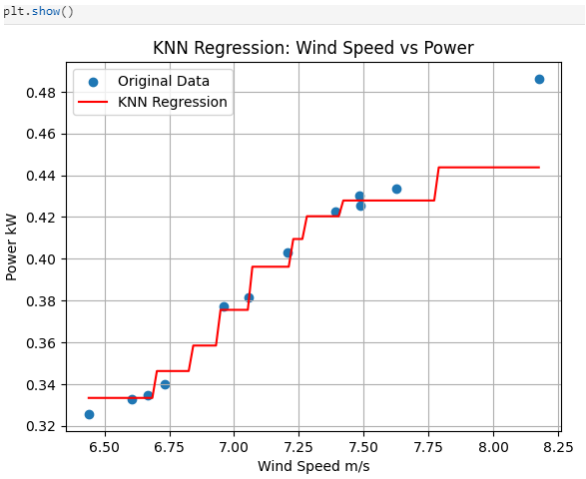
(a)



(b)



(c)



(d)

Figure 6. Shows the regression graphs and the predicted values by using the KNN technique at (a) k1, (b) k2, (c) k3, and (d) k4 values.

Figure 7 Shows accuracy scores for wind power dataset for Vetas V90 2000 wind turbine. The dataset practiced in KNN accuracy for different k values. The power threshold estimated at >0.4 kW which the turbine estimated to perform in daily power generation. The highest accuracy score obtained at 0.93 at k=1. Consequently, when the number of k increased the graph start to show various results as shown in figure below. Hence, the increasing of KNN value reduced the accuracy of the wind power generation and it is suggested to keep the k values between 1 and 3 for achieving the most accuracy of power

.

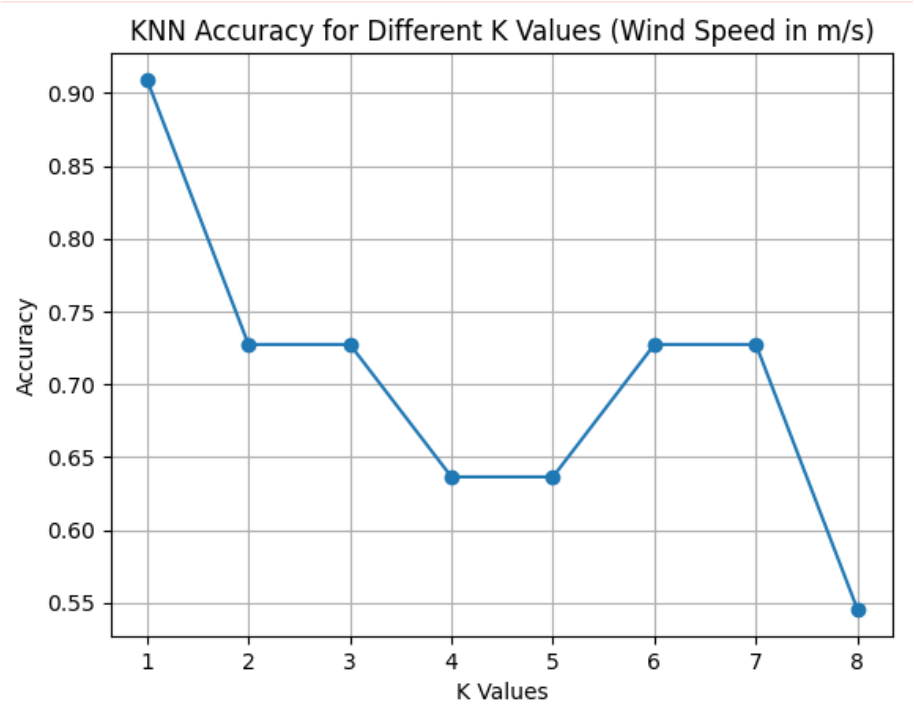


Figure 7. Shows the result for KNN accuracy for different k values.

The summary of the result demonstrate that KNN algorithm is one of the helpful machine learning technique that can help to attain optimum results and it is influence by value of k. A dataset can be varied to small or big number, and the value of k can predict the nearest value or predict accuracy of dataset.

5. CONCLUSION

To conclude, the research was about estimating wind energy using KNN nearest neighbor algorithm machine learning technique on Python programming. The research conducted important steps to find accuracy result and provide estimated wind power in monthly basis. A dataset was collected from Waterloo Wind Farm located in Southern Australia. The

dataset was collected from 1st January 2019 to 31st December 2019 every 1 hour. The mean value was taken for every month and KNN algorithm was applied to identify the type of model. The k value appeared near model 1 and the dataset for model 1 was selected for performing regression method. The regression value was obtained for k1, k2, k3, and k4 respectively (see. Figure 6). Python programming was used to plot the accuracy result and calculate for KNN values. Hence, the similar method and codes can be used for any dataset which related to wind energy and even the solar energy dataset can interact with this methodology to achieve expected results.

It is recommended for researchers to take dataset from different locations and study the method using KNN algorithm. This method can help them to predict more accuracy result and it helps them to find the geographical location is suitable for energy production.

REFERENCES

1. Y. Sugiawan and S. Managi, "New Evidence of Energy-Growth Nexus From Inclusive Wealth," *Renewable and Sustainable Energy Reviews*, vol. 103, pp. 40–48, Apr. 2019, doi: 10.1016/j.rser.2018.12.044.
2. N. S. Caetano, T. M. Mata, A. A. Martins, and M. C. Felgueiras, "New Trends in Energy Production and Utilization," vol. 107, pp. 7–14, Feb. 2017, doi: 10.1016/j.egypro.2016.12.122.
3. F. Martins, C. Felgueiras, M. Smitkova, and N. Caetano, "Analysis of Fossil Fuel Energy Consumption and Environmental Impacts in European Countries," *Energies*, vol. 12, no. 6, p. 964, Mar. 2019, doi: 10.3390/en12060964.
4. T. Kurbatova and T. Perederii, "Global Trends in Renewable Energy Development," in *2020 IEEE KhPI Week on Advanced Technology (KhPIWeek)*, 2020, pp. 260–263. doi: 10.1109/KhPIWeek51551.2020.9250098.
5. W. Lv, H. Huang, W. Tang, and T. Chen, "Research and Application of Intersection Similarity Algorithm Based on KNN Classification Model," in *2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA)*, Xi'an, China: IEEE, May 2021, pp. 141–144. doi: 10.1109/CAIBDA53561.2021.00037.
6. M. S. Mathew and M. L. Kolhe, "Performance Modelling of Renewable Energy Systems Using kNN Algorithm for Smart Grid Applications," in *2022 7th International Conference on Smart and Sustainable Technologies (SpliTech)*, Split / Bol, Croatia: IEEE, Jul. 2022, pp. 1–4. doi: 10.23919/SpliTech55088.2022.9854321.
7. "The big book of machine learning use cases – 2nd edition." [Online]. Available: https://www.databricks.com/resources/ebook/big-book-of-machine-learning-use-cases?scid=7018Y000001Fi0oQAC&utm_medium=paid%2Bsearch&utm_source=google&utm_campaign=17152786057&utm_adgroup=139126362849&utm_content=ebook&utm_offer=big-book-of-machine-learning-use-cases&utm_ad=666067181030&utm_term=machine+learning+algorithms&gad_source=1&gclid=Cj0KCQiAwP6sBhDAARIsAPfK_wbMFxpeKtIYwTAsAxhrVW417NWEJbPT3QCuTXc3lzNzYBvzwWlaCHcaAn8CEALw_wcB
8. A. Kumar and Supriya. P. Panda, "A Survey: How Python Pitches in IT-World," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India: IEEE, Feb. 2019, pp. 248–251. doi: 10.1109/COMITCon.2019.8862251.
9. X. Zhang, R. Yan, J. Yan, B. Cui, J. Yan, and J. Zhang, "ExcePy: A Python Benchmark for Bugs with Python Built-in Types," in *2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, Honolulu, HI, USA: IEEE, Mar. 2022, pp. 856–866. doi: 10.1109/SANER53432.2022.00104.
10. H. Jian and H. Chen, "A portable fall detection and alerting system based on k-NN algorithm and remote medicine," *China Commun.*, vol. 12, no. 4, pp. 23–31, Apr. 2015, doi: 10.1109/CC.2015.7114066.
11. K. Theresia Diah, A. Faqih, and B. Kusumoputro, "Exploring the Feature Selection of the EEG Signal Time and Frequency Domain Features for k - NN and Weighted k-NN," in *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)(47129)*, Depok, West Java, Indonesia: IEEE, Nov. 2019, pp. 196–199. doi: 10.1109/R10-HTC47129.2019.9042448.
12. Y. Yamasari and R. A. Ahmad, "Classifying Student Achievement Using K-NN Based on Feature Normalization Techniques," in *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia: IEEE, Sep. 2023, pp. 92–96. doi: 10.1109/iSemantic59612.2023.10295372.
13. A. M, A. K, A. K, A. M, and C. K. R, "Accident Prediction Using KNN Algorithm," in *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Mandya, India: IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/ICERECT56837.2022.10059746.
14. S. H. Adil, M. Ebrahim, K. Raza, and S. S. Azhar Ali, "Prediction of Eye State Using KNN Algorithm," in *2018 International Conference on*

Intelligent and Advanced System (ICIAS), Kuala Lumpur: IEEE, Aug. 2018, pp. 1–5. doi: 10.1109/ICIAS.2018.8540596.

15. S. G. A, R. Prabha, M. Razmah, T. Veeramakali, S. S, and Y. R, "Machine Learning Heart Disease Prediction Using KNN and RTC Algorithm," in *2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, Chennai, India: IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/ICPECTS56089.2022.10047501.
16. Y. Gong and P. Zhang, "Research and Realization of Air Quality Grade Prediction Based on KNN," in *2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*, Manchester, United Kingdom: IEEE, Oct. 2021, pp. 299–304. doi: 10.1109/AIAM54119.2021.00068.
17. X. Jia, "Detecting Water Quality Using KNN, Bayesian and Decision Tree," in *2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)*, Hangzhou, China: IEEE, Mar. 2022, pp. 323–327. doi: 10.1109/CACML55074.2022.00061.
18. Q. Yunneng, "A new stock price prediction model based on improved KNN," in *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, Changsha, China: IEEE, Dec. 2020, pp. 77–80. doi: 10.1109/ICISCE50968.2020.00026.
19. "Waterloo Wind Farm (Australia), the Wind Power." [Online]. Available: https://www.thewindpower.net/windfarm_en_10457_waterloo-wind-farm.php
20. "Wind power data." [Online]. Available: [1] Renewables.ninja, https://www.renewables.ninja/ (accessed Jan. 1, 2024).
21. "Waterloo wind farm locatin." [Online]. Available: https://www.google.com/maps/place/Waterloo+Wind+Farm/@-33.6747011,136.8091401,8z/data=!4m6!3m5!1s0x6ab971db7fee0c9f:0x3c1e00772d32e584!8m2!3d-34.0107713!4d138.9108365!16s%2Fg%2F11rnk797fb!5m1!1e4?hl=en&entry=ttu (accessed Jan. 1, 2024).