Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Turbine recommender: The selection of wind turbine type using one of a machine learning technique

Mayda Alrige *, Hind Bitar, Joud Aljaeed, Somaiah Alasmari

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 29 August 2021
Received in revised form
12 November 2021
Accepted 16 December 2021

Keywords:
Turbine recommender
Machine learning
Modelling
Wind turbine energy
KNN

ABSTRACT

This study aims to utilize the machine learning technique to build a model to recommend the suitable wind turbine type based on some variables, such as air speed and air density, as well as visualize the location of the recommended wind turbine selection on a 3D map. Particularly, we applied the K-nearest neighbor model (KNN) to determine the amount of energy produced by a single wind turbine. We applied it on 10 separate wind farms in Saudi Arabia. The results indicate that the model performs very well in predicting the best wind turbine type with the mean accuracy of 88%, where ten wind stations resulted from the optimized model with the suggested turbine type in each station. Adding more wind attributes and other factors may assist in increasing the model mean accuracy. The project's findings will assist decision-makers in Saudi Arabia to make informed decisions as to what kind of wind turbine is suitable for a specific location. In the long run, this will help to make wind energy-a sustainable source of energy-one of the main goals of the 2030 vision, specifically under National Industrial Development and Logistics Program.

1. Introduction

The kingdom of Saudi Arabia has a distinct geographical and climatic location that makes utilizing renewable energy sources plausible and economically attractive. For this reason, the 2030 vision has set investing in renewable energy as one of the main goals' emissions. In line with this goal, the National Renewable Energy Program (NREP) has been formed as a strategic initiative to increase the Kingdom's share of renewable resources. Since its launch back in 2015, Renewable and Sustainable Energy (RnSE) resources have recently been marked as a major contributing factor for a stable economy in the Kingdom of Saudi Arabia (KSA) (Amran et al., 2020). Indeed, wind energy is one of the main renewable energy sources due to its natural, cheap, and clean nature. It is possible to produce energy from wind turbines at any hour of the day and it is suitable for systems that continuously require energy (Demolli et al., 2019).

However, the efficient selection of a wind turbine at a given site is presently limited by the developer's knowledge of what turbines are available on the market, as well as their inability to test and compare available turbine designs before investing. Poor turbine selection results in a financially sub-optimal investment. There are many types of wind turbines. For a specific location, determining the most suitable type depends on many contextual variables. This makes the process to identify the proper type for a specific place lengthy and complicated. New advanced methods are needed to facilitate this process. Traditional methods, such as blade element momentum theory (BEM), allow all possible turbine designs to be analyzed (Perkin et al., 2015). This is a lengthy complicated process that takes an extended amount of time (up to several months) to determine the best suitable wind turbine.

The proposed research aims to accelerate and facilitate the process of establishing the best wind turbine type-specifically during the installation of new infrastructure in a specific wind turbine farm-by utilizing machines leaning on historical data. Our research aims to produce several results. This includes a (1) predictive model, or the building of a model that would recommend the best suitable wind turbine type that produces more energy based on relevant contextual variables for a certain wind turbine farm, such as wind speed and air density (W/m2). As there are many types of wind turbines,

* Corresponding Author.
Email Address: malraegi@kau.edu.sa (M. Alrige)
https://doi.org/10.21833/ijaas.2022.02.013
Corresponding author's ORCID profile:
https://orcid.org/0000-0002-1315-7053

our project will suggest the most suitable type based on how much energy this specific wind turbine produces. This research also aims to produce a (2) web-based map to visualize the location of the recommended wind turbine stations in a 3D map.

To achieve these two aims, we followed the cross-industry standard process for data mining (CRISP-DM) method, which consists of a six-phase process that naturally explains the data science life cycle. CRISP-DM is highly recommended to follow when the project is goal-directed and process-driven (Martínez-Plumed et al., 2019). We used two data sets after obtaining all relevant legal approvals. One of the research's aims is to utilize machine learning techniques to optimize the wind turbine selection process. After data preparation, we implemented the multiple regression techniques due to our utilization of continuous variables.

2. Literature review

Machine learning has been widely utilized in the domain of wind energy forecasting, specifically for forecasting long-term wind power values with respect to historical wind speed data. Furthermore, the results showed that machine learning-based models could be applied to a location different from model-trained locations. Researchers have discovered that machine learning algorithms could be successfully used before the establishment of wind plants in an unknown geographical location-and whether or not it is logical-by using the model of a base location (Demolli et al., 2019).

In this domain of wind energy forecasting, machine learning has shown the ability to support the optimization and estimation of renewable energy. Energy engineers and data scientists have previously used machine learning algorithms. They are suitable choices in comparison to other methods, including rule-based methods, due to the fact that this is a poly-parametric problem containing a large amount of data. Also, attributes take different values. For instance, this method would predict the turbine response for any combination of wind speed, turbulence intensity, and wind shear that might be expected at a turbine site (Vladislavleva et al., 2013). The accuracy of these kinds of algorithms for power predictions is three times higher than that of the traditional power curve methodology (Clifton et al., 2013).

A 2015 study estimated the value of energy produced using machine learning algorithms based on the temperature, wind speed, and direction values collected from the wind turbine. The estimation of these values prior to the installation of wind turbines helped determine the energy value to be generated by meteorological measurements and assisted in more efficient operation during the operation period. As a result, resources were correctly directed, and the wind turbine was installed in the most appropriate location. Moreover, a mathematical equation that correctly estimated the energy production value by 90% was used. In addition, a computer program was developed for other users to view the results of this mathematical equation (Aksoy and Selbaş, 2021).

Another study was conducted in rural areas over Switzerland to estimate wind speed. This study proposes a methodology combining machine learning, GIS, and wind models to estimate theoretical wind speed. This estimate was based on measurements of wind speed and several meteorological, topographic, and wind-specific features available across the country. The wind speed values were calculated at a typical height for rural commercial wind turbine installation, which is z=100m. However, due to the availability of data of interest, the methodology developed is applicable to any large region (Assouline et al., 2019).

In Saudi Arabia, wind energy production is still in its infancy. The Kingdom of Saudi Arabia recently set ambitious targets in its national transformation program and vision for 2030 to move away from oil dependence and redirect oil and gas exploration efforts to other higher-value uses, chiefly meeting 10% of its energy demand through renewable energy sources (Amran et al., 2020).

Brahimi (2019) utilized the artificial neural networks (ANNs) method as a means of predicting daily wind speed for wind energy conversion systems (WECS) in a number of locations in Saudi Arabia with the ultimate goal of monitoring, controlling, planning, and dispatching generated power while meeting customer needs. His algorithm is built based upon multiple local meteorological measurement data provided by King Abdullah City for Atomic and Renewable Energy (K.A.CARE). The suggested model is a feed-forward neural network model with the administered learning technique using a back-propagation algorithm. After comparing his model with four other machine learning models, he concluded that it is feasibly possible to predict wind speeds for executing sustainable integration of wind power into Saudi Arabia's electrical grid and assisting operators in efficiently managing generated power.

Another study by Brahimi and a group of researchers utilized KNN not only to predict wind speed but also to select the best site for wind turbine installation within a wind farm. This included ensuring a secure and reliable electrical power output and helping the operators in a wind farm to manage the generated power efficiently.

Based on our knowledge, most of the research that has been conducted in Saudi Arabia is in the domain of wind energy geared toward predicting wind speed (Aksoy and Selbaş, 2021). This study focuses on understanding other contributing wind energy variables and specifically wind turbine data combined with wind farm speed data.

3. Materials and methods

In this research, we have followed the cross-industry standard process for data mining (CRISP-DM) methodology. CRISP-DM is a six-phase process model that naturally explains the data science life cycle. It is highly recommended to follow when the data science project is goal-directed and process-driven (Martínez-Plumed et al., 2019). Our research aims to optimize the wind turbine selection process by recommending the best wind turbine that would produce the maximum wind energy in a specific wind farm. This section demonstrates the six stages that we performed following the CRISP_DM, as seen in Fig. 1.

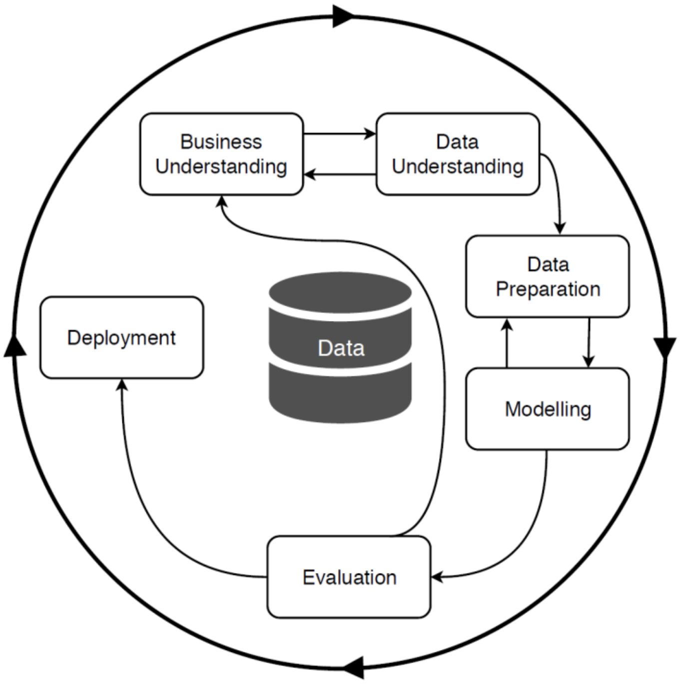


Fig. 1: Cross-industry standard process for data mining CRISP-DM (Martínez-Plumed et al., 2019)

3.1. Business understanding

All datasets obtained in this study are official and legal to use. The first data set is the turbine data. We utilized this data set from a scientific article and reproduced it (Rehman and Khan, 2016). This scientific article was chosen as it contains various types of turbines, in addition to the presence of key variables such as turbine diameter. Also, the wind turbine data in the article were collected from the site of Qassim, Saudi Arabia, which is a relevant region to our target geographic region.

The second data set concerns wind data for 10 locations in Saudi Arabia. We requested the wind data through the online renewable resource, Atlas portal. We signed up on the website, filled out the required information, and indicated our organizational affiliation. The option of downloading data was selected by choosing the free basic package of parameters. Within two days, the order of data was sent via email.

The Renewable Resource Atlas of Saudi Arabia is made available by K. A. CARE. The Atlas provided collected and historical solar and wind resource monitoring data to support power project developers, researchers, industries, academics, and the general public. The data had been collected from October 2013 to November 2016.

The main goal is to contribute to the wind turbine selection domain by mining specific location wind data in Saudi Arabia. The resultant machine learning model would be used to recommend the best location-matching turbine. As mentioned before, our data is collected from two different data sources (Rehman and Khan, 2016). The model would mine wind turbine specification data such as air temperature (C°, barometric pressure (hPa), wind speed (m/s), and air density (kg/); these variables affect the final energy product where wind energy equation is calculated based on these variables (Rehman and Al-Abbadi, 2008).

3.2. Data Understanding

We have two dataset sources. The first dataset source is wind turbines data that includes various types of wind turbines provided by multiple manufacturers. Some of these turbines and their characteristics are outlined in Table 1. Diameter is given and it will be divided into two to get the turbine's radius as it is essential in the energy equation. Also, the cut-in wind speed is determined by the manufacturer to protect the turbine from damages. It is the point at which the turbine starts to generate electricity from turning. Wind turbine efficiency is a useful parameter for comparing wind turbine performance among each other.

Table 1: Sample of turbine data set

Turbine	Diameter (m)	Cut-in Wind Speed (m/s)
Unison U93	93	3
REpower MM92	92	3
AAER A-2000-84	84	3.25
Vensys 62-1200	62	2.5
Vestas V90	90	4
Unison U57	57	3

The site-specific data on wind speed is obtained from the Atlas Renewable Resources website and outlined in Table 2. The request can be submitted via the Atlas Renewable Energy website to obtain data. It is an electronic portal containing a database of geographical data and modern climate information on weather phenomena in the various regions of the Kingdom. This portal can be used as an observatory to identify renewable energy resources and to support feasibility studies to find appropriate investment solutions by developers, researchers, and government institutions such as the Educational University.

The second source of our data is obtained from Atlas. The Atlas site data are the average monthly values for all wind resource data from 10 monitoring stations in 10 separate locations in the Kingdom of Saudi Arabia. These are 10 stations represented in 10 separate Excel files. The number of rows in each city represents the number of days observed. The measurements of wind energy resources are collected from a special wind energy monitoring and measurement device, installed at a height of between 100 and 200 meters. There are certain variables affecting the amount of energy produced which are average wind speed and air density.

Table 2: Site-specific wind speed

Name	Data type	Description
Data	Date	The recorded date reflects a full month during which the other variables were calculated.
Air Temperature	Integer	Represent the average air temperature in one month.
Avg Wind Speed	Integer	Wind speed at different altitudes from the surface.

Air density is an important factor in the equation used to calculate the amount of wind power, which will be calculated using the following where p dry air is: Density of dry air (kg/m3), p=air pressure (Pa), R=Specific gas constant for dry air, 287.05/(kg.k), and T=temperature (°k).

$$P_{dry\ air} = \frac{P}{R \times T}$$

To calculate the power of the wind (P), we use the following equation, where p=Density (kg/m3), A=Swept Area (m2), and v=wind speed (m/s).

$$P/A = 1/2\ pv\text{^}3$$

3.3. Data preparation

Data selection covers the attributes selection (columns) and the record selection (rows). We took all records from each region and selected several columns from the wind dataset as shown in Table 3.

Our wind dataset contained a lot of attributes. We only needed certain inputs to include in the modeling stage based on the aforementioned equation. These attributes include wind speed (m/s) and air density (kg/.). Finding the air density depends on several variables: Specific gas constant for dry air, air temperature, and barometric pressure. We loaded the 10 cities datasets in Jupyter Notebook. Then, we merged all of them into one dataset and selected the relevant attributes. These attributes included the site, air temperature, average wind speed (at 100m/s height), and barometric pressure.

On the other hand, our turbine dataset included the following: turbine type (20row/type), diameter, cut-in wind speed, and radius. There is not any specific selection for the turbine's dataset, and all columns are taken. These attributes are outlined in Table 4.

Table 3: Wind data for the 10 regions in Saudi Arabia

Variable	Unit	Description
Air temperature	Celsius (C°)	Temperature describes the kinetic energy of the gasses that make up the air.
Average wind direction	Meter (m)	The direction from which the wind is blowing. The direction comes in different heights: 37 m, 80 m, and 98 m
Relative humidity	Percent (%)	Percentage of the maximum amount that the air could hold at the specified temperature.
Barometric pressure	Hectopascal (hPa)	Pressure within the Earth's atmosphere.
Average wind speed	Meter per second (m/s)	The essential atmospheric quantity is associated with air moving from high to low pressure.
Average Battery	Volt (V)	Monitoring device's battery used in talking the data.
Logger Temperature	Celsius (C°)	The measuring device records the temperature separately over a defined period.

Table 4: Wind turbines data

Variable	Unit	Description
Diameter	Meter (m)	The distance from one point in a circle to another point in the same circle.
Cut-in Wind Speed	Meters per second (m/s)	This is when the blades begin to rotate and generate power.
Radius	Meter (m)	Half of the Diameter, which represents the blade length.

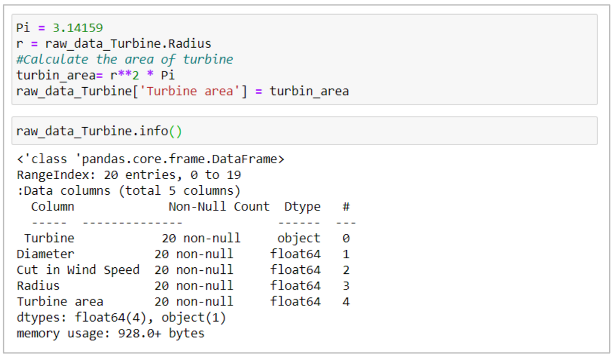
Constructing data includes activities such as derived attributes production, whole new records, or transformed values for existing attributes. We found three derived attributes. First, as shown in Fig. 2, we calculated the turbine area from Radius and π.

Before calculating air density, we had to convert barometric pressure from hectopascal to pascal (given that 1 hectopascal=100 pascals). Also, the air temperature must be converted to kelvin (given that 273+Celsius=Kelvin). These two units are standard in finding air density. Air density is a derived attribute produced from temperature and pressure.

The data sets were combined. For every region, we computed the potential amount of energy that will be produced for every wind turbine type. For instance, for the Unison U93 wind turbine, the following steps were performed to calculate the amount of energy produced for the Al-Wajh Wadi Al Seeh station (all the equation inputs shown in Table 5):

Calculate the area of turbine:
r=93/2=46.5
A=π r2=3.14×46.52=6792.9m2
Convert temperature unit from Celsius to Kalvin:
(273+Celsius=Kalvin)
K=19.2605+273
=292.26 K
Convert pressure unit from hectopascal to pascal:
(1 hectopascal=100 pascal)
=(1007.7087× 100)=100770.87 Ph
Calculate air density:
=100770.87/287.058×292.26
=1.20kg/m3

Apply wind energy equation:
P=½ ρv3 A
P=½ ×1.20×5.41083 × 6792.9=645,640.7 W
(1 Watt=1000 kW)
P=645,640.7÷1000=645.6407kW


```
Pi = 3.14159
r = raw_data_Turbine.Radius
#Calculate the area of turbine
turbin_area= r**2 * Pi
raw_data_Turbine['Turbine area'] = turbin_area

raw_data_Turbine.info()

<'class 'pandas.core.frame.DataFrame>
RangeIndex: 20 entries, 0 to 19
:Data columns (total 5 columns)
  Column              Non-Null Count  Dtype    #
  -----  --------------               ------  ---
  Turbine             20 non-null     object   0
Diameter              20 non-null     float64  1
Cut in Wind Speed     20 non-null     float64  2
Radius                20 non-null     float64  3
Turbine area          20 non-null     float64  4
dtypes: float64(4), object(1)
memory usage: 928.0+ bytes
```
Fig. 2: Calculating turbine area from radius and π

Table 5: Energy equation inputs (Al Wajh Wadi Al Seeh Station)

Turbine	Diameter (m)	Air temperature (C°)	Barometric Pressure (hPa)	Wind speed 100 (m/s)	Wind energy (kWh)
Unison U93	93	19.2605	1007.7087	5.4108	645.6407
UNISON u57	57	19.2605	1007.7087	5.4108	242.5299

3.4. Modeling

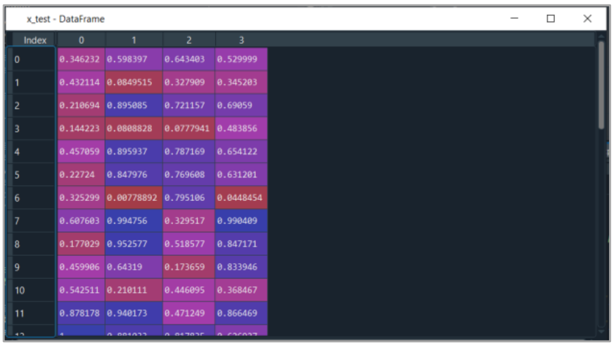
Modeling tasks include selecting the modeling technique, generating the test design, building a model, and assessing its performance. The first step in modeling is choosing the modeling technique that will be used. This research aims to predict the best wind turbine type based on the amount of energy it would produce in a certain station. The multi-regression technique is best for use in modeling this problem since we have continuous variables. In particular, the K-nearest neighbor model has been chosen (KNN) since it is nonparametric, which means it tends to make no assumptions regarding data distribution. We have implemented Min-Max Scaling for this purpose.

In this step, we have implemented the plan for splitting the data into both the training set and the testing set. We have chosen the 80:20 ratios for the data split. The random state was equal to 1 to control the shuffling and identify the K position that would give the optimal result.

Fig. 3 outlines the four required attributes—Avg Wind Speed, Barometric Pressure, Air Temperature, and Air Density—to calculate the amount of energy denoted by 0, 1, 2, and 3 respectively. These are the scaled input data that are used to feed the model. The dependent variable is the amount of energy produced by every wind turbine for every site. The data shown in Fig. 4 shows the amount of energy that can be produced by the Union U57 turbine for all 10 sites.

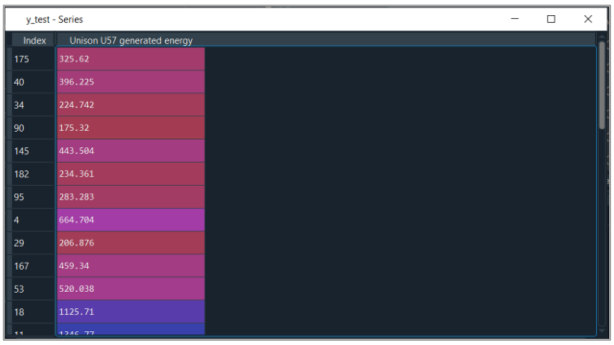
3.4.1. Build the model

This stage required the execution of the code on Python to implement the model which is the KNN Regressor. We also conducted k-folds cross validation and chose to the fold with less RMSE (root mean square error). The number of iterations in the loop to calculate RMSE for each k is shown in Fig. 5. Fig. 5 shows that the optima k-fold value is 14 (RMSE=82.56).


x_test - DataFrame

Index	0	1	2	3
0	0.346232	0.598397	0.643403	0.529999
1	0.432114	0.0849515	0.327909	0.345203
2	0.210694	0.895085	0.721157	0.69059
3	0.144223	0.0808828	0.0777941	0.483856
4	0.457059	0.895937	0.787169	0.654122
5	0.22724	0.847976	0.769608	0.631201
6	0.325299	0.00778892	0.795106	0.0448454
7	0.607603	0.994756	0.329517	0.990409
8	0.177029	0.952577	0.518577	0.847171
9	0.459906	0.64319	0.173659	0.833946
10	0.542511	0.210111	0.446095	0.368467
11	0.878178	0.940173	0.471249	0.866469

Fig. 3: Input data (wind data)


y_test - Series

Index	Unison U57 generated energy
175	325.62
40	396.225
34	224.742
90	175.32
145	443.504
182	234.361
95	283.283
4	664.704
29	206.876
167	459.34
53	520.038
18	1125.71

Fig. 4: Output data (energy produced by wind turbines)

Next, we ran the KNN model on the prepared data set as shown in Fig. 6. The applied K-nearest neighbor regression model has mined the wind data: Avg Wind Speed, Barometric Pressure, Air Temperature, and Air Density for ten locations with 199 as the total row number. K-nearest neighbor aids in predicting the relationship between independent variables and the continuous outcome.

It is useful when the target variable is continued, as well as when it is a direct algorithm that stores all available cases and does the prediction based on feature similarity. These are almost 20 energy amount values for 10 different sites.

```
model = neighbors.KNeighborsRegressor(n_neighbors = 14)

model.fit(x_train, y_train) #fit the model

pred=model.predict(x_test) #make prediction on test set
error = sqrt(mean_squared_error(y_test,pred)) #calculate rmse
```

Fig. 5: K values

```
RMSE value for k=  1 is: 151.04495544558324
RMSE value for k=  2 is: 134.16357626260773
RMSE value for k=  3 is: 108.3150862594823
RMSE value for k=  4 is: 118.12986553931832
RMSE value for k=  5 is: 113.92084631574792
RMSE value for k=  6 is: 128.0200353922823
RMSE value for k=  7 is: 121.96161172654804
RMSE value for k=  8 is: 115.13985670353921
RMSE value for k=  9 is: 113.49042309249778
RMSE value for k=  10 is: 103.1235133384384
RMSE value for k=  11 is: 97.38267258707971
RMSE value for k=  12 is: 92.5842759597977
RMSE value for k=  13 is: 87.70459759646486
RMSE value for k=  14 is: 82.56048381616812
RMSE value for k=  15 is: 83.43609143132326
RMSE value for k=  16 is: 82.66245343927183
RMSE value for k=  17 is: 85.51009066074883
RMSE value for k=  18 is: 87.7719991234552
RMSE value for k=  19 is: 91.42872450701509
RMSE value for k=  20 is: 93.71217131112351
82.56048381616812
```

Fig. 6: KNN model

We identified 80% of data for training, which amounted to 159 of 199, and 20% of data for testing, which amounted to 40 of 199. The y_test value is (40) since the model should predict the energy column for the selected turbine. Table 6 shows the difference between the predicted and the actual value for one turbine.

Table 6: Model prediction

Turbine_name	K-value	RSME	Predicted value	Actual Value
'Unison U57 generated energy'	14	82.56	467.883	312.15

3.5. Evaluation

Intervention studies involving animals or humans, and other studies that require ethical approval, must list the authority that provided approval and the corresponding ethical approval code.

During the evaluation stage, we ensured that the chosen model meets the business success criteria, which means it does in fact recommend the wind turbine type that produces the maximum amount of energy for a specific site. Fig. 7 shows OLS regression results. R-squared denotes the percentage of variance in dependent variables that can be explained by independent variables. Here, 97.6% variation in y is explained by Avg Wind Speed, Barometric Pressure, Air Temperature, and Air Density. As seen in Fig. 8 and Fig. 9, the mean absolute error=64.36 and the root mean square error=82.56, which is the same RMSE for k=14.

```
                              OLS Regression Results
==========================================================================================
Dep. Variable:     Unison U57 generated energy   R-squared (uncentered):             0.976
Model:                                     OLS   Adj. R-squared (uncentered):        0.975
.Method:                         Least Squares    F-statistic:                        1569
Date:                         Fri, 09 Apr 2021   Prob (F-statistic):             3.09e-124
Time:                                 20:24:22   Log-Likelihood:                   -930.65
.No. Observations:                          159    AIC:                               1869
.Df Residuals:                              155    BIC:                               1882
                                                  Df Model:                              4
                                                  Covariance Type:               nonrobust
==========================================================================================
coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
Avg Wind Speed at 100m (m/s)   241.9853      5.828     41.522      0.000     230.473     253.498
Barometric Pressure (Pa)         0.0270      0.005      5.897      0.000       0.018       0.036
Air Temperature (K)             -6.3677      0.547    -11.650      0.000      -7.447      -5.288
Air density                  -1616.3349    296.171     -5.457      0.000   -2201.388   -1031.282
==============================================================================
Omnibus:                       64.645   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              195.425
Skew:                           1.632   Prob(JB):                     3.66e-43
Kurtosis:                       7.341   Cond. No.                     4.15e+06
==============================================================================
```

Fig. 7: OLS regression results

```
In [2]: from sklearn.metrics import mean_absolute_error
   ...: print('MAE : '+ str(mean_absolute_error(predicitions,y)))
   ...: error = sqrt(mean_squared_error(y_test,pred))
   ...: print('RMSE :' , error)
MAE : 64.36827330422359
RMSE : 82.56048381616812
```

Fig. 8: Mean absolute error and root mean square error

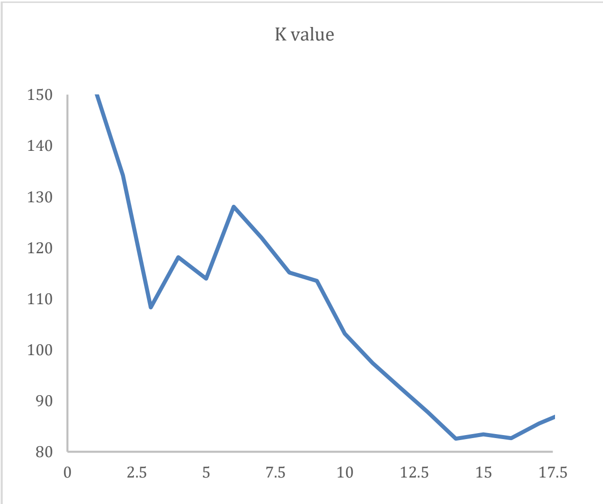


Fig. 9: Error rate visualization

Cross-validation is a method used to test the usefulness of machine learning models; it also is a re-sampling procedure used to evaluate a model if there is a limited amount of data. K-folds validation is one effective approach in the case of limited input data as in our project. We implemented it with 5 folds, as seen in Fig. 10 and Fig. 11.

```
from sklearn.model_selection import KFold
kf = KFold(n_splits=5)
kf

for train_index, test_index in kf.split('T1.csv'):
    print(train_index, test_index)

def get_score(model, x_train, x_test, y_train, y_test):
    model.fit(x_train, y_train)
    return model.score(x_test, y_test)

get_score(model, x_train,x_test, y_train, y_test)
```

Fig. 10: K-folds cross-validation

```
[2 3 4 5] [0 1]
[0 1 3 4 5] [2]
[0 1 2 4 5] [3]
[0 1 2 3 5] [4]
[0 1 2 3 4] [5]

In [18]: get_score(model, x_train,x_test, y_train, y_test)
Out[18]: 0.8868865070390102
```

Fig. 11: KNN model score

3.6. Deployment

One of the objectives of this project was to visualize the location of the recommended wind turbine stations on a map. We have built a satellite map that shows the 10 locations (Fig. 12). Each location's properties can be shown in a drop-down menu when the user hovers over the location, as shown in Fig. 13.

Fig. 12: Wind stations in satellite map

During the evaluation stage, it should be ensured that the chosen model meets the business success criteria, which means that it does in fact recommend the wind turbine type that produces the maximum amount of energy for a specific site. Fig. 8 shows OLS regression results. R-squared denotes the percentage of variance in dependent variables that can be explained by independent variables. Here, 97.6% variation in y is explained by Avg Wind Speed, Barometric Pressure, Air Temperature, and Air Density. As seen in Figs. 9 and 10, the mean absolute error=64.36 and the root mean square error=82.56, which is the same RMSE for k=14.

4. Findings and discussion

This research used a KNN model to optimize the wind turbine selection process by recommending the best location-marching wind turbine for a specific site (wind farm) based on produced energy in the Kingdom of Saudi Arabia. At k=14, our model shows a score of 0.88. This indicates that the model performs very well in predicting the best wind turbine type with a mean accuracy of 88%. Table 7 shows the wind stations that resulted from the optimized model with the suggested turbine type in each station. It can be seen that Unison U93 is the wind turbine recommended to produce the maximum amount of energy in all locations except for "Yanbu North 1". This is due to the fact that the wind speed in this location is higher than the maximum energy that Unison U93 can tolerate based on its cut-in and cut-out values.

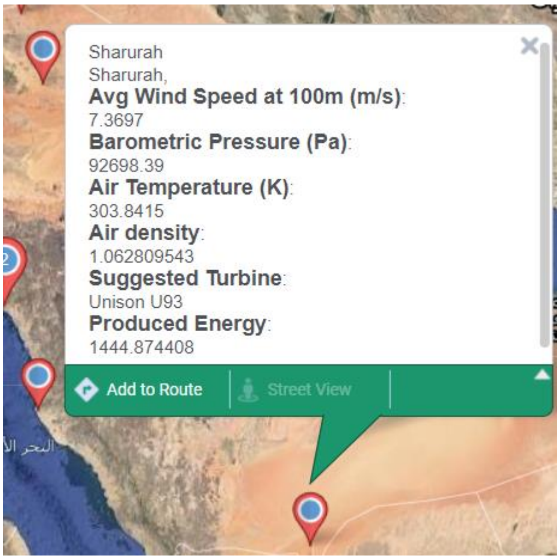


Fig. 13: Station specifications with suggested turbine

Table 7: The suggested wind turbines for the ten wind farm locations

Location	Suggested turbine	Energy generated amount in KW
Yanbu North 1	REpower MM92	5312.486
Yanbu South	Unison U93	809.2213
Turaif	Unison U93	1723.848
Sharurah	Unison U93	1444.874
Riyadh - City Site B	Unison U93	1127.835
Riyadh - City Site A	Unison U93	1362.307
Jeddah South	Unison U93	1355.579
Hafar Al Batin	Unison U93	2441.867
Al Wajh	Unison U93	900.2878
Al Jouf	Unison U93	1253.345

These findings prove that historical data is in fact valuable in the context of wind energy forecasting, and it has been studied previously in these studies (Demolli et al., 2019; Clifton et al., 2013; Brahimi, 2019). The KNN has been also widely used in previous wind power research studies with different focuses. For example, a 2017 research study applied random forests and KNN to complete the wind turbine data sets. The quantitative analysis results show that KNN provided superior satisfactory results, where the absolute deviation was 0.001 (Becker and Thrän, 2017). Another example is the usage of the KNN model to predict wind speed parameters using air temperature, relative humidity and atmospheric pressure, and wind direction parameters. The KNN model performed well and achieved the best prediction results when k=5 (Yesilbudak et al., 2013). As reported by the authors, the obtained NAE, MAPE, and NRMSE were 0.594m/s, 5.695%, and 8.696%, respectively. The

KNN model deployed in a geographic map can be used as an assistive tool for decision-makers to facilitate the process of selecting the best wind turbine based on the turbine specific characteristics and the geographic data about the location itself: air temperature (C°, barometric pressure (hPa), wind speed (m/s), and air density (kg/); for a certain wind farm station. Combining these two sources of data together proves useful for the machine learning model to determine the best wind turbine for a specific wind farm station.

5. Conclusion

One of the contributing factors for a stable economy in Saudi Arabia is renewable energy, including sources such as wind energy. According to Demolli and colleagues, it is possible to generate energy from wind turbines at any hour of the day (Demolli et al., 2019). Therefore, the National Renewable Energy Program (NREP) has been formed as a strategic initiative to increase the Kingdom's usage of renewable energy resources. This includes the wind energy recourse, which is the focus of this research. One of the issues that researchers and decision-makers may face is the efficient selection of the wind turbine types, and this may result in several consequences due to the lack of testing ability of these types before investing takes place (Perkin et al., 2015). This is what motivates this research which aims to achieve two goals. The first is to build a model that would recommend the best suitable wind turbine type. This would include identifying which type would produce more energy based on several relevant variables. Second, it aims to visualize the location of the recommended wind turbine stations in an easy-to-understand map. Using the machine learning technique specific, the KNN model has the potential to assist in the prediction of the most suitable wind turbine types based on the generated energy in one specific location coupled with wind turbine characteristics.

Although the results are promising, our KNN model exhibits some inevitable limitations. The historical data that was obtained from the Renewable Resource Atlas was recorded over the period of October 2013 to November 2016. This is the most up-to-date data available at the current time. During the subsequent period, many changes may have affected the weather. This may make it harder to draw conclusions to act upon in the current time.

Another limitation is the number of contributing factors that can be used to feed the model beyond the air density, wind speed, and barometric pressure. One possible direction for this research is to build on the KNN model to incorporate as many wind attributes as possible and other factors too, as well as build more than one model with different selections of attributes for every model. This would give the chance to compare these different models and select the one with the highest accuracy. Another limitation is that there were some differences in the parameters in the selected turbines, such as Unison turbine U93 and Unison U57. Thus, this must be considered as another future direction, which would assist in having a more accurate result.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Aksoy B and Selbaş R (2021). Estimation of wind turbine energy production value by using machine learning algorithms and development of implementation program. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 43(6): 692-704. https://doi.org/10.1080/15567036.2019.1631410

Amran YA, Amran YM, Alyousef R, and Alabduljabbar H (2020). Renewable and sustainable energy production in Saudi Arabia according to Saudi Vision 2030; Current status and future prospects. Journal of Cleaner Production, 247: 119602. https://doi.org/10.1016/j.jclepro.2019.119602

Assouline D, Mohajeri N, Mauree D, and Scartezzini JL (2019). Machine learning and geographic information systems for large-scale wind energy potential estimation in rural areas. Journal of Physics: Conference Series, IOP Publishing, EPFL Lausanne, Switzerland, 1343: 012036. https://doi.org/10.1088/1742-6596/1343/1/012036

Becker R and Thrän D (2017). Completion of wind turbine data sets for wind integration studies applying random forests and k-nearest neighbors. Applied Energy, 208: 252-262. https://doi.org/10.1016/j.apenergy.2017.10.044

Brahimi T (2019). Using artificial intelligence to predict wind speed for energy application in Saudi Arabia. Energies, 12(24): 4669. https://doi.org/10.3390/en12244669

Clifton A, Kilcher L, Lundquist JK, and Fleming P (2013). Using machine learning to predict wind turbine power output. Environmental Research Letters, 8: 024009. https://doi.org/10.1088/1748-9326/8/2/024009

Demolli H, Dokuz AS, Ecemis A, and Gokcek M (2019). Wind power forecasting based on daily wind speed data using machine learning algorithms. Energy Conversion and Management, 198: 111823. https://doi.org/10.1016/j.enconman.2019.111823

Martínez-Plumed F, Contreras-Ochando L, Ferri C, Orallo JH, Kull M, Lachiche N, and Flach PA (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. IEEE Transactions on Knowledge and Data Engineering, 33(8): 3048-3061. https://doi.org/10.1109/TKDE.2019.2962680

Perkin S, Garrett D, and Jensson P (2015). Optimal wind turbine selection methodology: A case-study for Búrfell, Iceland. Renewable Energy, 75: 165-172. https://doi.org/10.1016/j.renene.2014.09.043

Rehman S and Al-Abbadi NM (2008). Wind shear coefficient, turbulence intensity and wind power potential assessment for Dhulom, Saudi Arabia. Renewable Energy, 33(12): 2653-2660. https://doi.org/10.1016/j.renene.2008.02.012

Rehman S and Khan SA (2016). Fuzzy logic based multi-criteria wind turbine selection strategy-A case study of Qassim, Saudi Arabia. Energies, 9(11): 872. https://doi.org/10.3390/en9110872

Vladislavleva E, Friedrich T, Neumann F, and Wagner M (2013). Predicting the energy output of wind farms based on weather data: Important variables and their correlation. Renewable Energy, 50: 236-243. https://doi.org/10.1016/j.renene.2012.06.036

Yesilbudak M, Sagiroglu S, and Colak I (2013). A new approach to very short term wind speed prediction using k-nearest neighbor classification. Energy Conversion and Management, 69: 77-86. https://doi.org/10.1016/j.enconman.2013.01.033