Applying Machine Learning to Energy Usage

Andrew Latterner
Division of Science and Mathematics
University of Minnesota, Morris
Morris, MN 562367
latte048@morris.umn.edu

Abstract

With the vast increase in our access to data, there is a benefit to being able to analyze, model, and understand this data. The University of Minnesota - Morris (UMM) has two wind turbines on campus that generate 60% of electricity for the campus in order to strive toward energy sustainability. Due to this desire for sustainability, there has been a significant increase in data gathering on wind turbine production and campus energy usage. UMM hopes to use this data to better understand energy usage and production trends.

One way to try to better understand this data is to apply machine learning techniques, such as Neural Networks, decision trees, and Bayesian techniques, in order to create models that can help to explore relationships in the data. Machine learning entails creating models that can learn from past experiences to accurately classify or predict future situations. We have gathered large amounts of data related to energy production from various UMM campus sources. With this data, we plan to create models that are able to predict future wind turbine energy production using past production patterns and weather data. If the models are sufficiently accurate, these models could help the campus to understand energy usage and production trends. For example, members of the campus community could choose to schedule activities that consume large amounts of electricity on days when the wind turbines are producing large amounts of power. This could include small jobs like student laundry, or larger jobs related to campus heating. UMM would also be able to use the relationships found in the models to help better understand why the campus produces or uses large amounts of energy in hopes of improving energy production and conservation strategies.

1 Introduction

With topics such as global warming being discussed constantly, it has become increasingly clear that energy usage and production are extremely pertinent issues. Many companies and academic institutions alike have attempted to improve energy efficiency. There are many approaches to this problem including researching new energy solutions, improving storage of energy, and even optimizing energy usage. One important challenge is the prediction of energy production.

Understanding and being able to predict energy production from various energy sources would lead to a deeper understanding of energy conservation. Certain fields, such as Artificial Intelligence, have attempted to use data-oriented approaches to try to build models and systems that can predict energy production. In the paper, "Short term wind power forecasting using time series neural networks" by Mohammadsaleh Zakerinia and Seyed Farid Ghaderi, the authors explain how they used neural networks, a form of artificial intelligence, to predict wind power generation over hour intervals [6]. They were able to use large amounts of data to build predictive models.

The University of Minnesota - Morris is working to increase green, sustainable energy. This comes in the form of studying and using wind and solar energy. The wind energy at Morris is produced using two wind turbines situated on a close, off-campus location. In an effort to predict energy production from the wind turbines, we have applied Machine Learning techniques to the historical energy production data in order to build a series of predictive models. The goal would be to use data gathered from the wind turbines along with weather data collected from Morris weather stations to create a model that the University of Minnesota - Morris can use to help predict future energy production.

In this paper, we will discuss necessary Machine Learning background in Section 2. In Section 3, we will discuss the specific algorithms and experimental setup necessary for creating the models. We will then discuss the results of building the models in Section 4. Lastly, we will discuss future projects to use and manipulate the models in Section 5.

2 Machine Learning

The goal of our study is to create models that have predictive capabilities. This process entails using large amounts of data to train the model to make predictions regarding future data inputs. In the context of our research project, we would supply training data, or data that contains weather information along with associated energy production. These weather variables include values such as wind speed and direction and the desired output, energy production. Using this data, the model can be built by using a learning algorithm.

A learning algorithm takes in a set of training data, such as past weather and production data, and outputs a model. This model can take in new predictor variables (e.g. weather data for tomorrow) and predict an output for those values (e.g. energy production for tomorrow). This entails having the learning algorithm use the past data to create rules or patterns which can be used to analyze the data and hopefully generate an accurate output. In the energy production data example, the learning algorithm might notice that wind speeds above 20 miles per hour relate to extremely high energy production. The learning algorithm then uses these rules to create a model. This model uses the rules acquired during the learning process to take in new input data and generate a predicted output. In the production example, using the rules it learned in the training phase, the model might see new data showing that wind speed is suspected to be around 25 miles per hour. As such, the model would predict that the energy production would be high for that instance.

There are a wide variety of different learning algorithms, each with their own properties and settings. A key feature is the type of outputs an algorithm's model is capable of producing. Some can generate a numeric output, e.g., providing a direct prediction of the expected Kilowatts (Kw) given the predicted weather. Other algorithms, however, are classifiers, which instead output one of a finite set of categories (often called factors), such as "low energy production", "medium energy production", or "high energy production".

We used three different learning algorithms during our model building phase – Naïve Bayes, Decision Trees, and Neural Networks. Each of these algorithms decides on rules in very different ways but each produces a model that can predict our desired output – energy production. We used each of these learning algorithms to produce a series of models with the goal being to compare each algorithm's results and accuracy for our problem as well as to decide on which algorithm fits the needs of the project in the future. We built these models using WEKA, a software suite developed to simplify the process of building models using machine learning techniques [1].

2.1 Naïve Bayes

Naïve Bayes [5] is a learning algorithm that relies on the use of Bayes' theorem to weight each predictor variable's effectiveness in predicting the output. However, it should be noted that Naïve Bayes is a classifier rather than a numeric predictor.

In the context of classification, Bayes' theorem estimates the probability of a certain classification being the correct classification given the features' or predictor variables' values. Bayes' theorem for single predictor variables states that:

$$P(C|F) = \frac{P(C)P(F|C)}{P(F)} \tag{1}$$

In the above equation, $P(C)$ is the probability of an output being classified as a specific case C, independent of the predictor variables. $P(F)$ is the probability of a variable having a specific value F. $P(C|F)$ is the probability of an output being a specific classification C given the state of the variables F.

However, we will not only be working with one input variable. In order to take into account multiple input variables V_i, we must modify Bayes' theorem to the following:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \tag{2}$$

In the above equation, $P(F_1 \dots F_n)$ represents the probability of $V_i = F_i$ for all i. Lastly, we need to introduce an independence assumption. This assumes that every input variable

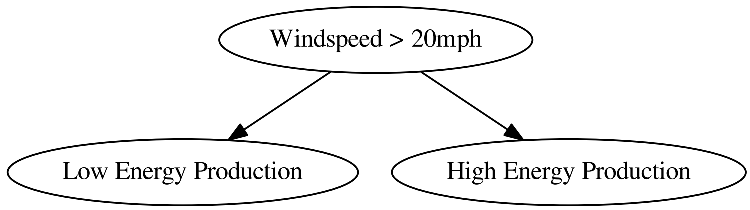

Figure 1: An example tree with a single parent node and two terminal nodes

F_i's chance of occurring is independent of every other variable F_j given a certain classification. Using the independence assumption, we can reduce this formula to the following:

$$P(C|F_1 \dots F_n) = KP(C)\prod_{i=1}^{n} P(F_i|C) \tag{3}$$

In the above equation, we have $K = 1/P(F_1 \dots F_n)$. This equation allows us to use multiple input variables to intelligently calculate the probability of those input variables producing a certain classification.

Naïve Bayes uses the modified Bayes' theorem stated above in order to create learned models for classifying instances of input variables. For each set of input variable samples $F_1 \dots F_n$, the chosen classification or output is the classification factor C that produces the highest probability $P(C|F_1 \dots F_n)$ in the above equation.

2.2 Decision Trees

Decision trees are a class of algorithms that create tree-like structures which can be used to classify new inputs. Like the Naïve Bayes algorithm, decision trees provide a classification as an output. The specific algorithm that was used in our experiment was called the C4.5 Tree learning algorithm [4]. In this algorithm, tree-like structures are created with two different types of nodes. The first kind is a parent node which has a variable on which to split input. The second kind is a terminal node which contains a classification.

In Figure 1, we see an example tree which will help explain the structure of decision trees. The parent node, "Windspeed $>$ 20mph" checks if the wind speed of the input is greater than 20 miles per hour. Connected to this parent node are two terminal nodes. The first terminal node, "Low Energy Production" is chosen if the wind speed is less than or equal to 20 mph. The second terminal node, "High Energy Production" is chosen if the wind speed is greater than 20 mph. Once a terminal node is reached, the algorithm classifies the input as the value of the terminal node.

The C4.5 algorithm creates a tree by splitting based on entropy, or how much information can be gained from a set of data. The algorithm will look through each predictor variable

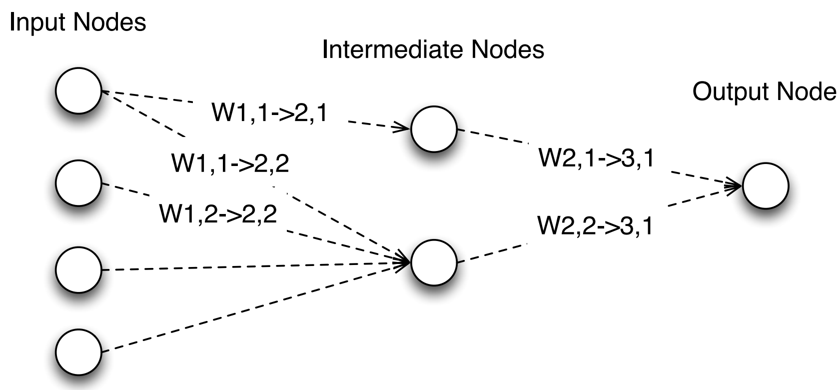

Figure 2: An example neural network with three layers

and calculate the change in entropy that it would achieve by splitting on that variable. When it finds a variable that has the highest change in entropy, it will split on that variable and continue the process on each side of the split. It will also look for base cases that might be fulfilled when deciding to split. One important base case that it will look for is when all samples or an overwhelmingly high proportion of samples are of the same classification; here it will create a terminal node with that classification instead of splitting.

2.3 Neural Networks

Artificial Neural Networks are inspired by the nervous systems found in animals and, as such, have a connected neuron-like structure. This entails having neurons as individual nodes with weighted connections [2].

In Figure 2, we see an example of a neural network. The first layer is the input layer; each node in this layer represents a different input. The last layer is called the output layer as these nodes will be the result of the model computation. Each intermediate and output node's values are calculated as follows:

$$x = \sum_{i=1}^{n} w_i y_i \tag{4}$$

where x is the current node's output value, y_i is a previous node's value, and w_i is the weight on the connection between a previous node and the current node.

In order to change the overall neural network to be more accurate, the weights are manipulated to minimize the error of the final output node layer. For instance, if the final node output is too high, the weights might be redistributed or edited so that the output node more closely reflects the actual expected output.

The specific Neural Network implementation used in our study was the Multilayer Perceptron which uses back propagation to learn the appropriate weights [3]. We used this algorithm due to the fact that it is one of the default neural network implementations on WEKA and that the output can be a real value rather than a classification. As such, the output of a model created using the Multilayer Perceptron can reflect a numeric energy production output rather than a classification like the Bayesian and Decision Tree algorithms.

3 Experimental Setup

In order to create the energy production models to be evaluated, decisions regarding data sources, program setup, and learning algorithms needed to be made. This includes finding reliable data sources for weather and production data, parsing that information into a usable, meaningful format, and deciding between a set of learning algorithms to create an optimal model.

The weather data source that we used was decided based on accessibility and reliability. The data was gathered from the University of Minnesota - Morris campus weather station. Weather conditions are recorded and stored reliably and are available for free online. From this source, we used average wind speed (mph), high wind speed (mph), wind chill (F), wind direction (cardinal), and temperature (F). However, there are nontrivial differences in the environmental conditions between the wind turbine and the campus weather station. We are using this data source but, when analyzing the resulting models, this issue should be kept in mind.

Regarding the production data, the data source was Otter Tail Power Company, the power company that handles the Morris campus energy needs. The data originating from the power company's records were both reliable and accurate. This data source gave us hourly production in kilowatt-hour (kWH).

After acquiring the data, a major processing task was to parse it into a usable format. This entailed converting the weather and production data sources into a set of comma-separated value (CSV) file that matched each other in terms of dates and times. In order to maintain consistency in dates, the year 2013 was used for both the weather and the production data as that year was the most recent and representative year. However, given that the weather data source and the production data source were gathering data on different intervals (15 minute intervals versus hour intervals), we parsed and averaged all gathered data for that hour. This entailed calculating average wind speed, wind chill, temperatures, and wind direction over an hour interval. Given that the wind direction variable was recorded on a cardinal direction scale, the most common wind direction over an hour interval was used as the average wind direction. The other variables were simple averages. Given that our model building software suite, WEKA, allowed for CSV files to be used, we maintained a CSV format for the data files during the entirety of the parsing process.

Another task that was necessary when parsing the data set was the factorization of the production data set for classification purposes. This required us to find ranges in the production data set and assign them to groupings. We decided to label these groupings as ”Very High”, ”High”, ”Medium”, and ”Low” production. The ranges for each factor are described in Table 1.

Table 1: Ranges on Production Factors

Data Set	Low Production	Medium Production	High Production	Very High Production
Power Company (kWH)	0-150.0	150.0-1123.5	1123.5-1882.7	1882.7-3263.4

For creating the models, we chose three learning algorithms, Naïve Bayes, decision trees, and neural networks, for many reasons. We chose Naïve Bayes due to success with the algorithm in past model-building initiatives. Decision trees were chosen due to their relatively high accuracy and intuitive output with useful model result visualizations. This algorithm would help us to build intuition as to the important features in our model building process. Lastly, we chose neural networks because they allowed us to predict power generation on a numeric scale rather than performing classification.

4 Results

In order to understand the accuracy of the model's ability to correctly predict output, one needs to understand several accuracy metrics. These metrics include accuracy, correlation coefficients, and confusion matrices. The accuracy is the percentage of all outputs that were correctly classified. Accuracy is difficult to use alone as it does not take into account classifications that were close to the expected classification. For example, if the model outputs "high energy production" and the expected output was "very high energy production", we'd consider it more accurate than if the expected output was "very low energy production." As such, we use a confusion matrix, or a matrix that counts occurrences of agreement and errors between expected outputs and actual outputs. This matrix will help us to analyze how inaccurate the model was. Correlation measures the relationship between two variables, and is used for models that output a real number rather than a classification. In the context of its use as a metric, it refers to the relationship between the expected output and the real output.

4.1 Naïve Bayes

The first learning algorithm that we used was the Naïve Bayes algorithm. As previously stated, Naïve Bayes classifies output rather than providing a real numeric output. As such, we will be evaluating the Naïve Bayes model using the accuracy metric and a confusion matrix.

This model had a somewhat low accuracy of 53.4%, implying that the model correctly classified the sample data about half of the time. However, given that there were four classification choices, it still did better than random chance. As such, it did learn some patterns for predicting power output. Likewise, as seen in the confusion matrix in Table 2, most misclassifications occurred close to the correct classification rather than far away. So while these misclassifications were incorrect, they were close to being correct.

a	b	c	d	<-Classified as
1486	121	398	68	a=Very High Production
500	129	658	179	b=High Production
266	145	1266	1003	c=Medium Production
13	12	499	1550	d=Low Production

Table 2: Confusion Matrix for Naïve Bayes

4.2 Decision Trees

The second learning algorithm that was used to create our models was the C4.5 Decision Tree algorithm. The C4.5 algorithm classifies output, like the Naïve Bayes algorithm. We will evaluate the C4.5 models using accuracy and a confusion matrix. Likewise, we can analyze what the C4.5 model found to be the most important variables to split on.

When analyzing the model, we had similar results to that of the Naïve Bayes model, with an accuracy of 54.5%. When looking at the confusion matrix, it is clear that there were issues classifying instances of high production as medium production. However, nearly every other case was handled reasonably well with misclassifications occurring near the correct classification.

a	b	c	d	<-Classified as
1563	206	252	52	a=Very High Production
524	312	501	129	b=High Production
299	361	1326	694	c=Medium Production
56	88	612	1318	d=Low Production

Table 3: Confusion Matrix for the Decision Tree

When analyzing the decision tree that is created, we can examine the variables the decision tree split on in the initial splits. This allows us to understand which variables, when split, produced the highest ability for pattern creation. We can look at the first three splits from the decision tree in Figure 3. It seems that high wind speed and average wind speed had the highest influence on the classifications of the model. The initial split was on High Wind Speed being greater than 13.75 mph. Even in the second and third splits, we only see Average Wind Speed and High Wind Speed being split on. This makes sense given that we are predicting wind energy output.

4.3 Neural Networks

When analyzing the results of the neural network models, we cannot use the accuracy metric and the confusion matrices. Instead, we are required to look at error and correlation

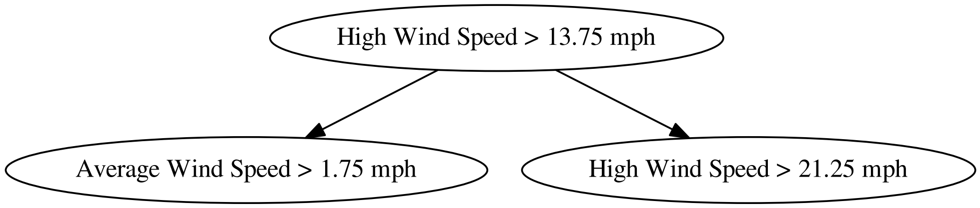

Figure 3: The first three splits of the decision tree model

coefficients. This is due to the fact that neural networks produce a numeric result rather than a classification.

The neural network model resulted in a correlation coefficient of 0.65, which implies that the correlation is positive with a moderate strength. This means that the predicted output and the expected output were moderately similar. However, the mean absolute error is 658.3. This implies that, for each prediction from the training data, the output was, on average, 658.3 kWH away from the expected output. Given that production data ranged from 0 to 3263.4 kWH, this error is fairly large and, as such, the model is not as accurate as we'd like.

4.4 Discussion

Given the similar accuracy between models, the process for deciding the best model is difficult. The error on each of the models is relatively high and, as such, we cannot make the decision based on performance alone. In fact, due to the fact that the model building project is in its infancy, we feel as though choosing one model and learning algorithm above the others would be a poor idea. Likewise, each of the models are able to provide us useful information. For example, decision trees allow us to visualize the data in a meaningful manner. This allows us to decide which variables are useful in predicting the power production. The neural networks provide a numeric output which is useful for analyzing the accuracy of the model and how we might change it in the future. As such, we have decided to keep all of the models in the pool of potential candidates as none of the models have proven to be reliable enough to use for decision-making purposes.

Lastly, we need to understand why the accuracy of the models are so low. On one hand, we noticed that the confusion matrices implied that there were many times when the model would choose a classification close to the expected output. This implies that accuracy maybe isn't the best metric for this project as accuracy does not take into account near-misses. This would require further analysis in the future.

The other source of inaccuracies might be related to the unpredictability of the production given our input variables. For example, in the decision tree, we noticed that average wind

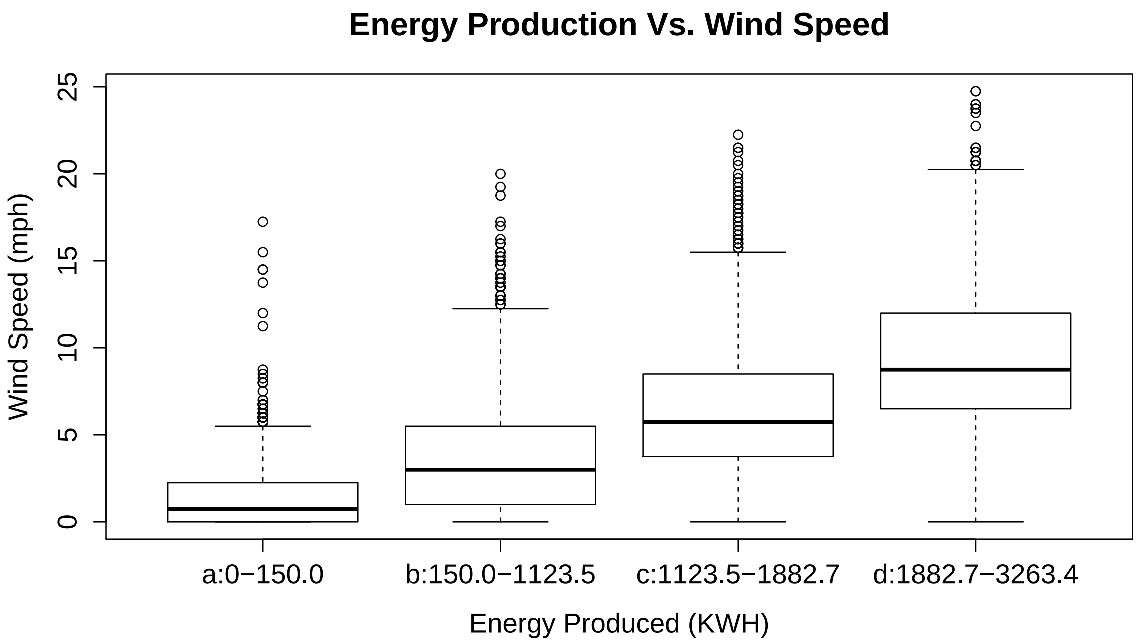

Figure 4: Analyzing Production and Wind Speed

speed and high wind speed tended to be the most common initial splits. We decided to analyze average wind speed to decide if splitting on it is very helpful. In Figure 4, we see the wind speed being plotted against the factors of the production. In this, we see that there is a large amount of overlap between the wind speeds of each factor. While there is usually an upward trend in the medians of wind speed as the production rises, the wind speed variable itself seems to be a poor predictor for production when considered alone. When considering that the models tend to value wind speed highly, it is understandable that the accuracies are relatively low. It will be important to potentially reconsider the bounds for each production factor or add more input variables in hopes of finding a more predictive mixture of variables.

5 Future Work

We have a number of ideas regarding the future of the project which could be explored further in an attempt to build more robust, reliable models. These ideas include enhancing the models for the energy production predictability as well as performing similar prediction analysis on campus energy consumption.

In regards to production predictability, there are many goals for future work relating to enhancing the data gathered from our production source as well as ensuring that the models that were built during the research can be used effectively. Ensuring that these goals are pursued is paramount to building effective models that would provide benefit to the

University of Minnesota - Morris and energy researchers alike.

The first goal related to production predictability entails finding a reliable source of weather data. As previously noted, the environmental setting of the weather station used for our weather data differs from the setting of the wind turbines. As such, we would like to find a weather data source that is as reliable as our current source, but also shares environmental conditions similar to that of the wind turbines. One option is to purchase and deploy sensors located near the wind turbines. This has the benefit of having wind-related data that can mimic the weather conditions experienced by the wind turbines. However, with this data, we still have the problem of not being able to have reliable predictions of future weather patterns as we would only be able to gather current data. In order to solve this problem, we can find a data source such as Weather Underground or the National Weather Service that has the ability to predict future weather conditions. However, weather prediction is not sufficiently localized to address the specific conditions for the turbines. This would require further research into how these weather services gather weather data.

A second goal related to production predictability relates to enhancing the data gathered from the production data source. There are times during the year when the University of Minnesota - Morris shuts down the wind turbines for maintenance and research purposes. As such, it would make the data potentially more predictable as occurrences of high wind but little-to-no production would be explainable. This would entail finding dates when the turbines are shut down and either removing or adding additional attributes to the data set in order to take into account these exceptions.

The second set of goals for future work relate to the campus consumption of energy. The initial goal of the research project was to predict both energy production and consumption, however given time constraints, we were unable to explore the energy consumption prediction problem. As such, a major goal for future work would be to find a way to model energy consumption using a similar process. This would entail gathering campus-wide energy usage data and attempting to predict the usage based on weather data, data related to time (e.g. days of week, whether or not class is in session), as well as the time of the year. This would hopefully allow for an accurate model to be built in order to predict future energy usage. The data required for this endeavor would not be difficult to obtain as most of the necessary data is either shared with the production data or is available using a mixture of a yearly calendar and the campus academic calendar.

When combining the production model and the consumption model, there are many interesting relationships and implications to explore. One such relationship would be how the high-production days relate to campus energy usage. This would allow us to explore how weather patterns change both production and usage on campus. Likewise, it would allow the University of Minnesota - Morris to use these models to influence decisions regarding activities and student life. For example, the campus could schedule large events on days when energy production is very high. Likewise, the campus could encourage students to perform tasks that require high amounts of energy, such as laundry, on days when production is expected to be high. This would allow the University of Minnesota - Morris access to another tool to advance a dynamic energy policy in order to further the goal of sustainability and efficient energy usage.

6 Conclusions

When using machine learning techniques to build our models, we've found that these models may not be accurate enough to be used for predictive purposes. Some reasons for this might be that we need to gather more predictive input variables, reconsider our factor boundaries for our classifications, or find a better source for weather-related data. These changes will hopefully lead to more accurate, reliable models.

Although our models are not as accurate as we'd like, the groundwork for developing this project further in the future is in place. We are confident that, with time and patience, we will be able to develop meaningful models that the University of Minnesota - Morris can use to further optimize energy usage and production on campus. Likewise, if we are able to achieve a reasonable set of models, we will be able to better understand wind energy production trends.

References

[1] HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., AND WITTEN, I. H. The weka data mining software: An update. *SIGKDD Explorations 11* (2009).

[2] WIKIPEDIA. Artificial neural network — wikipedia, the free encyclopedia, 2014. [Online; accessed 29-March-2014].

[3] WIKIPEDIA. Backpropagation — wikipedia, the free encyclopedia, 2014. [Online; accessed 30-March-2014].

[4] WIKIPEDIA. C4.5 algorithm — wikipedia, the free encyclopedia, 2014. [Online; accessed 29-March-2014].

[5] WIKIPEDIA. Naive bayes classifier — wikipedia, the free encyclopedia, 2014. [Online; accessed 29-March-2014].

[6] ZAKERINIA, M., AND GHADERI, S. F. Short term wind power forecasting using time series neural networks. In *Proceedings of the 2011 Emerging M&S Applications in Industry and Academia Symposium* (San Diego, CA, USA, 2011), EAIA '11, Society for Computer Simulation International, pp. 17–22.