UTILIZING DECISION FOREST REGRESSION MACHINE LEARNING ALGORITHM TO PREDICT FILLING LINE UTILIZATION FOR OPTIMAL MAINTENANCE SCHEDULING

L. N. Smith[1] & C.M. Zastron[2] & P.J. Vlok[3*]

[1]Department of Industrial and Systems Engineering
University of Tennessee, USA
lsmit194@vols.utk.edu

[2]Pragma South Africa (Pty) Ltd
mauritz.zastron@pragmaworld.net

[3]Department of Industrial Engineering
University of Stellenbosch
pjvlok@sun.ac.za

ABSTRACT

Small margins within the packaging industry mean financial success in this field relies on high equipment availability. To achieve this high equipment availability, maintenance schedules should be carefully planned to minimize downtime. A key component of maintenance schedule planning is predicting equipment utilization. This can prove very difficult as there are many variables such as market demand, seasonality of products, capability and diversity of equipment, and inherent reliability, to name a few. Even some of the leading players in the packaging industry treat the complexities and chaos involved with predicting equipment utilization as a topic best avoided. Current approaches to this problem range from no prediction at all to only a simple linear extrapolation.

This paper investigates the merits of using machine learning algorithms to predict equipment utilization in the packaging industry with the aim of optimizing maintenance schedules. Machine learning entails pattern recognition of past data and inclusion of pertinent variables in the present to forecast behaviour. This paper begins with a brief literature review of the field before using data, obtained from a multinational packaging company, to test some of the most promising methods of machine learning in a case study.

[1] The author was enrolled for an B Eng (Industrial) degree in the Department of Industrial and Systems Engineering, University of Tennessee
[2] Business Intelligence Engineer
[3] Extraordinary Professor, Department of Industrial Engineering, University of Stellenbosch
*Corresponding author

1. INTRODUCTION

1.1 1.1 Problem Statement

Successful packaging companies depend on high availability of their equipment for financial success. Breakdowns are detrimental to availability and maintenance must be performed regularly, in a minimally disruptive manner, to achieve high equipment availability. To ensure equipment availability, maintenance plans are created for machines in an attempt to replace parts before they break. These maintenance plans normally contain three types of tasks, namely: condition based, time/calendar based, and usage based. Condition based tasks usually have certain limits in which the condition of the part must be. If the part's condition falls outside the specified limits, the part is replaced. Common features which are measured include vibration, heat, wear and tear, as well as noise. Time or calendar based tasks are used when parts need to be cleaned or replaced periodically, regardless of their use or the time on the asset. Finally, usage based tasks are used to replace parts which should be replaced after it has been used for a specified number of cycles or running hours. Usage based scheduling will be the only maintenance type investigated in this paper. The time of work order (collection of tasks) execution needs to be predicted before the tasks are due so that the planning department can ensure all resources are available at the time of the service. Currently, the date at which a usage based service will be done is forecasted by looking at the asset's average usage over a number of past readings and then using that average usage per day when making a future prediction. The number of readings used to calculate the average usage can either be a predefined number or all of those that fall within a fixed time period. The current method employed by the packaging company uses a predefined number of meter readings to calculate the average daily usage. Machine learning, with its unique ability to adapt to different problems and harvest information from large data sets, has the potential to create more accurate usage predictions than the current method. While other solutions do exist, this paper will focus solely on the merits of machine learning when applied to machine usage prediction

With the creation of such software as Microsoft's Azure, IBM's Watson, and TensorFlow, machine learning integration has never been easier or more accessible. As computational resources are constantly becoming more readily available, numerous applications are still being discovered. Machine learning has been utilized by power plants to predict electricity usage and detect malicious energy usage [1]. It has been used to diagnose and classify cancers, with higher accuracy and precision than doctors [2]. Machine learning can also be used to optimize manufacturing processes while increasing product quality and decreasing process testing costs [3]. It has been used to create better marketing and pricing strategies within the steel industry by predicting raw steel prices [4]. Machine learning has even been used to predict stream flows, providing important information for hydrological studies [5]. The success machine learning has had in these and other applications warrants an investigation into how it can improve machine usage predictions.

This paper begins investigating the merits of predicting equipment usage with machine learning with the aim of optimizing maintenance schedules by conducting a literature review. Since the review revealed no sources that considered this use of machine learning, similar applications were researched. From this research, a generic method was constructed to use machine learning to predict equipment usage. This method is then tested in a real world case study to test the performance of machine learning and investigate if and by how much it can improve on the current prediction method.

1.2 Machine Learning: An Overview

Machine learning can be divided into two categories, namely: unsupervised and supervised. When given unlabeled data points, machine learning can find an underlying pattern. This is known as unsupervised machine learning. When given labeled data points, machine learning can detect anomalies, make classifications, or predict numerical values. These are types of supervised machine learning and are called anomaly detection, classification, and regression respectively. To create usage predictions, numerical values need to be predicted, so supervised regression machine learning will be used in this study.

Supervised regression begins with the collection and preparation of data. During the preparation phase, missing and incorrect values in the dataset are identified and corrected. This data is then split into a training set and a testing set. The training data is used to train the chosen algorithm. Every algorithm creates predictions differently and uses the training data to create, optimize, and validate these prediction decisions. Features can be selected before the algorithm is trained, known as the fitler method, or while training the algorithm, known as the wrapper method. Once the algorithm has been trained with the chosen features, the algorithm then makes prediction on the testing data set. The accuracy of predictions is calculated by a prediction error, or the difference between predicted and actual outcome. The lower the error, the more accurate the algorithm is in making predictions. The steps outlined above are summarized in the Figure 1, seen below.

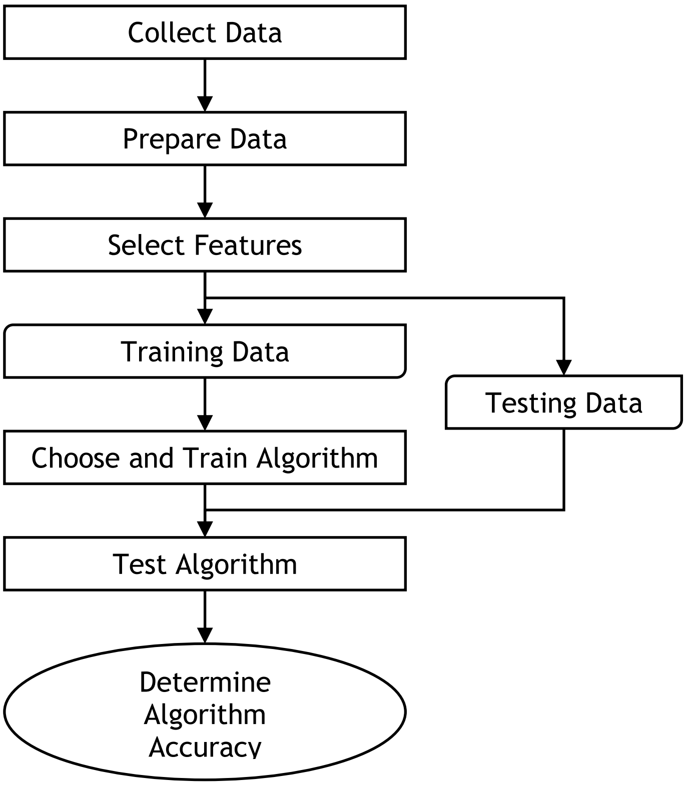

Figure 1: Overview of Machine Learning

1.3 Literature Review

There is currently very little research on using meter readings for predicting the future use of packaging machines using machine learning. However, the application of machine learning to improve prediction accuracies is prevalent in the literature. The most similar application found was the use of machine learning to predict electricity usage. These sources provided the framework for the literature review with other sources supplementing the specific considerations of machine learning.

After acquiring electricity usage data, one source identified incorrect values as those above the average of a continuous period of time [1]. The incorrect values were then replaced with that average [1]. Missing values were filled in with interpolation by one study [1] and linear extrapolation in another [6]. After preparing data, one study considered the distribution of the data with Q-Q plots and the Shapiro Wilks test [1]. This study considered using the Kolmogorov-Smirnov test as well, but did not because this test can only be used if the cumulative density function is known [7].

To divide the data into a training and testing set, the past 500 hours of data was used as the training data to predict the next 48 hours of testing data [1]. In another instance, the training set was composed of four years' worth of data and the testing set was composed of the most recent year of data [6]. Data sets can also be split using random splits [8] and K-folds [9]. If the data set is extremely large, learning curves can be used to limit the training data. As data set size increases, algorithm accuracy increases greatly at first. Eventually, algorithm accuracy sees marginal improvements as data set size continues to increase [10].

Feature selection for predicting electricity usage was done using the relieff algorithm [1] and the statistical properties of the features [6]. Both of these are examples of filter algorithms which are better suited for large datasets [11]. A wrapper method of feature selection is more time intensive, but provides information about the interaction of features [12].

The two main types of wrapper selection algorithms are forward selection and backward elimination [13]. Although these algorithms work quickly, they do not always return the most helpful subset of features [13]. The accuracy of these algorithms can be improved by using a combination of the two algorithms [12].

Boosted decision tree [6], support vector machine [1], and neural network [1] algorithms have all been used to predict electricity usage. The efficiency at which the decision forest regression method operates also makes a popular choice of algorithm for large data sets [14].

Although the literature review did not reveal any publications in which machine learning was used for the same objective as this paper, enough elements of the larger machine learning process were found to construct a methodology in which the problem under review can be addressed. In the next section a proposed solution is constructed with which the paper objectives can be achieved.

2. PROPOSED SOLUTION

The findings from the literature review were used to create a proposed solution for more accurate machine usage predictions with machine learning. Data will be prepared by filling missing values with linear extrapolation [6]. Incorrect values rarely occur as meters are maintained regularly and data is validated upon capturing to draw attention to possible mistakes. The most recent year of data will be withheld for testing and the rest will be used for training [6]. Learning curves based on data set size will be used to further reduce the testing set size and shorten training time [10] if necessary. Features will be selected with a combined forward selection and backward elimination wrapper algorithm [13]. The boosted decision forest regression algorithm will be used to create predictions because of its short training time, high accuracy, and insensitivity to parameters [14]. Success will be measured by reduction of total prediction error. This error can be further classified into over maintenance and under maintenance. When a maintenance task is executed before it is required, it is called over maintenance. This inaccuracy results in increased replacement part costs and increased labor costs. A maintenance task occuring after the scheduled meter reading is known as under maintenance. This results in a higher probability of a breakdown occurring. Breakdowns can be very costly as they result in lost production time, unavailability of artisans, costly expedition of replacement parts, and in some cases loss of the product in the machine at the time of the breakdown. The proposed solution is outlined in the following figure (figure 2)

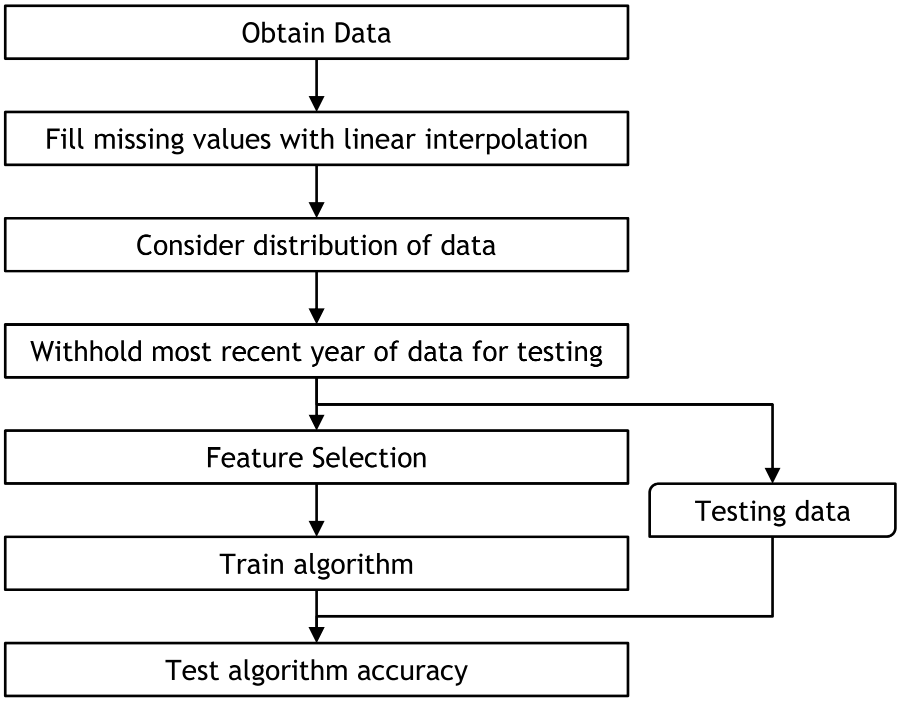

Figure 2: Flowchart of Proposed Solution

3. CASE STUDY

An algorithm can be trained to predict meter readings in a variety of ways. This paper will consider three different methods. First, a reading will be predicted for a certain future date. Secondly, the amount produced between the current date and future date will be predicted, then added to the current reading to construct the future reading. Lastly, daily rates will be predicted for every day between the current and future date, summed, and then added to the current reading to predict the future reading. These rates are referred to as Average Daily Rates (ADRs) within the industry.

3.1 Case Study Data

Data for this case study comes from a multinational packaging company with over 50,000 assets. These assets are fitted with meters which record the total hours that a machine has been operational. When a reading is taken, the date, unique meter code, and total hours of operation are recorded. Additional information such as meter location and brief description of product is obtained from the asset attached to the meter. The time span between readings varies by meter location and date. More recent readings have a shorter time span than older readings. If an asset is refurbished, the meter is often adjusted (reset) to reflect the age of the asset after the refurbishment. For the case study, 500 meters without meter resets were selected. These meters were evenly divided among the five cluster locations to ensure an adequate distribution.

Also used in this study was the most recent year of scheduled work order data. Work order data includes a work order code, meter code, date of work order creation, and date the work order was scheduled for (due date). The work orders are scheduled for certain hour readings and the due date of the work order is calculated by extrapolating the last x number of meter readings to predict when the meter will reach the reading of the work order.

3.2 Reading Predictions

To create a training data set for reading predictions, the span between between work order creation and the work order due date was calculated for each work order. These values were then averaged for each meter. The training data was constructed to reflect this average time span. For every reading and date, a previous reading and date were selected such that the difference between date and previous date closely matched the average span found from the work order data. This created two additional features of previous reading and time between. When a value was missing the preceding reading and date were used.

All features of the reading data were included as well as the days between and previous reading features. Combined forward selection and backward elimination was used to select the most helpful subset of features. They were determined to be days between and previous reading. This created a relatively small training set, so all data except for the most recent year was used for training.

The decision forest regression algorithm was then trained with this data. Parameters were selected using a partial grid sweep. This algorithm proved to be much less accurate than the current prediction method. One algorithm was then trained for each cluster to reflect locational differences. These cluster specific algorithms created more accurate results than the single algorithm, but were still not more accurate than the current prediction method.

3.3 Amount Predictions

The same data preparation technique used to construct the predict readings data was again used to construct the predict amounts training data. Instead of of using the previous reading feature, an amount feature was used. This amount feature was constructed by subtracting the previous reading from the reading. Using forward selection and backward elimination, days between and cluster were selected as the best subset of features. With only two features, the data set size was quite small, so all data after the most recent year was used for training.

After training one decision forest regression algorithm, no improvement in prediction accuracy was seen against the current method. Therefore, 5 algorithms were trained, one per cluster. Parameters were selected with a random grid sweep. Using these algorithms, the total absolute error of one out of the five clusters was reduced by 26 610 units. All other clusters saw an increase in total absolute error. As the total absolute error was not decreased for these algorithms, maintenance costs would not be reduced either.

3.4 ADR Predictions

The final prediction method in this case study was predicting average daily usage. Data was prepared by finding the ADR between each reading. As readings are not taken every day, missing values were filled with linear interpolation.

In addition to the eight features included with a reading, five additional features were mined from the data. To account for seasonality, the month of the reading was added. Additionally, ADRs were averaged based on day of the year, for each of the four location features. Using combined forward selection and backward elimination; site code, meter, reading on, customer plant description, description, and month were selected as the most helpful features.

The training data set size for this prediction type was quite large, so a learning curve was constructed for data set size. Using data from 100 meters, percentages of the meters were used to predict the most recent year of all meters. Ten meters, representing 10% of the data, were randomly selected and used to train the ML alogrithm. The resulting trained algorithm was then used to predict ADR for all 100 meters and the mean error was recorded. This process of random selection was then repeated five times for each 10% increment. The five mean error values for each increment were then averaged and plotted in Figure 3. From the graph, it was determined that a subset of 50% of the meters would provide sufficient accuracy.

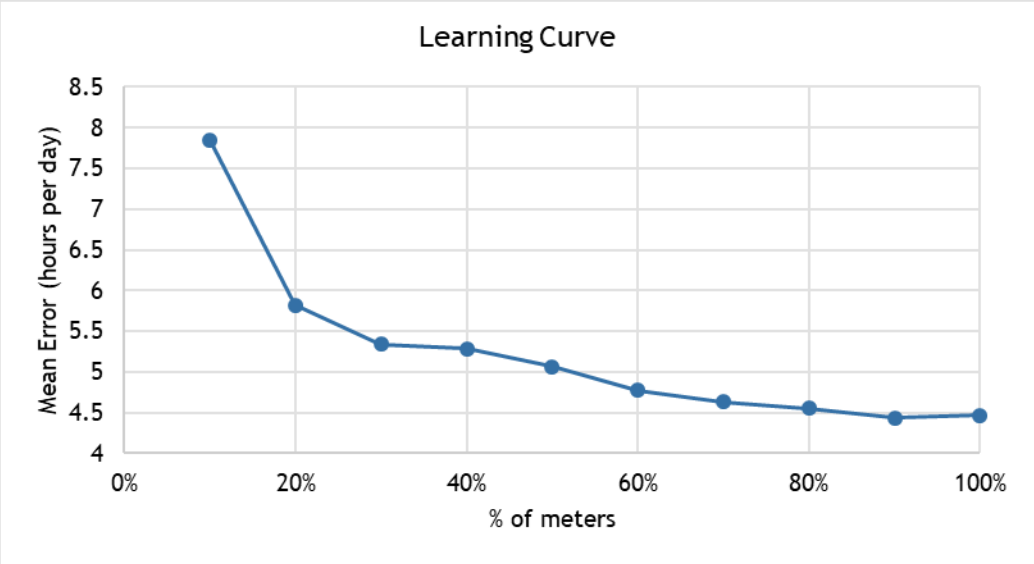

Figure 3: Mean error per day as a function of the percentage of available meters used to train the machine learning algorithm.

To increase accuracy, one algorithm was trained for each location cluster. Parameters were chosen for each algorithm with a partial grid sweep. The accuracy of the algorithms was then calculated with the most recent year of work order data. This was done by summing the ADRs between the most recent reading date to the work order creation date and work order due date, then adding this value to the most recent reading to the work order creation date.

The current method of meter reading prediction saw a total absolute error of 360 032 hours. If the machine learning method had been used over the past year, it would have had a total absolute error of 296 100 hours. Thus, machine learning was able to reduce the total absolute error by 63 931 hours (17.8%). These hours can be further divided into over maintenance and under maintenance hours. The current method had 272 811 hours of under maintenance, while the machine learning method had 113 087 hours of under maintenance, a reduction of 159 724 hours. Over maintenance hours for the current method were 87 221. The machine learning method had 183 031 hours, an increase of 95 810 hours. These errors are further summarized in the box and whisker plot in Figure 4.

Figure 4: Box and whisker plot showing the distributions of the mean error of the current prediction method and the machine learning algorithm.

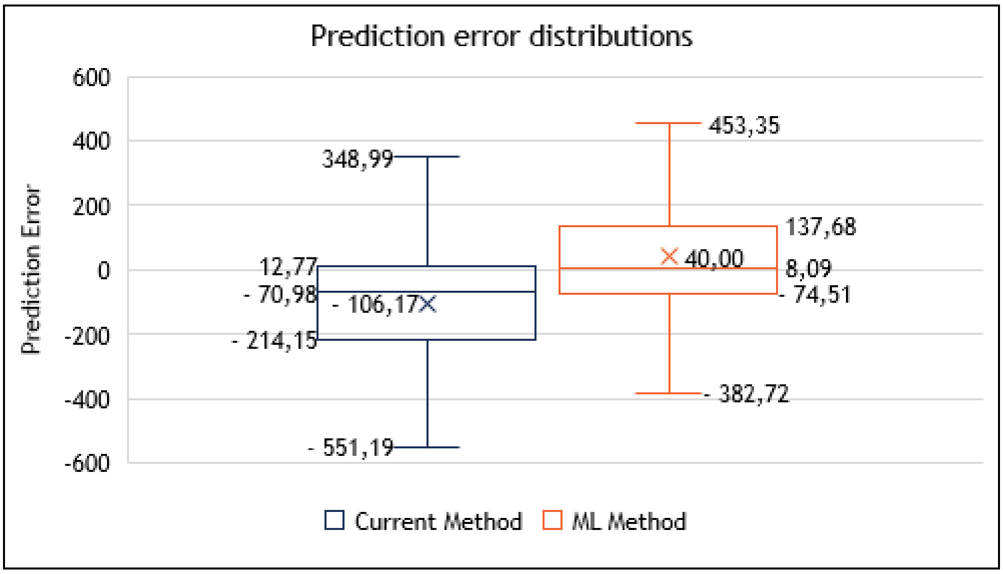

Figure 4: Box and whisker plot showing the distributions of the mean error of the current prediction method and the machine learning algorithm.

An ideal prediction method would have a total absolute error of zero. As no prediction method will be ideal, a reasonable accurate prediction method will have prediction errors centered around zero error, specifying accurate predictions, and a small spread from the zero error, translating to a precise error. The centering of the errors can be considered by the median and mean of the errors. The machine learning method had a mean error of 40,00 and a median error of 8,09. The current prediction method had a mean error of -106,17 and a median error of -70,98. As the absolute value of the mean and median error for the machine learning method is less than the absolute value of the mean and median error for the current method, machine learning creates more accurate predictions than the current prediction method.

The spread of the errors can be considered by adding the absolute values of the upper and lower whiskers on the box plot. The machine learning method had a spread of 836,07. The current prediction method had a spread of 900,18. As the machine learning method had a smaller spread, it is the more precise prediction method. Because the machine learning method was more closely centered on the zero error line and had a smaller spread, it is both more accurate and precise than the current prediction method.

The monetary savings of using machine learning instead of the current method are difficult to quantify, particularly for under maintenance. The costs associated with under maintenance vary greatly based on the part or parts that break, availability of those parts, and lost production, to name a few factors. However, a breakdown does not occur every time a machine is under maintained. Therefore, the true cost of under maintenance is the probability of a breakdown occurring for each hour under maintained multiplied by all associated costs of a breakdown. As breakdowns are not always recorded, and their costs are so variable, data was not available to construct an average cost for each hour of under maintenance.

Calculating over maintenance is much simpler as it only involves the cost of labor and parts. These values are known so the average cost of over maintenance can be calculated. However, without knowing the cost of under maintenance, this value does not provide an accurate picture of cost saved. Therefore, reduction of total error, error spread, and median error value should be used as the determiner for success.

4. CONCLUSION

Like most industries, financial success within the packaging industry relies on high equipment availability. In turn, this high availability relies on maintenance schedules to minimize downtimes. One way of creating these maintenance schedules is through machine usage predictions. The current method of predicting machine usage considers the last certain number of meter readings to create a future prediction. With a total prediction error of 360 032 hours for 500 meters over the past year with this method, there is great room for improvement.

Machine learning was considered as a viable option for creating improved usage predictions in this paper. A literature review was conducted and a proposed solution was created from that review. This proposed solution was then applied to three prediction types. First, readings were predicted. This method produced less accurate results than the current prediction method. Secondly, production amounts between two dates were predicted. This method created slightly more promising results. It did create improved predictions for one out of the five clusters, but for every other cluster it created less accurate predictions. The final method tested was ADR predictions based on the month. When scored on the most recent year of work order data, is reduced the total error by 63 931 hours. Extrapolating that across all 50 000 meters, that translates to a total amount of hour reduction of 6 393 100. The only drawback of the machine learning method was that it is more prone to under maintenance than the current method. Unfortunately, the costs associated with these hours was not able to be calculated due to unavailability of data.

While the total costs saved could not be calculated, machine learning still shows improvements from the current method of machine usage predictions. It reduced the total absolute error of predictions on the 500 meters selected for testing. Furthermore, it shows every indication of extending that prediction accuracy to all 50 000 meters. In conclusion, machine learning was able to create more accurate machine usage predictions than the current method, enabling the creation of more accurate maintenance schedules.

5. REFERENCES

[1] **Yu, Wei & An, Dou & Griffith, David & Yang, Qingyu & Xu, Goubin**. 2015. Towards Statistical Modeling and Machine Learning Based Energy Usage Forecasting in Smart Grid, Applied Computing Review, 15(1). pp 6-16.
[2] **Wong, Derek & Yip, Stephen**. 2018. Machine Learning Classifies Cancer, International Journal of Science, Nature, 555, pp. 446-447.
[3] **Lu, Y. & Rajora, M. & Zou, P. & Liang, S.Y.** 2017. Physics-Embedded Machine Learning: Case Study with Electrochemical Micro-Machining, Machines, 5, pp. 4.
[4] **Ou, Cheng, Chen, & Perng**. 2016. Dynamic cost forecasting model based on extreme learning machine - A case study in steel plant, Computers & Industrial Engineering, 101, pp. 544-553.
[5] **Yaseen, Zaher Mundher & Jaafar, Othman & Deo, Ravinesh C. & Kisi, Ozgur & Adamowski, Jan & Quilty, John & El-Shafie, Ahmed**. 2016. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq, Journal of Hydrology, 542, pp 603-614.
[6] **Bansal, Anshul & Rompikuntla, Susheel Kaushik & Gopinadhan, Jaganadh & Kaur, Amanpreet & Kazi, Zahoor Ahamed**. 2015. Energy Consumption Forecasting for Smart Meters. BAI Conference
[7] **Hazewinkel and Michiel**. 2001. Kolmogorov-Simirnov test. 1st edition. Springer
[8] **Bloice, M.D. & Holzinger, A**. 2016. Machine Learning for Health Informatics, 1st edition, Springer.
[9] **Fushiki, Tadayoshi**. 2011. Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, Springer. 21. pp. 137-146
[10] **Senior, Andrew & Heigold, Georg & Ranzato, Marc' Aurelio**. 2013. An Empirical Study of Learning Rates In Deep Neural Networks For Speech Recognition, Acoustics, Speech and Signal Processing (ICASSP) IEEE International Conference, pp. 6724 - 6728.
[11] **Hall, Mark A.** 2000. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning, Working Paper Series, ISSN 1170-487X, Working Paper 00/8
[12] **Bolón-Canedo, Verónica & Sánchez-Maroño, Noelia & Alonso-Betanzos, Amparo**. 2015. Feature Selection for High-Dimensional Data, 1st Edition, Springer.
[13] **Aboudi, N. El & Benhlima, L**. 2016. Review on Wrapper Feature Selection Approaches, International Conference on Engineering & MIS (ICEMIS), pp. 1-5.
[14] **Criminisi, Antonio & Shotton, Jamie & Konukoglu, Ender**. 2011. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning, Foundations and Trends in Computer Graphics and Vision., 7(2-3), pp. 81–227