Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2022) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 5th International Conference on Emerging Data and Industry 4.0 (EDI40)
March 22-25, 2022, Porto, Portugal

# Intelligent energy management, using data mining techniques at Bosch Car Multimedia, Portugal facilities

Nasim Sadat Mosavi[a], Francisco Freitas[a], Rogério Pires[b], César Rodrigues[b], Isabel Silva[b], Manuel Santos[a] , Paulo Novais[a]

[a]*Algoritmi research center, department of Information systems and technology, University of Minho, Guimareas,Portugal*
[b]*Bosch Car Multimedia,Braga,Portugal*

**Abstract**

Fusion of emerged technologies such as Artificial Intelligence, cloud computing, big data and Internet of Things in manufacturing has pioneered this industry to meet the fourth stage of industrial revolution (industry 4.0). One of the major approaches to keep this sector sustainable and productive is intelligent energy demand planning. Monitoring and controlling the consumption of energy under the industry 4.0 protocol directly results in minimizing the cost of operation and maximizing the efficiency. To advance the research on adoption of industry 4.0, this study examines CRISP-DM methodology to project data mining approach over data from 2020 to 2021 which was collected from industrial sensors to predict/ forecast future electrical consumption at Bosch car multimedia facilities located at Braga, Portugal. Moreover, the influence of indicators such as humidity and temperature on electrical energy consumption was investigated. This study employed five promising regression algorithms and FaceBook porphet (FB prophet) to apply over data belongs to two HVAC (heating, ventilation, and air conditioning) sensors (E333, 3260). Results indicate Random Forest (RF) algorithms as a potential regression approach for prediction and the outcome of FB prophet to forecast the demand of future usage of electrical energy associated with HVAC presented. Based on that, it was concluded that predicting the usage of electrical energy for both data points are time dependent. Where, “timestamp” was identified as the most effective feature to predict consume of electrical energy by regression technique (RF). The result of this study was integrated with Intelligent Industrial Management System (IIMS) at Bosch Portugal.



*Keywords:* Energy consumption; Prediction, Optimization, Data Mining, Machine Learning, Forecasting, Industry 4.0

## 1. Introduction

The widespread of adopting emerged technologies such as Artificial Intelligent (AI), big data, Internet of Things

(IoT), data analytics through the industry 4.0 platform, has caused manufacturing to be smarter in minimizing the cost of production/ operation with higher quality and efficiency. The increasing use of industrial sensors for self-sense, self-act and communication to provide real time data acquisition, has facilitated accurate and rapid decision making particularly in regard to the energy management 4.0 [1],[2].

Based on that, using AI applications to optimize energy consumption not only affects estimating of future energy demand in manufacturing, but also directly influences the development of green production system, where the lower energy consumption results in lower production of CO2 and cost of operation [3],[4], [5]. In this line of research, literature has demonstrated deep investigation using various techniques such as mathematical modelling and AI over industrial data to predict energy demand for policy makers and smart energy management 4.0. The main reason for such industrial energy modelling is to identify the required quantity of energy as an input parameter for regional, national or particular industrial sector. The outcomes help decision maker to control the consumption and supply[6], [7]. For example, in 2008, Adams and Shachmurove offered an econometric model to forecast Chinese energy consumption. Moreover, in 2010, Azadeh et al. applied a fuzzy regression algorithm to estimate the energy consumption of Iran. In 2013, Geem and Roper, estimated energy demand of South Korea with an ANN model using variables such as GDP, population and import/export. In total, energy consumption modeling relies on input data and the type of technique adopted [8].

The current study was a collaboration between University of Minho and Bosch Car Multimedia Portugal under the R&D project: Factory of Future (FoF). Bosch Car Multimedia Portugal is located in Portugal and is part of the audio & video equipment manufacturing industry with more than 1,780 company's corporate family. The major objective of this collaboration on project 52, characterised as the development of an Intelligent Industrial Operation Management framework based on AI technology with the focus on energy consumption management; a platform that employ data generated in IoT platform to support decision-making task for monitoring, controlling and optimizing the consumed electrical energy in operation. Based on that, this study took advantage of emerged technology for energy management 4.0 in manufacturing industry to estimate the demand of electrical energy associated with HVAC sensors. To perform this experiment, CRISP-DM methodology was adopted, and five potential regression algorithms were employed for predicting the demand of energy for two specific energy data points (E333, E260). Moreover, the influential variables were investigated. In addition, FB prophet was used to forecast the usage of energy. This paper is organized to present the outcome of this project by addressing materials and methods which is about the data, methodology, tools and techniques. Then the result of this experiment is presented in six sections: business understanding, data understanding, data preparation, modeling, evaluation and deployment. At the end, conclusion reviews the study, limitations and future works.

## 2. Materials and Methods

### *2.1. Methodology*

For the current study, we have used the CRISP-DM 0.1 approach. CRISP-DM stands for CRoss-Industry Standard Process for Data Mining. Data Mining (DM) refers to the process of applying intelligent techniques on data to extract patterns and to identify valid and useful information [9]. It is a multidisciplinary subject leveraging various techniques such as Artificial Intelligence/Machine Learning(ML), statistics and data analytics [10].

Whereas Fayyad considers Data Mining (DM) as one of the phases in Knowledge Discovery from Database (KDD) process for searching and discovering patterns [11], CRISP-DM guides people to know how DM can be applied in practice in real systems [12]. This is a standard methodology used to support translating business problems and objectives into data mining projects. Regardless of the type of industry, CRISP-DM helps the effectiveness of the outcome by extracting the knowledge from the raw data [13]. This methodology was introduced in the late 90s for Knowledge Discovery from Database (KDD) [14] and was developed by the means of the effort of a consortium initially composed with Daimler-Chrysler, SPSS, and NCR. The six phases of CRISP-DM 0.1 include (1) Business /Application Understanding, (2) Data Understanding, (3) Data Preparation, (4) Modelling, (5) Evaluation, and (6) Deployment. Each phase includes tasks and outcomes addressed in the following sections [13], [15].

### *2.2. Data:*

The dataset for the present study is part of a major research project on Factory of Future project involving University of Minho and Bosch Car Multimedia Portugal facilities. More than fifty million records of registered data in "Energy Platform" application were used from January 2020 to August 2021. The data employed by "Energy Platform" was gathered from more than 2000 IoT sensors (temperature, humidity, production, temperature, energies), at a frequency of 15 minutes. A copy of data was used via access to Hadoop platform.

### *2.3 Tools and Techniques*

The prediction phase uses historical data for training various algorithms [16] which underscores its functional role in learning how to estimate future behavior and event. This phase depends on the type of target, i.e., if the variable is categorical, the prediction methods are called classifications; if the variable is continuous, it is called regression; and in the case of time-dependent variables, the act of prediction is time-serious [17]. In this study, in order to predict/forecast the consumption of energy, we projected regression and forecasting algorithms.

Regression is a type of Machine Learning (ML) technique that allows delivering continuous estimates [18]. The general purpose of regression algorithms is to investigate and find the relationship between several independent variables or features and a dependent variable or target [16]. For this study, we have included the most promising ML methods for regression: Decision Tree (DT), Random Forest (RF), Linear Regression (LR), Support Vector Machines (SVM), and K-nearest neighbors (KNN). In addition, FB prophet was employed to forecast energy demand based on time dependent approach and python (various libraries) and MYSQL on cloud was used to perform the work.

Moreover, cross-validation was used for assessing the predictive models. This technique promotes the training of several ML models on K-fold subsets (folds) of the available input data and evaluating them on the complementary subsets of the data [19]. Finally, the performance of each algorithm was evaluated by using the Root Mean Squared Error (RMSE) which is the root of difference between estimated and actual value.

## **3. Results and discussion**

### *3.1. Business/ Application understanding*

The objective of data mining study was defined to predict/forecast the consumption of electrical energy for specific HVAC sensors. In addition, identify the impact of humidity and temperature on consuming electrical energy. The result of this investigation meets the business objectives which were defined as energy management 4.0. To estimate, monitor and control the consumption /demand of energy using data generated via IoT platform.

### *3.2. Data understanding and preparation*

In this phase and in order to understand data, we performed a data analysis pipeline, including describing, exploring, and verifying the quality of data. Data from January 2020 to August 2021 was extracted and aggregated based on daily/hourly circumstances for the two electrical HVAC data points (E333 and E260). Furthermore, datasets include 450 rows with no missing values and with 7 variables: "timestamp" (time and date of data registration), "EEN_Consumption" (value of electrical energy consumption), "Internal Humidity" (value of internal humidity),

“External Humidity” (presenting external humidity) “Internal temperature” (shows the internal temperature) and “External Temperature” (the value of external temperature). Finally, in order to understand the key relationship among variables, Pearson correlation method was performed.
According to figure 1, whereas “Internal Humidity” with the most positive relation affects electrical consumption for E333, “timestamp” was identified with the highest impact for energy consumption for E260. Based on this analysis, “timestamp” was identified as the second strong variable in positive relationship with the target value. In both datapoints, “External Temperature” was observed with least impact. Moreover, less “holidays” impact on more energy consumption. There is negative link between external humidity and external temperature (r=-45) in both datapoints and the strongest link among variables are identified between “External Temperature” and “Internal Humidity” (r =0.35).

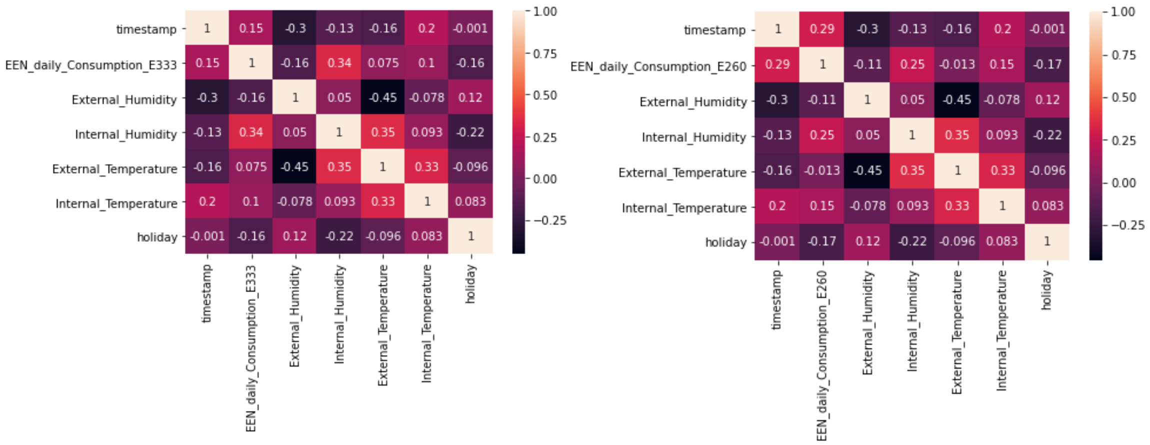


Figure 1. Pearson Correlation for datapoints (E333, E260)

## 3. Modelling

Modelling is a core step in the data mining process and includes tasks such as selecting modelling techniques, generating test design, building and assessing models [13]. As it was mentioned above, the default configuration of five regression algorithm resulted in a benchmarking conclusion to evaluate techniques based on the RMSE metric.
RF is a meta estimator that fits a number of decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control overfitting [20]. This technique uses bagging to increase the diversity of the trees by creating them from various training data subsets [21]. Furthermore, DT regression is a tree-based algorithm for predicting the numeric independent variable and is used to fit a sine curve with additional noisy observation [22]. SVM performs by finding a line of best fit that minimizes the error of a cost function [23] and uses a C parameter, called the complexity parameter, which controls how flexible the process for drawing the line to fit the data can be. In addition, LR is a predictive method to identify the link among the variables when there is a linear relationship between them [24], and KNN, as a regression algorithm, assign a weight (k nearest) to identify the contribution of neighbours based on the average value [25].
According to Table 1, for E333, among regression techniques, RF with the least error rate (RMSE= 293.01) performed better than the other four techniques. Moreover, DT with the RMSE =303.11 was identified as the second potential regression model for predicting the target value. SVM, LR and KNN presented more error rate, where RSME_SVM was observed as 405, RMSE_LR was 475.05 and this metric for KNN was 540.

Similarly, the evaluation method shows RF as the potential technique for predicting the electrical energy consumption associated with E260. Where RMSE for RF was 206.99, this rate for DT was identified as 360.61. RMSE_KNN was 480.50 and SVM and LR presented the highest error where RMSE_SVM=528 and RMSE_LR=508.00.

We have observed similar performance for E260. As table 1 demonstrates, RF with the lowest RMSE (206.99) was identified as the best performance among other regression algorithms. DT with RMSE=360.61 and LR (RMSE=480.50) were the second and the third performance. Accordingly, SVM (RMSE=528) and KNN (RMSE=508) carried the highest error rete in predicting energy consumption associated with E260.

Table1. Performance assessment- regression

| | RMSE scores | | | | |
|---|---|---|---|---|---|
| Data Points | RF | DT | SVM | LR | KNN |
| E333 | 293.01 | 303.11 | 405.00 | 540.00 | 475.05 |
| E260 | 206.99 | 360.61 | 528.00 | 480.50 | 508.00 |

To investigate the influential features associated with electrical energy usage for the HVAC data points, we have analyzed “feature Importance” by the best performance regression algorithms (RF). RF used Gini importance or Mean Decrease Impurity (MDI) method for feature importance to reduce variance. This method performs as the total decrease in node impurity over the all trees [26], [27] According to the bar charts in figure 2, “timestamp” was identified as the most significant feature to predict the value of energy consumption of E333 and E260. Moreover, “Internal Temperature” showed less impact on predicting the target value.

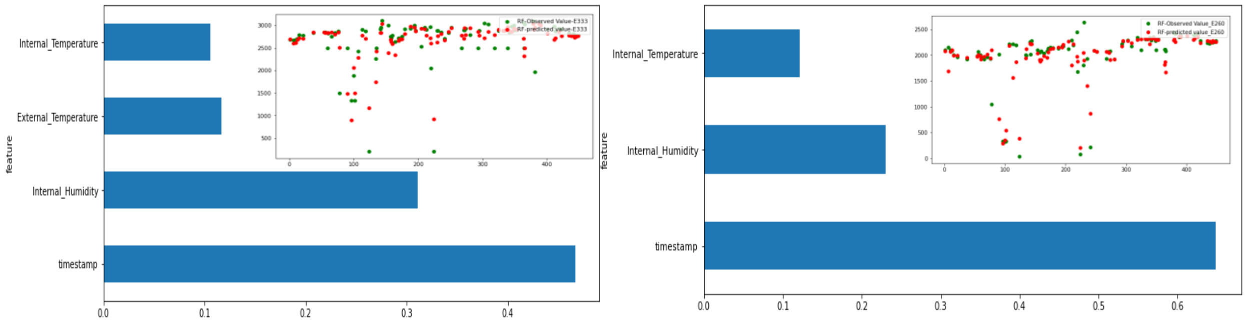


Figure 3. Feature importance by RF

Forecasting approach was applied by projecting FB prophet as the time-series technique to check for the stability of the relationship between the inputs and outputs along time using historical data. This algorithm performs by using additive model for forecasting time series data. It is to manage with non-linear trends with yearly, weekly, and daily seasonality and holiday impacts. FB Prophet works well on sub-daily time series data with a trend in it and is able to fit with missing data, outliers, sudden changes in the data series and creates reasonable forecasts on them [28].

Moreover, we used the upper and lower boundaries to define alarmistic dashboard.

In this experiment, 13965 records were used based on hourly aggregation and we have defined 2880 hours as a future forecasting.  Table 2 shows that RMSE for E333 was 8.07 and this metric for E260 was observed as 15.00.

Table2. Performance assessment- time series

| Data Points | RMSE- FB prophet |
|---|---|
| E333 | 8.07 |
| E260 | 15.00 |

As it is illustrated in figure 3, lines in light blue are the upper and lower boundaries. The black dots are real energy consumption values, and the blue line in the middle is the predictions. We can see on this figure that FB prophet were able to identify the trend line of energy consumption by time and at the same time give a boundary that delineates the values that are small outliers and shows that how far the extremes outliers are from the norm.

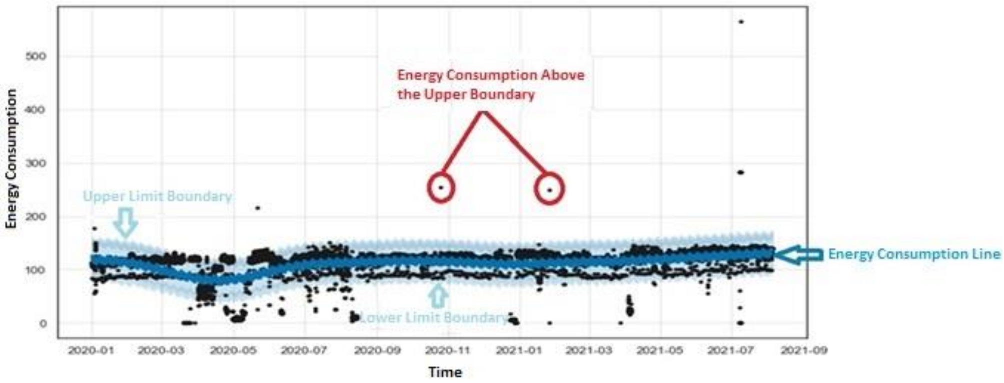


Figure 4. FBprophet forecasting

## Evaluation

While in the modeling phase the accuracy and generality of the models were assessed according to data mining objectives, in the evaluation phase we illustrated to which degree the model meets the business objectives. In other words, we evaluated the suitability of the model by considering the application objectives. This level includes three tasks: evaluation, review and deciding which algorithm meets the objectives. The final decision defines whether the selected model will be deployed or not [13].

Considering the results obtained in the modelling phase, we can observe in Table 3 that FB prophet is the most potential technique to estimate the future demand of electrical energy associated in E333 and E260 which were HVAC sensors. This algorithm performed better over data that belongs to E333 (RMSE=8.07) than data extracted for E260 (RMSE=15.00). Considering the objective of intelligent energy management application, for providing alarmistic capacities, we used upper and lower boundary performed by FB prophet to identify energy usage boundary. Based on that, FB prophet was identified as a potential approach for deployment phase.

Table3. Evaluation and model selection

| | RMSE score | | | | | |
|---|---|---|---|---|---|---|
| Technique | RF | DT | SVM | LR | KNN | FB prophet |
| E333 | 293.01 | 303.11 | 405.00 | 540.00 | 475.05 | 8.07 |
| E260 | 206.99 | 360.61 | 528.00 | 480.50 | 508.00 | 15.00 |

## 4. Deployment

In this phase, the functionality of the result was examined. The outcome of selected algorithm (fbprophrt) was integrated with IIMS at Bosch Car Multimedia. For this purpose, the ETL (Extract, Transformation & Load) process was automated, and the user was able to monitor energy consumption and forecasting based on specific time period and type of sensor. In figure 4, blue line is predicted value and green lines are actual value.


Figure 5. Integrated outcome with Intelligent Industrial Management System

## 5. Conclusions and future work

The increasing use of IoT and AI applications in manufacturing industry to meet the objective of energy management 4.0 pursuits many research and development works. This work was focused on the estimation of energy demand for future planning and policy making. In this line, using data generated via industrial sensors in an integrated platform for predicting/forecasting energy consumption significantly influences productivity and optimization via cutting the cost of energy usage. This paper proposed a framework for predicting electrical energy consumption for two specific HVAC data points at Bosch Car Multimedia Portugal facilities. To perform this experiment, we employed the six phases of CRISP-DM methodology to obtain the best possible performance of ML techniques.

Although using regression support the objective of study to identify importance of features in predicting the usage of energy, this study shows that this target is strongly dependent in time. Based on that observation, FB prophet was selected as a potential approach to forecast future consumption of electrical energy where, the upper and lower boundaries were adopted to define smart alarm system. Moreover, to facilitate energy monitoring for the end use, the result of the work was integrated with Intelligent Industrial Management System at Bosch car multimedia Portugal facilities. Thus the capability of such integrated system provides monitoring of consumed energy and by estimating the demand of energy, the firm is able to plan future policies to control the usage of energy based on each individual data points. Moreover, understanding the influence of each factors (humidity, temperature, timestamp) support Bosch Car Multimedia facilities to consider the level of effectiveness of those variables in future protocols in regards to energy management. Therefore, in long term, the outcome of this project will be resulted in optimization of electrical energy consumption that itself effect of sustainability and green approach to manufacturing.

As future work, a predictive maintenance approach for each individual data point can be a potential method to increase the quality of data registered in Energy Platform; that would be another way to control usage of energy.

## References


[1] X. X. Pai ZHENG, Honghui WANG, Zhiqian SANG, Ray Y.ZHONG, Yongkui LIU, Chao LIU, Khamdi MUBAROK, Shiqiang YU, "Smart manufacturing systems for Industry 4.0. Conceptual.pdf." Springer Berlin Heidelberg, Germany, 2018.

[2] L. Xu, C. Huang, C. Li, J. Wang, H. Liu, and X. Wang, "A novel intelligent reasoning system to estimate energy consumption and optimize cutting parameters toward sustainable machining," *J. Clean. Prod.*, vol. 261, p. 121160, 2020.

[3] P. M. B. M. B. J. Franke, "Reducing the Energy Consumption of industrial robots in manufacturing system." London, 2015.

[4] C. Wei, "Design of Energy Consumption Monitoring and Energy-saving Management System of Intelligent Building based on the Internet of Things," pp. 3650–3652, 2011.

[5] C. Schmidt, W. Li, S. Thiede, S. Kara, and C. Herrmann, "A Methodology for Customized Prediction of Energy Consumption in Manufacturing Industries," vol. 2, no. 2, pp. 163–172, 2015.

[6] P. Palensky, S. Member, D. Dietrich, and S. Member, "Demand Side Management : Demand Response , Intelligent Energy Systems , and Smart Loads," vol. 7, no. 3, pp. 381–388, 2011.

[7] X. Zhang and X. Ming, "An implementation for Smart Manufacturing Information System (SMIS) from an industrial practice survey," *Computers and Industrial Engineering*, vol. 151. 2021.

[8] A. Kialashaki and J. R. Reisel, "Development and validation of artificial neural network models of the energy demand in the industrial sector of the United States," *Energy*, vol. 76, pp. 749–760, 2014.

[9] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Appl. Artif. Intell.*, vol. 17, no. 5–6, pp. 375–381, 2003.

[10] J. Leprince, C. Miller, and W. Zeiler, "Data mining cubes for buildings, a generic framework for multidimensional analytics of building performance data," *Energy Build.*, vol. 248, p. 111195, 2021.

[11] U. Fayyad and R. Uthurusamy, "Data Mining and Knowledge Discovery in Databases," *Commun. ACM*, vol. 39, no. 11, pp. 24–26, 1996.

[12] C. I. Ipp, A. Azevedo, and M. F. Santos, "Double-Coated Bonding Material," *Appliance*, vol. 61, no. 1, p. 35, 2004.

[13] C. Pete *et al.*, "Crisp-Dm 1.0," *Cris. Consort.*, p. 76, 2000.

[14] N. W. Grady, "Knowledge Discovery in Data KDD Meets Big Data," *Arch. Civ. Eng.*, vol. 62, no. 2, pp. 217–228, 2016.

[15] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," *Procedia CIRP*, vol. 79, pp. 403–408, 2019.

[16] Z. Michalewicz, M. Schmidt, M. Michalewicz, and C. Chiriac, *Adaptive business intelligence*, vol. 52, no. 1. 2006.

[17] D. Delen, *Prescriptive Analytics The Final Frontier for Evidence-Based Management and Optimal Decision*. Pearson Education, Inc, 2020.

[18] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "International Journal of Information Management Artificial intelligence for decision making in the era of Big Data – evolution , challenges and research agenda," *Int. J. Inf. Manage.*, vol. 48, no. January, pp. 63–71, 2019.

[19] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Int. Jt. Conf. Artif. Intell.*, no. March 2001, 1995.

[20] S. L. A. Lee, A. Z. Kouzani, and E. J. Hu, "Random forest based lung nodule classification aided by clustering," *Comput. Med. Imaging Graph.*, vol. 34, no. 7, pp. 535–542, 2010.

[21] V. Rodriguez-galiano, M. Sanchez-castillo, M. Chica-olmo, and M. Chica-rivas, "Machine learning predictive models for mineral prospectivity : An evaluation of neural networks , random forest , regression trees and support vector machines," *Ore Geol. Rev.*, vol. 71, pp. 804–818, 2015.

[22] S. Gupta, A. K. Kar, A. Baabdullah, and W. A. A. Al-Khowaiter, "Big data with cognitive computing: A review for the future," *Int. J. Inf. Manage.*, vol. 42, no. April, pp. 78–89, 2018.

[23] M. A. and R. Khanna, *Efficient Learning Machine*. .

[24] S. Kavitha, S. Varuna, and R. Ramya, "A comparative analysis on linear regression and support vector regression," *Proc. 2016 Online Int. Conf. Green Eng. Technol. IC-GET 2016*, 2017.

[25] S. Bafandeh, I. And, and M. Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background," *J. Eng. Res. Appl. www.ijera.com*, vol. 3, pp. 605–610.

[26] A. P. Cassidy and F. A. Deviney, "Calculating feature importance in data streams with concept drift using Online Random Forest," *Proc. - 2014 IEEE Int. Conf. Big Data, IEEE Big Data 2014*, pp. 23–28, 2015.

[27] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in Forests of randomized trees," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2013.

[28] P. Chakraborty, M. Corici, and T. Magedanz, "A comparative study for Time Series Forecasting within software 5G networks," *2020 14th Int. Conf. Signal Process. Commun. Syst. ICSPCS 2020 - Proc.*, 2020.