The Scientific Temper (2023) Vol. 14 (2): 383-386
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2023.14.2.22
https://scientifictemper.com/

RESEARCH ARTICLE

Data analysis and machine learning-based modeling for real-time production

S. Chandra Prabha[1] and S. Kantha Lakshmi[2]

Abstract

This article primarily focuses on data analysis and real-time data modeling using linear regression and decision tree algorithms that might make revolutionary predictions on production data. Factual time data points include temperature, load, and warning, on all the presented axis are the dependent parameters which be contingent on the changes in the autonomous paraments like load. Monitoring and innovative prediction is very much needed in the industry as there are recurrent load changes that would create a data drift and in term of maintenance, that could impact the production side as continuous monitoring and control are needed. Machine learning based approaches would work better on these real-time production datasets.

Keywords: Machine Learning, Data Analysis, Manufacturing Industry, Real-time data modeling.

Introduction

The manufacturing industry has evolved a lot, and automation has impacted the production sector that focuses on quality, safety and automation, which paved the way in invoking multiple technologies entirely together in one sector. New evolving technologies like the Internet of Things, cloud computing and artificial intelligence paved wave in digitization of even the production data point that are continuously monitored and analyzed. Data monitoring could be very useful in preventive and predictive maintenance (Pech et al., 2021). Fault detection is easily pointed out by detecting irregularities. Data collection from the production machinery by means of various sensor and edge computing is easily possible depending on the type of analysis that could be made on a data set (Liu et al., 2021). Various attributes and features are cleaned using statistical and progressive approaches.

Modeling the real-time data using the advanced algorithms and deployment on cloud so that it could predict well for new data further machine learning operations commonly called as MLops can be developed (Data Science and MLOps). While designing such a process the main important points to consider are data drift and model drift are commonly stated issues and can be overcome by monitoring the deployed model and retraining it whenever necessary by means of fast APIs and micro frameworks. The main focus in this article would be proper data analysis and modeling such that it could be deployed well once tested with the test data set and if the accuracy is high then the model can be deployed on web page or on an environment.

Methodology

Real-time data analysis and modeling involves the following steps data collection, data mining, data visualization, machine learning modeling and validation are the standard steps involved in forecasting data points.

Data Collection

Digitization is part of which manufacturing data points are easily collected with the help of Fanuc and Siemens in the industrial sector usually so that the collected data could be analyzed and decisions could be made relevant to it. Data monitoring and big data analysis take wider steps in defining the accuracy depending on which forecasting and

[1]Department of Electronics and Communications Engineering, KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India

[2]Department of Electrical and Electronics Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India

***Corresponding Author:** S. Chandra Prabha, Department of Electronics and Communications Engineering, KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India, E-Mail: chandraprabhasbe@gmail.com

How to cite this article: Prabha, S.C., Lakshmi, S.K. (2023). Data analysis and machine learning-based modeling for real-time production. The Scientific Temper, **14**(2):383-386.

Doi: 10.58414/SCIENTIFICTEMPER.2023.14.2.22

Source of support: Nil

Conflict of interest: None.

Received: 13/03/2023 **Accepted:** 03/05/2023 **Published** : 30/06/2023

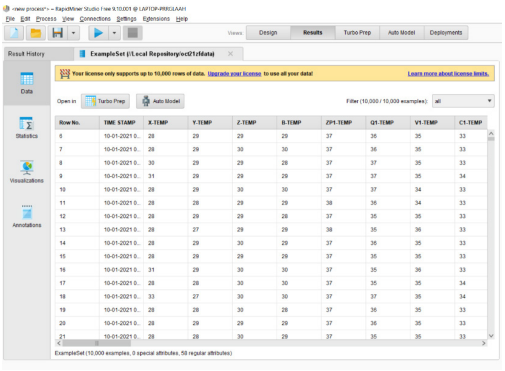

Figure 1: Monthly production data set imported

production vary. Monthly data sets as shown in Figure 1 are used and relevant analysis is done in this articles

Data Wrangling

Data wrangling or data mining is a significant and important step in creating an automated model for prediction (Racickas, 2023). The obtained data as shown in Figure 2 that could have outlier data points, duplications and missing values that could affect the precision of the model and the deviation between the actual and predicted values could be large enough removing and cleaning the data involves various tedious processes that includes exploratory data analysis after which the correlation of the data can be checked.

A correlation matrix that shows the relationship between each label's or variable in an entire data sheet that helps us find good correlation and bad correlations. With that important feature can be selected for training or modeling the algorithm, the remaining features can be dropped out. The correlation matrix as shown in below figure 3 shows that always the self-correlation of the attributes is one

Data Visualization

Visualization of data which can be mainly used in better data understanding and by visualizing the data points the relationship between various features can be analyzed. Various charts and plots are available in this method that invoke bar char, pie chart, histograms, line charts as shown in Figure 4, box plot, a correlation heat map and scatter plot.

The visualizations mainly used here are line plot the is plotted for the entire data set that is shown in Figure 4, and a scatter plot which is used to find the correlation between to particular data points in which each data points are represented in terms of the scattered dots the scatter plot as shown in Figure 5 is plotted for x temperature and y temperature values better visualizations by means of colors code replicas the recurrent analysis of the attributes used

Machine learning Modelling

Linear Regression

Linear regression is a type of supervised machine-learning algorithm which is commonly used for regression (Sen et al.,

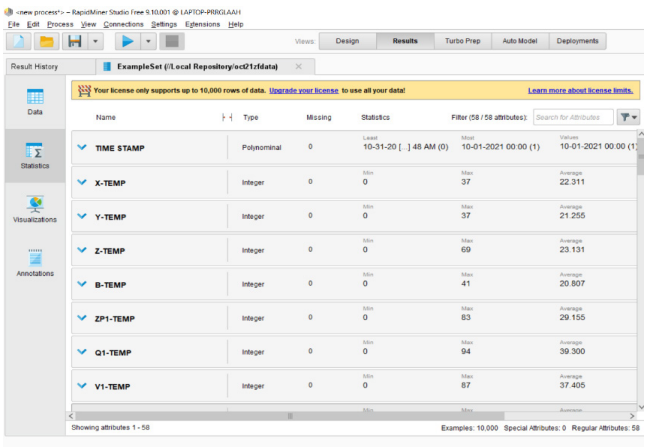

Figure 2: Statistics process in data cleaning

Figure 3: Representation of correlation matrix

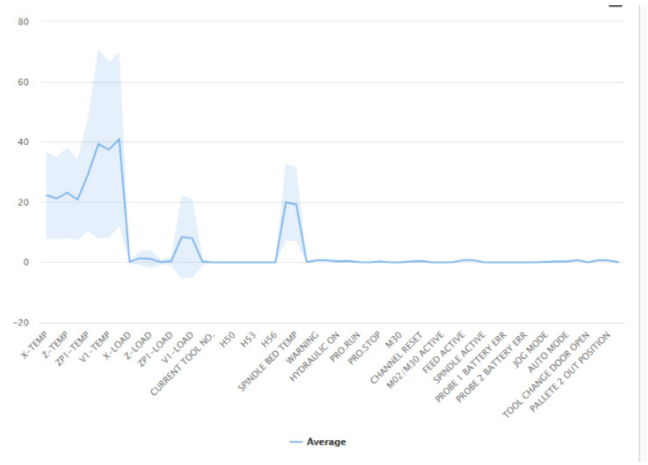

Figure 4: Line plot visualization of entire data points

2020). The x and y features are mapped such that it uses and label the database by formulating and hypothesis through hyperparameter tuning of the intercept and cost function in the gradient descent algorithm. The hypothesis line is iteratively tuned such that the deviation between the actual and predicted data points would be less, such that the cost function is minimum. After implying the hypothesis such that the independent and dependent parameters fits very well and the accuracy of the model can be based on the

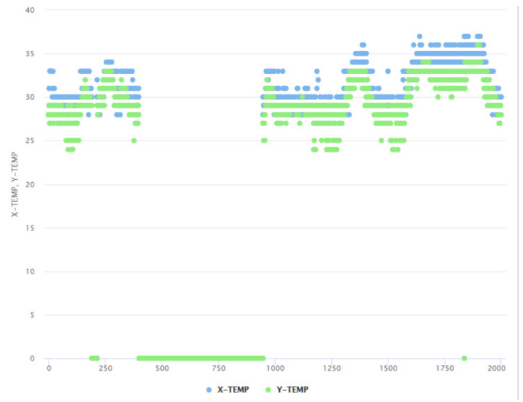

Figure 5: Scatter plot visualization of x temp and y temp

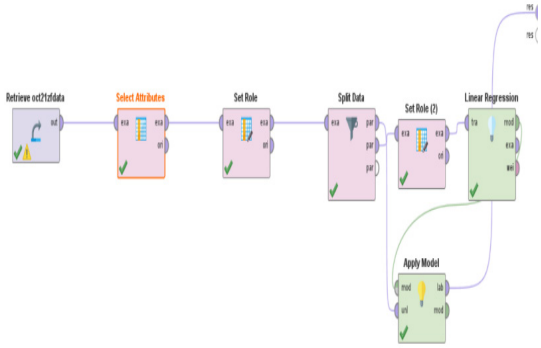

Figure 6: Block representation of linear regression in rapid minor

split size of the training and testing data points and also by means of tuning the hyperparameters below block of linear regression modeling is done using rapid minor software in which various block are interconnected in the process of linear regression modeling

Input data is imported and dropped into the design layout, after which the features are to be selected. The major feature selection part can be done by means of doing many processes the invokes ANOVA analysis and correlation analysis after which the entire required features are filtered out, and set role block, as shown in Figure 6 is extracted in order to set the independent and dependent feature x and y separately by setting their roles after which the entire data set has been split into training and testing and the exact modeling part that is done by linear regression block the model is simulated and such that the target parameters are predicted.

Decision Tree

Decision tree is a supervised machine learning and in regression, it impersonators of the tree. The architecture of the algorithm as shown in Figure 7 which starts with an root node from which the branches are splatted based on the rules and split is made such that it invokes sub split that corresponds to sub-branch's n number of splits and branches can be made deponing on the depth of the data set used in modeling the stopping condition of the algorithm corresponds such the further splits could not be

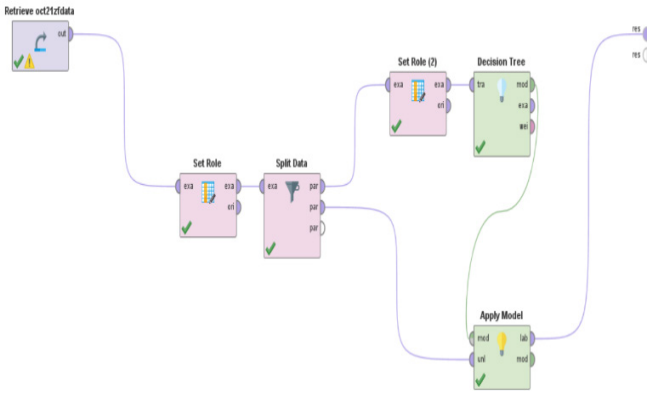

Figure 7: Block representation of linear regression in rapid minor

Row No.	prediction(Y-TEMP)	X-TEMP	Y-TEMP	Z-TEMP
1	26.114	28	29	30
2	26.251	29	28	32
3	26.251	28	29	29
4	26.251	28	29	30
5	26.251	30	29	29
6	26.114	28	29	30
7	26.688	28	28	29
8	26.963	28	27	29
9	26.251	28	29	30
10	26.251	28	29	29
11	26.389	31	29	30
12	26.251	28	28	30
13	26.251	28	28	30
14	26.251	28	29	29
15	26.251	28	28	30
16	26.389	28	28	30
17	26.389	28	29	30

Figure 8: X-axis, Y-axis and Z-axis temperature (Model -1)

Row No.	prediction(Y-TEMP)	X-TEMP	Y-TEMP	Z-TEMP	ZP1-TEMP
1	27.836	28	29	30	37
2	27.836	29	28	32	37
3	27.486	28	29	29	37
4	27.836	28	29	30	37
5	27.872	30	29	29	37
6	27.836	28	29	30	37
7	27.486	28	28	29	38
8	27.486	28	27	29	38
9	27.836	28	29	30	37
10	27.486	28	29	29	37
11	29.429	31	29	30	37
12	27.836	28	28	30	37
13	27.836	28	28	30	37
14	27.486	28	29	29	37
15	27.836	28	28	30	37
16	27.836	28	28	30	37
17	27.836	28	29	30	37

Figure 9: X-axis, Y-axis and Z-axis temperature (Model -2)

made and then the entire algorithm stops the training part (Rokach & Maimon, 2005).

Results and Discussion

Machine learning modeling is done using the linear regression and decision tree algorithm. Both are supervised machine learning algorithms that are mainly modeled for forecasting the y-axis temperature depending on the load parameters in all the axes. This predictive modeling could help in better planning in production. The upcoming temperature changes mainly depend on the amount of

load applied and circuital points that could affect the machine's performance. The relative model is validated in terms of accuracy. The decision tree algorithm comparatively performed well and the accuracy of the model is high than the linear regression modeling.

Linear Tree Predictions on Y Temperature

The decision tree model performed better as the deviation between actual and predicted variables is very low. The model could be further improved by performing hyperparameter tuning. It can be improved by performing data augmentation, retraining the model and improving the skewness in the model so that all data points are included in training and testing.

Decision Tree Predictions on Y Temperature

By comparing the two machine learning models, we observed that changes in load varies the temperature which is a critical parameter in production. The predictive modeling could help in better planning in production. Form the Figures 8 (Model 1) and 9 (Model 1) it can be concluded the Model 2 have batter performance

Conclusion

This article used real-time production data to forecast the fault or irregular patterns that could increase the machinery's downtime and affect the production rate. By comparing the two machine learning models, we could better visualize the changes caused by load changes. The main critical parameter in production is temperature. Further, the machine learning model, which was developed in the developer environment, could be deployed using micro frameworks like flask and fast APIs on the cloud so that machine learning operations would be a complete loop cycle and management of data drift and concept drift also can be handled further.

References

Data Science and MLOps use case. (2023). https://dataplatform.cloud.ibm.com/docs/content/wsj/getting-started/use-case-data-science.html

Liu, C., Su, X., & Li, C. (2021). Edge computing for data anomaly detection of multi-sensors in underground mining. Electronics, 10(3), 302.

Pech, M., Vrchota, J., & Bednář, J. (2021). Predictive Maintenance and Intelligent Sensors in Smart Factory: Review. Sensors, 21(4), 1470. https://doi.org/10.3390/s21041470

Racickas, L. (2023, Feb 10). Data Wrangling: Benefits, Processes, and Application in AI. https://coresignal.com/blog/data-wrangling/

Rokach, L., & Maimon, O. (2005). Decision trees. Data mining and knowledge discovery handbook, 165-192.

Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. In Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018 (pp. 99-111). Springer Singapore.