Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 5, May 2021: 2408 – 2418

Design and Development of a Neoteric Predictive Maintenance Technique Implemented through Comparative Analysis of ML Algorithms to reduce machine failure

M. Shyam[a], K. Srinithy[b], S. Srinidhi[c], S.Sivaranjani[d], J.Yogapriya[e]

[a]Assistant Professor, Department of Electronics & Communication Engineering, R.M.K Engineering College, Chennai, Tamil Nadu, smm.ece@rmkec.ac.in
[b]UG Scholar, Department of Electronics & Communication Engineering, R.M.K Engineering College, Chennai, Tamil Nadu, srin17416.ec@rmkec.ac.in
[c]UG Scholar, Department of Electronics & Communication Engineering, R.M.K Engineering College, Chennai, Tamil Nadu, srin17415.ec@rmkec.ac.in
[d]UG Scholar, Department of Electronics & Communication Engineering, R.M.K Engineering College, Chennai, Tamil Nadu, siva17413.ec@rmkec.ac.in
[e]Programmer Analyst, Cognizant, Chennai, Tamil Nadu, yogapriya.j@cognizant.com

Abstract

Predictive maintenance uses software and historical data collected using sensors to prevent machine failure by analyzing the production data, to identify the operating pattern and thus predict issues before they commence. Predictive maintenance is the process of continually gathering and transmitting the behavior data of the product , storing the data in a central storage hub, and applying analytics methodologies available in advanced big data to sort the massive amount of data so as to identify important data pattern. The data is fitted into machine learning models and trained with the past data to successfully predict the probable failure of the machine. Five machine learning algorithms such as Support Vector Machine, K-Nearest Neighbor, Random forest, Naive Bayes and the Stochastic Gradient Descent are implemented on the real time data obtained from the machines using sensors. Three models with highest efficiency of prediction are taken to predict machine status collectively. The machine status which is predicted by the majority of the machine learning models is taken as the final machine status. The data of whether machine failure will occur or not is constantly being updated in a real-time dashboard and is made accessible to the workers. The dashboard for the generated insights is created to allow maintenance engineers to perform corrective action.

Keywords: Accuracy, Classification algorithms, Efficiency, KNN, Machine Learning, Naive Bayes, Predictive maintenance, Random Forest, Remaining Useful Time, SVM, Stochastic Gradient descent, Sensor data, Training set.

1. Introduction

The nature of the process manufacturing industry is featured in terms of high capital investments involved mainly for the machines that are deployed in production. In order to ensure maximum turnover in the business for this investment, maximum and efficient usage of these equipment are necessary, for which maintenance of these are highly essential such that they are in their optimal performing state. Only then the machines can work continuously without any halt in the production lines. Maintenance activities are proved to be highly needed for a wide range of reasons. Until now, scheduled maintenance is carried out by the factory managers and machine operators to regularly perform maintenance of the machines and equipment for preventing downtime. This involves visual inspections, followed by regular asset inspections to learn and monitor more specific, objective information about the condition of the machine or system. This not only consumes unnecessary resources but also leads to productivity losses. Moreover, half of all preventive maintenance activities are ineffective and inadequate. Thus predictive maintenance serves as a superior approach to resolve the problems of preventive maintenance. The novelty is to improve the prediction accuracy of the previously existing predictive maintenance techniques and increase the 'hits to miss' ratio. Predictive maintenance is carried out using machine learning algorithms and analytic techniques to predict asset failure.

The root of AI application is the computerized self-learning and this ability is known as Machine Learning (ML). ML models involves pattern-learning abilities, adapts changes that can occur in the input data along with the ability to self-educate about the changes that takes place in the real time data feed. In Machine Learning algorithms the systems don't rely on explicit programming, but they are capable of improving their own performance by using the collected data with their experience. This experience includes trillions of observations, these machine learning algorithms have the ability to learn continuously, improve accuracy by recording "hits and misses" which therefore leads them to make predictions. This prediction making capability of Machine Learning leads to set its path towards the Predictive Maintenance, making maximum utilization of traditional yet sophisticated algorithms or newly created algorithms to optimize maintenance, improve model standard and further enhance the throughput of the production.

The models are first created using various machine learning algorithms using data obtained from sensors. This phase is generally called the training phase. The second phase is the testing phase where the model created is put into use. The model predicts the output classes, then the accuracy of predictions is calculated.

TRAINING PHASE:

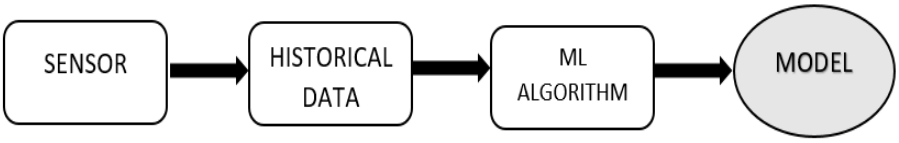

Fig.1-Training Phase

TESTING PHASE:

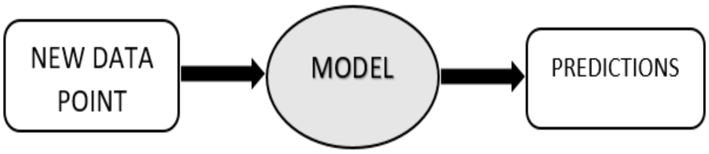

Fig.2-Testing Phase

2. Existing Model

The minimization of machine breakdowns and reducing the time taken to repair the machinery is the main objective. However, the maintenance strategy differs from one firm to another, depending on the operations they perform, manpower they possess and also the frequency of the maintenance activities required. The sensor data obtained varies from machine to machine depending on the function they perform and the scale of the unit. Some of the computations that are the basic needs for tracking the operating conditions of the equipment are obtained by non-invasive data collection activities such as sensors, transducers and some condition monitoring tools.

2.A Condition Monitoring Techniques

The commonly used techniques are:

- Vibration Analysis: Vibration analysis is commonly used for equipment like centrifugal pumps, motors, etc. which performs rotation-based operations. Installed vibration sensors can keep track of axial along with vertical or horizontal movement which is then sent as an alert when the desirable range exceeds.

● Lubricant analysis: Lubricant is also a non-invasive criterion for which analysis is require. In this analysis the number and size of debris such as iron, silicon, aluminum silicate, etc. observed in the samples of oil which is taken in order to determine the asset wear.

● Infrared Thermography: This monitoring technique is related to the radiation emitted by the object with respect to the temperature variation. Though this radiation is invisible for a human eye, infrared cameras are capable of detecting these radiations. The cameras present are set to constantly keep track of the temperature changes in energized equipment.

● Ultrasound Testing: Ultrasonic sensors are capable of detecting the minute noises produced by the malfunctioning equipment. It quickly alerts operators about the fault observed. Some of the issues addressed by the ultrasound testing are extensive corrosion, gas leakage, welding defects and over or under lubricated bearings.

The sensor data is analyzed to identify patterns among them so that with the behavior of one more sensor collectively a generalized notion for the machine status can be attained.

2.B. Maintenance Activities In Use

The manufacturing units adopt one of the following techniques to cope with the machine failure they come across.

1. BREAKDOWN OR CORRECTIVE MAINTENANCE:

In this strategy maintenance activities are carried out after the equipment is out of order and it cannot perform its normal function any longer. It is a traditional method of maintenance and less efficient. Only after there is a machine failure or breakdown the machine is repaired, else the machine is left to operate till it's breakdown time.

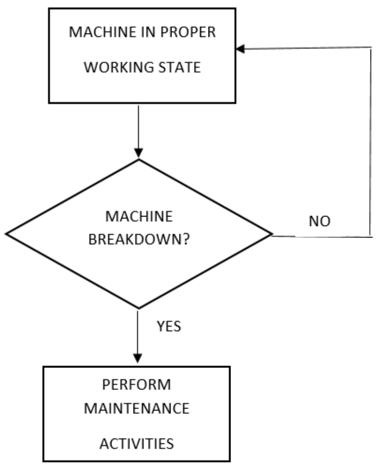

Fig.3-Breakdown Maintenance

The major disadvantage is that the production halts for a certain duration till the machine is restored back to its working state. It therefore becomes essential to have a backup machine but in reality, becomes highly impossible.

2. PREVENTIVE MAINTENANCE:

In this type of maintenance activity, maintenance is carried out either before the failure or before the breakdown of equipment. It is a safety measure designed to reduce the possibility of unexpected breakdowns and disruption in production. This becomes very cumbersome and time consuming in the case of huge production lines and also inefficient to perform maintenance for machines working in a good state.

3. SCHEDULED MAINTENANCE:

This maintenance activity involves the process of keeping the machineries in proper condition. The process includes frequent inspection, replacing defective parts, cleaning, repairing of fault identified segments, timely lubrication, etc. Usually, these operations are carried out when the machine is in use or by a pre-planned shtdown.

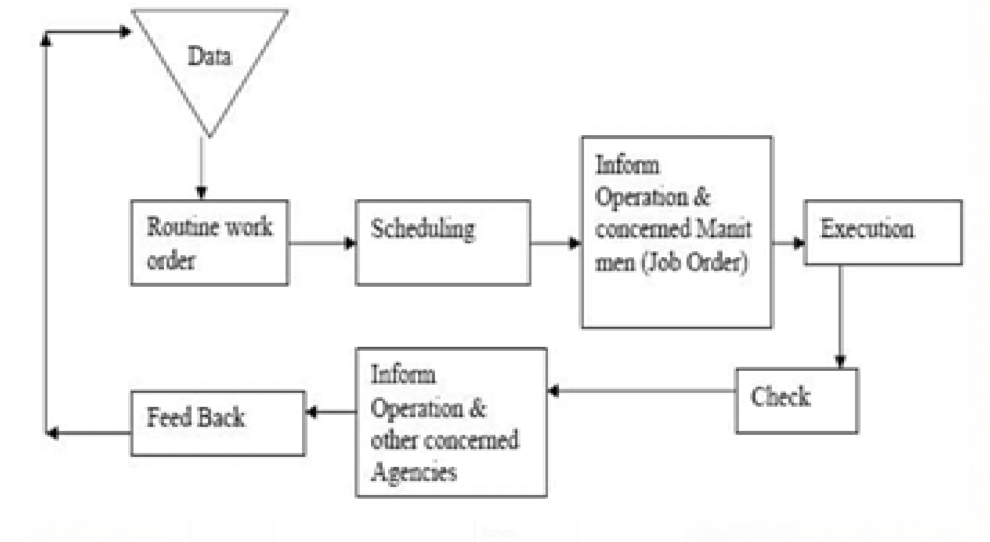

Fig.4-Scheduled Maintenance

This will be ineffective when the production line is huge and will also require more manpower.

All the traditional maintenance activities prove to be ineffective in some way or the other, hence the need for a more efficient way of maintenance activity becomes essential. Industry 4.0 paved the way for a new type of maintenance activity which involved software programming and analysis.

3. Proposed Model Of Evaluation

Depending on the company's maturity level, maintenance activities are being undertaken for many years by the manufacturers. Starting from the traditional methods of maintenance, the maintenance activities have evolved over the years. The disadvantages in the previously used maintenance activities have finally led us to adopt the most efficient maintenance technique. Industry 4.0 has witnessed the emergence of predictive maintenance strategies, which has proven to reduce the down time and capital required for unnecessary machinery repair costs of the company.

3.A. Predictive Maintenance Architecture

For performing predictive maintenance on any industrial assets, few functional blocks are necessary, those are:

● Sensors: Sensors that are usually installed in physical products play an important role in collection of real time data corresponding to the asset.

● Data communication: This acts as a bridge between the data collecting tool and the central data storage hub. It makes use of communication protocols and gateways.

● Central data store: It acts as a central hub which stores the received data. The data is then processed and analyzed. The data can be either be stored on cloud or premises. Database Management Systems like Hadoop can be brought into use for storing this large set of real time data.

● Predictive analysis: The data is aggregated to recognize patterns and generate insights. Machine learning algorithms are used on the historical data to predict machinery default. The insights are further conveyed to the operator in the form of dashboard and alert.

● Root cause analysis: Using the analytic tools on data, the insights are investigated and corrective measures are performed.

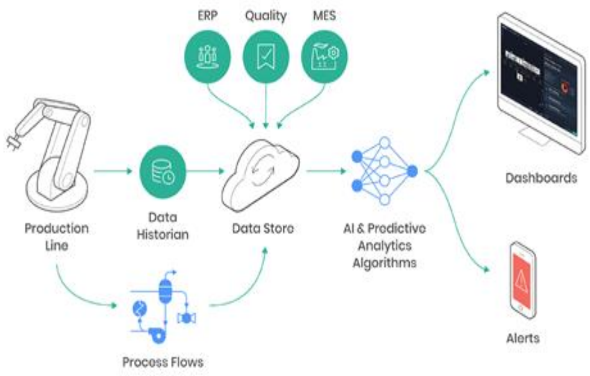

Fig.5-Predictive maintenance Architecture

The data for analysis is obtained from the sensors placed at different parts of the machinery. Different machinery requires the usage of different sensors. The data obtained will be large and streaming data as the machine status is recorded at every second. Hence the data is stored in a central data storage. Machine learning algorithms are used to create models. The models which give the highest efficiency in terms of prediction are used for further computation. A majority voting is taken between the predicted classes which then becomes the final output class. Dashboards act as an interactive interface for the workers from where they can easily monitor the working condition of the machinery. In case of any fault detection by the model, the dashboard quickly alerts the workers, who can perform corrective actions.

3.A.1. Training Module

The dataset obtained from the manufacturing unit contains a timestamp field which specifies the instant at which the sensor data was captured. The data is obtained for every second. Sensor data is obtained from many sensors placed at various machinery locations. Different machinery will have different condition monitoring sensors, some of them include temperature, vibration, ultrasonic, pressure sensor and so on.

The last column is the machine status field which gives the working condition of the machine. The data set is recorded with the previously monitored sensor conditions and the machine status. The machine learning model is trained with the data set inputs, and on the arrival of a new data point the model will be able to successfully predict on which class the output falls under. Depending on the machinery, two output classes can be present such as broken and normal. In some cases, the machinery can have three output classes namely, broken, normal and recovering. In either case, the categorical variables must be encoded into numerical values for performing training.

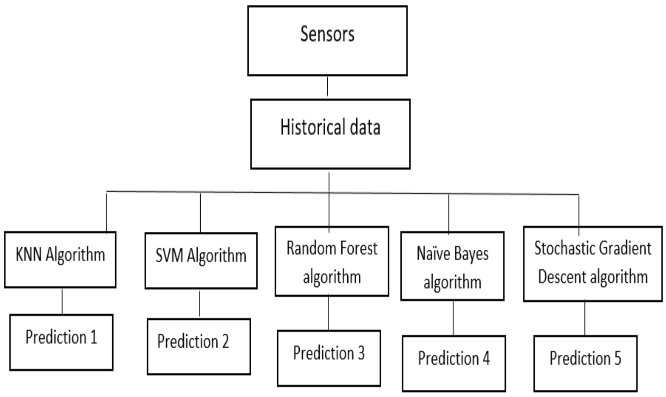

Fig.6-Training Module

Classification algorithms are used to predict the status of the machine. A classification model tries to draw some conclusions from the input values that are fed for training. It is then capable of predicting the class labels/categories for the new data. In our case the labels are Broken, Recovering and Normal. The sensor data is fitted into the models and the model is trained with the inputs and the corresponding outputs. The model will analyse patterns and understand correlations between the input and output variables.

The five machine learning algorithms used are:

1. K Nearest Neighbor

The KNN- k-nearest neighbor is a supervised machine learning algorithm that is simple to train and test is used for the prediction of machine failure. Though it is easy for execution it has a drawback of reduced performance rate with increase in data size upon time. The logic behind this algorithm is the finding of distance between the query and all the other data samples taken into account while training. Following this it selects the predefined number of examples stated as k by the programmer closest to the query, then with which voting is taken for the most frequent class. The distances from the new point and every other points are calculated using the Euclidean distance formula:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The distance between the new input point and all the labels are accumulated category wise. The new point is classified into that label which has the least cumulative distance. The right K value for this classification model is identified by trying various K values and choosing the best fit. The KNN algorithm is suitable for nonlinear data. It is relatively a simple algorithm to interpret and implement. It is considered to be very versatile. However, high memory is required as it stores all of the training data. The prediction stage might be slow when the data set is huge. The KNN algorithm is sensitive to irrelevant features and the scale of the data.

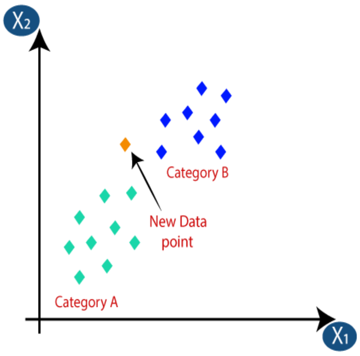

Fig.7-KNN

2.Support Vector Machine:

Support Vector Machine also belongs to supervised machine learning category. This has the ability to perform classification, regression and even outlier detection. The objective of the support vector machine algorithm is hyperplane formation in specified N-dimensional space where N denotes the number of features that is unique for data point classification. A straight line is drawn in between two classes in a linear SVM. All the data points that lie on one particular side of the line will be labelled as one class and the rest on the other side is labelled as the second. SVM assumes the input to be numerical instead of categorical. So, the data is converted using one of the most commonly used is One Hot Encoding. Support Vector Machine is really effective in handling higher dimensions. This algorithm is proven to be effective when the number of uniquely identified features are more than the number of training examples. The hyperplane thus created is affected only by the support vectors thus less impact is created for outliers. SVM is suited for extreme case classification.

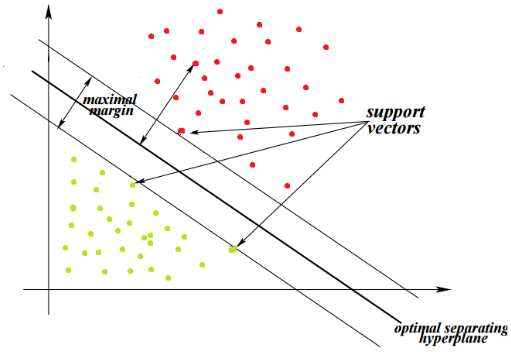

Fig.8-SVM

3.Random Forest Algorithm

Random Forest is one among the famous machine learning algorithm under supervised learning technique. It works on the principle called ensemble learning, which combines multiple sub-classifying models to solve a hard and complex problem in order to improve the performance of the created model. Random Forest is a type of classifier which takes the average to improve the accuracy of the prediction made for the dataset where it contains a number of decision trees under different subsets of that given dataset. Here the prediction is not relying on one decision tree, instead of that prediction of each tree is considered by the classifier based on the majority voting from which final output is decided. The accuracy can be higher if the trees count is increased in this algorithm. This algorithm can also prevent over fitting results thus maintaining the standard of the output.

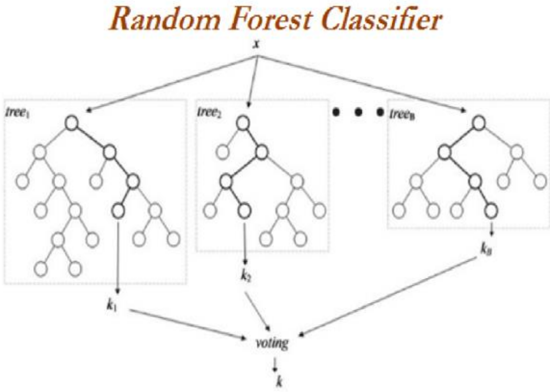

Fig.9-Random Forest

4.Naive Bayes Algorithm

Naïve Bayes is among the easiest methods to create classifiers by assigning labels to the problem's instances. These instances created further are represented as vectors of the feature data. These class labels are drawn from a finite set. This algorithm works on the assumption that the value for a certain feature is independent of the any other feature given in the class variable. For prediction this algorithm takes each feature to be independent. The requirement of small data set for training this model is its advantage, on the other hand this leads to emerging of overfitting model.

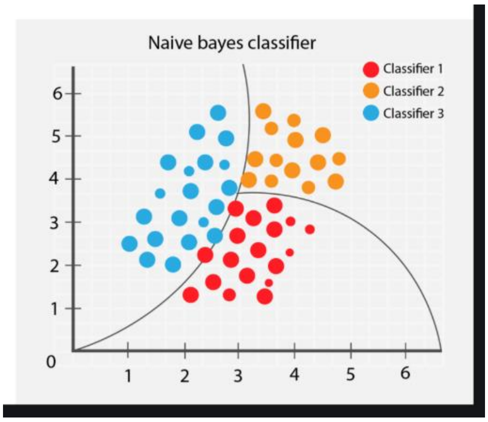

Fig.10-Naïve Bayes

5.Stochastic Gradient Descent Algorithm:

Stochastic Gradient Descent algorithm is one of the machine learning algorithms which is widely used for the classification problems. It is regarded as the continuation of the linear regression algorithm. Gradient is simple terms mean slope and hence gradient descent is to descend to the lowest point on the surface. Basically, the objective is to determine the slope of the objective function with respect to each parameter by picking a random value for the parameters. The gradient function is updated by plugging the parameter values. New parameters values are calculated on every step with respect to the old parameter values and the step size. These steps are repeated concurrently till the gradient is almost 0.

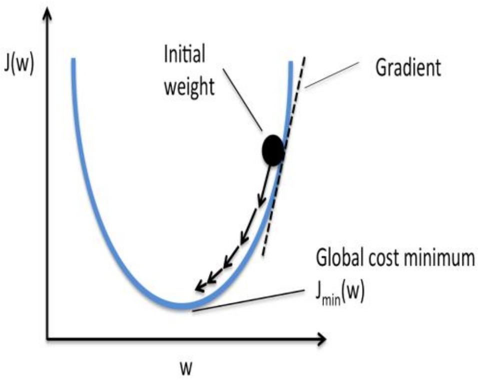

Fig.11-Stochastic Gradient Descent

III.A.2. Testing Module

Once the model is trained, testing data point is used for testing the model. The predicted output class is weighed against the true output class to compute the efficiency of the model. To provide the best result, five machine learning algorithms are used to create five models. Out of the five, the top three models which have higher efficiencies for prediction are taken for further computations.

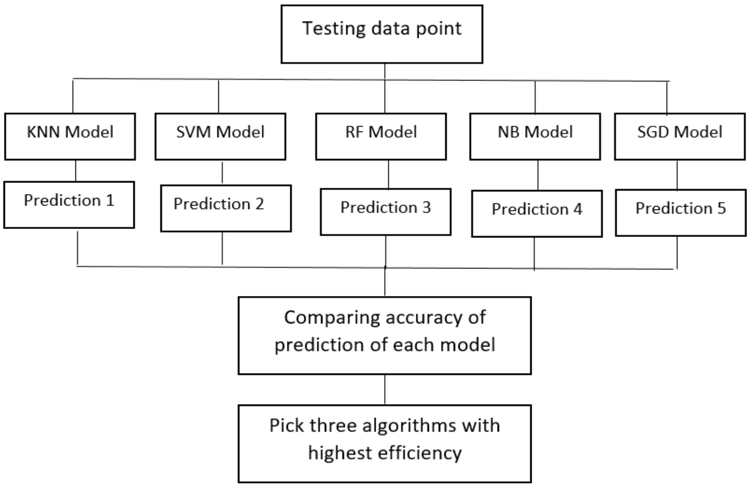

Fig-12-Testing Module

The efficiency of each of the used five models are given below. It is evident from the analysis that the Naive Bayes, SVM and KNN algorithms perform better. It is important to note here that the efficiency and accuracy of each model may vary with different datasets and also the different training approaches used.

S.NO	ALGORITHMS	ACCURACY
1.	NAIVE BAYES	91.6675%
2.	SUPPORT VECTOR MACHINE	89.234%
3.	K-NEAREST NEIGHBOURS	83.714%
4.	STOCHASTIC GRADIENT	75.10%
5.	RANDOM FOREST	70.6675%

Fig.13-Accuracy Table

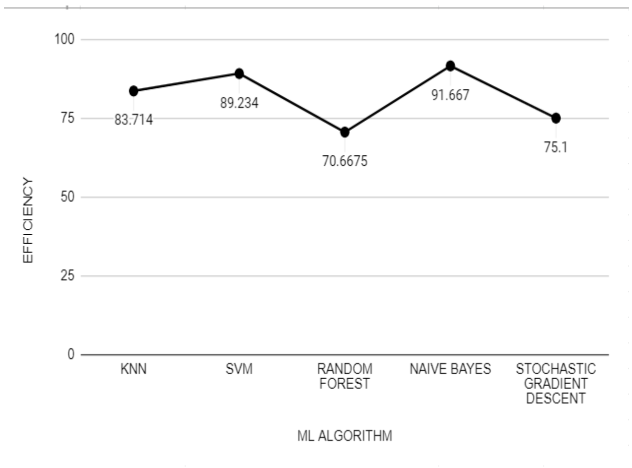

Fig.14-Efficiency Graph

4. Experiment and Result

The top three algorithms are used for further computations. At every point of time the models predict an output class, a majority vote is taken between the three models. The final machine status is the output class predicted by more than one of the machine learning models. With the predicted output classes from all the three algorithms in hand, the majorly occurring (i.e. predicted by more than 1 algorithm) class is taken as the final output machine status.

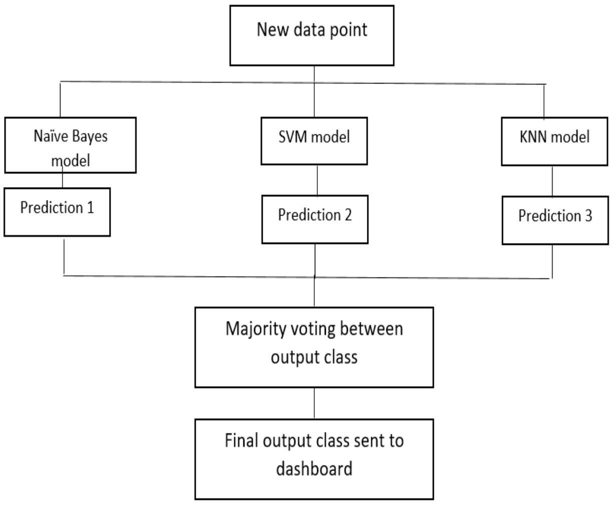

Fig.15 Implementation phase

The machine status is updated in the form of a dashboard as every new output class is predicted. In case more than one model predicts the output class as BROKEN, immediate update is done to the dashboard and alerts are given to the machine engineers to perform corrective actions.

5. Conclusion

From the above analysis it is found that the Naive Bayes, SVM and KNN algorithms produce the best efficiencies among the five machine learning algorithms used for the used data set. Hence, the prediction of the machine failure is performed by collectively using the predictions of all the three algorithms. The novelty of this proposed model is that three algorithms are used in combination to give more accurate results rather than relying on only a single algorithm to predict. It becomes very crucial to predict well in advance the machine status in machinery units for proper uninterrupted production. This not only reduces the time for repair drastically but also ensures that the machinery unit functions throughout without a halt in the production.

The proposed model can be further improved by making advancements in code by also adding regression techniques to predict the Remaining Useful Time(RUL). Also, we can further generalize the code to perform with all types of data sets obtained from different machineries. Further, many algorithms can be implemented to improve the efficiency.

References

[1] A. Kanawaday and A. Sane, "Machine learning for predictive maintenance of industrial machines using IoT sensor data," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 87-90, doi: 10.1109/ICSESS.2017.8342870.

[2] Khorsheed RM, Beyca OF. An integrated machine learning: Utility theory framework for real-time predictive maintenance in pumping systems. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*. November 2020. doi:10.1177/0954405420970517

[3] .Li, Z., Wang, Y. & Wang, KS. Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario. Adv. Manuf. 5, 377–387 (2017). https://doi.org/10.1007/s40436-017-0203-8

[4] Wu, J., Roy, J., & Stewart, W. (2010). Prediction Modeling Using EHR Data: Challenges, Strategies, and a Comparison of Machine Learning Approaches. *Medical Care, 48*(6), S106-S113. Retrieved January 28, 2021, from http://www.jstor.org/stable/20720782

[5] Çınar ZM, Abdussalam Nuhu A, Zeeshan Q, Korhan O, Asmael M, Safaei B. Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability*. 2020; 12(19):8211. https://doi.org/10.3390/su121982

[6] Gęca, J. (2020). Performance comparison of machine learning algorithms for predictive maintenance. *Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska*, *10*(3), 32-35. https://doi.org/10.35784/iapgos.1834

[7] Carvalho, T. P., Soares, F. A. A. de M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 106024. doi:10.1016/j.cie.2019.106024

[8] Silvestrin, Luis P., Mark Hoogendoorn, and Ger Koole. "A Comparative Study of State-of-the-Art Machine Learning Algorithms for Predictive Maintenance." In *SSCI*, pp. 760-767. 2019.

[9] Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2014). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, *11*(3), 812-820.

[10] Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018, July). Machine learning approach for predictive maintenance in industry 4.0. In 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA) (pp. 1-6). IEEE.

[11] Amruthnath, N., & Gupta, T. (2018, April). A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 355-361). IEEE.

[12] Candanedo, I. S., Nieves, E. H., González, S. R., Martín, M. T. S., & Briones, A. G. (2018, August). Machine learning predictive model for industry 4.0. In *International Conference on Knowledge Management in Organizations* (pp. 501-510). Springer, Cham.

[13] Zhao, P., Kurihara, M., Tanaka, J., Noda, T., Chikuma, S., & Suzuki, T. (2017, June). Advanced correlation-based anomaly detection method for predictive maintenance. In *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 78-83). IEEE.

[14] Senanayaka, J. S. L., Kandukuri, S. T., Van Khang, H., & Robbersmyr, K. G. (2017, April). Early detection and classification of bearing faults using support vector machine algorithm. In *2017 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD)* (pp. 250-255). IEEE.

[15] Rashid, Md Mamunur, Muhammad Amar, Iqbal Gondal, and Joarder Kamruzzaman. "A data mining approach for machine fault diagnosis based on associated frequency patterns." *Applied Intelligence* 45, no. 3 (2016): 638-651.

[16] Sakib, Nazmus, and Thorsten Wuest. "Challenges and opportunities of condition-based predictive maintenance: a review." *Procedia CIRP* 78 (2018): 267-272.

[17] Shimada, J., & Sakajo, S. (2016, July). A statistical approach to reduce failure facilities based on predictive maintenance. In *2016 International Joint Conference on Neural Networks (IJCNN)* (pp. 5156-5160). IEEE.x

[18] Cheng, Jack CP, Weiwei Chen, Keyu Chen, and Qian Wang. "Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms." *Automation in Construction* 112 (2020): 103087.

[19] Almstead, J.G., Morjaria, M.A. and Longtin, K.A., General Electric Co, 2002. *Remote diagnostic system and method collecting sensor data according to two storage techniques*. U.S. Patent 6,499,114.

[20] Kuo, C. J., Ting, K. C., Chen, Y. C., Yang, D. L., & Chen, H. M. (2017). Automatic machine status prediction in the era of Industry 4.0: Case study of machines in a spring factory. *Journal of Systems Architecture*, *81*, 44-53.

[21] Rish, Irina. "An empirical study of the naive Bayes classifier." *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3. No. 22. 2001.

[22] Liao, Yihua, and V. Rao Vemuri. "Use of k-nearest neighbor classifier for intrusion detection." *Computers & security* 21.5 (2002): 439-448.

[23] Pal, Mahesh. "Random forest classifier for remote sensing classification." *International journal of remote sensing* 26.1 (2005): 217-222.

[24] Ding, S. F., B. J. Qi, and H. Y. Tan. "An overview on theory and algorithm of support vector machines." *Journal of University of Electronic Science and Technology of China* 40.1 (2011): 2-10.

[25] Bottou, Léon. "Large-scale machine learning with stochastic gradient descent." *Proceedings of COMPSTAT'2010*. Physica-Verlag HD, 2010. 177-186.