JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Factory Downtime Prediction Using Machine Learning Algorithms

[1]Sonali Kadam, [2]Anjali Kadam, [3]Ziya Fatima Sayed, [4]Sakshi Priya, [5]Janhavi Pardeshi, [6]Satvika Nemanwar
[1]Department of Computer Engineering,
[1]Savitribai Phule Pune University, Pune, India

***Abstract**:* The time interval when the system doesn't work is downtime. Downtime duration is the amount of time when a system fails or develops a fault to perform its primary function. The main aim of factory downtime prediction is to identify the potential areas to quickly increase efficiency without operational changes. By doing the literature survey we got to know about the issues, tools available and algorithms like decision trees, Random Forest, Gradient Boosting, Ada Boosting, XGBoost, and SVM are important approaches for building a predictive model that can predict the downtime duration of the factory which can help engineers to quickly identify factories that would cause the most significant downtime. This paper talks about the prerequisite for utilizing quicker, less complex, and more powerful working techniques in predicting downtime duration so that the company can better handle downtimes and increase operational efficiency to better handle demands.

Keywords: downtime, predictive analytics, machine learning, data wrangling, and visualization.

I. INTRODUCTION

The manufacturing industry is crucial to the global economy, and any delay or downtime in production can cause significant financial losses for manufacturers. Predicting the duration of factory downtime is essential for manufacturers to effectively manage production schedules, allocate resources, and minimize losses. In recent years, there has been a growing interest in developing predictive models to forecast the duration of factory downtime, based on historical data and various factors affecting production.
There are several types of downtime prediction, including planned and unplanned downtime prediction. Planned downtime prediction is used to anticipate the need for scheduled maintenance or repairs, while unplanned downtime prediction is used to anticipate unforeseen equipment failures or other events that could disrupt production. Unplanned downtime is costly for industries and predicting factory downtime duration is a challenging task due to the complex nature of manufacturing processes, the large volume of data involved, and the various factors that can affect production.

In this paper, we focus on factory downtime prediction, which is an important area of research in industrial engineering. Downtime in a factory can have a significant impact on production schedules, lead times, and overall productivity. Being able to accurately predict whether the downtime is occurred or not can help managers make more informed decisions about maintenance schedules, staffing levels, and production planning. In the past, manufacturers have relied on historical data and expert knowledge to estimate downtime, but these approaches are often subjective and may not provide accurate predictions. With the advancement of machine learning algorithms and big data analytics, it has become possible to develop more accurate and reliable models for predicting factory downtime.

This paper presents a machine learning-based model for predicting downtime, which has the potential to improve the accuracy and reliability of downtime predictions in the manufacturing industry. In order to predict downtime, various algorithms have been developed by various researches. All these algorithms have been studied with their pros and cons, and an attempt to develop a reliable downtime prediction model is made. The paper will describe the data collection process, feature engineering, model selection, and performance evaluation of the proposed model for factory downtime.

II. Literature Review:

The literature survey is thoughtfully divided into two parts to provide a comprehensive understanding of the subject matter. In the first part, we extensively studied all types of downtime and critically reviewed various machine learning algorithms developed to date. The aim was to identify the most effective algorithms that can be leveraged for factory downtime prediction. The second part of the literature survey presents an in-depth case study of factory downtime prediction, which provides a practical understanding of how the proposed algorithms can be utilized to minimize downtime and enhance operational efficiency.

There are several factors that can impact the downtime in a factory. These factors can include the type of equipment being used, temperature, power, Speed and the skill level of the maintenance personnel. Additionally, there are several techniques that can be used to predict downtime in a factory. These techniques can range from simple statistical models to more complex machine learning algorithms. Some studies have focused on using sensor data and other real-time monitoring technologies to predict downtime events before they occur.

There are a variety of machine learning algorithms used for analyzing and predicting downtime based on different features. Each algorithm has its own advantages and limitations. After analyzing these algorithms, we can conclude Decision trees and random forests offer the ability to identify factors contributing to downtime and improve accuracy. Support vector machines are effective in handling high-dimensional data, while Ada Boosting and extreme gradient boosting can improve the performance of weak classifiers and handle complex interactions, respectively. Time series analysis can reveal patterns and trends over time, and linear and logistic regression can provide simple models for understanding downtime and its predictors. Naive Bayes is a simple and fast algorithm that can handle high-dimensional data but may not capture complex relationships. XGBoost has high predictive power, can handle missing data and imbalanced datasets, and has a fast training speed. The selection of the most suitable algorithm will depend on the specific goals and features of downtime data.

Domain	Description	Issues	Algorithms
Web Application	User-friendly and smooth-running web applications are essential for business success, but issues like slow response times, job failures, database deadlocks, and unexpected traffic can cause poor user experience	• Application issues can lead to unplanned downtime, resulting in blank pages, incorrect data, failed purchases, and more. • Factors responsible for downtime include memory leakage, configuration errors, thread pool exhaustion, and resource-intensive processes	• Decision tree [5] • Random forest • Linear Regression [5]

System Reliability	System reliability is crucial, but issues like overload, bad dependencies, and uneven sharding can cause problems.	• The first issue is the various factors that can contribute to reliability issues, including overload, noisy neighbours, bad dependencies, uneven sharding, bad deployment, and monitoring issues.
		• Development of predictive analytics tools, such as IBM's PASIR, which combines incident ticket information with server properties and utilization metrics to predict and recommend modernization strategies.
		• Determining optimal intervals for preventive maintenance operations, which may not be suitable.
			• Classifiers such as linear regression, decision tree.[1]
			• Framework Architecture
			  ○ Hypothesis
			  ○ Architecture [1]
			• Evaluation of effectiveness of approach running a set of experiments.

Software	Outdated software can lead to inefficient system performance and productivity loss, while software failures can corrupt the application and cause network problems.	• Poorly written code is also responsible for software failures.	• Time Series Analysis • Decision trees • SVM • Random Forest

Hardware	Outdated hardware can increase the chances of system outages and inefficient operations.	• Faulty connections and exceeding storage capacity can also cause hardware issues.	• Time Series Analysis [8] • Reliability Block Diagram [6] • Bayesian Network [6] • SVM
Factory	Factory downtime prediction aim to develop accurate models for predicting machine or equipment downtime in manufacturing settings. These models can help manufacturing companies optimize their production schedules and maintenance activities to minimize downtime and maximize efficiency.	• Limitations of traditional regression models • Impact of machine downtime on manufacturing processes • Need for accurate predictions to optimize production schedules and maintenance activities	• Decision trees [14] • Random forests [14] • Support vector machine [15] • Ada Boosting [16] • Extreme Gradient Boosting

Table 1: Literature Survey summary

Machine Learning Algorithms	Applicability to Factory Downtime	Pros	Cons
Decision trees	Can be used to identify factors contributing to downtime and classify downtime based on various features.	Easy to understand and interpret, can handle both categorical and numerical data.	Prone to overfitting, may not work well with small datasets.
Random forests	Can be used to improve decision tree performance by reducing overfitting and providing better accuracy.	Low risk of overfitting, can handle both categorical and numerical data, easy to parallelize.	Can be slow for real-time predictions, may not work well with imbalanced datasets.
Support vector machine	Can be used to classify downtime based on various features and identify factors contributing to downtime.	Effective in high-dimensional spaces, can handle both linear and non-linear data.	Can be sensitive to the choice of kernel function, may take longer to train with large datasets.
Ada Boosting	Can be used to improve classification accuracy by boosting the performance of weak classifiers.	Can improve the performance of weak classifiers, less prone to overfitting than decision trees.	Can be sensitive to noisy data, may not work well with imbalanced datasets, can be computationally expensive.
Extreme Gradient Boosting	Can be used to predict downtime, identify the root cause of downtime, and optimize maintenance schedules.	Can handle complex interactions between variables, less prone to overfitting than decision trees.	Can be sensitive to hyperparameter tuning, may take longer to train with large datasets.
Time series analysis	Can be used to predict and forecast downtime patterns over time.	Reveals patterns and trends in downtime data over time, allowing for predictions and forecasting.	Reveals patterns and trends in downtime data over time, allowing for predictions and forecasting.
Linear regression	Can be used to identify the most significant factors contributing to downtime and predict the duration of downtime.	Provides a simple and interpretable model for understanding the relationship between downtime and predictors.	Assumes a linear relationship between predictors and downtime and can be sensitive to outliers and extreme values.
Logistic regression	Can be used to identify whether a machine failure caused the downtime and classify	Can handle binary classification problems like identifying if a machine	Assumes a linear relationship between predictors and log odds

	downtime based on various features.	failure caused the downtime.	of downtime and may not capture non-linear relationships and interactions.
Naive Bayes	Can be used to classify downtime based on various features and handle high-dimensional data.	Simple and fast algorithm that can be used for identifying the cause of downtime based on predictors.	Assumes independence between predictors and may not capture complex relationships between them.
XGBoost classifier	Can be used to predict downtime, identify the root cause of downtime, and optimize maintenance schedules with improved performance over traditional gradient boosting.	High predictive power, can handle missing data and imbalanced datasets, fast training speed.	Simple and fast algorithm that can be used for identifying the cause of downtime based on predictors.

Table 2: Applicability Table

III. Methodology

We have used grid and cross validation search over hyperparameters with all the algorithms so that we can estimate the performance of our model on data and find the combination of hyperparameters that gives the best performance. After tuning, we evaluated the performance of these models on the training and testing data and found that XGBoost had the best performance with the highest F1-score.

A. Methodology

The proposed methodology for implementing a system to predict downtime of a factory involves defining problem statement and objectives, collecting and pre-processing data, extracting and selecting relevant features, creating training and testing datasets, selecting and training a suitable machine learning model, validating the model's performance, deploying the model in a real-time environment, generating alerts for impending downtime events, continuously monitoring the model's performance, and evaluating and improving the system. This methodology can help ensure successful implementation and achievement of objectives.

B. Proposed System

Here we are using Predictive maintenance to predict the downtime and type of downtime. We are using algorithms like Random Forest, XGBoost, SVM, Logistic regression.

We analyze machine maintenance prediction using various parameters such as air temperature, process temperature, relational speed, torque, and tool wear. The prediction focuses on five types of failures: Tool Wear Failure (TWF), Heat Dissipation Failure (HDF), Power Failure (PWF), Overstrain Failure (OSF), and Random Failure (RNF). We utilize the Python library pandas for data analysis.

C. Procedure

i. Data Collection: This step involves collecting data from various sensors and equipment in the factory, such as temperature sensors, pressure sensors, flow meters, and other relevant data sources. The collected data can also include historical data on past downtime events

ii. Data Pre-processing: The collected data needs to be cleaned and pre-processed to ensure it is accurate and ready for analysis. This involves removing outliers, handling missing data, and normalizing the data

iii. Feature Extraction: Relevant features need to be extracted from the pre-processed data that can be used for predicting downtime. This includes identifying patterns and trends in the data that may indicate downtime. Feature extraction can involve techniques such as Principal Component Analysis (PCA), Wavelet Transform, or Fast Fourier Transform (FFT).

iv. Training Data Creation: The pre-processed and feature-extracted data needs to be split into training and testing datasets. The training dataset is used to train the machine learning model on how to predict downtime based on the extracted features.

v. Machine Learning Model Selection: The appropriate machine learning model needs to be selected for the prediction task. Decision trees, random forest, XGBoost, and SVM are some examples of machine learning algorithms that can be used for predicting downtime duration

vi. Model Training: The selected machine learning model needs to be trained on the training dataset using the extracted features. The model is trained to learn the patterns and trends in the data that can be used to predict downtime.

vii. Model Validation: The performance of the trained machine learning model needs to be evaluated on the testing dataset. This helps ensure that the model can accurately predict downtime. The performance of the model can be evaluated using metrics such as accuracy, precision, recall, and F1-score

viii. Deployment: Once the machine learning model has been trained and validated, it needs to be deployed in a real-time environment to predict downtime duration based on the incoming sensor data. The model can be deployed on a cloud-based platform or on-premise infrastructure.

ix. Alert Generation: When the model predicts an impending downtime event, an alert can be generated to notify relevant personnel to take appropriate action. The alert can be sent via email, SMS, or through an automated notification system.

x. Continuous Monitoring: The performance of the model needs to be continuously monitored to ensure its accuracy in predicting downtime duration. The model can be retrained periodically using new data to improve its performance. Hyperparameters of the model such as learning rate, regularization, and number of features can be tuned to optimize the model's performance.

xi. Overall, this system can help factory owners to Predict the downtime and also can help to predict the type of downtime.

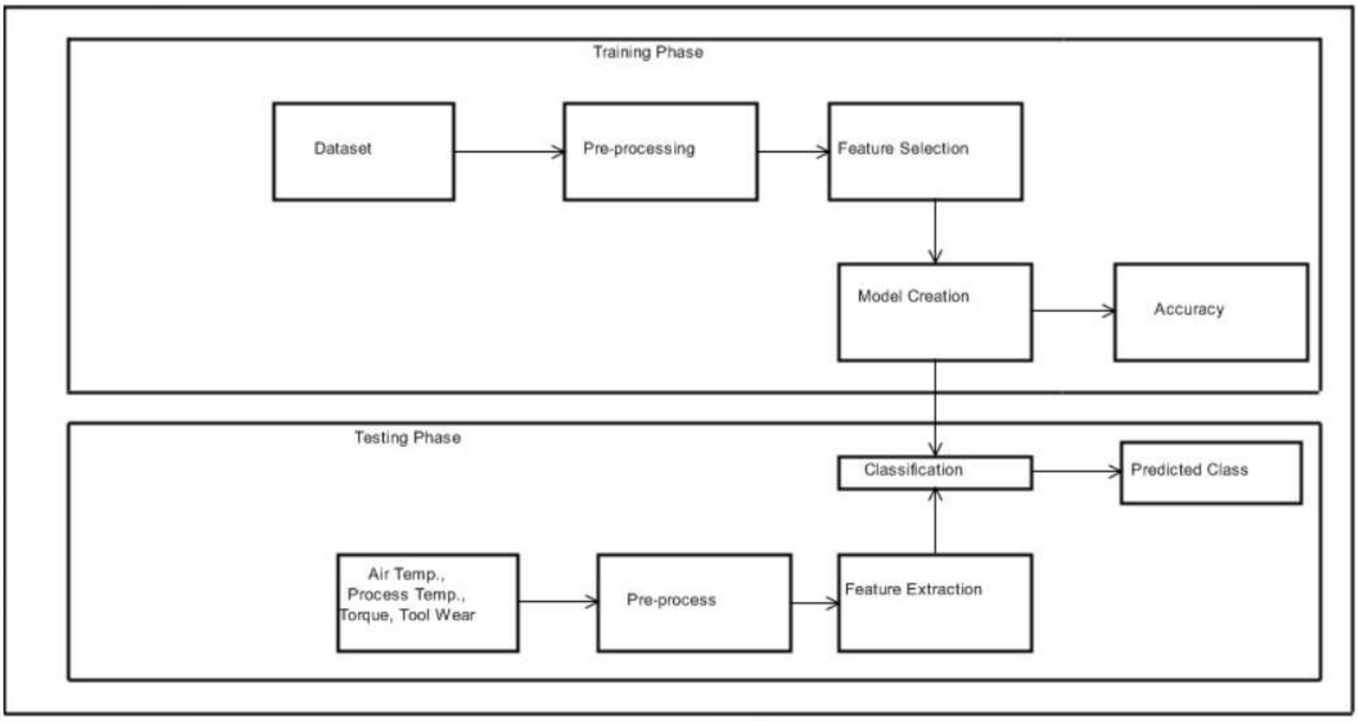

Figure 1: System Architecture

IV. Factory Downtime

The project of system downtime prediction uses Random Forest Algorithm, XG Boost algorithm, SVM and KNN algorithms.

• Random forest classifier is a popular machine learning algorithm that can be used for predicting factory downtime duration. It works by creating multiple decision trees on subsets of the data and then combining their predictions to make a final decision.

• The algorithm uses a random subset of features and data samples to train each decision tree, which reduces overfitting and increases accuracy. Each decision tree predicts the downtime duration based on a set of input features, such as the type of machinery, maintenance history, production volume, and temperature.

• The predictions of all the decision trees are combined by taking the average (in regression problems) or the majority vote (in classification problems) of their predictions. This ensemble method results in a more accurate and robust model that can generalize well to new data.

• In XGBoost, multiple decision trees are created sequentially, with each subsequent tree trying to correct the errors of the previous tree. The final prediction is obtained by aggregating the predictions of all the trees.

• XGBoost offers many hyperparameters that can be tuned to optimize model performance, including learning rate, maximum tree depth, and number of trees. It also provides a built-in feature selection method that can identify the most important features for predicting the downtime duration.

• To use XGBoost for factory downtime duration prediction, one would first need to prepare the data by selecting appropriate features and cleaning the data. Next, the data would be split into training and testing sets. Then, the XGBoost classifier would be trained on the training set using the training data and validated using the testing data. Finally, the trained classifier could be used to predict the downtime duration for new instances.

- Support Vector Machines (SVM) is a machine learning algorithm that can be used for classification and regression tasks. It works by finding the optimal hyperplane that separates different classes in the feature space. In the case of system downtime prediction, SVM can be trained to classify downtime durations into different categories based on relevant features.
- K-Nearest Neighbours (KNN) is another machine learning algorithm commonly used for classification. It classifies new instances based on the majority vote of their k nearest neighbours in the feature space. For system downtime prediction, KNN can be trained on historical data to classify new instances into appropriate downtime categories.
- In summary, the project utilizes Random Forest, SVM, KNN, and XGBoost algorithms for system downtime prediction. Random Forest and XGBoost are ensemble algorithms that combine multiple decision trees, while SVM and KNN are classification algorithms that classify downtime into different categories based on relevant features.

V. Accuracy of the model:

Validation scores:

	KNN	SVC	RFC	XGB
ACC	0.959	0.966	0.977	0.987
AUC	0.954	0.987	0.997	0.999
F1	0.902	0.916	0.943	0.969
F2	0.928	0.931	0.954	0.970

Test scores:

	KNN	SVC	RFC	XGB
ACC	0.966	0.973	0.972	0.983
AUC	0.954	0.992	0.997	0.998
F1	0.916	0.934	0.931	0.958
F2	0.927	0.941	0.945	0.956

Figure 2: Accuracy of models used

To assess the accuracy of machine learning models and determine their suitability for deployment in production via a web application, we employ a confusion matrix. We have used four different machine learning models and will compare them to determine which model is more appropriate.

V. Conclusion

Factory downtime prediction is an important aspect of modern-day manufacturing as it allows engineers to quickly identify potential areas of inefficiency without making any operational changes. Predictive maintenance techniques, including advanced analytics and machine learning algorithms, are proving to be highly effective in predicting downtime duration, enabling companies to optimize maintenance schedules and reduce unplanned downtime. By utilizing these techniques, manufacturers can improve operational efficiency, automate maintenance workflows, and reduce the need for manual intervention, resulting in increased productivity and profitability. So, we have built a model which will tell that downtime occurred of not and if yes then what type of failure it is along with the prevention.

VI. References:

[1] Javier Alonso, Jordi Torres, Ricard Gavalda, "Predicting web server crashes: A case study in comparing prediction algorithms", 2009.

[2] Jasmina Bogojeska, Ioana Giurgiu, George Stark, Dorothea Wiesmann, "IBM Predictive Analytics Reduces Server Downtime", 2020.

[3] Fuad S. Al-Duais,A.-B. A. Mohamed,Taghreed M. Jawa,Neveen Sayed-Ahmed, "Optimal Periods of Conducting Preventive Maintenance to Reduce Expected Downtime and Its Impact on Improving Reliability", 2022.

[4] Soila Pertet , Priya Narasimhan, "Causes of Failure in Web Applications", 2005.

[5] Rushba Javed,Sidra Anwar,Khadija Bibi,, M. Usman Ashraf,Samia Siddique, "Prediction and Monitoring Agents using Weblogs for improved Disaster Recovery in Cloud", 2019.

[6] Chiranjiv Roy, Sourov Moitra, Rashika Malhotra, Subramaniyan Srinivasan, Mainak Das, "IT Infrastructure Downtime Preemption using Hybrid Machine Learning and NLP", 2015.

[7] Joey Oostenbrink, University of Twente,Netherland, "Financial impact of downtime decrease and performance increase of IT services", 2015.

[8] Aaron Jaech, Baosen Zhang, Member, IEEE, Mari Ostendorf, Fellow, IEEE and Daniel S. Kirschen, Fellow, IEEE, "Real-Time Prediction of the Duration of Distribution System Outages", 2018.

[9] Dan Henriksson, Ying Lu, Tarek Abdelzaher, "Improved Prediction for Web server Delay Control", 2004.

Splunk, "Preventing Outages with Predictive Analytics".

[10] Zheng Huang, Jiajun Peng, Huijuan Lian,Jie Guo, and Weidong Qiu, "Advances in Processing, Mining, and Learning Complex Data: From Foundations to Real-World Applications", 2017.

[11] Bulent Oral,Regina M. Cullen,Danny L. Diaz, Eldad A. Hod, MD and Alexander Kratz, "Downtime Procedures for the 21st Century: Using a Fully Integrated Health Record for Uninterrupted Electronic Reporting of Laboratory Results During Laboratory Information System Downtimes", 2015.

[12] Edward J. Williams, "Downtime Data - its collection , analysis and importance", 1995.

Sunita A Yadwad,Dr V. Valli Kumari, Dr. S.Venkata Lakshmi, "Service Outages Prediction through Logs and Tickets Analysis", 2021.

[13] Jim Martin, Shobhana Kirtane, "Service Management using Application Performance Prediction".

[14] B.W. Boehmke and A. Das, "Predicting machine downtime in manufacturing: A comparison of regression and machine learning approaches".

[15] Giobergia, F.; Baralis, E.; Camuglia, M.; Cerquitelli, T.; Mellia, M.; Neri, A.; Tricarico, D.; Tuninetti, A. "Mining sensor data for predictive maintenance in the automotive industry". In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA, Turin, Italy, 1–3 October 2018; pp. 351–360.

[16] Patil, S.S.; Bewoor, A.K.; Kumar, R.; Ahmadi, M.H.; Sharifpur, M.; PraveenKumar, S. "Development of Optimized Maintenance Program for a Steam Boiler System Using Reliability-Centered Maintenance Approach." Sustainability 2022, 14, 10073

[17] Joel Anto Williams N, Vishnu Kumar M, Selva Kumar V, Dr. Sathya Sofia, "Machine predictive maintenance system for industrial applications", 2020

[18] Stephan Matzka, "Explainable Artificial Intelligence for Predictive Maintenance Applications", 2020

[19] Aaron Jaech, Baosen Zhang, Mari Ostendorf, Daniel S. Kirschen, "Real-Time Prediction of the Duration of Distribution System Outages", July 2018

[20] Bita Ghasemkhani, Ozlem Aktas, Derya Birant, "Balanced K-Star: An Explainable Machine Learning Method for Internet-of-Things-Enabled Predictive Maintenance in Manufacturing", February 2023