# Machine failure forecasting using sensor data on cloud

Kaustubh Patil, Arshee Rizvi
M.Tech. (Technology & Development), CTARA
Indian Institute of Technology Bombay (IITB), Mumbai

Abstract- The Industrial Internet of Things (IIoT) in manufacturing involves using machine data from various sensors to be analysed to gain helpful information. The time series component and other machine parameters collected is used to forecast the failure of the machine.

This paper explores the method of storing the data from machine sensors in Google spreadsheet. The Google script editor is used to create an API that forwards data in a spreadsheet. Finally, the machine learning models are applied to the data for data prediction using the Google collaborator. The use of machine learning helps in quality management and reduce maintenance cost. Furthermore, overall manufacturing process is improved on account of data analysis.

Index Terms Machine Learning, Machine failure forecasting, Data Analysis

## I. Introduction

The collection and sharing of data between computers and physical machines have enabled on account of IoT. The processing of data is easily possible once it is available in digital form. The manufacturing industry is exploring how IoT technology can benefit them. The industry is entering into the fourth phase of industrialisation in terms of the cyber-physical system [1]. The availability and reliability of production equipment enhance the competitiveness of the industry, especially in the continuous production industry [2]. However, the maintenance cost to product value cost is greater than 25% [2]. This makes predictive machine failure forecasting a required field.

Challenges. The machines are attached with sensors that measure the pressure changes. This data is stored in a log file which needs to be manually analysed to point out the abnormalities in values. The data is nonlinear, containing unstructured data as well. The challenge is storing, modelling and analysing this data in real-time [4].

Contribution. The expertise in the field and knowledge of underlying manufacturing system is not necessary while modelling using machine learning. However, training data and test data is required to calibrate and validate the constructed model. The steps involved are two. In the first step, the data is inserted into the google spreadsheet using the API in the script editor. The second step is applying the machine learning models to acquired data.

## II. Technologies Used

### A. Google spreadsheet

This easily accessible free platform helps to store the data acquired from the machines with sensors. The spreadsheet has a unique URL and is accessible on the cloud from anywhere. Moreover, it is integrated with tools that enable interaction with content remotely.

### B. Google script editor

This integrated tool in google spreadsheet. In order to create an API that can be used to insert the sensor values in the the spreadsheet, this tool is used.

### C. Postman

This desktop application helps to make a request to the APIs. The request can either retrieve data from the source, or it can be used to send data. The HTTP methods are used to process the APIs request on a web server.

### D. Google collaborator

This computational environment helps to apply machine learning models and visualise the data by carrying out code execution.

## III. Solution Overview

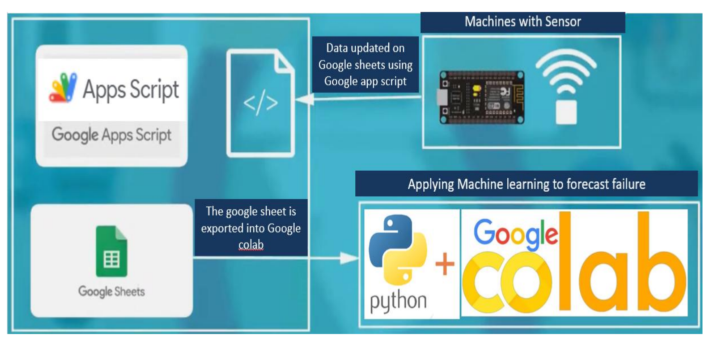


Fig. 1. Architecture of the solution for machine failure forecasting

Fig. 1 shows the architecture of the solution for machine failure forecasting. In this project, the data from machines fitted with sensors is stored in Google spreadsheet. The Google Apps Script is used to create an API that adds data to a spreadsheet. Here JavaScript code is written to write API. The Postman desktop app enables to test the API by sending the data request.

The google sheet with data is exported into Google colab. Finally, the python programming language is used to apply machine learning to forecast the data.

## IV. Building Forecasting Model for Failure Detection

In this section, the steps used to build the forecasting models for failure detection from the data collected are described in detail. Then, the Support Vector Machine (SVM) and Random Forest models predict the failure.

### A. Creating Google spreadsheet in drive

The google spreadsheet can be created on google drive with just in few clicks. First, right-click in blank space and select "Blank spreadsheet under Google sheets. The blank sheet obtained is renamed for easy identification.

### B. Writing API JavaScript

Under tools in google spreadsheet, there is "Script editor" tool". There are two functions written in the script editor. The "HereGet" function triggers the add_data function. The "add_data" function appends and adds the data in the spreadsheet whose URL is present in the script accordingly.

### C. Deploy as web app

The API is created once it is deployed as web app. The API has a unique web app URL. The app access can be set to "anyone", and the app is executed in the authorised user's name.

scored,
dataset
a real-

### D. Testing using Postman

The data in a spreadsheet can be added by passing parameters through a desktop application named "Postman". Then, the API is triggered, and values are sent to the spreadsheet as request.

### E. Exporting the spreadsheet

The interactive environment in Google Colab helps to import the .csv file and process the data. Finally, the data is read and visualised.

### *F. Setting up the environment*

The libraries imported as pandas, NumPy and seaborn. In addition, some of the ML libraries from sklearn are imported. These are train_test_split, LogisticRegression, RandonForestRegressor and metrics.

### G. Creating functions for models

Functions to extract column names from the dataset, train logistic regression model, train random forest model, and return the predicted values are created. In addition, model evaluation functions to calculate prediction accuracy, to construct a confusion matrix are created.

### H. Algorithm for machine failure prediction

The data is split into training and testing data. The Logistic Regression model training and testing are carried out. Scoring of test data is done by constructing the confusion matrix for test data. The absolute error is also calculated, and Mean Absolute Error (MAE) is found out. The prediction accuracy analysis is done for actual fail and predicted fail. The scoring and accuracy calculation is also done for Random Forest Regressor and compared with Logistic regression.

## V. Result

The supervised models were trained on the data set split into 70% training and 30% testing, showed the following accuracy and MAE (Table 1)

TABLE I
Comparison of Different Models

| Model | Prediction Accuracy (%) | Mean Absolute Error (MAE) |
|---|---|---|
| Logistic Regression | 90.91 | 0.07 |
| Random Forest Regressor | 95.27 | 0.13 |

It can be inferred that the Random Forest Regressor is more efficient than the other model compared. However, the mean absolute error observed was more in the case of Random Forest regressor than the Logistic Regression model.

The devised forecasting model can be used to help to reduce the inefficiencies in production cycles and help in putting in place maintenance activities efficiently [3]

## VI. Discussion and Limitations

The ideas and algorithm discussed in this paper are
and confusion matrix is constructed. However, the
processed has certain limitations because it is not from

world manufacturer and generated. The Postman desktop application is used to test the insertion of data into google spreadsheet by creating API in Google script editor and deploying the web app. However, the actual method to transfer sensor data in the real world to the google spreadsheet needs to sort. The PLC connected to the sensor must be able to process the data. The real-time monitoring can be put in place when the inflow of data can also be made in real-time. There are privacy and legal limitation to obtain the actual manufacturer data. In general, the algorithm put forth can forecast machine accuracy with reasonable accuracy.

## VII. Conclusion

In this paper, machine failure prediction from the data obtained from a machine sensor is addressed. The proposed solution makes use of tools available from Google. The casual relationship between the feature values responsible for machine failure is analysed from collected data. The major limitation in maintenance cost and productivity can be overcome by deploying such algorithms. Furthermore, the insights obtained from the data can be used to make the production process more efficient with minimum cost spent for maintenance and plan the maintenance activities in time.

## Acknowledgement

The author would like to thank Prof. Amit Sethi and all the Teaching Assistants for EE769 (Introduction to Machine learning course) course conducted in the spring semester of

2020-21, for their unparalleled support throughout the course invaluable learning.

## REFERENCES

[1] R. Ackermann, "M2M, Internet of Things and Industry 4.0 An Industry Perspective", [Online]. Available: http://www.acatech.de/fileadmin/user upload/BaumstrukturnachWebsite/Acatech/root/de/Material fuer Sonderseiten/SecondGerman-Indian-Workshop/India0113

[2] C. Wang, H. T. Vo, and P. Ni, "An IoT Application for Fault Diagnosis and Prediction," presented at the 2015 IEEE International Conference on Data Science and Data Intensive Systems (DSDIS), Dec. 2015, doi: 10.1109/dsdis.2015.97.

[3] J. Shimada and S. Sakajo, "A statistical approach to reduce failure facilities based on predictive maintenance," presented at the 2016 International Joint Conference on Neural Networks (IJCNN), Jul. 2016, doi: 10.1109/ijcnn.2016.7727880.

[4] V. Gupta, "Applying Machine Learning on IoT Data", [Online]. Available: https://iotmonk.com/courses/applying-machine-learning-on-iot-data