

A TIME SERIES FORECASTING MODEL FOR EQUIPMENT FAILURE PREDICTION USING IOT SENSOR DATA

SATISH KUMAR SOGI¹ AND SUDESH KUMAR MITTAL

ABSTRACT. The manuscript presents a conceptual IoT transformation model to transform grain industry into an IoT compliant entity. The cost factor in grain industry is directly dependent upon the grain storage available and an optimal inventory management. The grain industry operators are always focusing on the cost reduction in the supply chain. Understanding connection between receipt, outturn, inside storage capacity and between storage site movements can give us bits of knowledge that can be helpful in getting ready for the next harvest season, assessing the throughput limit of the framework, connection among the inventory. The study dives into the important pillars of grain inventory management and propose a forecasting model.

1. INTRODUCTION

IoT (internet of things) concepts have been pushed into the industries world-wide for enhancing communication between entities embedded with electronics (intelligent sensors, RFID, etc. based devices) and IT entities (software) which, both, have already attained super intelligence over the years of their technological advancements. The real potential of IoT (Internet of Things) lies in the fact that it can bind all the things, tangible and intangible, around us in an IoT ecosystem where they could have seamless communication (data exchange) and

¹corresponding author

2010 *Mathematics Subject Classification.* 62M10, 68M11.

Key words and phrases. Internet of Things, Anomaly Detection, Forecasting, Outliers, Auto Regressive Integrated Moving Average .

evolve as well simultaneously. The other two dominant technologies which are building solid foundations for the IoT ecosystem to progress are advanced networks and cloud computing. With all these advanced concepts and technologies forming cohesive and pragmatic bonds, the industries today are standing at a point of harvesting profits from the convergence of advanced networks, cloud computing, and the internet of things.

The convergence of blockchains, cloud computing, networks, and the Internet of Things (IoT) has promised all industries a sizeable business shift from the tangent of improved supply chains, product functioning, tracking, persistence and durability of products. This convergence will confer dual benefits on the businesses as in their toolkits will not only have speedy business logic, intelligence, security etc and but also will attain virtues of self-governing, progressive, and learning environment [1]. Quicker response time indicates that business processes should be directly related to business product which further implies that a better product is an outcome of a superior process.

A business product, tangible or intangible, by itself, is a lifeless entity but when it's attached to a business process in a system it gets its vitality. The process provides it a lifespan by first labelling it with characteristics like a model number, shape, type etc, and then following its performance for defects, malfunctioning, feedbacks etc. So far, the business processes mostly keep gathering the data about the products until the product reaches its last lap of a supply chain run. Once it has seen the outside world with the consumer, there are generally no intelligent systems to communicate with it anymore. Though it's complicated at this moment, the IoT ecosystems have the potential to let businesses achieve communications with the sold products for as long as their processes desire [2]. Not only the IoT has opened room for after communication possible between both, processes and products, but also the processes can control the product functioning to an extent as in correcting it in real-time. The sensor-enabled digital twin is designed to have the ability to serve the customer better by detecting faults sooner, quickly improving them, and predicting outcomes over an entire life cycle with greater accuracy [3].

The manuscript presents a framework for anomaly detection and prediction in IoT devices. The data is collected for finding the various insights namely monitoring health of devices, production processes, usage statistics, and load prediction. The time series analysis of the chosen measurements is done to find

a deviation pattern of the measurement compared to the normal pattern. The analysis can be done to predict future anomalies, enabling real time maintenance of the IoT devices.

2. PROPOSED FRAMEWORK

IoT Devices, automated equipments like robots, tools, machinery etc. gather and communicate the data incessantly and the data is a timeseries data i.e. collected at regular intervals of time [4]. The timeseries data generated by single IoT sensor device is analyzed to identify any considerable change in the performance of the system [5]. The step wise demonstration of the proposed framework is given in Fig.1 and the details of the steps followed in the proposed framework are given below:

- (i) The first step in the framework is the data acquisition and data preprocessing. Data acquisition is done using Redflows and DB2.
- (ii) The next step in the framework is the data retrieval and statistical analysis using R-Jupyter notebooks to analyze and detect the change points in the data. R-Jupyter is a statistical platform to analyze the data.
- (iii) The next step in the framework is the data retrieval and statistical analysis using R-Jupyter notebooks to analyze and detect the change points in the data. Before statistical analysis can be performed, data preprocessing is done to handle missing values and outliers.
- (iv) In this step, the time series visualization is done. A time series contains three essential components namely trend, seasonality, and residual. Time series decomposition is done to visualize the sub components of the time series. The anomalies of the time series are analyzed using the residual component whereas the prediction is done using the trend and seasonality component [6].
- (v) A time series can be modelled if it is stationary meaning the mean across various time points is constant. Fourth step involves checking the time series for stationarity. L-Jung and Augmented Dickey-Fuller (ADF) tests are done to check the stationarity of the series [7].
- (vi) The final step is model fitting where the model is fit to trend and seasonality part to predict future occurrences. Model is applied on the residual

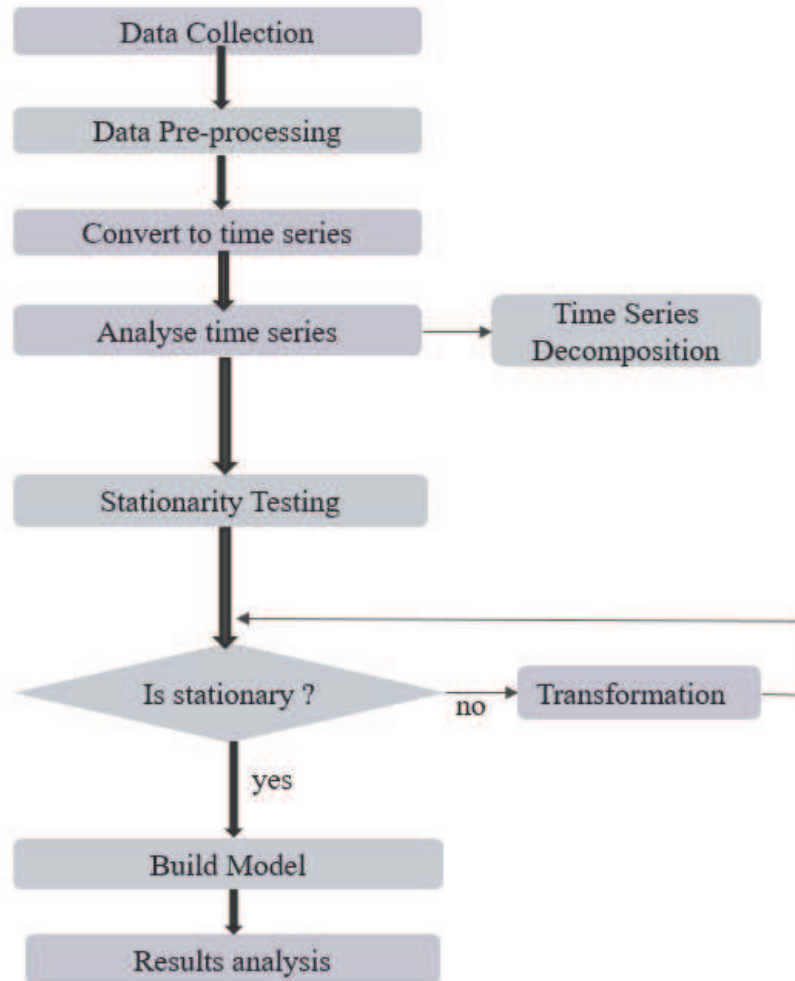


FIGURE 1. Stepwise Demonstration of the Proposed Framework

- part to check for the outliers. ARIMA(Auto Regressive Integrated Moving Average) Model is applied to fit the data and do the forecasting [8].
- (vii) After model fitting, the results are verified using several accuracy measures viz. RMSE(Root Mean Squared Error), MAPE(Mean Absolute Percentage Error), Auto Correlation Function(ACF) error, MAE(Mean Absolute Error), and ACF(Auto Correlation Function) [9].

TABLE 1. IoT Sensor Data

	MoteId	Temp	humidity	Light	Voltage
03:38:15.75	1.0	122.153	-3.919	11.04	2.03
00:59:16.02	2.0	19.984	37.093	45.08	2.69
01:03:16.33	3.0	19.302	38.462	45.08	2.68
01:06:16.01	4.0	19.175	38.837	45.08	2.69

3. EMPIRICAL EVALUATION

Data is collected from online repository maintained at Kaggle [10]. Data consists of the the time series of 28 sensors deployed in various IoT devices for industry automation. The temperature, humidity, light, and voltage reading sensors are installed for continuous monitoring of the devices. The sample of the data is shown in Table 1. The first step involves the time series visualization of the data which is shown in Fig. 2. Next step involves the decomposition of the time series into its constituent parts namely trend, seasonality, and the remainder component. The decomposition of the series is shown in the Fig. 3. From the above decomposition graph, it can be safely concluded that some part of the above plotted time series has trend and seasonality i.e. time series is non-stationary. So, it is important to convert this time series in stationary form [11]. Here this is done by taking log of the time series and by this non-stationarity and variance is handled [12]. The fitted and predicted graph of auto-arima model on actual training data whose order is given as ARIMA (1,1,2)(1,0,0) with drift, the forecasted graph shows that there is not much difference in actual and predicted the result. It can be seen that ARIMA model fit accurately with respect to the selected data as the prediction graph line is not drifting considerably from the actual series. The inverse AR(Auto Regressive) roots and MA(Moving Average) roots are depicted in Fig. 4.

4. RESULTS AND DISCUSSION

The results of the model fitting is depicted in Fig. 5 From the graph, it can be concluded that the model is able to fit the graph quite well and is able fit the trend accurately. Grey lines represents the pattern of the sensors chosen at different data points, whereas the red line represents the fitted trend of the ARIMA Model.

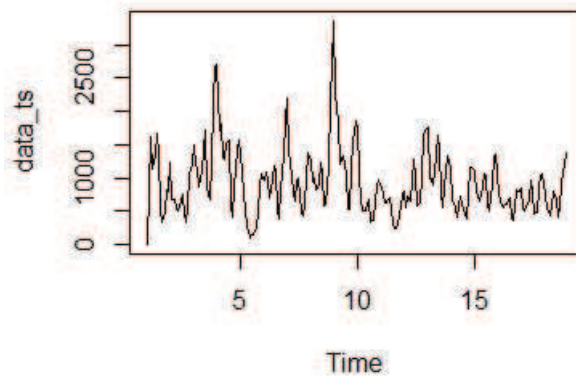


FIGURE 2. Time Series Visualization of the IoT Sensor Data

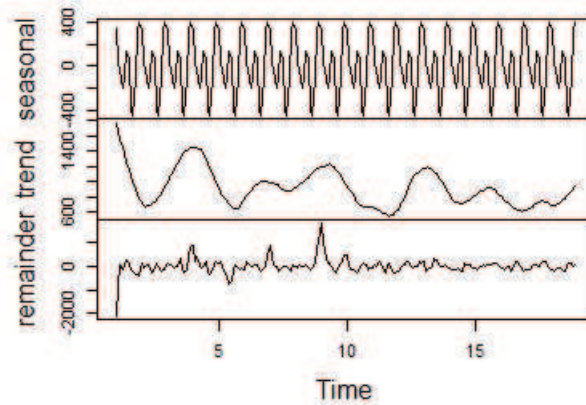


FIGURE 3. Time Series Decomposition

The forecasting of the anomalies on the basis of the past observations is given in Fig.6. The graph here is created using the remainder component of the time series. The remainder component represents the outliers or anomalies in the IoT system. From the graph, it can be concluded that the selected model fits the sytem accurately and the forecast follows the same trend as the past observations.

Further, the model is compared with other state-of art forecasting techniques namely NNAR(Neural Network Auto Regression), ETS (Exponential Smoothing), and Holt Winter. Table 1 shows the error result of above forecasting base and ARIMA model. Error Metrics namely RMSE, MAE, MAPE, ACF1 are reduced significantly in case of selected model. RMSE metrics for ARIMA Model

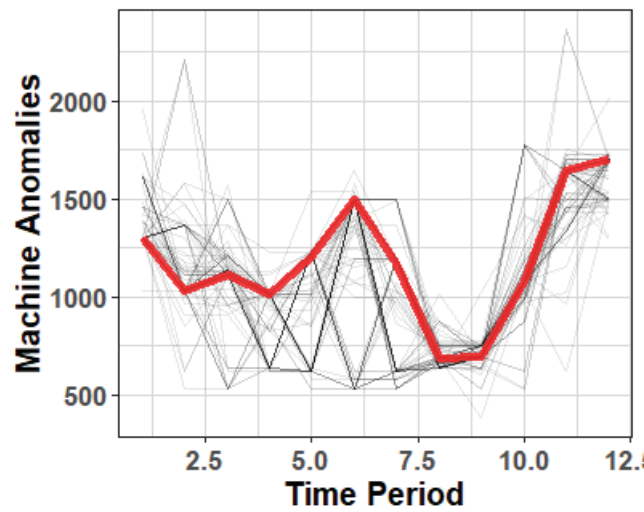


FIGURE 4. Modelling the Trend of IoT Sensor Data

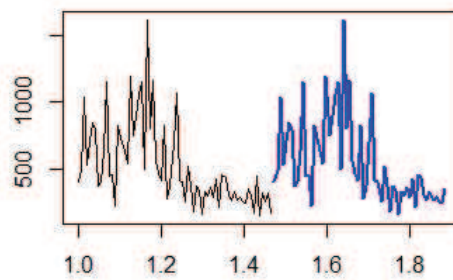


FIGURE 5. Modelling the Trend of IoT Sensor Data

is 0.237 which is 9.12%, 20.48%, and 17.23% less as compared to NNAR, ETS, and Holt Winter model respectively. MAE, MAPE, and ACF1 value calculated for the ARIMA is 0.184, 5.178, 0.359 respectively. All the error metrics are reduced significantly, hence improving the prediction accuracy.

5. CONCLUSION

The manuscript presents the forecasting model for anomalies detection in IoT devices. The data is collected for 28 sensors deployed at various locations in industry. The available forecasting model (auto-arima, Neural Network Autoregression, Exponential Smoothing, and Holt Winter) are applied and the comparison of trend depicted is drawn. Here on the comparison, it can be safely

TABLE 2. Errors Metrics of Forecasting Models

	RMSE	MAE	MAPE	ACF1
ARIMA	0.237	0.184	5.178	0.359
NNAR	0.260	0.206	5.829	0.410
ETS	0.298	0.25	6.822	0.495
Holt Winter	0.286	0.242	6.640	0.483

concluded that the ARIMA model has better prediction trend and less errors. Considering the RMSE for instance, it is 0.237 for ARIMA model which is 9.12%, 20.48%, and 17.23% less as compared to Neural Network Autoregression, Exponential Smoothing, and Holt Winter model respectively. In future, this work can be extended by analysis result after including environmental factors like temperature, electricity fluctuations, etc.

REFERENCES

- [1] M. HASAN, M.M. ISLAM, M.I.I. ZARIF, M.M.A. HASHEM: *Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches*, IEEE Internet of Things Journal, **7** 2019, 100059.
- [2] A. COOK, G. MISIRLI, Z. FAN: *Anomaly detection for IoT time-series data: A survey*, IEEE Internet of Things Journal, **7**(7) (2020), 6481–6494.
- [3] S.A. ALJAWARNEH, R. VANGIPURAM: *GARUDA: Gaussian dissimilarity measure for feature representation and anomaly detection in Internet of things*, The Journal of Supercomputing. Springer, **76** (2020), 4376—4413.
- [4] I. ALRASHDI, A. ALQAZAZ, E. ALOUFI, R. ALHARTHI, M. ZOHDY, H. MING: *Ad-iot: Anomaly detection of iot cyberattacks in smart city using machine learning*, IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), (2019), 0305–0310), doi: 10.1109/CCWC.2019.8666450.
- [5] P.J. BROCKWELL, R.A. DAVIS: *Introduction to time series and forecasting*, Springer, 2016.
- [6] V. ZARNOWITZ, A. OZYILDIRIM: *Time series decomposition and measurement of business, trends and growth cycles*, Journal of Monetary Economics, **53**(7) (2006), 1717–1739.
- [7] S. VERMA, N. SHARMA: *Statistical Models for Predicting Swine Flu Incidences in India*, IEEE First International Conference on Secure Cyber Computing and Communication (ICSCCC), 134–138, (2018), doi: 10.1109/ICSCCC.2018.8703300.
- [8] Y. ZHANG, T. WANG, K. LIU, Y. XIA, Y. LU, Q. JING, J. LU: *Developing a time series predictive model for dengue in Zhongshan, China based on weather and Guangzhou*

- dengue surveillance data. PLoS neglected tropical diseases.* **10**(2) (2016), e0004473. doi:10.1371/journal.pntd.0004473
- [9] <https://www.kaggle.com/>
- [10] H. WANG, C.W. TIAN, W.M. WANG, X.M. LUO: *Time-series analysis of tuberculosis from 2005 to 2017 in China*, *Epidemiology & Infection*, **146**(8) (2018), 935–939.
- [11] J.F. DE OLIVEIRA, T.B. LUDERMIR: *A hybrid evolutionary decomposition system for time series forecasting*. *Neurocomputing*, **180** (2018), 27–34.
- [12] L. LYU, J. JIN, S. RAJASEGARAR, X. HE, M. PALANISWAMI: *Fog-empowered anomaly detection in IoT using hyperellipsoidal clustering*, *IEEE Internet of Things Journal*, **4**(5) (2017), 1174–1184. doi: 10.1109/JIOT.2017.2709942.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
 RAYAT BAHRA UNIVERSITY
 MOHALI, INDIA
Email address: gauravsogi@gmail.com

INSTITUTE OF ENGINEERING AND TECHNOLOGY
 CHITKARA UNIVERSITY
 PUNJAB, INDIA.
Email address: sudesh.mittal@chitkara.edu.in