IMPLEMENTATION OF DIGITALIZATION IN FOOD INDUSTRY

SEN, K. O.; DURAKBASA, N.; ERDOL, H.; BERBER, T.;BAS, G. & SEVIK, U.

Abstract: *Due to the increase in world population along with changing customer demands and market dynamics, new challenges are arising for the food production industry. Novel approaches based on digitalization, smart manufacturing and the cognitive factory that are announced for transition to Industry 4.0., are compatibly required in the food production to meet the present and future needs effectively and reliably. Therefore, in this study, we modeled and implemented an integrated system of wireless sensor network (WSN), cloud and big data in the scope of future food production process, which will collect, transform and process the sensor data of the production line of a real food production facility. This approach targets higher efficiency, productivity, quality in addition to improving asset utilization by optimized resource management and waste. Furthermore, the addition of statistical methodologies and presenting the results via a web-based real-time data display applications will offer a high level of user interaction and monitoring in digital smart manufacturing that provides flexibility, efficiency and quality of production process.*

Key words: *Digitalization, statistics, IoT, cloud computing, WSN*

Authors´ data: MSc. **Sen,** K[amil] O[ncu]*, **; Ao.Univ.Prof. Prof.h.c. Dipl.-Ing. Dr.techn. Dr.h.c **Durakbasa,** N[uman] M[uhammet]*; MSc. **Erdol,** H[akan]***; Asst.Prof. **Berber,** T[olga]***; PhD. **Bas,** G[okcen]*; Asst.Prof. **Sevik,** U[gur]***, *University of Technology, Karlsplatz 13, 1040, Vienna, Austria, ** The Scientific and Technological Research Council of Turkey (TUBITAK), Tunus cad. No: 80, 06100, Ankara, TURKEY, *** Karadeniz Technical University, Trabzon, Turkey, e0626662@student.tuwien.ac.at, numan.durakbasa@tuwien.ac.at, hakanerdol@ktu.edu.tr, tolga.berber@fen.ktu.edu.tr, goekcen.bas@tuwien.ac.at, usevik@fen.ktu.edu.tr

1. Introduction

Food demand in correlation with population has increased all over the world over the last decades. To meet this increasing demand and to supply enough products to the market, producers require new and disruptive techniques and methodologies. There are several factors that affect food production and final products. Environmental factors and productivity have to be considered hand in hand using minimum resources with a result of minimum waste.

More efficient, cost-effective and less waste manufacturing has been demanded by production facilities throughout decades. Several academic studies and commercial solutions are provided in the literature to increase the efficiency and quality of production. One of the latest approaches to this problem is the "Industry 4.0". Zhou et. al. (Zhou, Liu, & Zhou, 2015) published in 2011 as a study of the German government. It was named as fourth stage of industrialization after mechanization, electrification and information stage. This change has influenced not only the German industry, but also international industrial development. In recent years, this change has been discussed by the industry and academics all over Europe. Main focus lies on the improvement of today's production system and on increasing the awareness of digitalization (Smit, Kreutzer, Moeller, & Carlberg, 2016). The study provides some general information about digitalization. It defines the main steps of digitalization for every stage of the production. Once these steps are applied to the production line, machine-to-human and machine-to-machine communication infrastructure can be developed and this interaction model may change the production model in the industry while creating an "Intelligent manufacturing Systems".

New approaches, such as digitalization, smart manufacturing, and the cognitive factory, which are announced with transition to the Industry 4.0, affect not only efficiency and quality of production process but also the environmental management. In this framework, the factories need to remotely monitor the critical production locations in terms of environmental conditions. In this study, we worked on the implementation of digitalization in a confectionery factory. The factory has critical places on the production line that they have to monitor and keep under control. At the moment, the production line is controlled by human workers. However, our implementation of remote monitoring and control will bring the factory's technology level to a step closer as the target of European Union strategy for digitalization.

In our study, we researched for the best performance. The study of Lin et. al. (Lin, Liu, & Fang, 2007)suggested the ZigBee technology which is a wireless protocol that is used in the industry and it is believed that it provides the best performance in reliability, power profile, capability, flexibility and cost. Therefore, in our setup, we embedded the ZigBee system into the WSN system using the mesh topology.

To implement an adaptive intelligent system, a literature research was carried out. Bas et. al. (Bas, Stoev, & Durakbasa, 2015) worked on increasing product quality in the production line by integrating a high precision measurement approach. Yerofeyev et. al. (Yerofeyev, Ipatov, Markov, Potekhin, Sulerova, & Shkodyrev, 2015) worked on adaptive intelligent manufacturing control systems and the design of processes. In their study, they also worked on the problem of automating the lifecycle

of the product. Gregor et. al. (Gregor, Krajcovic, Hnát, & Hancinsky, 2015) referenced the cyber physical systems (CPS) and used it for providing machine-to-machine (M2M) connectivity, planning and adaptation of personal behavior according to ambient conditions, learning of new models and ways of behavior and interpreting information about the environment. It provides an effective release of products. We have used a similar approach for implementation of the digitalization steps to the real production system. The results of the implementation can be used for M2M communication for the transformation to the smart production system. The system can provide a self-reaction for the changing environmental conditions and it can be more efficient and faster than traditional management systems.

For decision making algorithms from computer science such as the one developed by Miroslav et. al. (Miroslav, Milan, & Tomás, 2015) in the productive manufacturing has been studied. Intelligent systems can contribute to significant reduction of time necessary to design and verify the manufacturing system within the Digital Factory concept. In our study, statistical prediction models were used since we need to work on time series analysis. According to experiments, our prediction models obtained 99% of accuracy rate. Initial models are generated by using R statistical software (https://www.r-project.org/, 2017). The statistical models are used not only for the sensors data but also for the nodes. These results of the study contributed to make production planning and control more accurate.

Other studies that have been considered included high-end communication and data transfer for monitoring the production of factories. Miroslav et. al. (Miroslav, Milan, & Tomás, 2015) were created experiments on the mobile technologies with real time monitoring and control for manufacturing facilities. Their work shows that the mobile monitoring and control system is not suitable for fast real systems on GSM GPRE/EDGE networks. In mobile networks UMTS and especially LTE can be used monitoring and control systems faster. So we try to minimize the disadvantage of the mobile monitoring by releasing a report of information by means of a webpage. This reporting system is accessible not only from the mobile phone but also from the desktop/laptop/Tablet. Osanna et. al. (Osanna, Durakbasa, Si, & Afjehi-Sadat, 2001) worked on quality improvement models designed for intelligent manufacturing, which includes metrology methodologies.

They offered multi-functions integrated factory phenomenon which is an innovative concept and a new model. The model was proposed for similar facilities to switch to a cost-effective and customer-driven manufacturing. Industry 4.0 represents the digital transformation of production systems with advances in information and communication technologies. Digitalization and the intelligent system can help to achieve lower manufacturing costs and at the same time higher quality, accuracy and efficiency in present production. This approach was used in our work for the improvement of the TQM model. Total quality management (TQM) is consisting of process management, process improvement, resource management, process planning (Morath & Doluschitz, 2009).

In our study, we worked on speeding up the manufacturing, with minimum resources, best quality and at lower prices with using the information of the environmental. There are various methodologies to enhance the efficiency of

production systems where quality management is one of the most used applications.

We have used WSNs, industrial internet of things (IIoT) and statistical prediction model techniques to increase the effectiveness, speed and reliability of quality management methodologies used for production systems.

To increase the efficiency and quality of production process, we have used digitalization and industry 4.0 components such as WSN, IoT, cyber security, big data, cloud, and statistical prediction models; and managed to create a system that processes the sensor data which is transformed and collected from the candy production line. The data is collected by the sensor nodes that are portable, smart and ready to make an M2M communication. The collected data sent to the cloud to be stored, processed and evaluated. The data is processed into sensible information by a central mechanism, which uses known statistical methodologies such as auto regressive models and box-cox normalization.

To benefit from this information, we have published the real time data via a web-based portal, which is designed as user friendly as possible. As a sample, the result of environmental condition of Node 4 (packaging) given at the discussion part. Results show that duration of optimal production environment is only 4% of total monitoring period. Moreover, we found that environmental condition of packaging division is usually above the expected conditions. In this study, the smart production steps have been done on the real candy factory and running the production line. Auto-regressive models were chosen for each sensor and nodes to make predictions about future information predictions. Since each node can contain multiple sensors, we applied statistical normalization methods to generate information about overall node status and compared performance of the methods in this study. As a future work, we will study on new prediction models to compare the results of the performance with auto regressive models.

This document is organized as follows: After the Introduction section and related studies, methodology section is provided. Conclusion section is given at the end.

2. Methodology

The increase in global food demand affected the production rate of factories. Manufacturing systems became more complicated over time to meet the increasing production rate. On the other hand, environmental factors have always been a problem that affects the food production quality and stability. To solve such problems, new concepts and solutions are added to the processes.

Digitalization of the food production is one of the assisting solutions for the problems discussed above. In this study, we have used digitalization components to detect and predict the environmental problems of actual confectionery factories in the critical point on the production line (raw candy mixer), storage and packaging units based on environmental factors. The selected locations cover only 20% of the production line, so additional work is required to model whole production process.

The components that are used are WSN, cyber-security, big data, cloud, statistical modeling and web-based monitoring to assist in solving or early diagnosis of the environmental problems. Components of industry 4.0 such as robotics, 3D

printers, etc. are excluded from this work; hence we only focused on monitorization of a food production line. These include robotic and so on. Four critical locations mixer, wind tunnel, aging room and packaging from the production line which they are the parts of the automation system were chosen. This digitalization implementation steps are given below;

2.1 Internet of Things (Iot) and Wireless Sensor Network (WSN):
IoT and WSN are used for gathering the large amounts of environmental data from the sensor networks to use for monitoring and the prediction models. The Atmel based microcontroller (MCU) has been used for the development of the embedded system. Four nodes which consist of sensors and ZigBee model was developed on it. Atmel based embedded system may have various features depending on application. In this study, wireless communication was required since there were moving parts in the setup. The ZigBee module is chosen for this study because of its communication protocol developed for industrial IoT (Communication protocol). Five nodes, including both its hardware and software, are designed and integrated with the sensors within the ZigBee module. Four nodes collect data from the environment and the node 1(coordinator) directs the data to the cloud to monitor and process in real time. Sensor nodes and coordinator communicates with each other via ZigBee communication protocol which uses IEEE 802.15.4 standard. There are several reasons to choose ZigBee and mesh networking for this project and some of them are;

- Carry data at far distances compared to the same propagation power
- Reliable link (High SNR-BER performance)
- Low power requirement (In terms of high range)
- Mesh network support (Alternative link support from node to sink)
  - multiple data paths
  - multi-hop,
  - self-configuration and self-healing
- Easily accessible network authentication (Joining network is relatively fast)

Because of the noise generated by machines and closed area of the production line, small data transmission delays between sensors are detected. We solve this issue by using a new node board capable to cache previous data. More accurate results could be obtained by using wired structure or open space instead of WSN. ZigBee protocol uses direct Sequence Spread Spectrum encoding. With this coding, 4-bit message data is encoded to 32 bits chip value. This encoding and decoding algorithm causes gain of about 9 dB signal-to-noise ratios while reducing the data rate. However, sensors data output rate per node is approximately 15 bytes per second and ZigBee can support up to 250 kbps data rate. Thus, XBEE pro transceiver modules are chosen for this project.

Low data rate is not a bottleneck in this study because the data received from the sensor used in the plant is very low compared to this data rate. XBee Pro S2C (63 mW), which communicates on 2400 MHz – 2483.5 MHz Industrial Scientific Medical ISM band identified by IEEE 802.15.4 standard (IEEE, 2016). The ISM band is the band allowed to use without a license. This band can communicate with RF signals without paying any fee. IEEE 802.15.4 PHY frequency definitions are listed in Tab. 1 below:

Frequency band identifier	Frequency band (MHz)	Band designation
0	169.400–169.475	169 MHz
1	433.050–434.790	433 MHz
2	450–470	450 MHz
3	470–510	470 MHz
4	779–787	780 MHz
5	863–876	863 MHz
6	896–901	896 MHz
7	901–902	901 MHz
8	902–928	915 MHz
9	915–921	918 MHz
10	917–923.5	917 MHz
11	928–960	928 MHz
12	1427–1518	1427 MHz
13	2400–2483.5	2450 MHz

Tab. 1. IEEE 802.15.4 PHY frequency definitions.

For the microcontroller units, Atmel 32-bit and 16-bit are used for design of the 5 nodes to data gathering. Four of which collect data from the environment and 1 (coordinator) directs the data to the cloud. XBee Pro information indicates that the range of communication is almost 1200 meters in line of sight outdoor and drops down to 60 meters indoor. That's why mesh network topology is used to communicate the nodes. The cooperative communication enables us to reach higher ranges, with the help of mesh network topology, allowing the sensor to increase its range by using other sensors closed to the coordinator node as the distance of the sensor does not reach the coordinator node. Designed network structure is consisted of 1 coordinator module and 4 router modules. The leaf type node structure is not used instead of router nodes for collecting data from sensors because router nodes support the cooperative communication. So, this configuration extends network range. Thus, data can be transmitted to the coordinator by passing over other nodes and this reduces the energy consumption resulting a higher duration. The designed structure is given at Fig. 1:
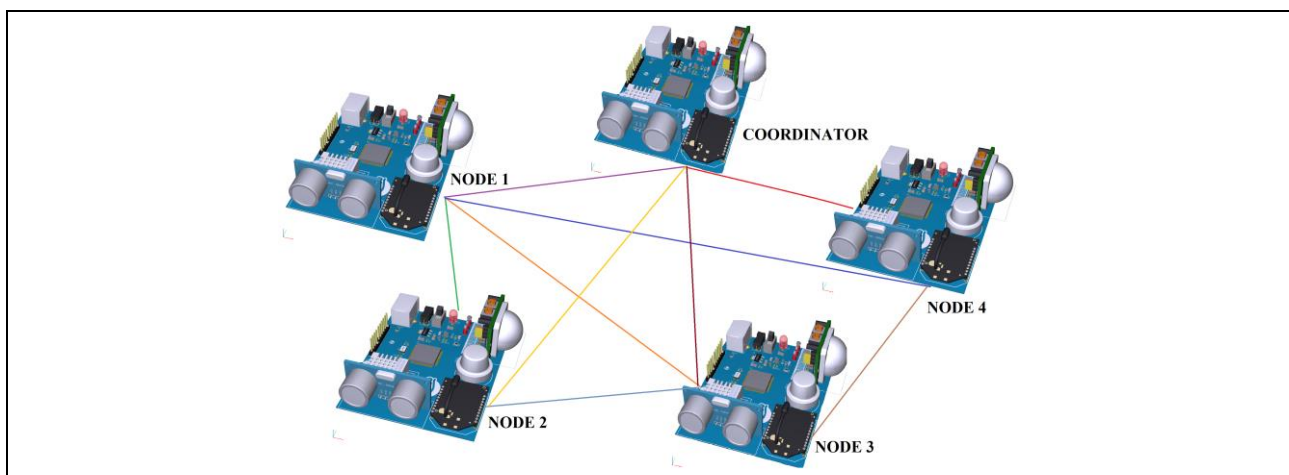

Fig. 1. Node's Mesh Network topology.

Sensor types according to nodes are given in Tab. 2.

Node 1	Node 2	Node 3	Node 4
Pressure	Fire	Fire	Fire
Humidity	Distance	Humidity	Humidity
Temperature	Light	Temperature	Temperature
	Noise		Movement
	Product Temperature		Light
	Mixer Temperature		Noise

Tab. 2. Sensors of Nodes.

All the nodes are designed and created with the same principles and methodologies. However, exact requirements and solutions for specific locations in the production line differentiate the modules integrated into each individual node. As an example, Node 4 has flame sensor, noise sensor and mainboard (on board sensors: motion, light, humidity and temperature) components. The structure of Node 4 is given in Fig. 2 and the communication structure of node 4 is given in Fig. 3 as well as the real application and image of node 4 in Fig. 4 respectively.
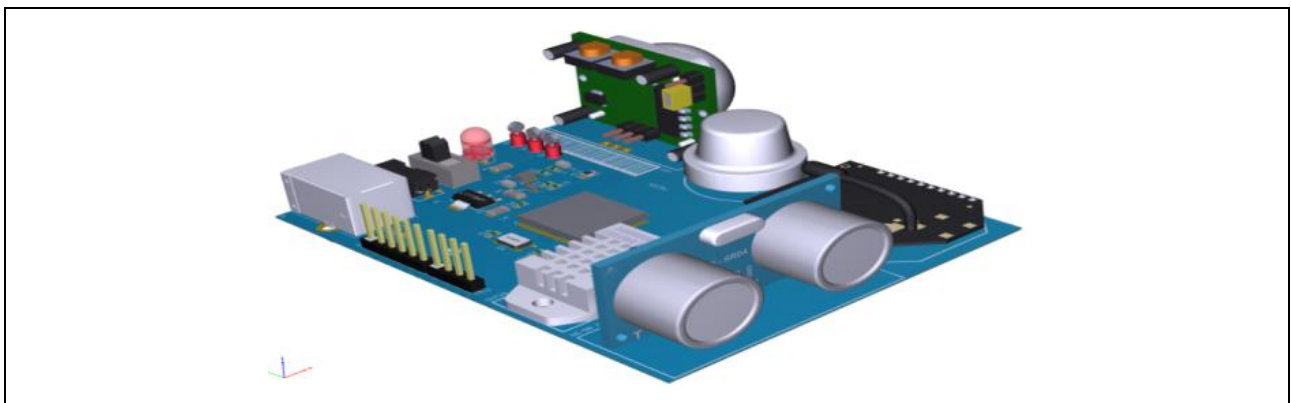

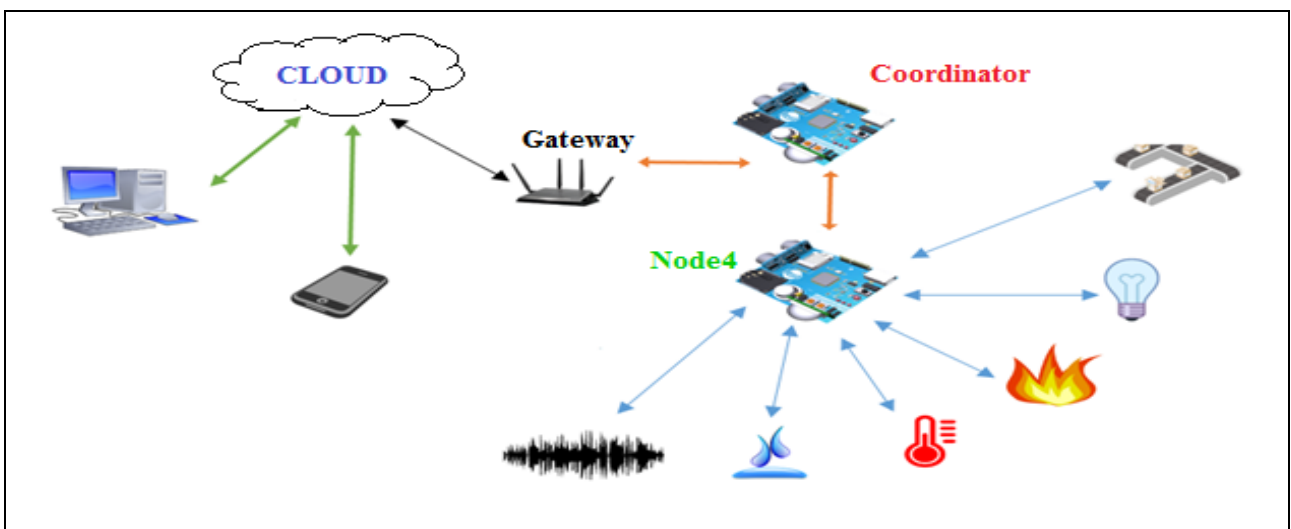
Fig. 2. Node 4's the mainboard.

Fig. 3. Communication structure for node 4.

Fig. 4. Node 4 settled on the production line.

2.2 Cyber Security

All collected data are encrypted in the node by the ZigBee module and forwarded to the coordinator. Coordinator uses 128-bit hardware level encryption algorithm for the security of the collected data. After encryption, the data is send to the cloud. Thus, the confidentiality of the collected data is protected against attack from both inside and outside.

2.3 Cloud and Web-Based Monitoring

In this study, the Google Cloud Platform has been used Infrastructure as a service (IaaS). Google Cloud provides mobility, storage and analyze of data collected via IoT. This data can then be systematically analyzed via statistical prediction models that are located in the cloud. For this target, 2.869.379 data values which are gathered from the sensors representing the status and condition of each node are collected and stored in the cloud as big data. This big data is used by the statistical model for the prediction and generation of sensible information.

The web-based monitoring system is used for displaying instant status, historical status information and future estimation information. Data are received from the cloud with equal intervals and reported periodically. The information is published via a webpage and it can be accessible not only from the mobile phone but also from the desktop. Thus, the accessibility problems are minimized (Miroslav, Milan, & Tomás, 2015). Status information consists of three different kinds of data which are gathered

from 17 sensors of 4 major nodes in the system. Three different groups of data are shown on the monitoring system. The first group has data representing raw (not processed) sensor status information, the second group has data containing meaningful (processed data) sensor status information and the third group has data that has general status of the nodes. The pages provide general status information about last 10 seconds of the system as well as the nodes' meaningful (processed data) status information. This meaningful status information from the nodes shows the trend of the system.

The real-time monitoring system displays the status of each individual node in separate pages. Every page has different number of components which reflects the sensors embedded in the node. As an example, "Node 4" (Packaging) page displays the data of 6 sensors and the general situation information and the page is visualized in Fig. 5 below.
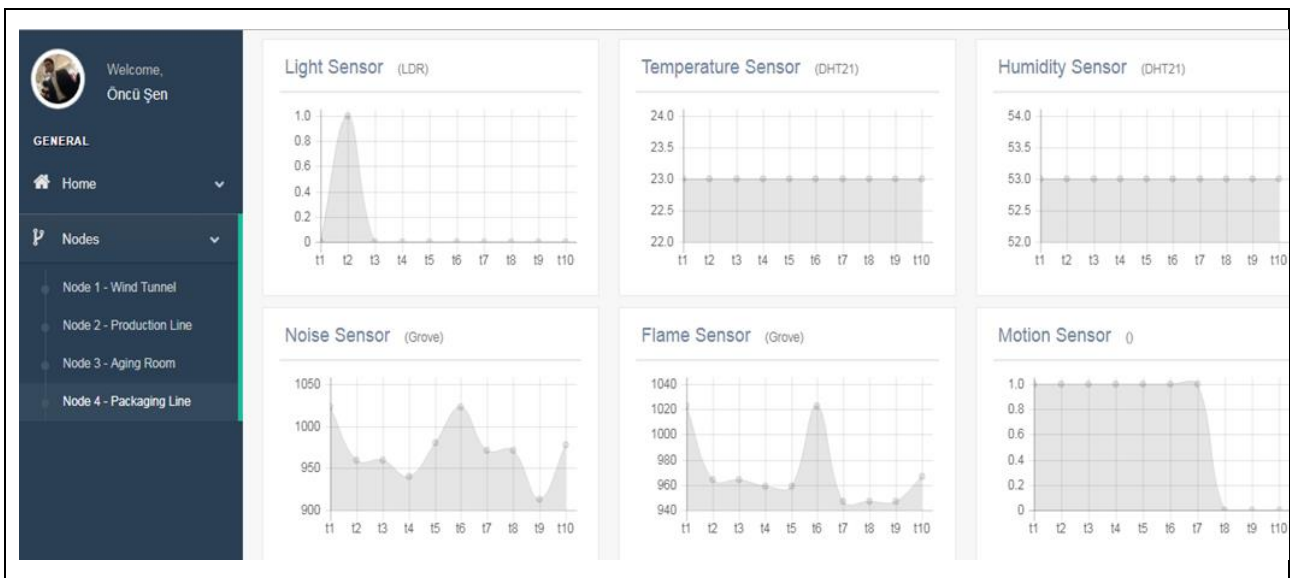

Fig. 5. Node 4 Web-based demonstration page.

2.4 Statistical Prediction Modeling

All collected sensor data systematically analyzing via statistical prediction models. The simple linear regression model is used for statistical prediction model which implemented to each sensor data to make condition prediction and to make sensible information of the sensor.

Our models are statistically meaningful according to the p-values of F-test ($p<0.05$). Depending on R-squared values of 15 models are above 90% that means they explain 90% of the variation of the following values which is given below Tab. 3. So, we may say that the models can be used for making predictions.
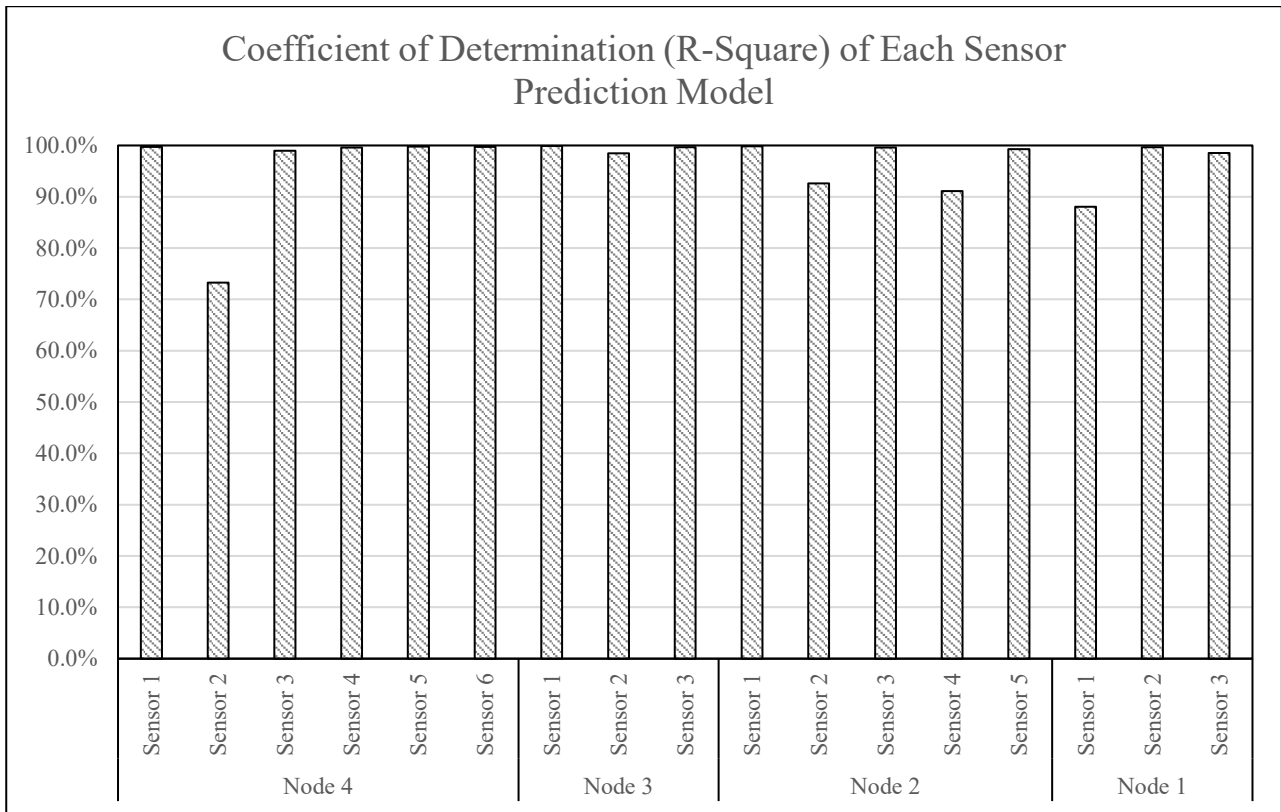
Tab. 3. Model's sensor performances.

Each node of our system contains multiple sensors and each sensor has its own measurement scales. Hence, it is hard to measure overall environmental status of a node. We applied statistical normalization and weighting to infer environmental information about nodes.

First, we need determined the weights of the sensor for each node to summarize node status. However, there is no standardized weights exists for candy production line environment. Environmental status of the production line is also considered as a trade secret for a factory. So, we used sensor weights for each node determined by factory experts. Later, we aim to determine effects of each sensor on final product quality. Although we obtained sensor weights for each node, we are still unable to use them because of sensor scale differences. Thus, we applied statistical normalization methods to make sensor measurements have same statistical distribution, which is chosen standard normal distribution (X~N(0,1)). We also measured effects of two different normalization approaches. First, we convert raw data to standard normal distribution as follows;

$$N_X = \frac{X - \mu_X}{\sigma_X}$$

where N_X is normalized data, X is raw data, μ_X is mean of X and σ_X is standard deviation of X. This conversion does not change distribution shape of original data, and ensures that mean and standard deviation of the data becomes 0 and 1 respectively. We also use power transformation (Box, G. E. P.; Cox, D. R., 1964) to normalize data. Power transformation is a monotonic transformation for raw data using power

functions. This normalization method stabilizes the standard deviation and aims to change the distribution of data to a normal-like distribution. Power transform defined as follows;

$$N_X^{(\lambda)} = \begin{cases} \dfrac{X^\lambda - 1}{\lambda(GM(X))^{\lambda-1}}, & \lambda \neq 0 \\ GM(X)\ln(X), & \lambda = 0 \end{cases}$$

and

$$GM(X) = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

where λ is the normalization parameter. However, λ parameter of the power transforms needs to be determined for each sensor data. We used maximum-likelihood approach to determine λ parameter. As a result, we have multiple normal-like distributions of sensor data. However, parameters of each sensor data distribution after power transform are different. To solve this issue, we applied standard normalization to power transformed values to equalize distribution parameters of each sensor data.

When we achieved equality of distribution of each sensor data by applying normalization approaches, we calculate overall node status values based on following formula;

$$S_{Node} = \sum_{i=0}^{n} w_i N_{X,i}$$

where S_{Node} is the status value of node, n is the sensor count of the node, w_i is the weight of sensor and $N_{X,i}$ is the normalized value of the sensor. Additionally, we calculate optimal range for S_{Node} using range values for each sensor.

As an example, raw expected values and normalized expected values for node 4 is given below Tab. 4:

Sensor	EV_{min}	EV_{max}	Standard Normalization		Power Transform Normalization	
			EV_{min}	EV_{max}	EV_{min}	EV_{max}
Humidity	30%	40%	-0.3108	-0.1506	-2.5948	-1.1307
Temperature	20°	25°	-0.1560	0.1407	-1.3386	1.1952

Tab. 4. Optimal Values for Sensors of Node 4.

After applying scale formula (S_{Node}), we determine optimal levels for Node 4. Results of our experiments are given in Tab. 5 and depicted in Fig. 6.

Parameter	Value	
Total Number of Observations	2539569	
	Standard Normalization	**Power Transform Normalization**
Node Status Lower Threshold	-0.2953	-2.4692
Node Status Upper Threshold	-0.1215	-0.8981
Number of Observations Below Lower Threshold	9453 (0.37%)	10082 (0.40%)
Number of Observations Above Upper Threshold	2442388 (96.17 %)	2441884 (96.15 %)
Total Percentage of Unexpected Condition Duration	96.55 %	96.55 %

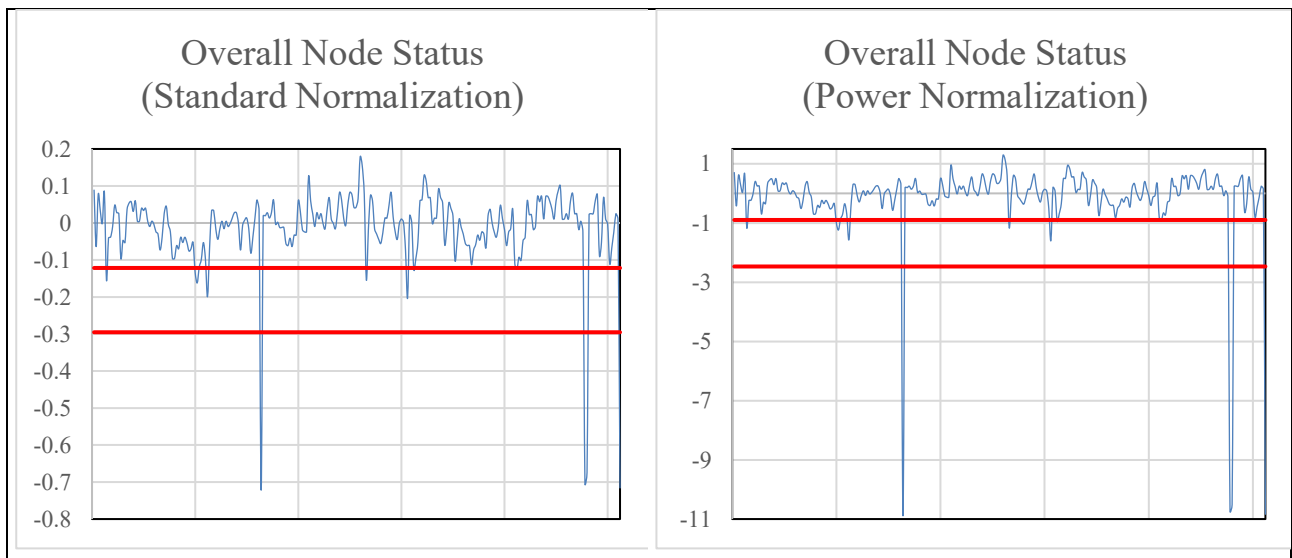Tab. 5. Results of Node 4 Environmental Status.

Fig. 6.   Overall environmental status of Node4.

According to the results, we could say that environmental status packaging section of the production line remains above expected conditions during the observation period. This situation could affect final product quality. Moreover, results of two normalization approaches produces nearly same results. However, more sensors need to be measured to select best approach for environmental scale.

The statistical prediction systems allow the company's production line units can able to make machine to human and machine to machine communication. Also, the substantive information is presented in a visual form to the company's authorities as monitoring. This study worked on that approach with using digitalization components and also high quality food products with minimum resources and waste.

3. Conclusion

Minimizing resource requirements and waste is a major goal of food production to meet increasing demand. However, final product quality is affected by several

environmental factors of the production line. In this study, we develop a new environmental monitorization approach to assess environmental status of a food production facility. Moreover, our approach could also be used to take necessary precautions for ensuring final product quality. For this purpose, today's technologies such as WSN's, IIoT and statistical failure prediction models were implemented to a real confectionery factory as digitalization. The WSN based 4 intelligent nodes had been placed on the critical locations of the production line. A coordinator node had also been placed to collect the data from the nodes and send them to cloud. The data was gathered in a cloud-based data store with the aim of archiving and processing. This processing was a type of analyzes which uses statistical methodologies to assess current and future environmental status of nodes. At the end of analyze, information was generated for monitoring the environmental conditions as well as providing data for M2M communication.

We proposed a new statistical environmental scale for combining multiple sensor data in one value to overcome difficulty of making decisions based on multiple information. Results of our scale were verified by factory experts. Our scale is also implemented by using R Statistical Software to ensure cloud compatibility. Then, overall node status scale results are also obtained using our implementation. Obtained scale and limits are presented to experts via web-based monitoring system which is also developed in the context of this work. With this approach, we have managed to digitalize the production process for the environmental condition and put apply it in a confectionery factory. At the end, real life data, prediction data and overall node status information have been provided to the use of decision makers as a business intelligence system. The system can be advised for facilities that has similar setup and production facilities.

As a result, we have developed assistant technologies to ensure suitability of production environment for final product quality. In future, performance of our approach will be evaluated on whole production line using different statistical models to choose best models for each part of the production line.

4. Acknowledgements

5. References

Bas, G., Stoev, L., & Durakbasa, N. M. (2015). Assessment of the Production Quality in Machining by Integrating a System of High Precision Measurement. Procedia Engineering 100 (2015), 1616 – 1624.

Box, G. E. P.; Cox, D. R. (1964). An Analysis of Transformations. Journal of the Royal Statistical Society, Series B. 26 (2): 211–252.

Gregor, M., Krajcovic, M., Hnát, J., & Hancinsky, V. (2015). Genetic Algorthims in the Design and Planning of Production System. Annals of DAAAM & Proceedings, 26, 494-500.

https://www.r-project.org/. (2017). R:The R Project for Statistical Computing. https://www.r-project.org/

IEEE. (2016, 4). IEEE Standard for Low-Rate Wireless Networks -Amendment 2: Ultra-Low Power Physical Layer. IEEE Std 802.15.4q-2016 (Amendment to IEEE Std 802.15.4-2015 as amended by IEEE Std 802.15.4n-2016), 1-52.

Lin, S., Liu, J., & Fang, Y. (2007, 8). ZigBee Based Wireless Sensor Networks and Its Applications in Industrial. 2007 IEEE International Conference on Automation and Logistics, (s. 1979-1983).

Miroslav, M., Milan, A., & Tomás, K. (2015). The Mobile Monitoring and Control of Real System. Annals of DAAAM & Proceedings, 26, 104-108.

Morath, C., & Doluschitz, R. (2009, 1). Total Quality Management in the food industry – Current situation and potential in Germany. Internetional Information System for the Agricultural Science and Technoogy, 3.

Osanna, P. H., Durakbasa, M. N., Si, L., & Afjehi-Sadat, A. (2001). Quality Management and Metrology in Intelligent Manufacturing. IFAC Proceedings Volumes, 34, 129-134.

Smit, J., Kreutzer, S., Moeller, C., & Carlberg, M. (2016). Industry 4.0. F. Gouarderes, Policy Department A: Economic and Scientific Policy (s. 1).

Yerofeyev, S. A., Ipatov, O. S., Markov, S. A., Potekhin, V. V., Sulerova, A. S., & Shkodyrev, V. P. (2015). Adaptive Intelligent Manufacturing Control Systems. Annals of DAAAM & Proceedings, 26, 1016-1024.

Zhou, K., Liu, T., & Zhou, L. (2015, 8). Industry 4.0: Towards future industrial opportunities and challenges. 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), (s. 2147-2152).