Advanced Process Control with Smart Smelter Sensors at EGA

Najeeba Aljabri[1], Edouard Mofor[2], Rawa Raheem[3], Rajwinder Kaur[4], Khursid Ahemad Sayed[5], Asmaa Ali Salem Ibrahim ALHosani[6], Abdallah A. Abdallah Rahbar[7] and Vinod Nair[8]
1. Vice President - Technical
2. Associate Manager - Process Efficiency Carbon
3. Associate Manager - Process Efficiency Potroom
4. Area Engineer - Process Control
5. Senior Technician – Process Efficiency Potroom
6. Engineer 1 – Process Efficiency Potroom
7. Senior Manager – Reduction
8. Lead Engineer - Reduction
Emirates Global Aluminum (EGA), Al Taweelah, Abu Dhabi, United Arab Emirates
Corresponding author: naljabri@ega.ae

Abstract

EGA's Midstream vision was to start operating the smelter in a more predictive than reactive mode. This was supported by financial analysis of potential gains of advanced process control with big data and Industry 4.0. In the first step, EGA successfully built a prediction model for anode spikes detection in 2020, using pot-measured data. Anode spikes are a critical anode problem in the smelters, which significantly decreases current efficiency and associated business benefits. It has always been a challenge to timely identify the spikes with manual checks as they have many causes, and delayed action rapidly increases pot performance. Big and reliable data is fundamental for artificial intelligence and machine learning (AI/ML) models. Thus, high frequency (minutes) pot control data have been the base for the spike prediction model. EGA has successfully developed the spike detection model for each pot technology and has been using the model since 2021. The models have brought the spike detection time from 2-3 days to less than a shift using bath temperature prediction as a soft sensor. The model includes more than 30 base parameters summarised each hour. Several iterative versions of the model were developed to reach the desired accuracy and continuous use of the model. The model has a user-friendly dashboard displayed in the control room of every potline. In addition, an audio announcement and SMS to the relevant operator are also used to communicate the prediction and ensure rapid spike removal. The spike detection tool has proven that AI/ML is the way to unleash the smelter's hidden opportunities. Other prediction models are being developed to enhance smelter performance as part of EGA's journey to Industry 4.0. The focus is on enriching the data and ensuring advanced usage of big data for prediction and prescription systems.

Keywords: Industry 4.0 in EGA smelters, Big Data, Machine learning, time series, data science, soft sensor, Spike prediction model, Bath temperature

1. Introduction

EGA is the world's biggest ‘premium aluminium’ producer with business from mining and refinery to smelting and casting. EGA produces around 2.5 million tonnes of hot metal annually from its two production sites, Jebel Ali and Al Taweelah. Potlines in both production sites comprise seven technologies (D18+, CD20, D20, D20+, DX, DX+ and DX ultra) with 2843 pots. This paper presents EGA's journey to develop soft sensors to solve pressing operational issues like anode spike detection [1].

At EGA, all technologies have increased amperage to much higher than originally designed for, resulting in smaller anode-cathode distance which makes the pot operation susceptible to anode problems. Historical data indicate that anode spikes are the highest contributor to anode problems in EGA potlines [Figure 1b] for various reasons, one of which is high anode current density. Additionally, once the spike starts, it becomes a chain reaction which burdens the potline operation for removal of the spiked anode and increases anode consumption.

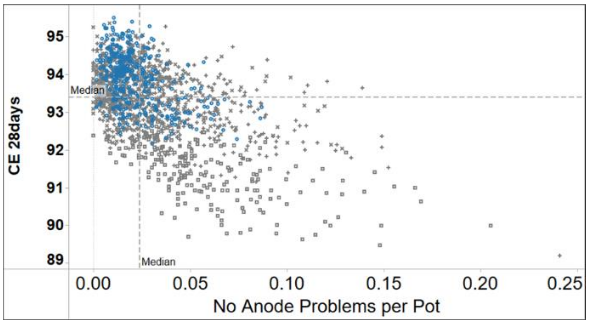

Figure 1a. Impact of anode spike on Current Efficiency [2].

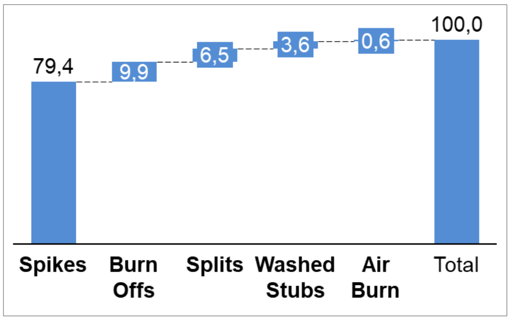

Figure 1b. Anode Problems 2019.

As illustrated in Figure 1 c, the loss of current efficiency is directly proportional to the time the spike spent in the pot. This shows the importance of identifying the abnormal condition earliest to minimise the current efficiency loss. This observation was further studied at EGA and it was found out that anode spikes affect alumina feeding in pots before spike removal. Change in Alumina dumps from baseline (Loss of current efficiency) is directly proportional to the time the spike spent in the pot. Early spike detection minimizes the current efficiency loss

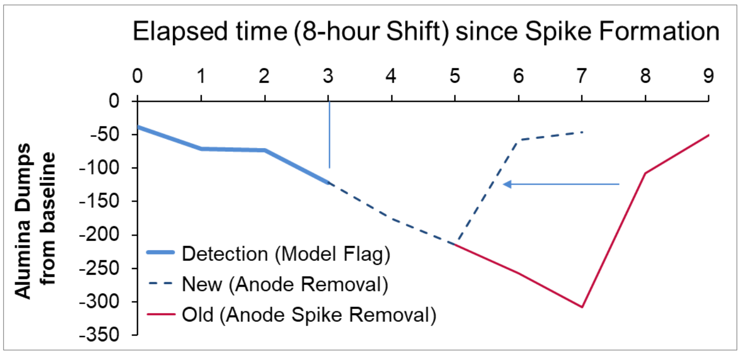

Figure 1c. Impact of early detection anode spike on alumina feed (proxy CE).

EGA has tested various traditional approaches to find spikes, such as high bath temperature and low alumina dumps, but these are delayed responses and depend on rigorous manual checks. Therefore, in 2020, EGA technical division decided to experiment with upcoming technologies of artificial intelligence/machine learning (AI/ML) and develop an innovative solution for spike detection. Furthermore, with the significant growth of data science and big data technologies, it has been possible to build advanced detection/prediction systems for proactive actions based on the data.

Although spike detection sounded like a value-adding business case that could be solved with AI/ML, another critical aspect for success is the availability of reliable data since it is the foundation of any successful data science project. Unfortunately, in potlines, data is available at the frequency of seconds and minutes and only kept for a few weeks for investigation in some potlines. Therefore, first, the parameters in Pot Control Systems were proactively stored for 12 months before the start of the project in parallel to the governance set-up, technology and vendor selection process.

2. Project Management Approach

Although Big Data and AI/ML technologies are being embraced worldwide, their use in the smelting industry, to say the least, in the Middle East was only beginning to grow. Therefore, the project was managed by a dedicated executive committee. Agile methodology was used as the execution and management approach to ensure high flexibility and iterative development in shorter sprints, considering the unknown nature of the project and the high investment involved. Anode spike model development started in Al Taweelah Potline 3, the longest potline in the world with 444 pots, by a dedicated core team of business, process and data science experts using a sprint-based agile delivery mechanism.

EGA's management opted to leverage the expertise of reputed solution providers in the field of data science, which were selected through a rigorous process blending typical Request for Quotes (RFQ) based on Request for Proposals, developed jointly by Operations and IT personnel.

3. Big Data Architecture Set-up

A preliminary short-listing of the proposals from the best vendors partnering with the leading cloud solutions industry was conducted. One of the selection criteria was open platform cloud technologies and the best fit data architecture in line with the best practice permitted by the regulation from the government.

4. Solution Framework Based on Process Understanding

It was imperative to develop a solution framework that is effective and sustainable, proactively solves the end user problem, and provides value to the end user. In addition, relevant data availability was necessary to understand the opportunities and limitations of developing the solution.

Approaches to detect anode spike with continuous individual anode current distribution (ACD) sensors could not be used, since only a few EGA potlines were equipped with such sensors. However, this shortcoming or challenge did not deter the project start-up. The project team believed that using lower frequency information from the programmable logic controllers (PLCs) for different events could unlock the mechanisms behind the formation of anode spikes in reduction pots.

In addition, developing the solution was aligned with the technical understanding of the pot behaviour. The subject matter experts, i.e., process engineers with deep knowledge, provided guidance about the spike formation in the pot. Specific process indicators start to deviate, e.g., noise, alumina dumps and pot resistance start to decrease due to short-circuiting between the anode spike and the metal, resulting in a thermal and chemical imbalance. The current passing through the spike is not producing metal which causes current efficiency loss and generates localised heating. Over a few shifts, it leads to the bath temperature increase. Bath temperature is measured only each 48 hours, which is too long to detect the temperature change; as a result, the identification of the spike is delayed. Therefore, it was vital to first develop a reliable bath temperature soft sensor at a higher frequency which can be fed as an input to the spike model.

Further development is explained in two steps:

- Bath temperature soft sensor development

- Spike model development.

4.1 Bath Temperature Soft Sensor

Bath temperature plays a critical role in understanding the thermal balance of the pot. Therefore, it is also known as the 'heart' of all pot measurements. Hence, it is necessary to predict the bath temperature at the lowest possible time interval. Current 48 h is too long for understanding the bath temperature trend. Therefore, a decision was made to develop an hourly soft sensor for bath temperatures.

4.1.1Data Engineering

An analytical dataset **(**ADS) was developed by including all the relevant data in one table before starting to investigate and visualise each of the variables to spot outliers and patterns, as a necessary stage called Exploratory Data Analysis (EDA). First, EDA was conducted using statistical tools to build concrete understanding and context of the data, which would help train the model. Then, data such as the Alarm and Events table, Cell-day table (summarising all the process data of a day), and Manual Operations history table were merged, considering the time-lapse between events. Finally, a complete ADS table featuring all relevant variables with lags mapped as per the process was achieved, ready for the bath temperature soft sensor training.

4.1.2 Modelling Approach

Multiple iterations, named versions, were developed to create bath temperature soft sensors, including V2 Random Forest, V4 stacking classification and V5 multiclassification. V5, a three classification-regression model pair, was used for the final model based on the results. The classification model involved different bath temperature ranges, which we call 'buckets': Low, Medium and High. Then, using the regression technique, these classified 'buckets' were used to estimate bath temperature values hourly.

Standard model performance metrics such as precision and recall, and confusion matrix were used to assess the model performance.

Precision: It quantifies the number of positive class predictions that belong to the positive class, Equation (1):

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{1}$$

Recall: It quantifies the number of positive class predictions made from all positive examples in the dataset, Equation (2):

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2}$$

Confusion Matrix: The confusion matrix provides more insight into not only the performance of a classification model but also the degree of correct/incorrect classification of each class along with the type of errors being made (Table 1)

Table 1. Model confusion matrix.

	Positive	**Negative**
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Pot 3A112 was selected for validation, and all the different models were compared for estimated vs measured bath temperature. The fifth version was the best-performing model and was chosen for live pot deployment. Initial validation was done in a few pots by measuring bath temperature at different times, and the results were satisfactory, (Figure 2)

Furthermore, we involved potroom teams in validating the bath temperature soft sensor and measuring bath temperature every 48 h. One of the metrics used to assess the model performance and share it with potroom employees was bath temperature 'Mean Absolute Error' (MAE). The MAE was 4.4 °C for the period from May 2020 to December 2020, after the deployment of the model in the potline. (Figure 3) shows the comparison of measured and predicted bath temperatures of a pot.

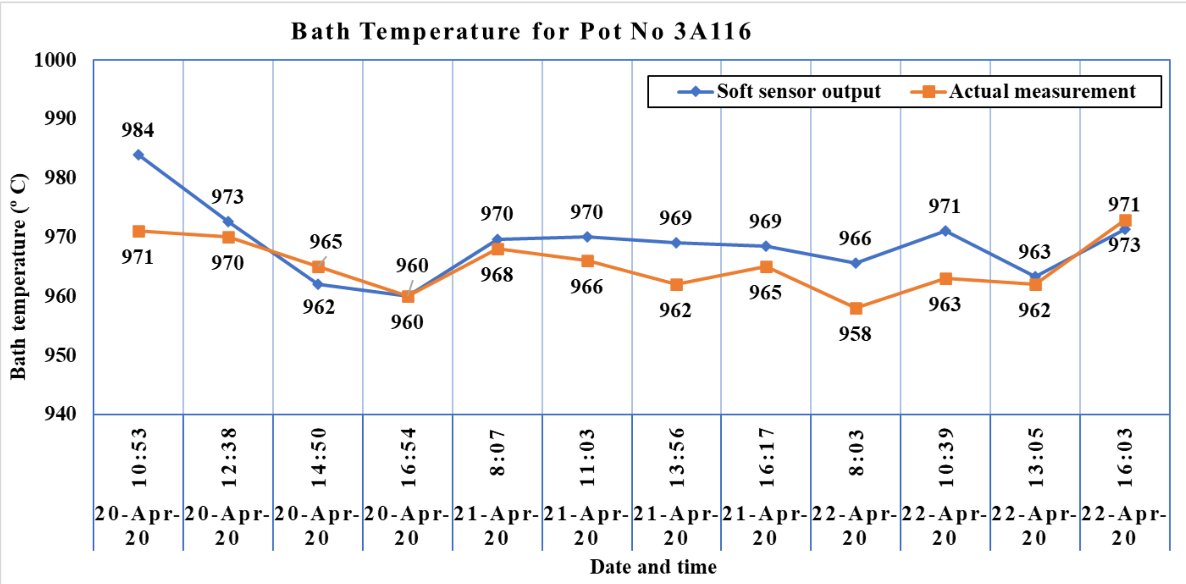

Figure 2. Validation of soft sensor prediction with measured bath temperature.

4.1.3 Bath Temperature Soft Sensor Utilisation

The bath temperature sensor was implemented in all pots in Jebel Ali and Al Taweelah sites, and here are the outcomes

- Bath temperature soft sensor was used as input for spike detection in Spike Detection Model.
- Soft sensor-based bath temperature prediction controls pot energy input and helps the Pot Control System by activating voltage adder (± adder) to maintain the right thermal balance. New program logic was developed on top of additional validations of the soft sensor accuracy to ensure high confidence before activating the adders.

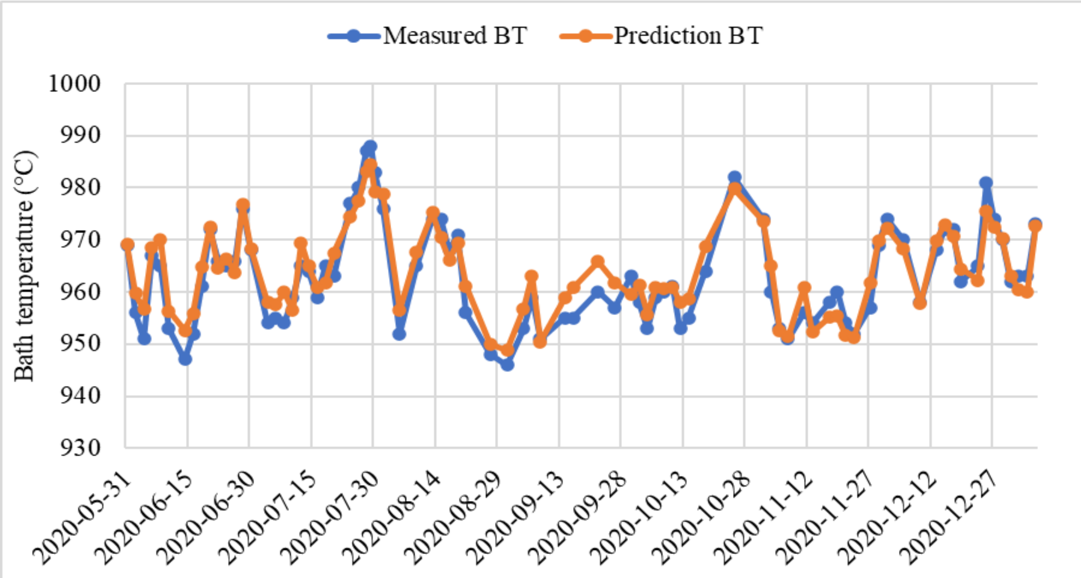

Figure 3. Bath Temperature (BT) predicted by the soft sensor and measured at 48 h intervals in a pot in EGA Al Taweelah Potline 3.

4.2 Spike Detection Model

The objective of the anode spike detection model is to predict the occurrence of an anode spike in a pot at shift level – bringing down the detection time of existing practice from 2-3 days to one shift. The anode spike detection model uses bath temperature soft sensor as one of its inputs.

Currently, the operation teams are using a heuristic approach for spike detection, which includes checking the actual values of the metrics like bath temperature (every 48 h), alumina dumps, etc. Then if the trends are out of the optimal operational ranges, the operator is only prompted to check the pot for spikes or any anode problem. In addition, each team relied on experts checking physically individual pots which is quite challenging for such a long potline. Therefore, the currently developed framework introduces a structured methodology, which analyses the characteristics of the metrics at several steps at a higher frequency and accordingly classifies a pot within a given production shift as anomalous or non-anomalous, thus streamlining the operation team's work on the detection of anode problems.

4.2.1 Model Development

Anode spike detection framework (Figure 4) followed a model development cycle in 7 steps:

- Step 1: Data stagging was created hourly with data for almost 1.5 years. The data was split into training, practical validation and testing groups.
- Step 2: Exploratory Data Analysis was conducted using statistical methods to select the features based on the technical understanding of individual metrics contributing to spikes.
- Step 3: Once the features were finalised from step 2, added columns were introduced to understand the metric disparity between anomalous and non-anomalous cases. This step is referred to as feature engineering. Apart from bath temperature, other critical metrics were analysed like alumina feeding, excess AlF_3, noise, the difference between smooth resistance and control resistance, bath resistance change, beam movements, etc. Additionally, a few random cases of spikes were deep-dived to understand the behaviour.

- Step 4: Time series analysis for each pot was mapped against the anode setting cycle, and several time series characteristics were analysed to identify the anomalous pattern of the spike metrics. This analysis was performed for multiple anode setting cycles together and for an individual anode setting cycle.
- Step 5: Post-time series analysis, the production shift metric behaviour was also analysed for spikes and typical cases to make the framework more consumable for actions required to be taken on shift basis.
- Step 6: Bayesian hyperparameter tuning was devised to determine the threshold values needed to balance the model framework's precision and recall. Step 7: Practical validation was the last but crucial step. It consisted of all model versions tested on the last three months data. It was needed to check the cases of overfitting that may have been developed while finalising the thresholds via Bayesian tuning. These validations were performed for the 57 test group pots using Randomised Control Trial (RCT) procedures. For example, Cathode type, collector bar type, lining Paste type & Paste supplier, Pot age, and Geographical location (Gas Treatment Plant, corner/middle pots) factors were uniformly distributed in the sample population. Pot selection was made as three tests vs three control to minimise the operational bias and ensure the pots received equal treatment from operators.

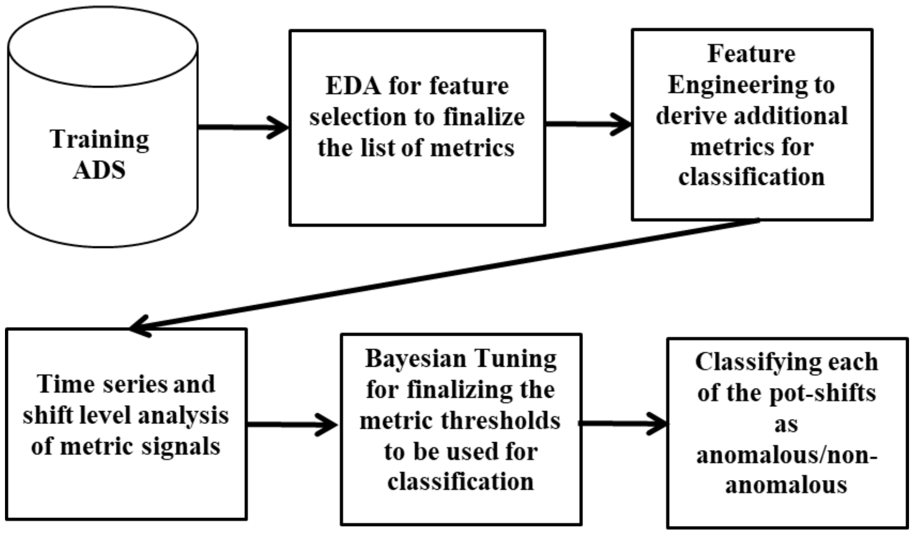

Figure 4. Flow of anode spike detection framework.

4.2.2 Model Validation

The idea was to simulate the historical data with the selected models and then check the alignment in the efforts and coverage numbers obtained during development. Effort and coverage were used as the key metric to measure the success of practical validation (Figure 5).

Effort per shift is computed as the ratio of the total effort (the number of times the operator has to check the pot for spikes) divided by the unique date-shifts available.
Coverage is the total of spike cases out of all the actual spike cases detected.

Anode Spike detection | Model validation for 11th Aug – 24th Aug,2020						
Model Version	**#Pots**	**Alarms Generated**	**Total Cases**	**Cases Predicted**	**Coverage**	**Effort Estimated**
v2.2	57	39	16	12	75 %	0.9

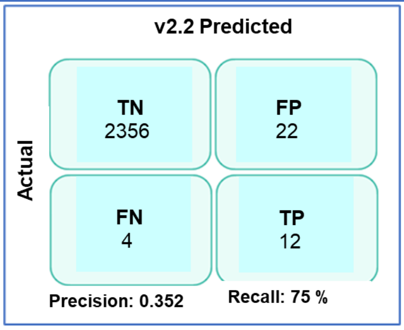

Figure 5. Model performance of version 2.2. FN = false negative, FP = false positive, TN = true negative, TP = true positive.

In data science theory, recall and precision are inversely proportional: In our case, we would like to have zero FN (predict all spikes), but this outcome usually increases FP, which means that we are checking more (all FP) anodes for suspected spikes, but do not find them on these. It is challenging to strike the right balance between FN and FP in the model. The questions to be answered in such a situation are: Have we found all spikes, or are we missing some? Should we accept more FP (if its cost is low) and get all TP (more net benefits)? The best compromise observed, as shown in Figure 5 was a recall of 75 % and a precision of 35 %.

4.2.3 Visualisation and Consumption Dashboard

Dashboards were developed as a web application and made available to the end users for faster response to the spike alerts. Deep dive pages were designed at a pot level to view the main metrics behaviour for scoring spikes. (Figure 6)

4.2.4 Implementation Journey

Successful deployment results in Al Taweelah Potline 3 were presented to management, and approval was obtained to implement the model for all technologies in EGA potlines. However, each technology had specific challenges with respect to availability of parameters, practices and severity of spikes. As a result, a model was developed specifically for each pot technology.

A few challenges listed below were encountered while using the time series model in other potline technologies, which resulted in exploring the opportunity of an alternative method (classification model). Firstly, historical data was insufficient for training the time series model due to conversion from 64 h to 32 h anode setting cycle and vice versa. Finally, variation in the time series behaviour across potlines affected the spike detection model developed at technology level. Considering the above factors and the vast availability of spike data for other technologies, the classification model seems to be the right fit.

Classification model adoption for other technologies helped get the final anode spike predictions at the 8-hour shift level. The model was dynamic and independent of the anode setting cycles; thus, handling any conversions in future will not affect the model. Also, such a model includes shorter development and experimentation time.

The essence of the alternate methodology of the classification model was to categorise spikes generated in a pot, which were classified according to their anomalous behaviour as Class 1-Less, Class 2-Medium and Class 3-High.

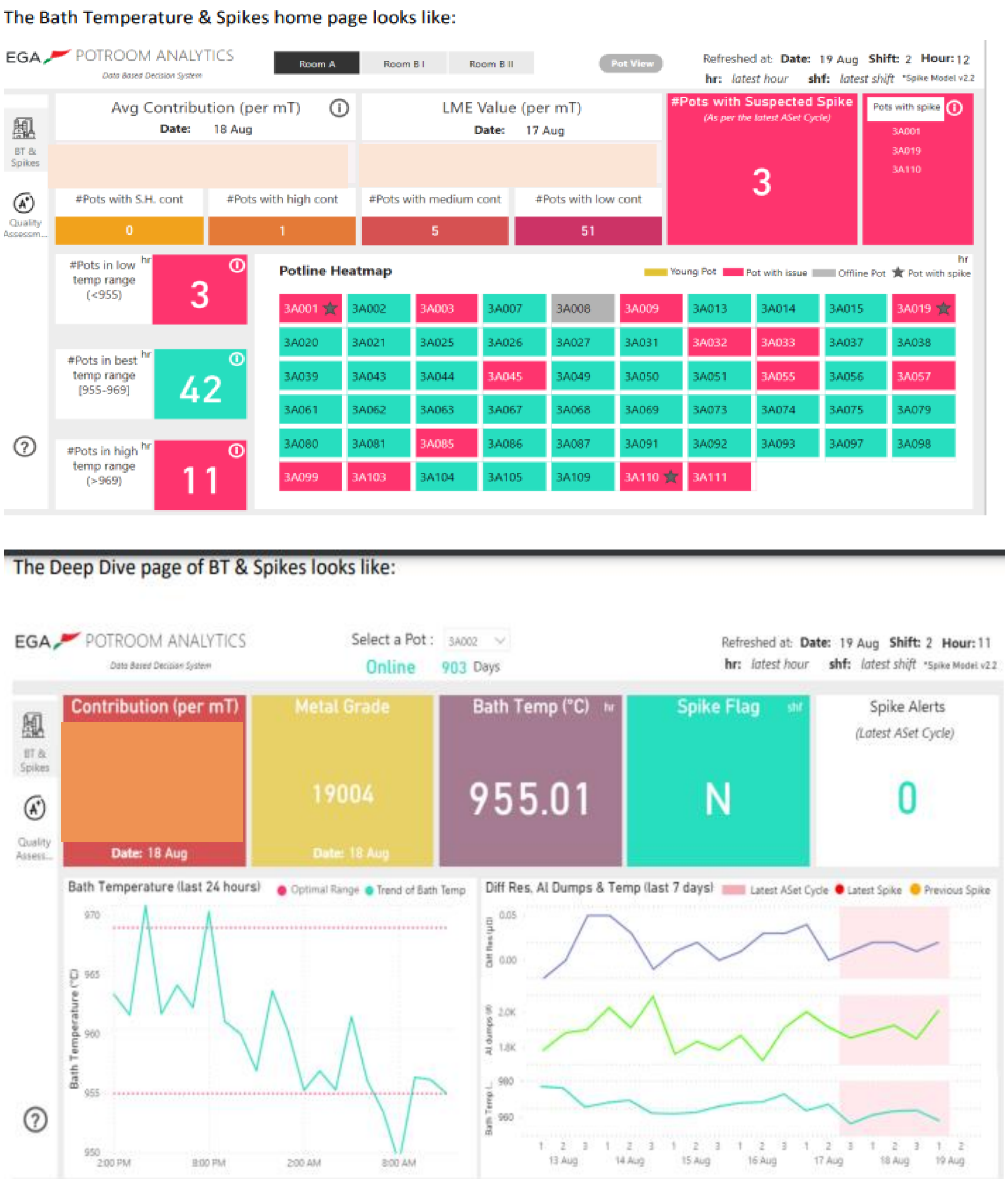

Figure 6. Consumption dashboard for soft sensors.

Identifying class 1 spikes was the highest priority as they act as standalone spikes. Class 1 spikes could form without showing extreme anomalous behaviour in the pot, making it difficult for the operators to identify them correctly. Thus, increasing the model coverage of Class 1 spikes plays key role; otherwise, there are chances that the Class 1 spike could be converted into a series of class 3 spikes.

Similar to the above, several other concepts were developed based on an in-depth process understanding of the spiking behaviour in the pot to improve spike detection. Figure 7 is an example of the final results of the deployed model in JA line 7 and line 9.

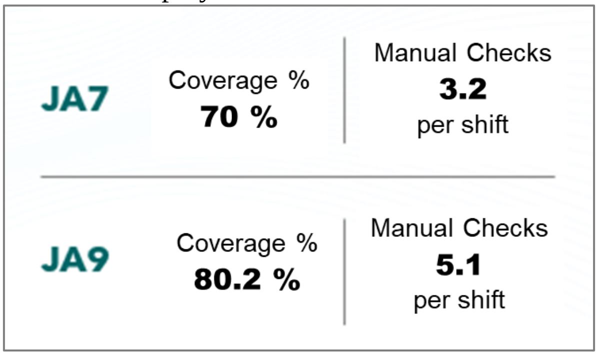

Figure 7. Performance of final deployed model. JA7 = Jebel Ali Potline 7, JA9 = Jebel Ali Potline 9.

Bath temperature soft sensor models and anode spike models were implemented and used in 1572 pots in Jebel Ali and 1020 pots in Al Taweelah. In addition, 30 % of the pots were kept for reference to establish financial benefits estimates and later deployed across the entire smelter once finance department confirmed performance.

5. Change Management

EGA had initiated its first advanced automated detection system, deployed across multiple potlines of varying operational environments, which depended on the involvement of more than 4000 employees to utilise the output. Therefore, a robust change management program was needed with respective smelter operations in charge as change champions.

The following key actions were taken for full-scale utilisation of the model.

5.1 Improve User Experience (UX)

- Integrated the predictive features of the soft sensor model into iPots potline information system in all Jebel Ali and Al Taweelah lines (Figure 8)

Figure 8. Spike alert in pot control system.

- Installed five additional monitors in supervisor rooms across several Jebel Ali lines to ensure that all operations personnel across all lines can monitor line and pot status at any time during a shift

5.2 Coaching

- Coaching more than 25 supervisors and almost 50 technicians and senior technicians with more than 60 walkthroughs and site visits

5.3 Efficacy

- Build enhanced mechanisms to track model accuracy and adoption by operations
- Upgraded the technician work order to display both scheduled and unscheduled tasks, including probing activities, pending activities and pots with abnormalities
- Developed an improved set of metrics covering various aspects of pot manual checks/probing and action adherence and accuracy
- Created a tracking mechanism in a database to capture all probing activities performed by operations (Table 2)

Table 2. Spike model usage and performance monitoring system.

AT L3					
Date	Spike Flags(Nos)	Spike Removed(Nos)	Detection Of Spike (%)	Checking Compliance (%)	Response Time(day)
5-Feb-22	173	140	77	74	1.5
12-Feb-22	146	133	76	77	1.5
19-Feb-22	140	100	63	74	1.6
26-Feb-22	156	121	77	74	1.6
5-Mar-22	162	120	78	71	1.7
12-Mar-22	147	92	66	74	1.5
19-Mar-22	196	103	78	69	1.4
26-Mar-22	144	82	76	70	1.6

6. Value Capture from the Development

Tangible benefits:
The Spike detection model gave EGA an edge to capture spikes early, which yielded a positive return. The audited gains in different technologies ranges from 3050 to 9070 mT of additional hot metal production on problem severity, model accuracy and time to act.

Intangible benefits:
The development has helped in reduction of carbon footprints. It has introduced EGA to data science and big data platforms. Developing the spike model based on operational metrics with good precision and recall without using anode current distribution measurements has brought the process to new possibilities. The journey resulted in the development of advanced analytical skills at EGA.

7. Conclusions and Way Forward

The journey has been exciting and full of challenges. Some challenges included the first-time development of the data science model, taking up the most critical operational issues for resolution. Others were the technical challenge of not having sensors, such as individual anode current distribution measurements, and the data engineering to develop fit-for-use datasets. Moreover, due to agile project execution and management methodology, the process was full of new learnings, strictly driven by time. Ultimately, the much rewarding success boosted the confidence to progress with other AI/ML uses in the company. To sum up, below are the key takeaways from this prestigious project.

- EGA has well demonstrated the use of Big Data/ AI/ ML with limited hard sensors and its scalability different technologies at EGA.
- The end goal should be always clear with defined business impact and aimed at actionable predictions.
- Data science must be coupled with in-depth process expertise to tailor the use to the desired outcome.

EGA has set a path to develop other soft sensors based on the successful development of bath temperature prediction and anode spike detection model. In addition, other initiatives are being developed and tested, using AI/ML approach:

- Excess AlF_3 soft sensor, shifting the excess AlF3 updating from 8 days (measurement) to one hour (soft sensor)
- Pot failure prediction model to predict end burst/side burst failures.

Acknowledgement

The team would like to acknowledge the encouragement, endorsement and trust of EGA leadership and the collaboration of all the smelter, IT, Finance, Procurement, Legal, transformation teams and and external vendors (C2V, Musigma, Siemens) with whom we partnered in this journey. Our special thanks to the IT team – Solaimalai, Mamatha Reddy Shyamala, Rama Mohan, Mohamed Shafi, Muhammad Sajith for their tireless support in execution. Without such support from multiple stakeholders, these fantastic results would not have been possible.

8. References

1. Najeeba Aljabri et al., Industry 4.0 and data-based decision system at EGA, *International Aluminum Journal*, Volume 97 January-February 2021, 48-49.
2. Arthur Martel, Anode spike detection using advanced analytics and data analysis, *Proceedings of 35th International ICSOBA Conference*, Hamburg, Germany, 2 – 5 October, 2017, *Travaux* 46, 815-822.