A Data Cleaning Approach for a Structural Health Monitoring System in a 75 MW Electric Arc Ferronickel Furnace

Jaiber Camacho-Olarte [1],[†] ⓘ, Julián Esteban Salomón Torres [1] ⓘ, Daniel A. Garavito Jimenez [1], Jersson X. Leon Medina [1] ⓘ, Ricardo C. Gomez Vargas [1] ⓘ, Diego A. Velandia Cardenas[1], Camilo Gutierrez-Osorio[1] ⓘ, Bernardo Rueda [2], Whilmar Vargas [2], Diego Alexander Tibaduiza Burgos [1],* ⓘ, Cesar Augusto Pedraza Bonilla [1], Jorge Sofrony Esmeral [1], Felipe Restrepo-Calle[1] ⓘ

[1]   Universidad Nacional de Colombia, Bogotá, Colombia.
[2]   Cerro Matoso S.A., Montelíbano, Colombia.
*   Correspondence: dtibaduizab@unal.edu.co
[†]   Presented at the 7th Electronic Conference on Sensors and Applications, 15–30 November 2020; Available online: https://ecsa-7.sciforum.net/.

El conocimiento es de todos — Minciencias

UNIVERSIDAD NACIONAL DE COLOMBIA

SOUTH32

7th International Electronic Conference on Sensors and Applications
ECSA-7
15 – 30 November 2020

sensors   *chemosensors*   MDPI

Outline

01

INTRODUCTION

Cerro Matoso S.A.
Smelting Company

South32 - Cerro Matoso

- Mining, Production and export of nickel
- Operating since 1982
- Open pit mine

Location

Montelibano

La Apartada

Montelibano

Puerto
Libertador

San José
de Uré

1880
Employees

40.000 Tonnes Ni
Average production last 3 years

1st Producer
FeNi LATAM

Total revenue FY20
500 US$M
10% South32

South32 - Cerro Matoso

Ferronickel Production Process

| Mining and Homogenization | Drying and Sorting | Calcination | Smelting | Refining | Finished Product |

Furnaces

Arc Electric Furnace
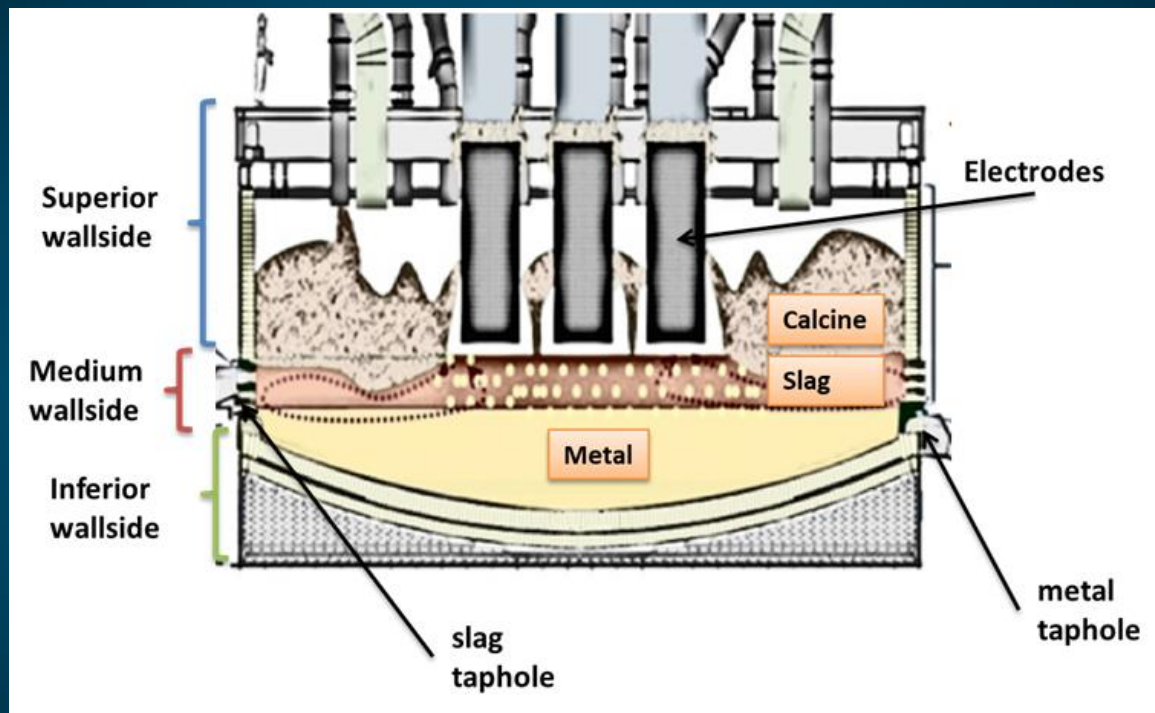
Adapted from Forero Caro, J.C. , 2019

02

LITERATURE BACKGROUND.

Different Approaches for Data Cleaning Process.

Statistical Analysis

Algorithms for time series

Integrity Constraints

Neuronal Networks

Machine Learning Models

What to detect?

How to detect?

Where to detect?

03
DATASET

DATASET

- **Hundreds of measurements per minute.**

  Experts → **15-minutes Frequency 4 years**

- **CSV file with 1180 variables and 175297 rows.**

- **Size of file: 2.8GB**

04

DATA CLEANING PROCESS.

Workflow for Data Cleaning Process.

Remove duplicates

— TAG x data   — TAG y data

TAG value

t

Empty and null values

☐ No data   — TAG data

TAG value

t

1.2 months

Unique values

TAG value

t

Strings

Time	Raw data	Processed
12:30	ON	1
12:45	ON	1
13:00	Failure	null
13:15	OFF	0
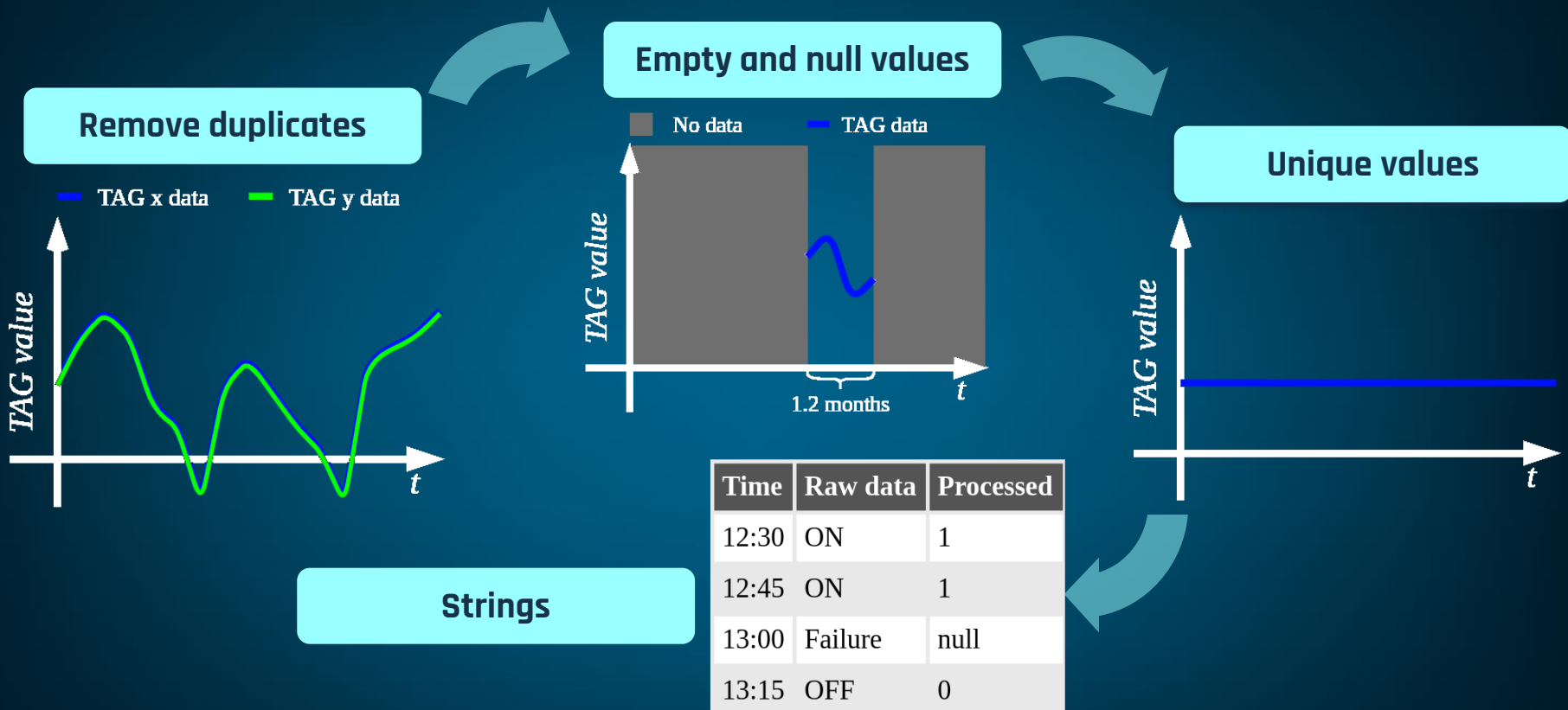
Workflow for Data Cleaning Process.
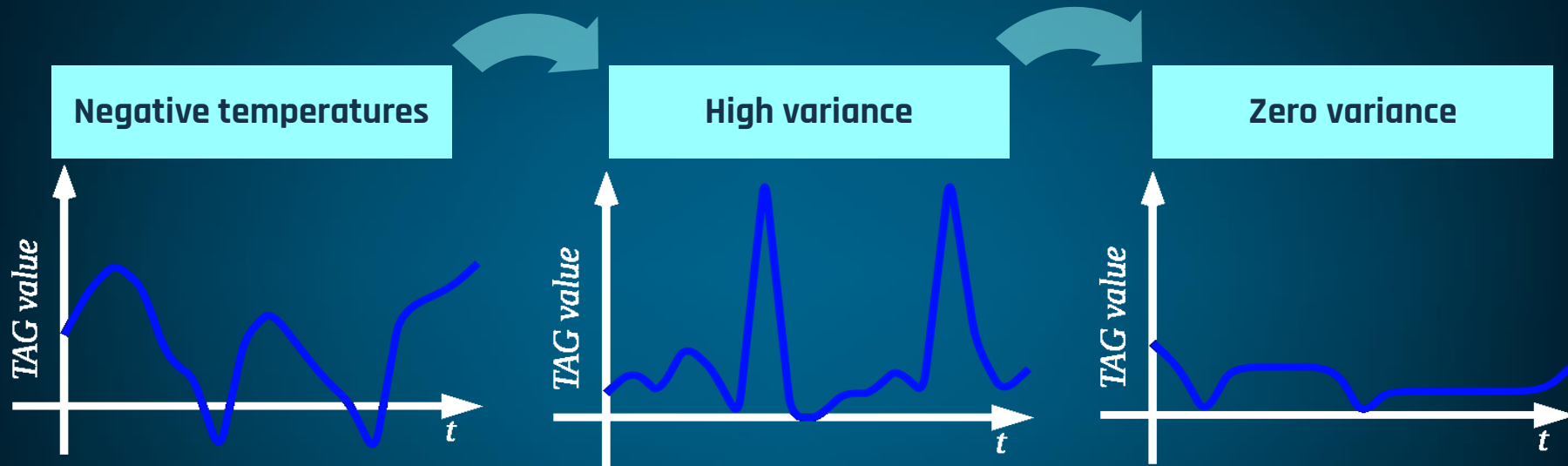
05

RESULTS

Amount Detected Variables in Data Cleaning Process.

Remove duplicates	Empty and null values	Unique values
↓	↓	↓
80 duplicated TAGs	2 empty TAGs	60 TAGs were found

Strings	Negative temperatures	High variance
↓	↓	↓
5 TAGs were eliminated	74 TAGs were eliminated	97 TAGs were found

Zero variance

↓

22 TAGs were found
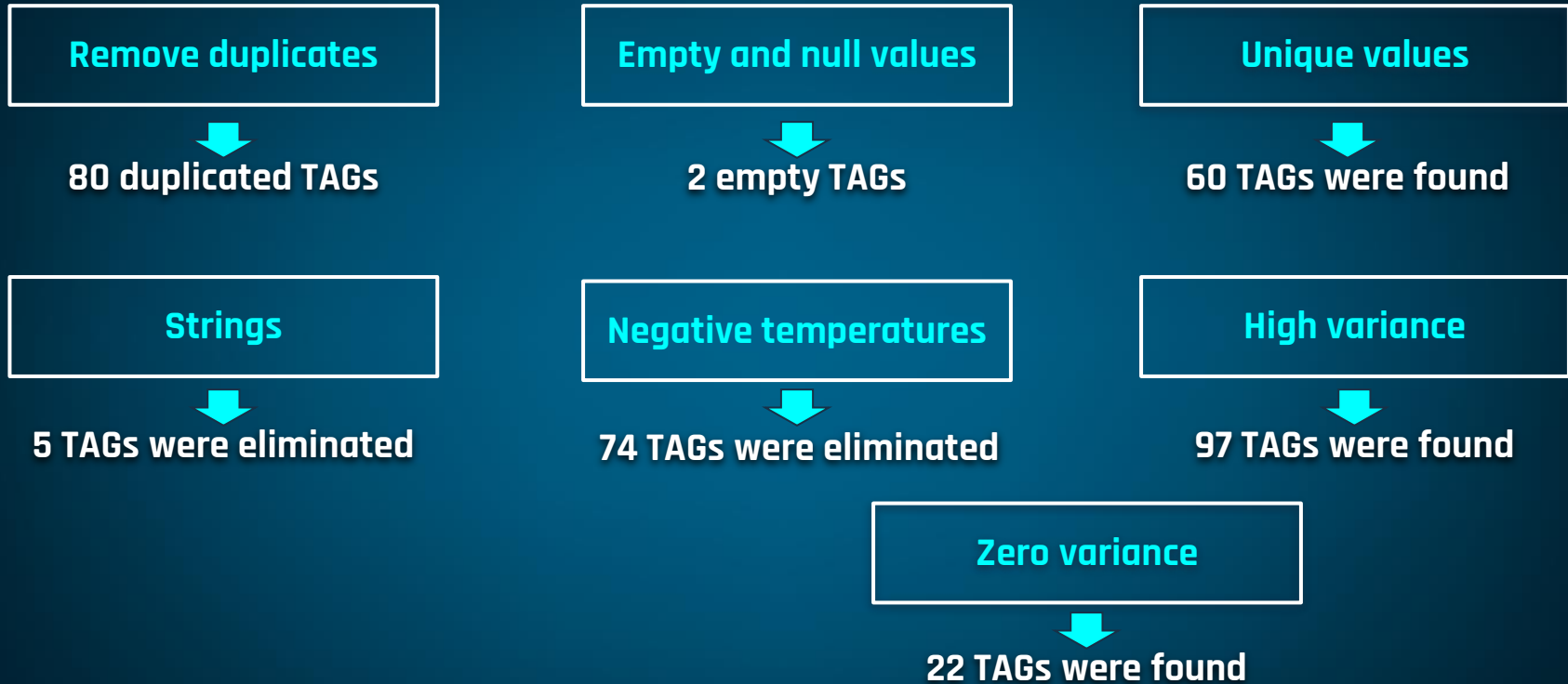
Principal Results

314
Eliminated Variables

28%
Dirty Data

1.1 GB
Size of file reduced

05

CONCLUSIONS

CONCLUSIONS

- Workflow for a data cleaning process.

- The expert judgment is very important.

- Data quality was improved.

REFERENCES

[1]. Chu, X.; Ilyas, I.F.; Krishnan, S.; Wang, J. Data Cleaning: Overview and Emerging Challenges. Proceedings of the 2016 International Conference on Management of Data; Association for Computing Machinery: New York, NY, USA, 2016; SIGMOD '16, p. 2201–2206. doi:10.1145/2882903.2912574.

[2]. Pearson, R.K. Data cleaning for dynamic modeling and control. 1999 European Control Conference (ECC). IEEE, 1999, pp. 2584–2589. doi:10.23919/ECC.1999.7099714.

[3]. Tang, N. Big RDF data cleaning. 2015 31st IEEE International Conference on Data Engineering Workshops. IEEE, 2015, Vol. 2015-June, pp. 77–79. doi:10.1109/ICDEW.2015.7129549.

[4]. Wang, X.; Wang, C. Time Series Data Cleaning with Regular and Irregular Time Intervals, 2020, [arXiv:cs.DB/2004.08284].

[5]. Wang, X.; Wang, C. Time Series Data Cleaning: A Survey. IEEE Access 2020, 8, 1866–1881. doi:10.1109/ACCESS.2019.2962152.

[6]. Hu, K.; Li, L.; Hu, C.; Xie, J.; Lu, Z. A dynamic path data cleaning algorithm based on constraints for RFID data cleaning. 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). IEEE, 2014, pp. 537–541. doi:10.1109/FSKD.2014.6980891.

[7]. Lin, J.; Sheng, G.; Yan, Y.; Zhang, Q.; Jiang, X. Online Monitoring Data Cleaning of Transformer Considering Time Series Correlation. 2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D). IEEE, 2018, Vol. 2018-April, pp. 1–9. doi:10.1109/TDC.2018.8440521.

[8]. Dai, J.; Song, H.; Sheng, G.; Jiang, X. Cleaning Method for Status Monitoring Data of Power Equipment Based on Stacked Denoising Autoencoders. IEEE Access 2017, 5, 22863–22870. doi:10.1109/ACCESS.2017.2740968.

[9]. Ge, C.; Gao, Y.; Miao, X.; Yao, B.; Wang, H. A Hybrid Data Cleaning Framework Using Markov Logic Networks. IEEE Transactions on Knowledge and Data Engineering 2020, XX, 1–1, [1903.05826]. doi:10.1109/TKDE.2020.3012472.

[10]. Lv, Z.; Deng, W.; Zhang, Z.; Guo, N.; Yan, G. A Data Fusion and Data Cleaning System for Smart Grids Big Data. 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications (ISPA/BDCloud/SocialCom/SustainCom). IEEE, 2019, pp. 802–807. doi:10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00119.

[11]. Alipour-Langouri, M.; Zheng, Z.; Chiang, F.; Golab, L.; Szlichta, J. Contextual Data Cleaning. IEEE 34th Int Conf Data Engineering Workshops (ICDEW). IEEE, 2018, pp. 21–24. doi:10.1109/ICDEW.2018.00010.

c

Questions?

jfcamachoo@unal.edu.co
Universidad Nacional de Colombia
2020

South32 - Cerro Matoso