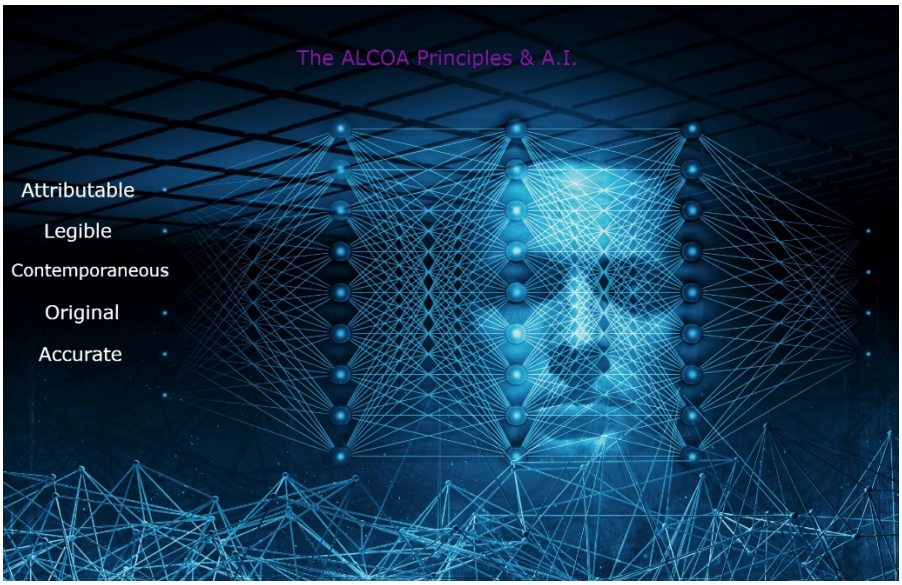
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
Θηβών 250, Αθήνα-Αιγάλεω 12241
Τηλ: +30 210 538-1614
Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
Τεχνητή Νοημοσύνη και Βαθιά Μάθηση
https://aidl.uniwa.gr/

UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING
DEPARTMENT OF INDUSTRIAL DESIGN AND PRODUCTION ENGINEERING
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
250, Thivon Str., Athens, GR-12241, Greece
Tel: +30 210 538-1614
Master of Science in
Artificial Intelligence and Deep Learning
https://aidl.uniwa.gr/

Master of Science Thesis

Artificial Intelligence in Pharmaceutical Domain (with emphasis on the data quality).
ALCOA Prediction from Pharmaceutical Industry Line



Student: Karidas Dimitrios
Registration Number: MSCAIDL-0005

MSc Thesis Supervisor

Helen-Catherine Leligou
Assoc. Professor

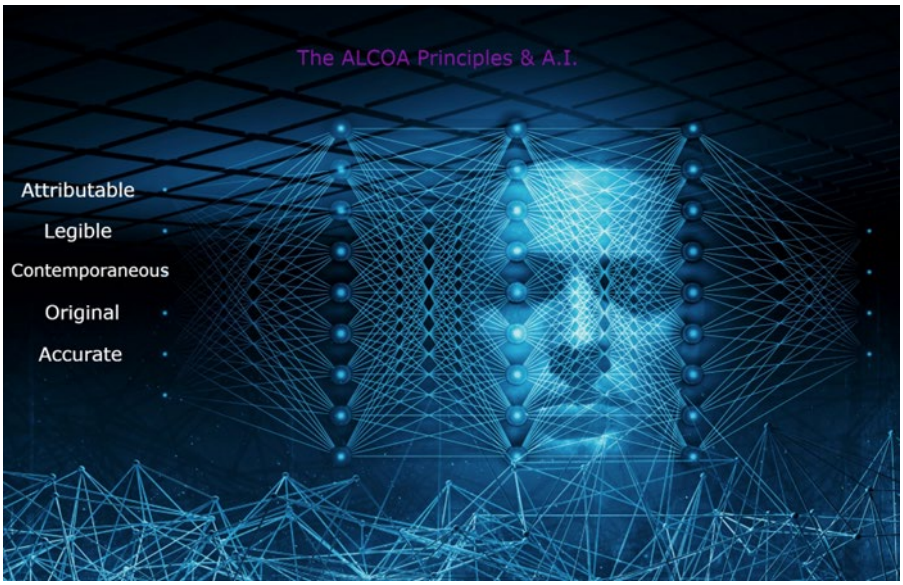
ATHENS-EGALEO, 7/2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
Θηβών 250, Αθήνα-Αιγάλεω 12241
Τηλ: +30 210 538-1614
Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
Τεχνητή Νοημοσύνη και Βαθιά Μάθηση
https://aidl.uniwa.gr/

UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING
DEPARTMENT OF INDUSTRIAL DESIGN AND PRODUCTION ENGINEERING
http://www.eee.uniwa.gr
http://www.idpe.uniwa.gr
250, Thivon Str., Athens, GR-12241, Greece
Tel: +30 210 538-1614
Master of Science in
Artificial Intelligence and Deep Learning
https://aidl.uniwa.gr/

Μεταπτυχιακή Διπλωματική Εργασία

Τεχνητή Νοημοσύνη σε Φαρμακευτικό Τομέα (με έμφαση στην ποιότητα των δεδομένων)
Πρόβλεψη ALCOA από Γραμμή Παραγωγής Φαρμακοβιομηχανίας



Φοιτητής: Καρύδας Δημήτριος
ΑΜ: AIDL-0005

Επιβλέπουσα Καθηγήτρια

Ελένη- Αικατερίνη Λελίγκου
Αναπληρώτριά Καθηγήτρια

ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, 7/ 2022

This MSc Thesis has been accepted, evaluated and graded by the following committee:

Supervisor	Member	Member
Eleni Aikaterini Leligkou Digitally signed by Eleni Aikaterini Leligkou Date: 2022.07.28 10:40:12 +03'00'	ANTONIOS KARAGEORGOS 29.07.2022 09:08	EFTHYMIOS LALLAS EFTHYMIOS LALLAS Jul 28 2022 12:40 PM
(Leligou, Nelly)	(Karageorgos, Anthony)	(Lallas, Efthimios)
(Assoc. Professor)	(Professor)	(Assistant Professor)
(Department of Industrial Design and Production Engineering)	(Department of Forestry, Wood Sciences and Design)	(Department of Forestry, Wood Sciences and Design)
(University of West Attica)	(University of Thessaly)	(University of Thessaly)

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Καρύδας Δημήτριος του Αχιλλέα , με αριθμό μητρώου MscAIDL-0005 μεταπτυχιακός φοιτητής του ΔΠΜΣ «Τεχνητή Νοημοσύνη και Βαθιά Μάθηση» του Τμήματος Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών και του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής, της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής διπλωματικής εργασίας και κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Η εργασία δεν έχει κατατεθεί στο πλαίσιο των απαιτήσεων για τη λήψη άλλου τίτλου σπουδών ή επαγγελματικής πιστοποίησης πλην του παρόντος.
Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.
Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου μέχρι ……1/2023…… και έπειτα από αίτησή μου στη Βιβλιοθήκη και έγκριση της επιβλέπουσας καθηγήτριας.»

Ο Δηλών
Καρύδας Δημήτριος

DIMITRIOS KARYDAS
27.07.2022 18:45
(Υπογραφή φοιτητή)

Eleni Aikaterini Leligkou
Digitally signed by Eleni Aikaterini Leligkou
Date: 2022.07.29 11:27:34 +03'00'

Declaration of the author of this MSc thesis

I, Karidas Dimitris of Achilles with the following student registration number: MSCAIDL-0005, postgraduate student of the MSc programme in "Artificial Intelligence and Deep Learning", which is organized by the Department of Electrical and Electronic Engineering and the Department of Industrial Design and Production Engineering of the Faculty of Engineering of the University of West Attica, hereby declare that:
I am the author of this MSc thesis and any help I may have received is clearly mentioned in the thesis. Additionally, all the sources I have used (e.g., to extract data, ideas, words or phrases) are cited with full reference to the corresponding authors, the publishing house or the journal; this also applies to the Internet sources that I have used. I also confirm that I have personally written this thesis and the intellectual property rights belong to myself and to the University of West Attica. This work has not been submitted for any other degree or professional qualification except as specified in it.
Any violations of my academic responsibilities, as stated above, constitutes substantial reason for the cancellation of the conferred MSc degree.
I wish to deny access to the full text of my MSc thesis until …1/2023……., following my application to the Library of UNIWA and the approval from my supervisor.

The author
Karidas Dimitrios

DIMITRIOS KARYDAS
27.07.2022 18:46

(Signature)

This master thesis is dedicated to those who contributed to its completion and especially to my family and friends for the precious time I deprived them of.

I especially want to thank my professor Mrs. Leligou Aikaterini, Mr. Isaac Kavasidis, my fellow student Mr. Economidis George and all my professors in the post-graduate program who without their valuable help this dissertation would not be possible.

Abstract

We can all imagine the amount of data generated during the procedure of the production of a medicine in a pharmaceutical industry. Data are taken from the import of the raw material to the factory, its analysis until it is used in the production of the drug, from the production line, from the warehouses but also from the distribution lines up to the final consumer, the patient. All this data must ensure traceability, if possible, from the raw material production plant to the final consumer. The most recent example is the example of the pandemic of COVID 19 vaccines. Therefore, before a drug can be used, the pharmaceutical company must prove that the drug is effective and safe. That is why the pharmaceutical companies are conducting many tests, and numerous studies in quality control.

Over the years, the pharmaceutical companies have adopted the concept of ALCOA as a framework for ensuring the observance, preservation, security and accuracy of data. The term ALCOA is an acronym that means Attributable, Legible, Contemporaneous, Original and Accurate. From the meaning of these words, we can easily understand why this acronym is so important and why it was adopted by the pharmaceutical industry. The term ALCOA is about the quality and integrity of the data, which has a direct impact on the quality of the drug.

This master thesis is an attempt to classify the values obtained from different sensors (from two production lines of a well-known Italian pharmaceutical company) as data that are Attributable and Contemporaneous. This attempt was done by using three deep learning models. We also tried to find out if there is a possibility to predict the next ALCOAs from the previous ones. These words came from the acronym ALCOA that mentioned above. The three deep learning models used are the LSTM Model, Bi-LSTM Model and GRU Model.

Unfortunately, the above deep learning models failed to predict the next ALCOA the **Attributable** and the **Contemporaneous** from the previous ones. The three models used showed better performance in the **Attributable** than in the **Contemporaneous** yet again this performance does not allow us to use them as models for predicting this ALCOA in a pharmaceutical industry.

However, this should not disappoint us as it is the first attempt to use such models in the prediction of ALCOA. After all, this dissertation focused only on the **Attributable** and the **Contemporaneous**. There are three other letters in the acronym as well and the dataset was only from two production lines. There are so many deep learning models, machine learning algorithms, so many more letters remaining letters in the acronym, other and more improved datasets that can only give promise for the future.

Keywords

Pharmaceutical Industry, A.I and M.L., ALCOA, Big data, Data Quality.

Περίληψη

Όλοι μπορούμε να φανταστούμε τον όγκο των δεδομένων που παράγονται κατά τη διαδικασία παραγωγής ενός φαρμάκου σε μια φαρμακευτική βιομηχανία. Λαμβάνονται δεδομένα από την εισαγωγή της πρώτης ύλης στο εργοστάσιο, την ανάλυσή της έως ότου χρησιμοποιηθεί στην παραγωγή του φαρμάκου, από τη γραμμή παραγωγής, από τις αποθήκες αλλά και από τις γραμμές διανομής μέχρι τον τελικό καταναλωτή, τον ασθενή. Όλα αυτά τα δεδομένα πρέπει να διασφαλίζουν την ιχνηλασιμότητα, ει δυνατόν, από τη μονάδα παραγωγής πρώτων υλών έως τον τελικό καταναλωτή. Το πιο πρόσφατο παράδειγμα είναι το παράδειγμα της πανδημίας των εμβολίων κατά του COVID 19. Επομένως, πριν χρησιμοποιηθεί ένα φάρμακο, η φαρμακευτική εταιρεία πρέπει να αποδείξει ότι το φάρμακο είναι αποτελεσματικό και ασφαλές. Γι' αυτό οι φαρμακευτικές εταιρείες πραγματοποιούν πολλές δοκιμές και πολυάριθμες μελέτες στον ποιοτικό έλεγχο.

Με τα χρόνια, οι φαρμακευτικές εταιρείες έχουν υιοθετήσει την έννοια της ALCOA ως πλαίσιο για τη διασφάλιση της τήρησης, της διατήρησης, της ασφάλειας και της ακρίβειας των δεδομένων. Ο όρος ALCOA είναι ένα αρκτικόλεξο που σημαίνει Αποδοτέο, Ευανάγνωστο, Σύγχρονο, Πρωτότυπο και Ακριβές. Από τη σημασία αυτών των λέξεων μπορούμε εύκολα να καταλάβουμε γιατί αυτό το αρκτικόλεξο είναι τόσο σημαντικό και γιατί υιοθετήθηκε από τη φαρμακοβιομηχανία. Ο όρος ALCOA αφορά την ποιότητα και την ακεραιότητα των δεδομένων, τα οποία έχουν άμεσο αντίκτυπο στην ποιότητα του φαρμάκου.

Η παρούσα μεταπτυχιακή διατριβή είναι μια προσπάθεια ταξινόμησης των τιμών που λαμβάνονται από διαφορετικούς αισθητήρες (από δύο γραμμές παραγωγής γνωστής ιταλικής φαρμακευτικής εταιρείας) ως δεδομένα που αποδίδονται και είναι σύγχρονα. Αυτή η προσπάθεια έγινε με τη χρήση τριών μοντέλων βαθιάς μάθησης. Προσπαθήσαμε επίσης να μάθουμε αν υπάρχει δυνατότητα πρόβλεψης των επόμενων ALCOA από τα προηγούμενα. Αυτές οι λέξεις προήλθαν από το αρκτικόλεξο ALCOA που αναφέρθηκε παραπάνω. Τα τρία μοντέλα βαθιάς μάθησης που χρησιμοποιούνται είναι το μοντέλο LSTM, το μοντέλο Bi-LSTM και το μοντέλο GRU.

Δυστυχώς, τα παραπάνω μοντέλα βαθιάς μάθησης δεν κατάφεραν να προβλέψουν το επόμενο ALCOA το Attributable και το Contemporaneous από τα προηγούμενα. Τα τρία μοντέλα που χρησιμοποιήθηκαν έδειξαν καλύτερες επιδόσεις στο Attributable από ότι στο Contemporaneous, και πάλι αυτή η απόδοση δεν μας επιτρέπει να τα χρησιμοποιήσουμε ως μοντέλα για την πρόβλεψη αυτού του ALCOA σε μια φαρμακευτική βιομηχανία.

Ωστόσο, αυτό δεν πρέπει να μας απογοητεύσει καθώς είναι η πρώτη προσπάθεια χρήσης τέτοιων μοντέλων στην πρόβλεψη της ALCOA. Άλλωστε, αυτή η πτυχιακή επικεντρώθηκε μόνο στο Αποδοτέο και στο Σύγχρονο. Υπάρχουν και άλλα τρία γράμματα στο ακρωνύμιο και το σύνολο δεδομένων προέρχεται μόνο από δύο γραμμές παραγωγής. Υπάρχουν τόσα πολλά μοντέλα βαθιάς μάθησης, αλγόριθμοι μηχανικής μάθησης, τόσα άλλα γράμματα που απομένουν στο ακρωνύμιο, άλλα και περισσότερα βελτιωμένα σύνολα δεδομένων που μπορούν να δώσουν υπόσχεση για το μέλλον.

Λέξεις – κλειδιά

Φαρμακευτική Βιομηχανία, A.I και M.L., ALCOA, Μεγάλα δεδομένα, Ποιότητα δεδομένων.

Table of Contents

List of Tables. **13**

List of figures. **14**

Acronym Index **19**

INTRODUCTION **20**

The subject of this thesis **20**
Aim and objectives **20**
Methodology **20**
Innovation **21**
Structure **21**

1 A.I. in Pharmaceutical Industry **22**
1.1 A.I. in development of new medicines **22**
1.2 Sensors and agents in pharmaceutical industry lines **24**
1.3 From Good Manufacturing Practices to Smart Pharmaceutical Practices. **24**
1.3.1 Maintaining the integrity of the data and GMP 25
1.3.2 Big Data and Pharma 4.0 26
1.3.3 Cloud 26
1.3.4 Big Data and A.I. 27
1.3.5 Security 28
1.3.6 Blockchain in pharmaceutical sector. 29
1.3.7 Minimum level of security and Minimum level of services 29

2 CHAPTER 2: SPUMONI & ALCOA **31**
2.1 SPuMoNI, the European project about Big Data and process modelling for smart industry [32] **31**
2.2 ALCOA principles **31**

3 CHAPTER 3: Machine Learning & Deep Learning **33**
3.1 Machine Learning **33**
3.2 Deep Learning **33**
3.3 RNN **34**
3.3.1 LSTM 35
3.3.2 Bi-LSTM 36
3.3.3 GRU 37

4 Data Prepossessing **38**
4.1 I 1000 Dataset description **38**
4.2 I 600 Dataset description **39**

5 Results and Discussion **42**
5.1 Attributable principle and I 600 Dataset **42**
5.1.1 Minmax Scaler 44
5.1.2 Standard Scaler 47
5.1.3 Prediction from previous Attributable prices for I 600 dataset. 49
5.2 Attributable principle and I 1000 dataset **53**
5.2.1 Minmax Scaler 55
5.2.2 Standard Scaler 58
5.2.3 Prediction from previous Attributable prices I 1000 dataset 60
5.3 Contemporaneous principle and I 600 Dataset **64**
5.3.1 Minmax Scaler 66

5.3.2 Standard Scaler 69
5.3.3 Prediction from previous Contemporaneous prices I 600 dataset. 71
5.4 Contemporaneous principle and I 1000 Dataset 75
5.4.1 Minmax Scaler 77
5.4.2 Standard Scaler 80
5.4.3 Prediction from previous Contemporaneous prices I 1000 dataset. 82
6 Conclusion 86
7 Future Research 88
Bibliography – References – Online sources 90
8 Bibliography 90
Appendix A 96
Appendix B 96
Appendix C 96
Appendix D 96

List of Tables.

Table 1. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 600 dataset for with no scaling. 42

Table 2. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 600 dataset with no scaling. 43

Table 3. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 600 dataset for with Minmax Scaler algorithm applied. 45

Table 4. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 600 dataset with Minmax Scaler algorithm applied. 45

Table 5. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 600 dataset for with Standard Scaler algorithm applied. 47

Table 6. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 600 dataset with Standard Scaler algorithm applied. 48

Table 7. Performance comparison of the four regression models used for prediction of Attributable for I 600 dataset. 53

Table 8. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 1000 dataset for with no scaling. 53

Table 9. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 1000 dataset with no scaling. 54

Table 10. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 1000 dataset for with Minmax Scaler algorithm applied. 56

Table 11. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 1000 dataset with Minmax Scaler algorithm applied. 56

Table 12. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 1000 dataset for with Standard Scaler algorithm applied. 58

Table 13. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 1000 dataset with Standard Scaler algorithm applied. 59

Table 14. Performance comparison of the four regression models used for prediction of Attributable for I 1000 dataset. 64

Table 15. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 600 dataset for with no scaling. 65

Table 16. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 600 dataset with no scaling. 65

Table 17. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 600 dataset for with Minmax Scaler. 67

Table 18. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 600 dataset with Minmax Scaler. 67

Table 19. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 600 dataset for with Standard scaler. 69

Table 20. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 600 dataset with standard Scaler. 70

Table 21. Performance comparison of the four regression models used for prediction of Contemporaneous for I 600 dataset. 75

Table 22. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 1000 dataset for with no scaling. 75

Table 23. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 1000 dataset with no scaling. 76

Table 24. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 1000 dataset for with Minmax Scaler. 78

Table 25. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 1000 dataset with Minmax Scaler. 78

Table 26. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 1000 dataset for with Standard Scaler. 80

Table 27. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 1000 dataset with Standard scaler. 81

Table 28. Performance comparison of the four regression models used for prediction of Contemporaneous for I 1000 dataset. 86

List of figures.

Figure 1**.** Representation of a mathematical artificial neuron model. The input to the neuron is summed up and filtered by activation function. (B) Simplified Representation of an artificial neuron model. Only the key elements are depicted, i.e., the input, the output, and the weights. https://www.frontiersin.org/articles/10.3389/frai.2020.00004/full 34

Figure 2 Comparison between normal and recurrent neural networks input vector. Adapted from Patterson, J. Gibson, A. Deep learning: a practitioner's approach. O'Reilly Media, Inc.; 2007. 34

Figure 3. Feed forward flow for recurrent neural networks. Adapted from Patterson, J. Gibson, A. Deep learning: a practitioner's approach. O'Reilly Media, Inc.; 2007. 35

Figure 4. Unrolling for recurrent neural networks. Adapted from Patterson, J. Gibson, A. Deep learning: a practitioner's approach. O'Reilly Media, Inc.; 2007. 35

Figure 5. LSTM input outputs and the corresponding equations for a single timestep. https://towardsai.net/p/machine-learning/tutorial-on-lstm-a-computational-perspective-f3417442c2cd 36

Figure 6. Bi-LSTM neural network structure deployed in time direction. https://www.researchgate.net/publication/339679582_Multi-time_scale_wind_speed_prediction_based_on_WT-bi-LSTM 37

Figure 7. Depth Gated RNNs by Yao, et al. (2015). There's also some completely different approach to tackling long-term dependencies, like Clockwork RNNs by Koutnik, et al. (2014). http://colah.github.io/posts/2015-08-Understanding-LSTMs/ 37

Figure 8. I 1000 Production Line dataset description..... 38

Figure 9. I 1000 Production Line Contemporaneous time series data..... 39

Figure 10. I 1000 Production Line Attributable time series data. 39

Figure 11. I 1000 Production Line dataset description..... 40

Figure 12. I 600 Production Line Contemporaneous time series data..... 40

Figure 13. I 600 Production Line Attributable time series data. 41

Figure 14. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 600 dataset with no scaling. 43

Figure 15. Loss/ Val loss curve of GRU network for Attributable and I 600 dataset with no scaling... 44

Figure 16. Loss/ Val loss curve of LSTM network for Attributable and I 600 dataset with no scaling. 44

Figure 17. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 600 dataset with with Minmax Scaler algorithm applied. 46

Figure 18. Loss/ Val loss curve of GRU network for Attributable and I 600 dataset with with Minmax Scaler algorithm applied..... 46

Figure 19. Loss/ Val loss curve of LSTM network for Attributable and I 600 dataset with Minmax Scaler algorithm applied..... 47

Figure 20. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 600 dataset with Standard Scaler algorithm applied..... 48

Figure 21. Loss/ Val loss curve of GRU network for Attributable and I 600 dataset with Standard Scaler algorithm applied..... 49

Figure 22. Loss/ Val loss curve of LSTM network for Attributable and I 600 dataset with Standard Scaler algorithm applied..... 49

Figure 23. Prediction of Attributable for I 600 dataset with random forest regression model. The performance of the model was: Test error (MSE): 2.83422..... 50

Figure 24. Prediction of Attributable for I 600 dataset with random forest regression model with grid search. The performance of the model was: Test error (MSE): 2.9538063879668854. 51

Figure 25. Prediction of Attributable for I 600 dataset with auto regressor model with Lasso Penalty. The performance of the model was: Test error (MSE) 3.5926376838689946..... 52

Figure 26. Prediction of Attributable for I 600 dataset with a linear regression model. The figure shows the last 30 batches which used for evaluate the prediction. The performance of the model was: Test error (MSE): 3.395375850595643. 52

Figure 27. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 1000 dataset with no scaling. 54

Figure 28. Loss/ Val loss curve of GRU network for Attributable and I 1000 dataset with no scaling. 55

Figure 29. Loss/ Val loss curve of GRU network for Attributable and I 1000 dataset with no scaling. 55

Figure 30. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 1000 dataset with Minmax Scaling..... 57

Figure 31. Loss/ Val loss curve of GRU network for Attributable and I 1000 dataset with Minmax Scaling..... 57

Figure 32. Loss/ Val loss curve of LSTM network for Attributable and I 1000 dataset with Minmax Scaling..... 58

Figure 33. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 1000 dataset with Standard Scaling..... 59

Figure 34. Loss/ Val loss curve of GRU network for Attributable and I 1000 dataset with Standard Scaling..... 60

Figure 35. Loss/ Val loss curve of LSTM network for Attributable and I 1000 dataset with Standard scaling. 60

Figure 36. Prediction of Attributable for I 1000 dataset with random forest regression model. The performance of the model was: Test error (MSE): 1.3849400000000012..... 61

Figure 37. Prediction of Attributable for I 600 dataset with random forest regression model with grid search. The performance of the model was: Test error (MSE) 1.7344624646202926..... 62

Figure 38. Prediction of Attributable for I 600 dataset with auto regressor model with Lasso Penalty. The performance of the model was: Test error (MSE) 1.6038020049755768..... 63

Figure 39. Prediction of Attributable for I 600 dataset with a linear regression model. The figure shows the last 30 batches which used for evaluate the prediction. The performance of the model was: The test error (MSE): 1.5267208846444105. 63

Figure 40. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 600 dataset with no scaling. 66

Figure 41. Loss/ Val loss curve of GRU network for Contemporaneous and I 600 dataset with no scaling. 66

Figure 42. Loss/ Val loss curve of LSTM network for Contemporaneous and I 600 dataset with no scaling. 66

Figure 43. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 600 dataset with Minmax Scaling. 68

Figure 44. Loss/ Val loss curve of GRU network for Contemporaneous and I 600 dataset with Minmax scaling. 68

Figure 45. Loss/ Val loss curve of LSTM network for Contemporaneous and I 600 dataset with Minmax scaling. 69

Figure 46. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 600 dataset with Standard scaler. 70

Figure 47. Loss/ Val loss curve of GRU network for Contemporaneous and I 600 dataset with Standard scaler. 71

Figure 48. Loss/ val loss curve of LSTM network for Contemporaneous and I 600 dataset with Standard Scaler. 71

Figure 49. Prediction of Contemporaneous for I 600 dataset with random forest regression model. The performance of the model was: Test error (MSE): 45.466243333333324. 72

Figure 50. Prediction of Contemporaneous for I 600 dataset with random forest regression model with grid search. The performance of the model was: (MSE) of this model is: 40.6529824518856. 73

Figure 51. Prediction of Contemporaneous for I 600 dataset with auto regressor model with Lasso Penalty. The performance of the model was: Test error (MSE) 40.59479708636836. 74

Figure 52. Prediction of Contemporaneous for I 600 dataset with a linear regression model. The figure shows the last 30 batches which used for evaluate the prediction. The performance of the model was: The test error (MSE): 40.05476184483597. 74

Figure 53. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 1000 dataset with no Scaling. 76

Figure 54. Loss/ Val loss curve of GRU network for Contemporaneous and I 1000 dataset with no scaling. 77

Figure 55. Loss/ Val loss curve of LSTM network for Contemporaneous and I 1000 dataset with no scaling. 77

Figure 56. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 1000 dataset with Minmax Scaler. 79

Figure 57. Loss/ Val loss curve of GRU network for Contemporaneous and I 1000 dataset with Minmax Scaler. 79

Figure 58. Loss/ Val loss curve of LSTM network for Contemporaneous and I 1000 dataset with Minmax Scaler. 80

Figure 59. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 1000 dataset with Standard Scaler..... 81

Figure 60. Loss/ Val loss curve of GRU network for Contemporaneous and I 1000 dataset with Standard Scaler..... 82

Figure 61. Loss/ Val loss curve of LSTM network for Contemporaneous and I 1000 dataset with Standard Scaler..... 82

Figure 62. Prediction of Contemporaneous for I 1000 dataset with random forest regression model. The performance of the model was: Test error (MSE): 77.08506000000001..... 83

Figure 63. Prediction of Contemporaneous for I 1000 dataset with random forest regression model with grid search. The performance of the model was: Test error (MSE): 70.67621716463478. 84

Figure 64. Prediction of Contemporaneous for I 1000 dataset with auto regressor model with Lasso Penalty. The performance of the model was: Test error (MSE) 65.75462106717687..... 85

Figure 65. Prediction of Contemporaneous for I 1000 dataset with a linear regression model. The figure shows the last 30 batches which used for evaluate the prediction. The performance of the model was: Test error (MSE): 60.135593776713236. 85

Acronym Index

A.I: Artificial Intelligence.

M.L: Machine Learning

Bi- LSTM Network: Bidirectional- Long Short-Term Memory Network.

LSTM Network: Long Short-Term Memory Network.

GRU Network: Gated recurrent units Network.

ALCOA: Attributable, Legible, Contemporaneous, Original, Accurate.

G.M.P: Good manufacturing practice.

G.D.P: Good Documentation Practice.

F.D.A: The Food and Drug Administration.

G.D.P.R: General Data Protection Regulation.

I.o.T: Internet of Things.

I.I.o.T: Industrial Internet of Things.

MAE: Mean Absolut Error

VMAE: Val Mean Absolut Error.

MSE: Mean Square Error.

INTRODUCTION

The definition of ALCOA has been around since 1990 and is being adopted by industries that produce large amounts of data as an attempt at good documentation practice. ALCOA concerns both written and electronic data, and according to the FDA and the European Medicines Agency these are: Attributable, Legible, Contemporaneous, Original and Accurate. These simple principles should be part of the whole data lifecycle. Data integrity and access control issues have arisen to be heard daily in newspapers and magazines to a large extent. Even since 2015 there have been warning letters issued by the FDA so a solution is sought immediately.

The ALCOA is a solution. A pharmaceutical company also gives an ALCOA value to any data generated during the coding of a drug as an attempt to ensure the accuracy of this data. When we hear the word precision and integrity, our mind goes to models of artificial intelligence and machine learning. Thus, was born the idea of this dissertation which is the prediction of whether possible prices of ALCOA with models of artificial intelligence from pharmaceutical production lines.

The subject of this thesis

Artificial intelligence today is the biggest tool where all the applications that run now and, in the future, will click. ALCOA is a system that will be pinned on by all good manufacturing practice (GMP) and many pharmaceutical industries are already working on it. If this tool called artificial intelligence clicks on this model called ALCOA it will give an extra edge to this model and even more dynamics where needed. This, in addition to being a big boost, will give the ALCOA-based system the ability to anticipate and correct future errors that may occur. Surely in the future these two tools will be used as a means of the data integrity thus ensuring the Good Documentation Practice (GDP).

Aim and objectives

The purpose of this master thesis is to see if artificial intelligence and in particular some models of deep learning and machine learning can be used as a platform for predicting future ALCOA's. Attempts were also made to see which models might be used to predict future ALCOA values in our case the Attributable and Contemporaneous. These letters are come from the acronym ALCOA of course.

Methodology

For the purposes of this dissertation, three models of artificial intelligence were used: the LSTM, the GRU and Bi - LSTM. Initially the datasets were improved as it presented many problems so that it could be introduced into an artificial intelligence model. After the datasets from the two production lines were entered in our models, an attempt was made to adjust them correctly with finetuning. For further investigation at the end, four machine learning models were used to compare and attempt to predict future Contemporaneous and Attributable.

Innovation

As far as artificial intelligence and the ALCOA acronym are concerned, they are innovative. The innovation is that an attempt was made to marry artificial intelligence and deep learning with ALCOA. This is evident from the fact that there is nothing like bibliography on this subject and thus paves the way for new research on this subject. Surely the pieces that can be explored are really infinity and maybe this degree will be the beginning for research on this topic and the subject.

Structure

At the beginning of this degree is presented the use of artificial intelligence and machine learning in the pharmaceutical sector, how it already helps in the development of new pharmaceutical products and in the entire production chain of a drug. Then a short paragraph analyzes the role of sensors in a pharmaceutical industry. After that, the good pharmaceutical practices are analyzed and from these we move on to the smart pharmaceutical practices. How data integrity ensures good pharmaceutical practices and the quality of medicines in general. Big data and the pharmaceutical industry, the cloud where data is stored, data security, blockchain as security and data integrity are tools that are analyzed and are part of the system that ensures data integrity.

SPUMONI is a mechanism of the European Union that using ALCOA to ensures the accuracy of the data using the tools mentioned above. There is a lot of talk about ALCOA and so they are also mentioned below. Then a few words about the artificial intelligence models used and their advantages.

Then we moved on to the main part of our dissertation with the pre-preparation of the data as well as the results of the models we used. The attempt to predict with machine learning algorithms.

Finally, the conclusions are analyzed, as well as the future challenges that arise from this master thesis.

CHAPTER 1: M.L. and Deep Learning in Pharmaceutical Domain

1 A.I. in Pharmaceutical Industry

More and more nowadays we hear the term Industry 4.0 [1] and digital transformation. It will affect all businesses and industry. Of course, the pharmaceutical industry can't and will not remain unaffected. Artificial intelligence and machine learning will play a key role in this and will be, if not already, an important tool in realizing the digital transformation and what we call industry 4.0. Artificial intelligence and machine learning are basically one of the key components of industry 4.0. In every functional unit of a company from the production lines, to the development of new products, to the assurance of quality even in the marketing and distribution of products. [2]

1.1 A.I. in development of new medicines

Artificial intelligence is expected to revolutionize the design and development of new pharmaceutical products. [3] Every new drug especially patents are products that can bring big profits to the pharmaceutical industry. Machine learning and artificial intelligence are already being used to discover new substances that could potentially be new drugs.

Evaluating the properties of a substance from different data that can be used as a medicine is very important. It is not a few times that a very promising substance was discovered after many years of research but in the end, it did not have such a long lifespan, its half-life, if it is a radiopharmaceutical, it did not have the appropriate physicochemical properties, it turned out to be very toxic to the human body. Molecule of the substance was not so effective as to bind to the appropriate receptors in the human body. Even the substance in its preparation was not easy to use on a large industrial scale, it was not easy to store, with the typical example of the first vaccines of Covid 19 where for their storage refrigerators with a capacity of -50 to -15 degrees Celsius were needed. [4]

Improving existing biological and chemical substances. Not all experimental data are available. Laboratory data are also very expensive and time consuming. With a simple generalization one can take advantage of the existing knowledge in the design of engineering learning models or artificial intelligence. This could lead to the creation of new cheaper drugs using simply the correlations of existing ones and their effects. An interesting field of research is how we can evaluate and utilize research results, as well as the results of gene surgeries. From there, a lot of data on cellular cauterization emerge, such as genetic manifestations and cellular images of cells that are undergoing cellular stress. Artificial intelligence and machine learning could help in combination with existing experiments and future to suggest new experiments based on these results. One such example is the prediction of the effect of a drug on the human brain. This can help predict the side effects of a drug that is the body's resistance or dependence on it. [5]

Active learning is interposed between very expensive and often ineffective experiments and machine learning and artificial intelligence can help to solve these problems. This is more necessary than ever to move from approximate models to real decisions using repeatable experimental data. Knowing that machine learning predictions can be used to generate applications e.g., substances, not just once but many times. This diversity of candidate

substances can be beneficial from using only the best initial conjecture. Algorithms can be adjusted to optimize and suggest new experiments by asking questions, thus reducing scientific uncertainty. Gaining relevant knowledge by capturing the causal structure we can fully explore many therapeutic interventions of many molecules or drug combinations. [6]

Also combining all the misconceptions, we can ask questions about the position of the molecules and their position in space. In the context of supportive learning, we can put additional questions of how we can move in the space of these molecules by adding new molecules or atoms or even whole blocks of molecules and atoms to discover the proper representation of the design of these molecules.

The speed with which we can have experimental data continues to be the key to success and rely on experiments outside the field of artificial intelligence and machine learning. The advancement of robotics could automate this whole process by incorporating both artificial intelligence and machine learning. Robotics combined with artificial intelligence is essential to guide the development of more relevant biotechnology tools for the synthesis and screening of performance-related substances. [7]

We can also integrate patient information in real time, we can do individualized treatment by loading it into machine learning models such as gene expression gene information but also molecular and tissue pathological data, from various vital points. All this presupposes the development of new pin sensors and new portable recording devices. All of these devices will continuously keep the patient history in real time. This whole piece is very important because the appropriate treatment to be given depends on the course of the disease and the data drawn from the patient's sensors. Special combinations of drugs for the patient could also be tested using mechanical learning and artificial intelligence algorithms. All this in the context of personalized therapy to predict what is the right combination for patient data. For example, if it is an antiviral drug, as we saw with the pandemic Covid 19, drugs that aim to inhibit the multiplication of the virus are more effective during the period when the virus is active, i.e., the incubation period. Before the symptoms of the disease pay the patient. While medicines such as anti-inflammatory drugs and analgesics should be introduced later and in the right order. This presupposes the online monitoring of the disease and the complications that occur in the patient as soon as possible. [8]

The drugs as we saw with the vaccines of Covid 19 require time for clinical trials. [9] Artificial intelligence and machine learning can help reduce costs and the duration of clinical trials. This can be done by better targeting volunteers than a wider population of infected patients. The use of machine learning and artificial intelligence can classify side effects reports with distorted trials from models trained from a background population. Such trials could reduce costly and time consuming adaptive clinical trials and extract information from them more quickly. Artificial intelligence can also help to generalize the results to the population from those included in the clinical trial to a population with the same or similar characteristics. [10]

There are still many clinical trials and documented in animal research. Many preclinical studies are performed on laboratory animals to predict the side effects and effects of a compound in later clinical trials but also on other higher animals or humans. Artificial intelligence and machine learning can help us move from an in-vivo and in-vitro model to a new, in-silicon human model. [11]

There is a lot of talk nowadays about developing privacy methods. Medical confidentiality, medical information is subject to many privacy risks and corresponding protection regulatory principles such as G.D.P.R. All of these data sources could give researchers great impetus and information if they could access them safely. Federal learning as we know it enhances learning without requiring access to individual files. The way privacy is not violated and so we can access more sensitive data. [12]

Artificial intelligence and machine learning could help automate drug development in the context of personalized therapy. In the current legislative framework, medicines are approved only once. As we know, a virus, for example, can mutate very quickly. This can make the approval process of a new drug very costly and time consuming. However, the adaptations required to be made are too few to adapt to the genome and the pathogenic microorganism. The procedure can be applied to the adaptation of the drug, for example, to a cancer that is metastatic. This ensures that the drugs or vaccines continue to be effective in the context of a mutation. However, in order for this to happen at the level of regulatory approval until the competent bodies approve the new drug as a procedure and not as a matter of fact, developers will still have the flexibility to adapt the drug based on patient population and pathogen composition of each patient. [13]

Artificial intelligence and machine learning is expected to revolutionize the design and development of new products. Every new drug especially patents are products that can bring big profits to the pharmaceutical industry. Machine learning and artificial intelligence are already being used to discover new substances that could potentially be new drugs. A typical example is the drug Halisin. Researchers from MIT have discovered a substance, an antibiotic that is capable of killing germs that are also resistant to existing antibiotics. This was made possible by a machine learning model that processed millions of compounds in just a few days and envisioned possible compounds that could be used as antibiotics using different mechanisms than existing antibiotics. [14]

1.2 Sensors and agents in pharmaceutical industry lines

In industry 4.0, which will be adopted by the pharmaceutical industry, there is a need for a new smart vertical network from which it is moving in the direction of intelligent production. The sensors capture data in real time to make decisions with the help of machine learning algorithms and artificial intelligence to optimize production. Self-optimization creates the maximum result in terms of production and that is what is required. Factory automation finds the best solutions individually for each product. The sensors will have critical growth prospects in the pharmaceutical industry with the use of I.o.T. The sensors in such a case are shifted from sensors during the process to sensors for monitoring processes and conditions with the ability to correct and predict. So, we see that artificial intelligence is much more than having machines that do some work. It completely changes the way we make medicines. [15]

1.3 From Good Manufacturing Practices to Smart Pharmaceutical Practices.

Good Manufacturing Practices (GMP), describes the minimum requirements and standards that a drug manufacturer must have during its production processes. The role of the European Medicines Agency is to coordinate inspections under these strict standards and to play

an active role in harmonizing all the activities of good pharmaceutical practice within the European Union. [16]

The drug manufacturer does not have to be based in the European Union but if the drugs are intended for consumption in the European Union they must be in compliance with this standard. Good Manufacturing Practice (GMP) requires that the medicines be of a consistently high quality suitable for their intended, use and meet the requirements of the marketing authorization or clinical trial authorization for which they have been approved. [17]

Quality Assurance is the core of the pharmaceutical industry all based on good manufacturing practices. The concept of good manufacturing practice has been shaped by the need to protect end users and create a reliable drug production system history. It is known through various crises when something went wrong finally Good Manufacturing Practices is an idea is to make medicines. How to measure the risk of human error as much as possible and to ensure the safety and effectiveness of medicines. [18]

The volume of data generated by a pharmaceutical industry is surprisingly large but research shows that very little of this data is used to predict a failure and most of it is only used for compliance. For compliance and if only the damage has been done. [19]

New technologies such as artificial intelligence and machine learning using this big data can analyze and predict situations before they are created. This can be done throughout the production of a drug. Using these technologies, they can better understand some processes and improve them. [20]

1.3.1 Maintaining the integrity of the data and GMP

The big data environments and the algorithms that follow them, must follow the principle of data integrity. So, we have to make a clear and well-coordinated effort to implement best practices in system design. The architecture of the system, how the data is collected and stored, as well as how it is disseminated for analysis. Not all data works for all jobs and we must not forget that algorithms are trained by them.

All of these technologies, however, allow for an automated and integrated control line. This allows the artificial intelligence to analyze the data that has been collected and to activate various alarms regarding non-compliances or the timeliness of the data. [21]

The pharmaceutical production process nowadays has a variety of systems for the application and management of GMP, as well as the ability to gather information in real time. However, 70% of the data collected is not used anywhere and there is generally a waste of data resources. But at a time when optimization is in demand, many manufacturers are seeing a lot of opportunities and benefits of using this data in relation to GMP functions and processes.

We must not forget, of course, that the possibility of using this data, however, given the plethora of independent systems, each with its own proprietary and unique form, is not insignificant. Industry 4.0 promises to solve many of these problems with technologies such as big data and Industrial Internet of Things (I.I.o.T.), which will have a single protocol that everyone will follow.

As we saw in the preface, the term Industry 4.0 will completely change the industries as it is believed. The pharmaceutical industry will not be left out. Thus, the term Pharma 4.0 has appeared in terms of the production of medicinal products. In the definition but also in the form, the importance of the big data but also of the artificial intelligence is shown. [22]

1.3.2 Big Data and Pharma 4.0

In industry 4.0 and consequently in Pharma 4.0 a production environment is fully connected. Every operation is connected in every equipment and transmits and receives data in real time. This happens at the full factory level from the operating systems to the units and production lines. The result as it is perceived, the volume of data collected is huge and varies from time series data to batches of drugs. Even a medium-sized installation can produce data of even tens of Petabytes of data. If these numbers seem excessive then we can imagine what will happen even with the introduction of new equipment. For and for this reason the storage in the form of big data is necessary so that there is direct access to the history in real time, and immediately. [23]

Managing this information and variety of data is a very important task and the smart systems we have are not easy to do effectively. The cost in particular but also the effort to maintain all this information throughout the flow of good pharmaceutical practice such as how to access this data, how to access it, how to create security data even and how to delete it, can to quickly make this process unmanageable. These quantities of data constantly require new investments in data centers, security data centers, and high-end IT services. All this is growing exponentially as more and more data emerges and new and new upgrades involve new costs. Of course, instead of investing in new indoor installations we can have data installations based on the cloud. [24]

1.3.3 Cloud

A company that chooses to keep its data online has the following options from the services offered. Infrastructure as a Service. External partners there have the physical material required, as well as the software to maintain the archive of big data. Computers, hard drives, servers, software as a site, access, data security, data lifecycle and backup, are offered as a service by external partners. The Service Platform as a Service is also offered by external partners without the need for a company to buy it. The company offers the software and its development. Such companies are amazon IBM and Microsoft. Finally, the Software as a Service which is a software which is not installed locally in some computers but in the cloud. This is something very different from the applications we have installed on our computer. All of these solutions provide new and new solutions. [25]

All of these alternatives can provide methods that can cover all industries and all types of businesses at a reasonable cost. These services provide a new way of storing big data, with access from anywhere in the world. Thus, the storage station can be on one continent and the

processing can be done in another even from another provider. All this safely with the control route guaranteed. They can and do provide the required computing power but also the ability to store data, which can be processed by artificial intelligence algorithms, because they were developed for the use and processing of large amounts of big data, a prerequisite for Industry 4.0 and by extension Pharma 4.0. [26]

1.3.4 Big Data and A.I.

Simple data collection alone can't produce knowledge and most industries do just that. They simply collect data required by the standard. Data requires processing and information must be turned into knowledge. This requires methodologies and a continuous plan. Big data is getting bigger and bigger all the time. Older 1Terrabyte of data was considered a huge number. Today 1 Petabyte is something completely normal. With this huge volume of data, the classical statistical methods such as student test or x^2 have no significance. Most are unstructured in various forms such as images, numbers, comments and even sounds. There the classical methods of analysis become useless. Another problem is that the processing power needed to analyze data in Petabytes is impossible to have at the server level. [27]

Artificial Intelligence and machine learning models can manage these vast amounts of unstructured data with simple algorithms such as random forums or more complex algorithms. Although most of these algorithms are not new (all discovered in 60's and 70's), they become incredibly efficient when combined with high computing power and huge amounts of data when combined in cloud. These three elements (big data, computational power and artificial intelligence or machine learning), can apply a new way of applying science. We can monitor production processes where the equipment, the operators, the systems are all interconnected and constantly produce data. They lay the groundwork for validating processes more easily and efficiently. [28]

The current best pharmaceutical practices of pharmaceutical production use its statistics and methods for specific control process parameters and quality control. With these it is possible to monitor the quality of the product. However, all these factors that are monitored depend on internal and external factors but also on the way they behave. In order to really have the best quality of a product, we must take into account, in addition to the control process and quality control, all aspects of production and how the factors of this process interact. [29]

Big data and artificial intelligence can do the whole process. They can find effects and interdependencies beyond what classical statistics as a science can support. So, this new model has been adopted by many companies that find solutions through it. The pharmaceutical industry is already promoting artificial intelligence as a tool for research and development and the production process. Neural networks, which are a part of artificial intelligence, have recently been proposed by European pharmacopoeias as valid chemical assay measurement techniques for advanced chemical methods. [30]

Today, there is no common framework, an integrated information model system, for the construction of artificial intelligence solutions. As a result, the data is not in a common format

and is not structured. Therefore, that framework that exists in automation and manufacturing systems is lost.

Of course, unstructured data does not necessarily require a common format. There is no need for a complete model. As we have seen, artificial intelligence and machine learning algorithms perform better on unstructured data. This happens if the big data is from different sources and different formats. But because having a framework and a regulation is useful for interpreting data and creating artificial intelligence models, future solutions must be prioritized by sensors from automation systems. Essentially this is how big data overlaps. Overlay does not substantially restructure the data. It just makes them show and be bounded. Overlays can provide multiple labels to big data sets. These can be, for example, the model of the equipment, the batch, the type of medicine, the prescription. Thus, by performing artificial intelligence algorithms we can reveal patterns and hierarchical motifs that even indicate relationships between big data that under other circumstances would not have been possible. [31]

1.3.5 Security

All the traditional applications we know, the operating system, storage units, system architecture and more, are installed within the infrastructure of a company. In such an environment where all services are hosted on the cloud, there are no physical servers or storage devices installed on site. There is not even a server serial number registered on the network. Thus, there is no need to develop a security data retrieval process. But the most important of all is that the data collected is stored and processed and what can be deleted after it has been processed, can't be managed with classic methods of computer systems approaches. Cloud technology can provide any storage computing capacity at any time. Depending on the production of the machines that are at all times active and recording. This whole system is called flexible computing and offers theoretically infinite storage space and computing power, something that artificial intelligence needs. All this at a much lower and negligible cost in relation to the need to install such systems locally in a pharmaceutical industry. [32] [33]

So, since everything is a matter of cost-benefit in the final form, it also introduces new competing tools such as full-flow encryption, on-demand processing systems, well-protected infrastructure, pay-per-user, day and time. Most estimates are that many companies in the coming years will transfer many of their operations to the cloud. All their systems as well as subsystems will have been transferred to connected infrastructures. Cloud-based solutions have shown great robustness and security compared to traditional firmware solutions that are stored locally. Some security issues that may arise tend to be resolved with new encryption systems. This can be evidenced by the rapid adoption of such systems by the banking sector, the automotive industry, the health services, but also the financial sector. All of these areas require a high degree of flexibility, trust and immediacy in such technological infrastructures.

Compared to other industries and those mentioned above, the adoption of such solutions is significantly delayed for application in the pharmaceutical industry. Adoption of cloud-based solutions is delayed due to quality and safety concerns. The adoption of such solutions by the pharmaceutical industry due to the GMP often requires risk analysis. This requires risk analysis

and for the provider, and the risk of firmware dependence on it. This calls into question the full benefit of switching all processes to the cloud. This often challenges the traditional mentality of the pharmaceutical industry and its entire material existence.

Unfortunately, there is no naturalness in these solutions provided by the cloud. There is no system hardware, no servers, no serial numbers and systems from operators and operators for verification. There are no operational aspects of data backup or backup procedures, such as management and compliance and information and their storage should be provided by the service provider providing the associated. The pharmaceutical company must accept that data storage and computing power systems are provided on demand in a server-free environment. This is very difficult and often contradicts the quality procedures of the pharmaceutical industry. [34]

1.3.6 Blockchain in pharmaceutical sector.

Blockchain could be considered a public catholic network that all committed transactions are stored in a block list. This chain is growing as new blocks are constantly added to it. Technologies such as asymmetric cryptography and distributed aggregation algorithms are implemented for the security of users so that the universal secure element is consistent. Today cryptocurrencies have become something household e.g. in both industry and academia. As one of the most successful cryptocurrencies, Bitcoin has been hugely successful with its capital market reaching the latest all-time high over $ 68,000 with specially designed data storage structure, Bitcoin trading network could happen without a third party and kernel. The technology for making Bitcoin is blockchain, which was the first was proposed in 2008 and implemented in 2009 by Satoshi Nakamoto. [35]

The uses of blockchain technology are practically unlimited from the financial sector, to transport, to the health sector. In the pharmaceutical industry it can be applied to the entire drug supply chain and is very powerful. Drugs are started from the raw material, active substances and excipients and even packaging materials, developed, produced, and often transported to drugstores before being further distributed to pharmacies, hospitals, and retail companies before being dispensed to patients. Having blockchain technology at our disposal we can verify the integrity of the drug supply chain and enhance the development of new drugs, utilizing blockchain technology to support and manage the drug development process. While currently counterfeit and degraded pharmaceuticals or drugs that do not meet good pharmaceutical standards entering the legal supply chain pose a significant threat to public health, blockchain technology can improve current processes by accounting for, smart contracts ensuring to the fullest the possibilities provided by this new technology on its own or in combination with other existing ones. [32] [36]

1.3.7 Minimum level of security and Minimum level of services

In case cloud services are adopted where the hardware and software are outsourced, the minimum level of services that must be provided is the end-user control mechanism. Quality can be ensured, with supplier certification, inspections and periodic inspections. The use of suppliers that already have some certifications and compliance practices has a significant advantage. Procedures are transferred as is and there are fewer quality risks. Of course, it is equally important that the pharmaceutical companies do not depend entirely on the supplier but that they themselves have a consistent program of controls to confirm compliance.

Cybersecurity and all the security issues around it is a major concern and often make businesses not trust the services and systems that are in the cloud. However, the services in the continuum have now entered a phase of maturation and are often considered safer than internal infrastructure. The most secure security systems have been adopted, the most secure protocols are applied. At the same time, cybersecurity issues are dealt with by many specialists who strive for continuous improvement. Many of the methods exist and are already applied to companies that take data security very seriously, such as the banking sector and hospitals that have adopted cloud-based solutions. [37]

2 CHAPTER 2: SPUMONI & ALCOA

In this paragraph an attempt is made to present SPUMONI. The SPUMONI is a European program to ensure the reliability and the end-to-end traceability of data produced by an industry. The main component of this program to achieve its purpose is ALCOA acronym. The ALCOA acronym is an abbreviation that means Attributable, Legible, Contemporaneous, Original and Accurate. The ALCOA it is used by industries because it ensures that data achieves the essential elements of quality and also helps to ensure data integrity.

2.1 SPUMONI, the European project about Big Data and process modelling for smart industry [32]

The pharmaceutical industry as we have seen through good pharmaceutical practice requires possessive and effective techniques for controlling and locating the production process of a drug. Today with the existing techniques there is no guarantee that the data can't be falsified. The resulting data is large and this makes the process even more complicated and vulnerable. All of this requires a more holistic approach to data integrity. It must be based on the principles of ALCOA — an abbreviation for data properties rendered readable, up-to-date, original, and accurate — has been proposed by the pharmaceutical industry as a framework for ensuring data integrity.

The pharmaceutical industry needs a new autonomous mechanism for collecting the resulting data. A mechanism that can guarantee governance, compliance, transparency and data traceability. This mechanism must include quality techniques that ensure falsified and non-falsified data. The mechanism must also be in a position to be able to detect random or systematic errors arising from data flows. One such mechanism is the SPUMONI. This mechanism is funded by the European Union. Using open-source tools, using best practices ensures the traceability of data based on ALCOA principles. [33] [34] [35] [38] [39] [40]

2.2 ALCOA principles.

The acronym ALCOA was first used as a definition by Stan W. Woolen of the FDA Bureau in early 1990s. ALCOA is used by various industries as the framework and regulation for ensuring data integrity. It is necessary to ensure Good Documentation Practices (GDP) which helps (GMP). The acronym ALCOA is not limited to electronic data generated by electronic devices. It also applies to printed data. ALCOA principles are necessary for the life cycle e. g. paper complaints and / or electronic data management, complies with GDP, complies with GMPs and leads data integrity initiatives.

Attributable

Electronic data should be able to be distorted and attributed only to the person who is the data producer, including how that person performed an action and when. E.g. a change in the file or in a physicochemical constant that affects the file. Attributable can be done manually by initialing and documenting a document on paper or by a control path in an electronic system.

Legible

All electronic data must be easy to read, legible and permanent. Once the files are legible and permanent this helps to make them accessible for as long as the electronic data is available,

including the storage of the electronic data in time. That is, theoretically for as long as they exist in time, we need them even in infinity.

Contemporary

To call something modern means to record electronic data at the time it is executed. The date and time must be marked in the order in which the data is executed so that we can say that the data is reliable. The data should never have an older date than the actual one with the expected results before execution.

Original

The original data is an electronic medium e.g. an industry sensor in which the data point is initially recorded, including production protocol, format to be taken, notes taken, calculations made as in an excel sheet, database or any software application used. It should be easy to understand exactly where the original data is being created to ensure that the content and meaning are kept intact as at the time they were created.

Accurate

For electronic data to be accurate, all data must be error-free and bug free, complete, 100% true, and reflect the order of observation that ensures the production line. Processing should only take place using the principles of GDP. [36] [37]

The structure of the ALCOA mechanism includes:

1. Preventing data falsification using quality control.
2. If there are concerns about confidentiality, compliance with standards and the source - ownership of the data there is assurance of traceability and these concerns cease to exist.
3. There is smart control. The data are collected and processed in various contexts and environments as they exist in the pharmaceutical industry.

The documentation of the clumsy mechanism comes as a consequence of:

- Ensuring the accuracy and reliability of product production includes a method of analyzing how this data changes over time from multiple input sources.
- The system is secured in terms of data management, data collection, handling, as well as flexible data integrity checks close to the source that produces the data, which allows valid prediction for possible deviations. This greatly reduces the costs and delays that arise during the production process.
- Taking advantage of blockchain technology, which ensures end-to-end traceability and immutability of the pharmaceutical industry. Exploiting a network of Ethereum protects the SPUMONI consortium from security threats. [38]

3 CHAPTER 3: Machine Learning & Deep Learning

This section summarizes the three known models of artificial intelligence used in this dissertation. The presentation was made briefly since the material that exists for these models in various papers scientific journals and articles is truly infinite. These models are the LSTM the GRU and the Bi-LSTM.

3.1 Machine Learning

Machine learning is a subset of artificial intelligence that can make machines learn without programming. The practice of artificial intelligence is mechanical learning. Computer systems can, through machine learning, perform functions such as grouping, reduction, regression, and the ability of computers to recognize patterns. The arithmetic process is achieved using various machine learning algorithms or arithmetic structures for the analysis of information from data. This information is classified by various characteristics called feats. Machine learning is used to find the relationship between features and values of an output called labels. This technique is considered to be ideal for problems such as regression data, data classification, and setting rules for data collection and correlation. The most important are K-means, linear regression but also our well-known neural networks and many more.

The two main challenges and common problems that remain to be solved are:

- Data storage, as neural networks and memory networks require a large working memory to process and store data.
- Natural language processing there is still a lot of work to be done to achieve natural language processing as well as comprehension. Of course, in recent years, very important steps have been taken. [41]

3.2 Deep Learning

Deep learning models represent a relatively new model for the artificial intelligence industry. Of course, as a definition and as an idea it is not new. In 1943, a new mathematical model of a neuron was proposed by McCulloch and Pitts (1943). This model provided an abstract formulation for the function of a neuron without interfering with the function of a real biological neuron. The most interesting thing is that this neuron model was not proposed for learning. In 1949, Hebb proposed a learning model named Hebbian as the first idea for learning with biological motives, in neural networks. Artificial neural networks are mathematical models that have mimicked the function of the human brain. The models we use are not aimed at producing biologically realistic models. The purpose of artificial neural networks is to analyze data. The basic entity of such a neural network is the model of an artificial neuron. [42]

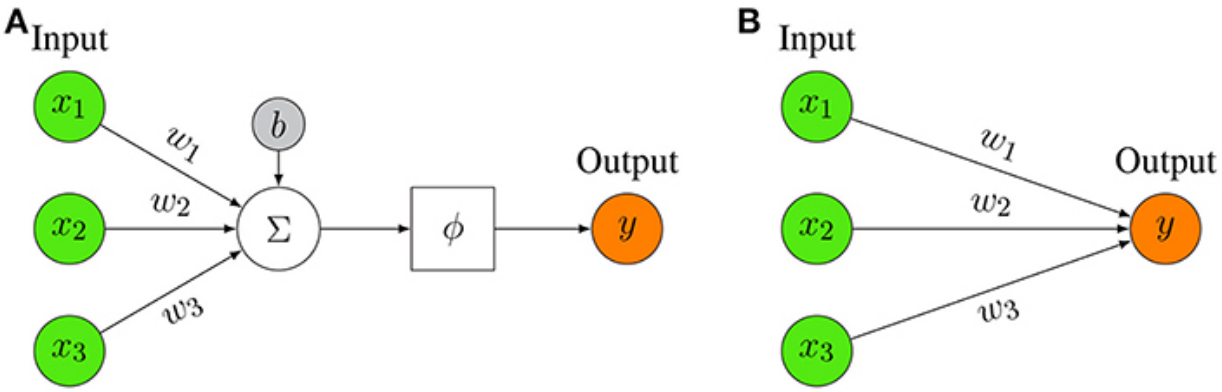


Figure 1. Representation of a mathematical artificial neuron model. The input to the neuron is summed up and filtered by activation function. (B) Simplified Representation of an artificial neuron model. Only the key elements are depicted, i.e., the input, the output, and the weights. https://www.frontiersin.org/articles/10.3389/frai.2020.00004/full

3.3 RNN

RNN networks are a class of neural networks that are suitable for processing time series and other data that are in sequence. So RNN networks are deep learning algorithms that feedback its output for time-based memory production. This internal memory in relation to other neural networks allows the RNN network to process dynamic input sequences.

So, their feature of RNN networks is that they send information over time. Their structure which has an additional port for parameters for time sequence connections and so they can enter and train in the unit of time. They take advantage of this extra port. RNN networks can, once trained, produce output that at any given time is based on the port receiving information from previous time steps. Thus, the data remain classified and are influenced by the sensitivity of the environment in which they have been received in the unit of time. The data for a moment is related to the data of a previous moment. The difference between an RNN network and a neural network can be seen in the figure. [43]

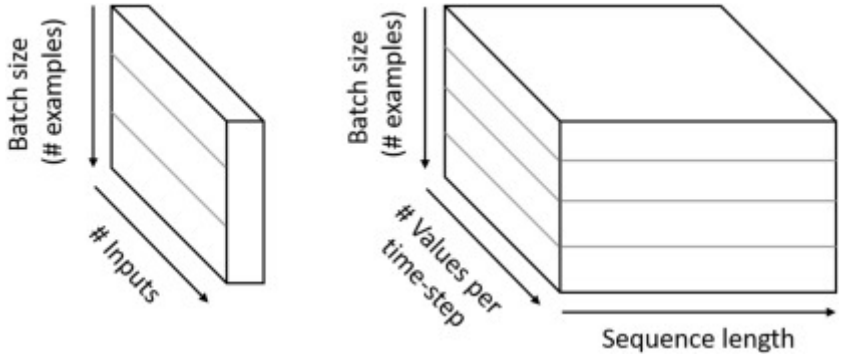


Figure 2. Comparison between normal and recurrent neural networks input vector. Adapted from Patterson, J. Gibson, A. Deep learning: a practitioner's approach. O'Reilly Media, Inc.; 2007.

Because RNN networks introduce both the time and temporal sequence parameter, we have a connection of a neuron to a hidden layer, as a feed stream to the hidden layer neuron. Repeated connections can be illustrated in the following figure.

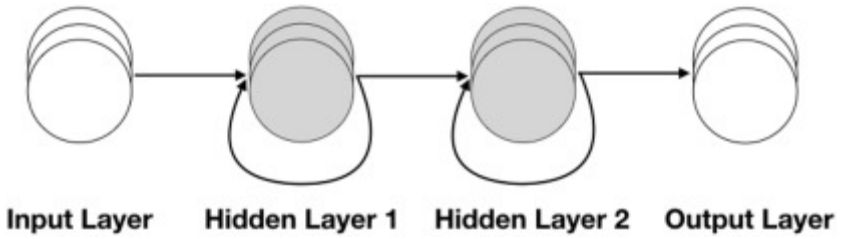


Figure 3. Feed forward flow for recurrent neural networks. Adapted from Patterson, J. Gibson, A. Deep learning: a practitioner's approach. O'Reilly Media, Inc.; 2007.

At each step in the unit of time each neuron of the RNN network is activated taking as input the current vector input but also from the previous states. Thus, the output also takes into account the previous input vectors. [44]

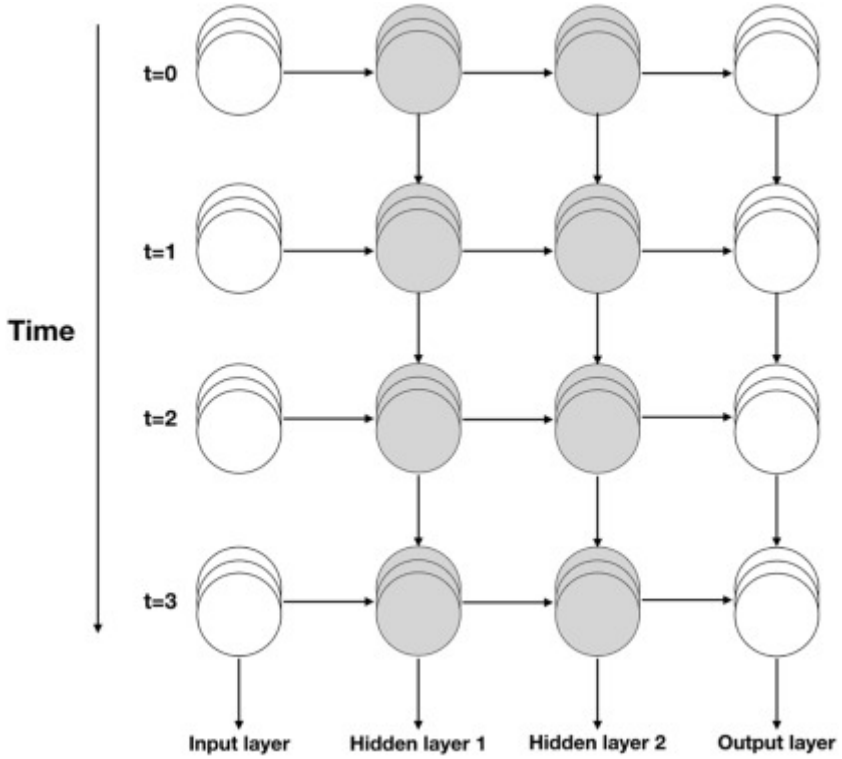


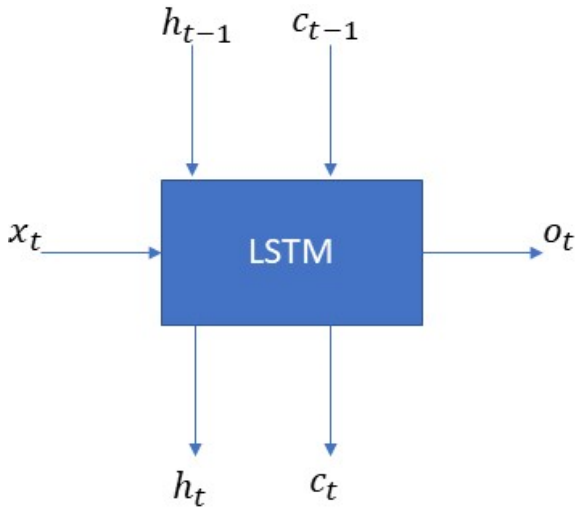
Figure 4. Unrolling for recurrent neural networks. Adapted from Patterson, J. Gibson, A. Deep learning: a practitioner's approach. O'Reilly Media, Inc.; 2007.

3.3.1 LSTM

LSTM networks are a variation of RNN networks of which the firebox is the memory unit. LSTM networks can solve the gradient explosion and the disappearance of inclination that RNN cannot do. In addition, LSTM networks can capture long distance dependence. The LSTM

memory module consists of memory ports, input ports, and output ports, which are used to discard or store information. [45]

An LSTM network can be described by the following mathematical formulas:



$$f_t = \sigma_g\,(W_f \times x_t + U_f \times h_{t-1} + b_f)$$
$$i_t = \sigma_g\,(W_i \times x_t + U_i \times h_{t-1} + b_i)$$
$$o_t = \sigma_g\,(W_o \times x_t + U_o \times h_{t-1} + b_o)$$
$$c'_t = \sigma_c\,(W_c \times x_t + U_c \times h_{t-1} + b_c)$$
$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t$$
$$h_t = o_t \cdot \sigma_c(c_t)$$

f_t *is the forget gate*
i_t *is the input gate*
o_t *is the output gate*
c_t *is the cell state*
h_t *is the hidden state*

Figure 5. LSTM input outputs and the corresponding equations for a single timestep. https://towardsai.net/p/machine-learning/tutorial-on-lstm-a-computational-perspective-f3417442c2cd

3.3.2 Bi-LSTM

Bi-LSTM is a sequence processing model consisting of two LSTM networks. One enters in one direction forward and the other in one direction backwards. The Bi-LSTM network effectively increases the volume of information added to this network, and thus improving the algorithm. For example, a Bi-LSTM network of natural language processing knows in this way which word follows and which word precedes a sentence. [46]

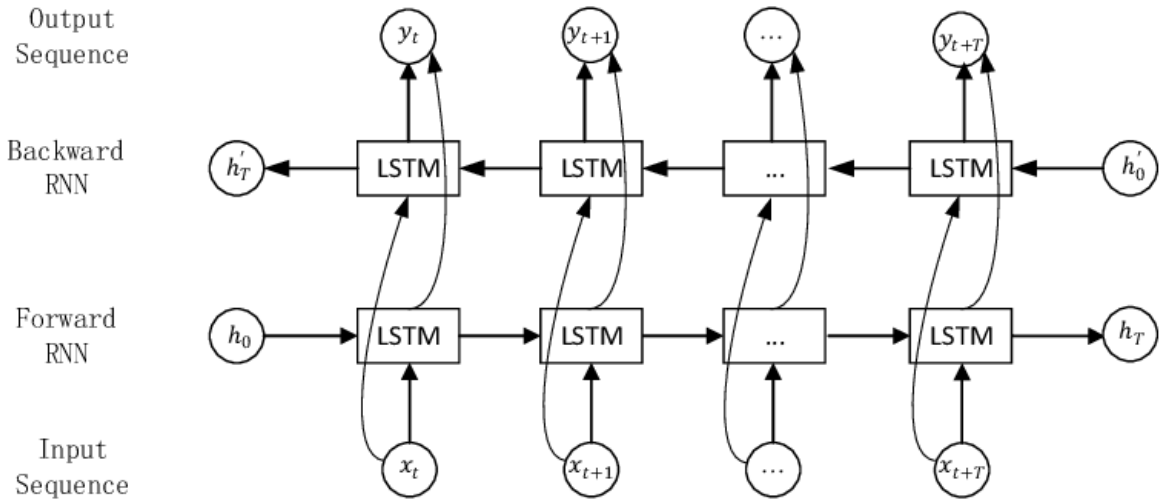
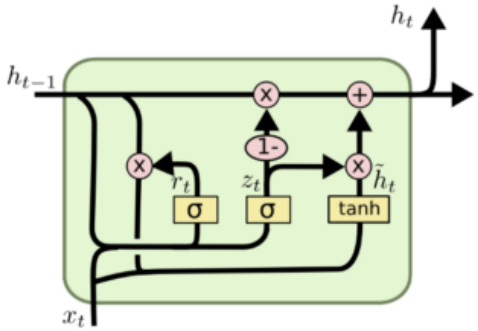


Figure 6. Bi-LSTM neural network structure deployed in time direction. https://www.researchgate.net/publication/339679582_Multi-time_scale_wind_speed_prediction_based_on_WT-bi-LSTM

3.3.3 GRU

The GRU network was first proposed by Kyunghyun Cho in a work on translating texts using neural networks. It is essentially a network. The only difference is that it contains two gates. The update gateway and the recovery gateway. The update gate checks the information flowing to the memory and the reset memory checks the information flowing out of the memory. The update gate helps the model determine how much of the information from the past (from previous time steps) should be transmitted in the future. This is really powerful because the model can decide to copy all the information from the past and eliminate the risk of the gradient problem disappearing. As with LSTM networks, the GRU network has gateway units that regulate the flow of information inside the unit but without a memory cell. [47]



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Figure 7. Depth Gated RNNs by Yao, et al. (2015). There's also some completely different approach to tackling long-term dependencies, like Clockwork RNNs by Koutnik, et al. (2014). http://colah.github.io/posts/2015-08-Understanding-LSTMs/

4 Data Prepossessing

The data for this postgraduate thesis were obtained from a well-known Italian pharmaceutical company participating in the program Spumoni. The data were taken from two production lines of its pharmaceutical industry. No other data given because the data is classified. The main data is divided into two files, the I 600 production line and the I 1000 production line.

These files are .tsv files as they produced from the production line in raw data. These files contain the batch numbers of the drugs produced. During the production of the drug, some alarms were detected, they were recorded in time series for each batch separately. Alarms are the speed, pressure, temperature, humidity and other alarms that can occur during the process of a drug production line. This was done for both production lines so that there is a large sample of alarms and more batches.

4.1 I 1000 Production Line Dataset description

The dataset from the I 1000 production line has 176 batches of drugs. For the attributable the values we take through the sensors are from 25.0 as the lowest value to 50.0 which is the highest value. The average price is 48.17. The standard deviation is 2.139. We have data for all 176 batches and no batch is missing. For the contemporaneous we have 176 batches also and no price is missing too for the production line I 1000. The minimum price is 62.00 and the maximum price is 95.00. The average price is 87.64 and the standard deviation of the prices is 6.32.

	Legible	Attributable	Contemporaneous	Original	Accurate	Complete	Consistent	Available	Enduring
count	176.000000	176.000000	176.000000	176.0	176.000000	176.000000	176.0	176.0	176.0
mean	83.102273	48.170455	87.647727	100.0	37.221591	96.761364	100.0	0.0	0.0
std	4.000114	2.139274	6.316490	0.0	6.427778	2.846529	0.0	0.0	0.0
min	75.000000	25.000000	62.000000	100.0	0.000000	84.000000	100.0	0.0	0.0
25%	80.000000	48.000000	86.000000	100.0	36.000000	97.000000	100.0	0.0	0.0
50%	84.000000	48.000000	89.000000	100.0	37.000000	98.000000	100.0	0.0	0.0
75%	86.000000	49.000000	91.000000	100.0	39.000000	98.000000	100.0	0.0	0.0
max	94.000000	50.000000	95.000000	100.0	50.000000	98.000000	100.0	0.0	0.0

Figure 8. I 1000 Production Line dataset description.

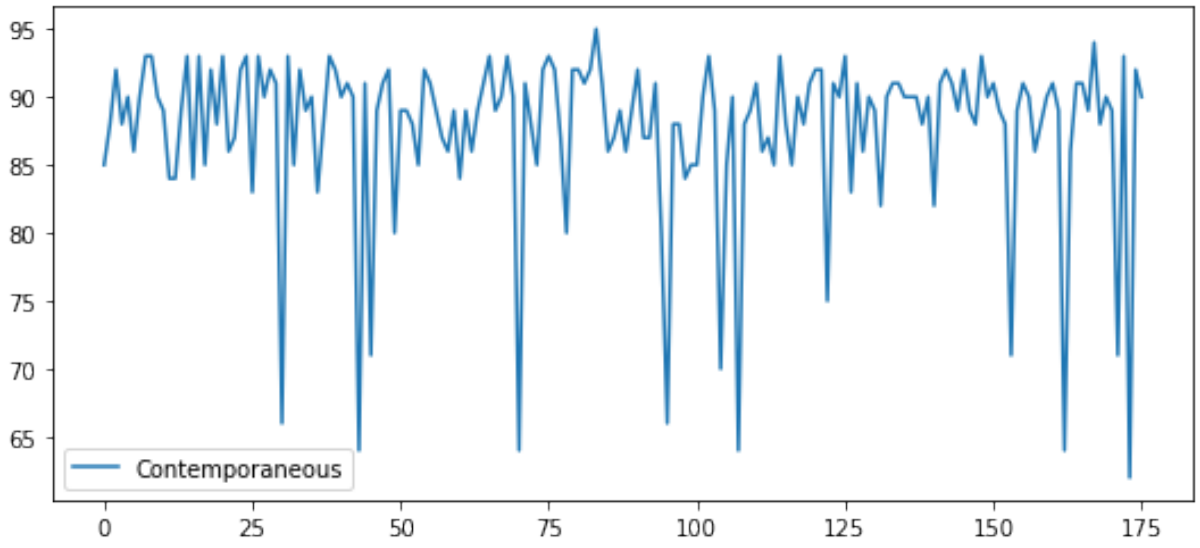


Figure 9. I 1000 Production Line Contemporaneous time series data.

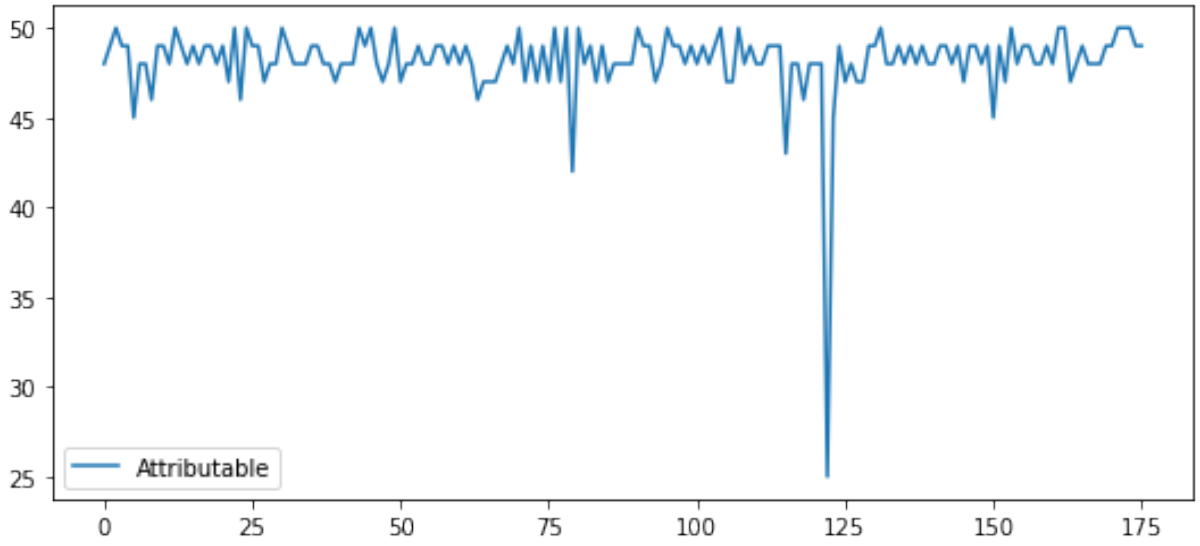


Figure 10. I 1000 Production Line Attributable time series data.

4.2 I 600 Production Line Dataset description

The dataset with the I 600 production line has 296 batches of drugs instead of 176 batches from I 1000 production line dataset. No other information given if it is the same or different drugs like I 1000 dataset for the attributable the values we take through the sensors are from 0.0 as the lowest value to 50.0 which is the highest value. The average price is 47.80 the standard deviation is 3.34. As we can see the dataset is quite imbalanced too as the I 1000 production

line. We have data for all 296 batches, no batch is missing too. For the contemporaneous and for the I 600 production line we have a minimum price 50.0 and the maximum price is 94.0 the mean is 87.44 and the standard deviation is 4.62. No data is missing too and all the 296 prices for batches are available.

	Legible	Attributable	Contemporaneous	Original	Accurate	Complete	Consistent	Available	Enduring
count	296.000000	296.000000	296.000000	296.0	296.000000	296.000000	296.000000	296.0	296.0
mean	84.658784	47.804054	87.449324	100.0	35.405405	97.152027	99.662162	0.0	0.0
std	1.711910	3.344936	4.615831	0.0	5.738762	1.820932	5.812382	0.0	0.0
min	79.000000	0.000000	50.000000	100.0	0.000000	83.000000	0.000000	0.0	0.0
25%	84.000000	48.000000	86.000000	100.0	34.000000	97.000000	100.000000	0.0	0.0
50%	84.000000	48.000000	88.000000	100.0	35.000000	98.000000	100.000000	0.0	0.0
75%	85.000000	49.000000	90.000000	100.0	38.000000	98.000000	100.000000	0.0	0.0
max	91.000000	50.000000	94.000000	100.0	100.000000	98.000000	100.000000	0.0	0.0

Figure 11. I 600 Production Line dataset description

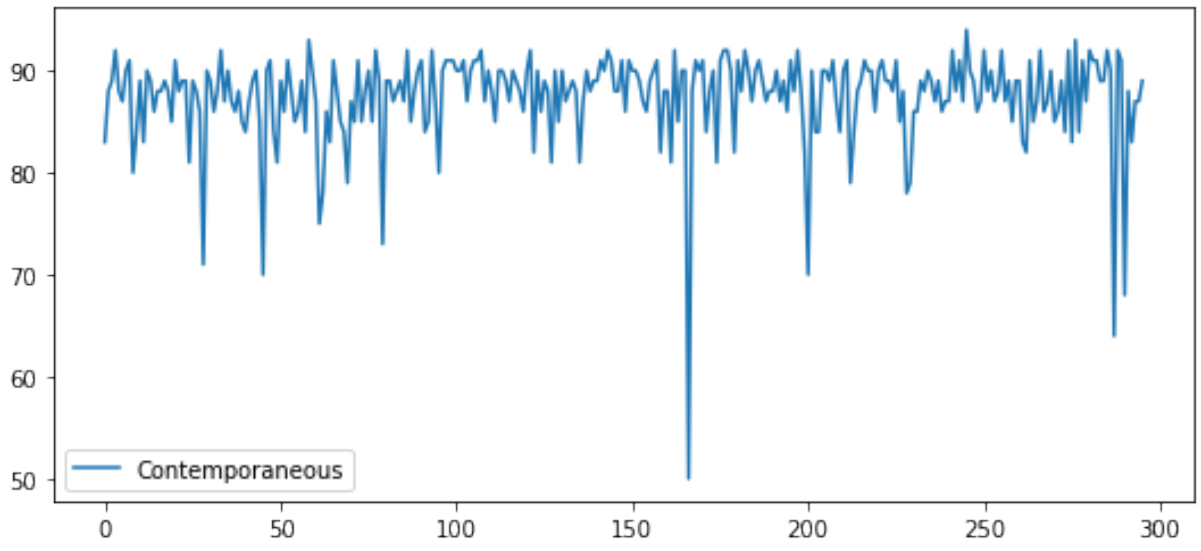


Figure 12. I 600 Production Line Contemporaneous time series data.

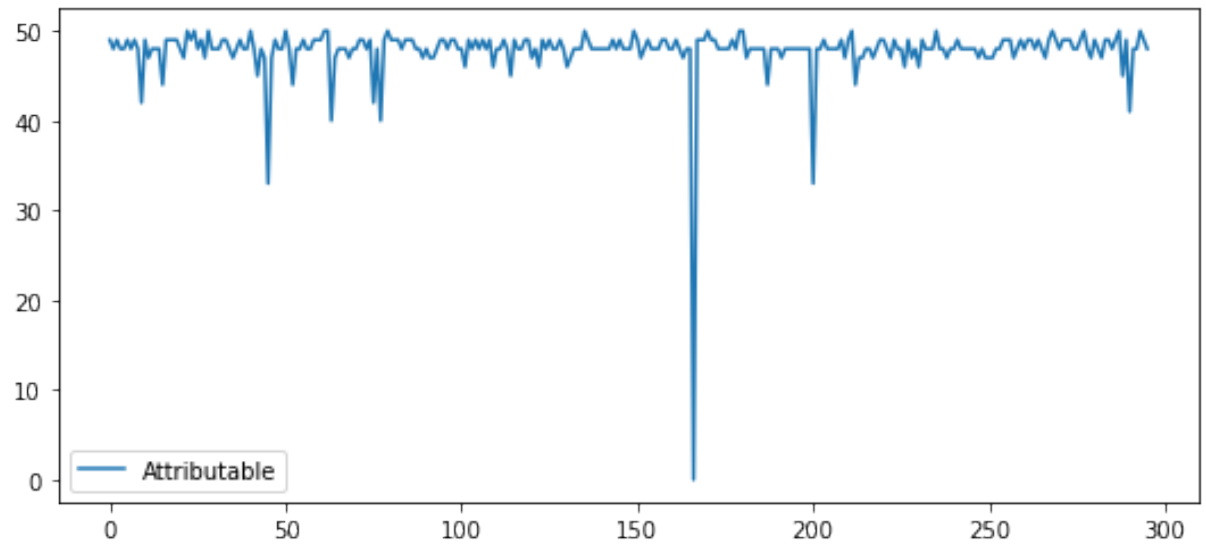


Figure 13. I 600 Production Line Attributable time series data.

The other main part of the data is that of the ALCOA's. For each batch of drug that was made, there is a .json file which contains the ALCOA values for each alarm generated by the sensors. For this reason, the archives of the ALCOA and the archives of the alarms were merged into one archive where each alarm passes some ALCOA values. After that the files resulting from the merging of the alarms and the ALCOA were organized into dictionaries. This improved the readability of the code and allowed it to detect errors. It also allowed the models to be positioned more efficiently and improved the speed of data analysis.

Another problem that arose is that many of the values from the sensors were missing. This did not allow the data to take the form we want so that it could be inserted into a neural network. The way we handled this is to replace the empty values with zeros. In neural networks it is generally safe to enter the missing values as zero provided that zero is not a significant value. The network will learn as it analyzes the prices that the value zero does not mean that data is missing and will start ignoring the prices. [48]. It should be noted that if the network is trained in data without missing values the network will not learn to ignore them. In this case, scanty training samples must be created with the missing values. In this case we can copy some values many times for training and discard some of the functions that the test data is expected to have.

As we can see from the table above, the data for the attributable and the contemporaneous are very imbalanced. Many machine learning algorithms such RNNs and LSTM Networks perform better or converge faster when the features are on a relatively similar scale and / or close to the normal distribution. To solve this problem and in order to normalize the data we will use two scaling algorithms for continuous variables. The Minmax Scaler and the Standard Scaler algorithms. The Minmax Scaler is a scaling algorithm that scales the maximum and minimum values to 0 and 1 respectively. The Standard Scaler scales the values between min and max so that they fall within a range from min to max. [49]

5 Results and Discussion

For the needs of this diploma, 3 models it is used for the prediction of **Attributable** and **Contemporaneous** of the ALCOA acronym. These models are LSTM Model, the GRU Model, and Bi-LSTM Model. The reason these three LSTM models' networks were chosen instead of a simple RNN is because these networks are good at finding relationships between continuous data points often at different lengths of time frames. As are the prices from the sensors in our dataset. They are able to recognize very good patterns that result from seasonality. They have a big advantage over other regression models as they look at the recent past in relation to for example a random forest which works well categorizing the data and detects some seasonality but examines them in series regardless of time. In order to be able to compare the performance of the 3 neural networks we will compare the mean absolute error (MAE) and the validation mean absolute error (VMAE) in all three networks. The LSTM the GRU and the Bi-LSTM.

5.1 Attributable principle and I 600 Dataset

Using the above artificial intelligence algorithms, we will detect if we can see if the data is collected and attributed to the person who collects it. For the I 600 dataset and for the attributable the Bi-LSTM Network the MAE is 1.5878 and VMAE is 1.3723. For the GRU Network the MAE is 1.8811 and the VMAE is 1.3154 and for the LSTM Network the MAE is 1.9603 and for the LSTM Network the VMAE is 1.3235. The price 1.5878 of Bi-LSTM Network is the best performance number for all networks.

Network	**MAE**	**VMAE**
Bi-LSTM	**1.5878**	1.3723
GRU	1.8811	1.3154
LSTM	1.9603	1.3235

Table 1. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 600 dataset for with no scaling.

The optimal hyperparameters was loss: mean_squared_error, the optimizer: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-3 for the Bi-LSTM and learning rate set to: 1e-4 for GRU and LSTM Network, lstm_units: 200, epochs: 20, the batch_size: 8, es_patience : 0.5.

HYPERPARAMETERS I 600 dataset no scaling

NETWORK	LOSS	OPTIMIZER	DROPOUT	LR	LSTM UNITS	EPOCHS	BATCH SIZE	ES PATIENCE
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	20	8	0.5
GRU	MSE	ADAM	0,2	1e-4	200	20	8	0.5
LSTM	MSE	ADAM	0,2	1e-4	200	20	8	0.5

Table 2. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 600 dataset with no scaling.

As we see above, the best results were obtained by the Bi-LSTM network with MAE **1.5878** the second was GRU Network with MAE 1.8811 and the third network was the simple LSTM with 1.9603 of MAE. The above results are also confirmed experimentally with the loss and val loss curves for each network separately. As we can see in the diagrams below, Bi- LSTM is superior to the other two neural networks used.

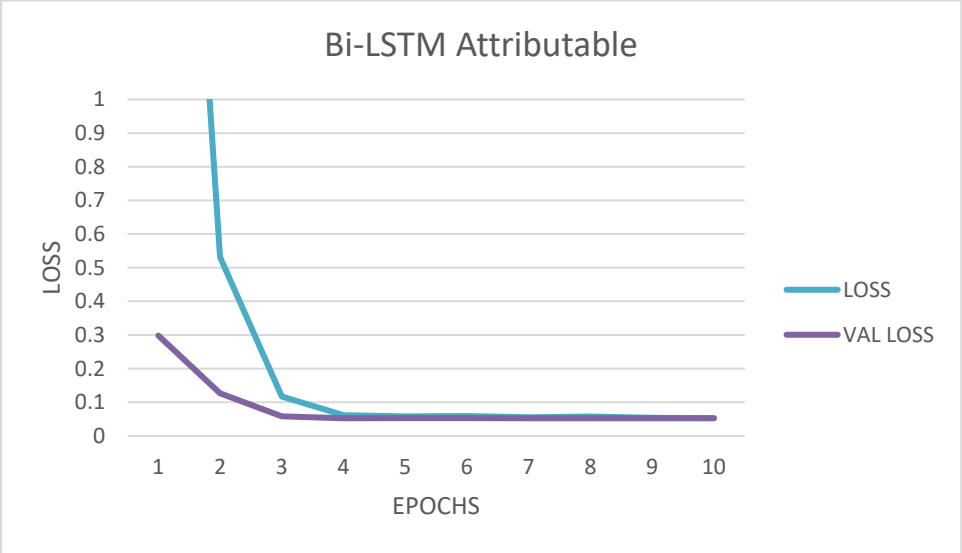


Figure 14. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 600 dataset with no scaling.

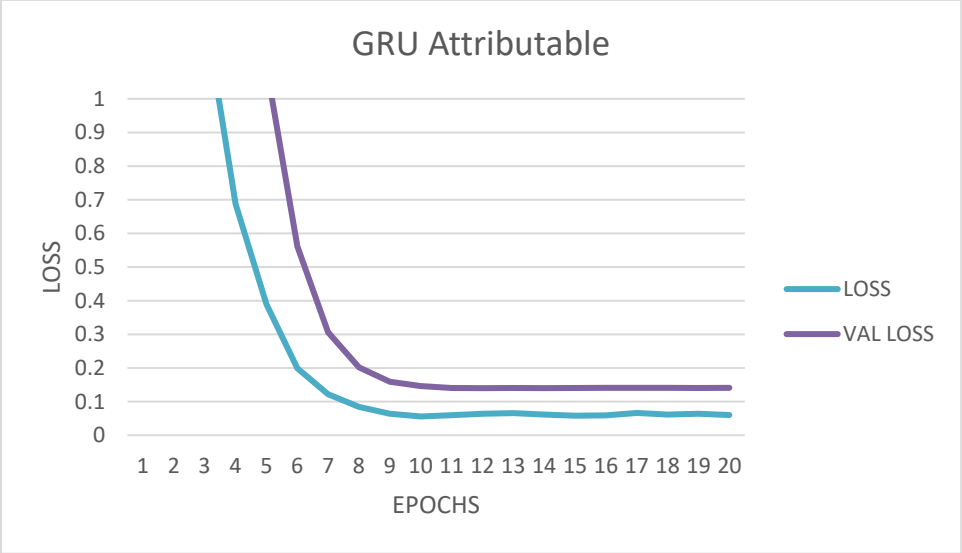


Figure 15. Loss/ Val loss curve of GRU network for Attributable and I 600 dataset with no scaling.

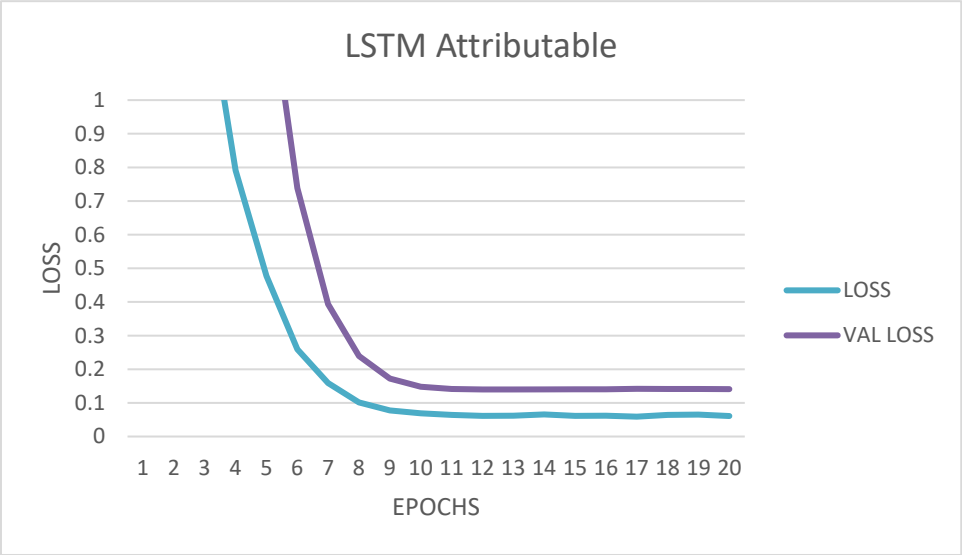


Figure 16. Loss/ Val loss curve of LSTM network for Attributable and I 600 dataset with no scaling.

5.1.1 Minmax Scaler

In order to take better performance to our neural networks we applied Minmax Scaler algorithm to our data. So, for the attributable and the I 600 dataset with Minmax Scaler the Bi-LSTM Network the MAE is 1.6078 and the VMAE for the Bi-LSTM 1.3773. The GRU Network

the MAE is 1.8911 and the VMAE is 1.3454 with Minmax Scaler and for the LSTM Network the MAE is 1.9503 and the VMAE is 1.3435 with the same algorithm.

Network	**MAE**	**VMAE**
Bi-LSTM	**1.6078**	1.3773
GRU	1.8911	1.3454
LSTM	1.9503	1.3435

Table 3. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 600 dataset for with Minmax Scaler algorithm applied.

The hyperparameters for the network in Minmax Scaler was loss: mean_squared_error, the optimizer: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-4 for GRU and LSTM Network and for Bi-LSTM the learning rate is 1e-3, lstm_units set to: 200, epochs:40, the batch_size: 8, es_patience : 0.5.

HYPERPARAMETERS With MinmaxScaler								
NETWORK	LOSS	OPTIMIZER	DROPOUT	LR	LSTM UNITS	EPOCHS	BATCH SIZE	ES PATIENCE
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	20	8	0.5
GRU	MSE	ADAM	0,2	1e-4	200	20	8	0.5
LSTM	MSE	ADAM	0,2	1e-4	200	20	8	0.5

Table 4. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 600 dataset with Minmax Scaler algorithm applied.

As we said before we applied Minmax Scaler Algorithm to improve the performance of our Networks. The best performance with Minmax Scaler Algorithm was again the Bi-LSTM Network. But we see no improvement with the best result was MAE with 1,6078 instead of 1.5878 MAE with no scaler applied. The Minmax Scaling algorithm has no significant effect in performance improvement.

This can be seen experimentally in the graphs that exist below where the Bi-LSTM network is slightly superior but without a significant difference from the other two. The GRU and single LSTM networks are almost identical in performance.

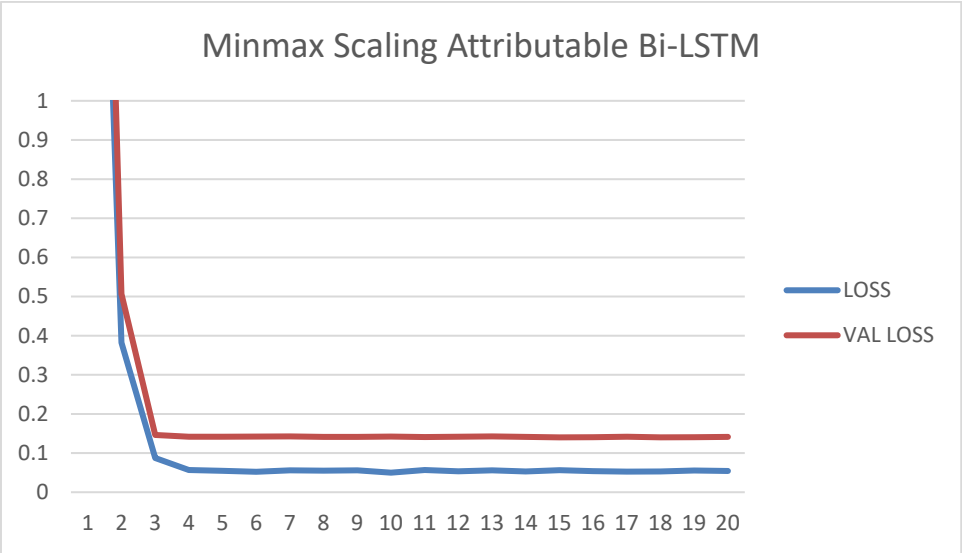


Figure 17. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 600 dataset with with Minmax Scaler algorithm applied.

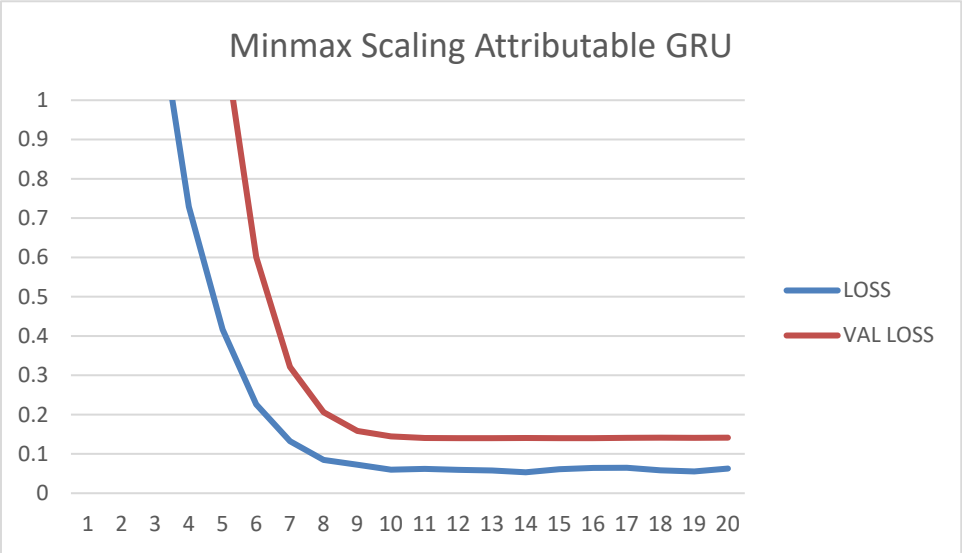


Figure 18. Loss/ Val loss curve of GRU network for Attributable and I 600 dataset with with Minmax Scaler algorithm applied.

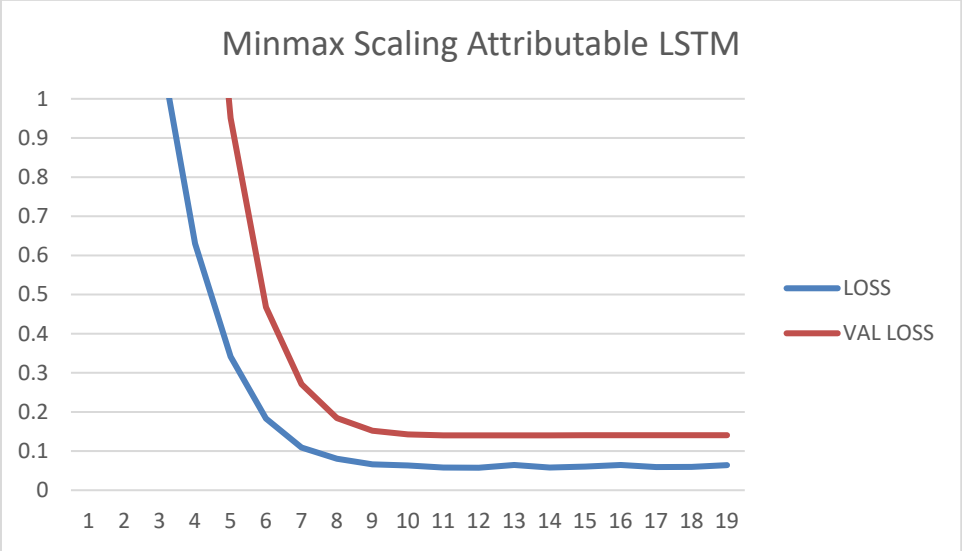


Figure 19. Loss/ Val loss curve of LSTM network for Attributable and I 600 dataset with Minmax Scaler algorithm applied.

5.1.2 Standard Scaler

With Standard Scaler algorithm applied in our dataset the Bi-LSTM has a performance 1.6630 in MAE and 1.3032 in VMAE. The GRU Network has MAE 1.9337 and VMAE 1.2904 with Standard Scaler applied and the LSTM Network the MAE is 1.9970 and the VMAE is 1.2872 with the same algorithm applied.

Network	**MAE**	**VMAE**
Bi-LSTM	**1.6630**	1.3032
GRU	1.9337	1.2904
LSTM	1.9970	1.2872

Table 5. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 600 dataset for with Standard Scaler algorithm applied.

The hyperparameters for the networks in Standard Scaler was loss: mean_squared_error, the optimizer set: adam, the dropout: 0.2, the optimal learning rate set to: 1e-4 for GRU and LSTM and 1e-3 for Bi-LSTM, the lstm_units: 200, epochs:40, the batch_size: 8, es_patience : 0.5.

HYPERPARAMETERS With StandardScaler

NETWORK	LOSS	OPTIMIZER	DROPOUT	LR	LSTM UNITS	EPOCHS	BATCH SIZE	ES PATIENCE
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	20	8	0.5
GRU	MSE	ADAM	0,2	1e-4	200	20	8	0.5
LSTM	MSE	ADAM	0,2	1e-4	200	20	8	0.5

Table 6. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 600 dataset with Standard Scaler algorithm applied.

With Standard Scaler applied to our dataset the best results were achieved with Bi-LSTM Network with 1.6630 MAE but we did not see any good improvement in our results as in Minmax Scaler.

The similarity in the loss and val/ loss graphs where the minmax algorithm was applied is obvious. The Bi-LSTM network shows slightly better results and the GRU and LSTM networks are almost identical in performance.

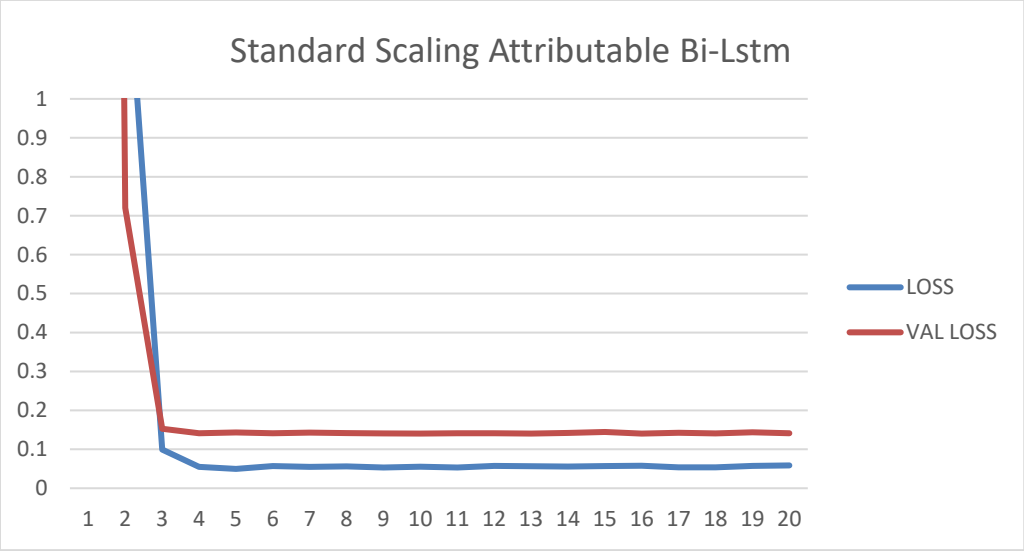


Figure 20. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 600 dataset with Standard Scaler algorithm applied.

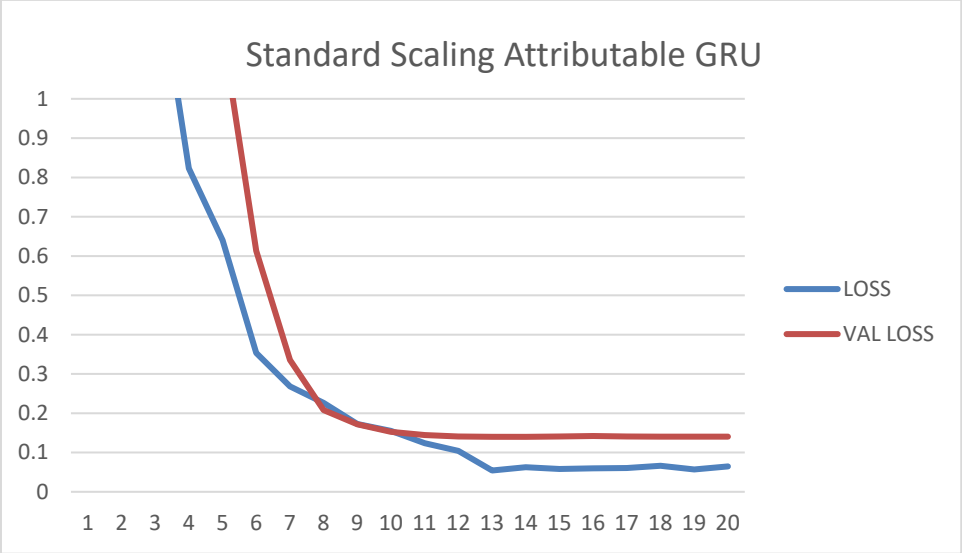


Figure 21. Loss/ Val loss curve of GRU network for Attributable and I 600 dataset with Standard Scaler algorithm applied.

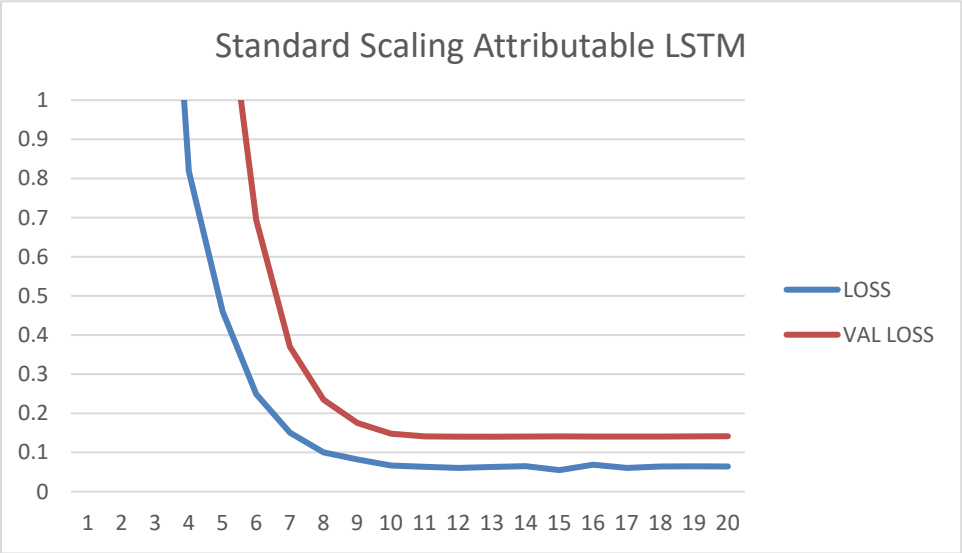


Figure 22. Loss/ Val loss curve of LSTM network for Attributable and I 600 dataset with Standard Scaler algorithm applied.

5.1.3 Prediction from previous Attributable prices for I 600 dataset.

The purpose of this master thesis is to be able to detect if we can predict future ALCOA prices from previous ones. [50] So, we made a class the Regressor Class of it and created and trained a model with a random forest regressor with reference to the 30 previous ALCOA. This means that the model uses the previous thirty ALCOA values for prediction. Since we want to

predict the attributable it makes sense to use the previous attributable 30 values. We can see the performance of the model to the following curve. This model has Test error (MSE): 2.83422

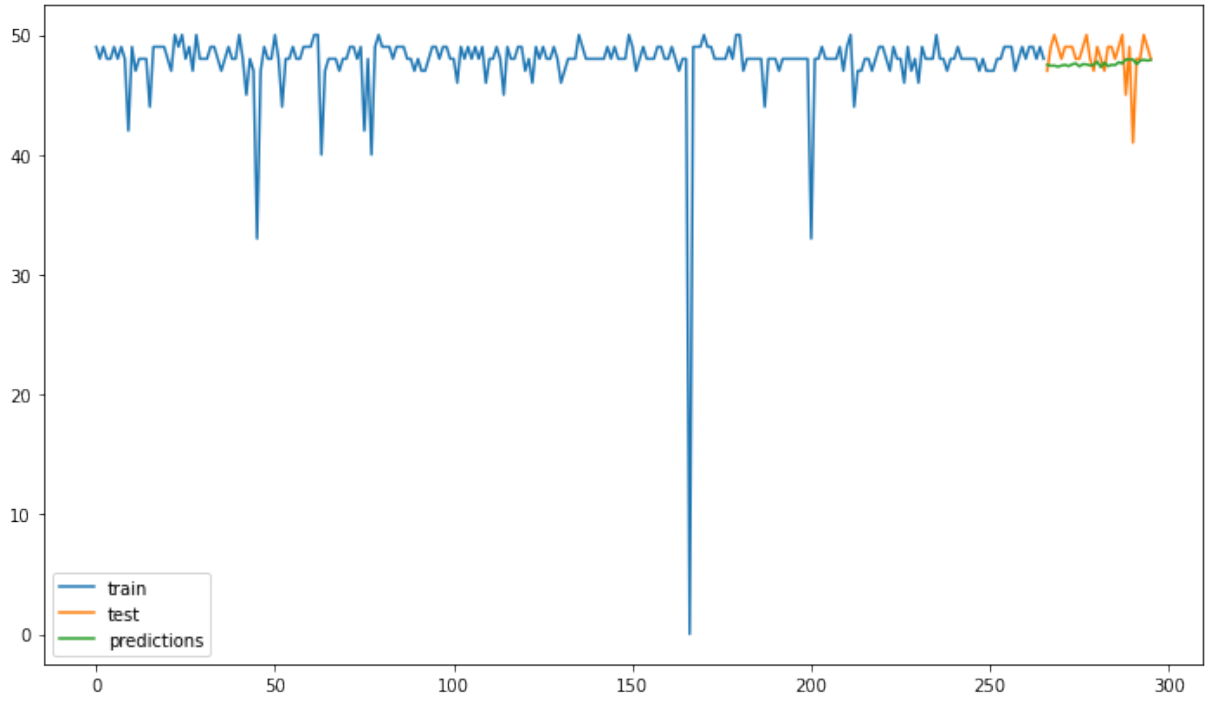


Figure 23. Prediction of Attributable for I 600 dataset with random forest regression model. The performance of the model was: Test error (MSE): 2.83422

As we have seen, the Autoregressive class took into account the last 30 attributable ALCOA values. But no one is sure if these prices are the best. Thus, a new class was created, the search grid search class in a random forest model which is looking for the best parameters. This model with grid search has Test error (MSE): 2.9538063879668854.

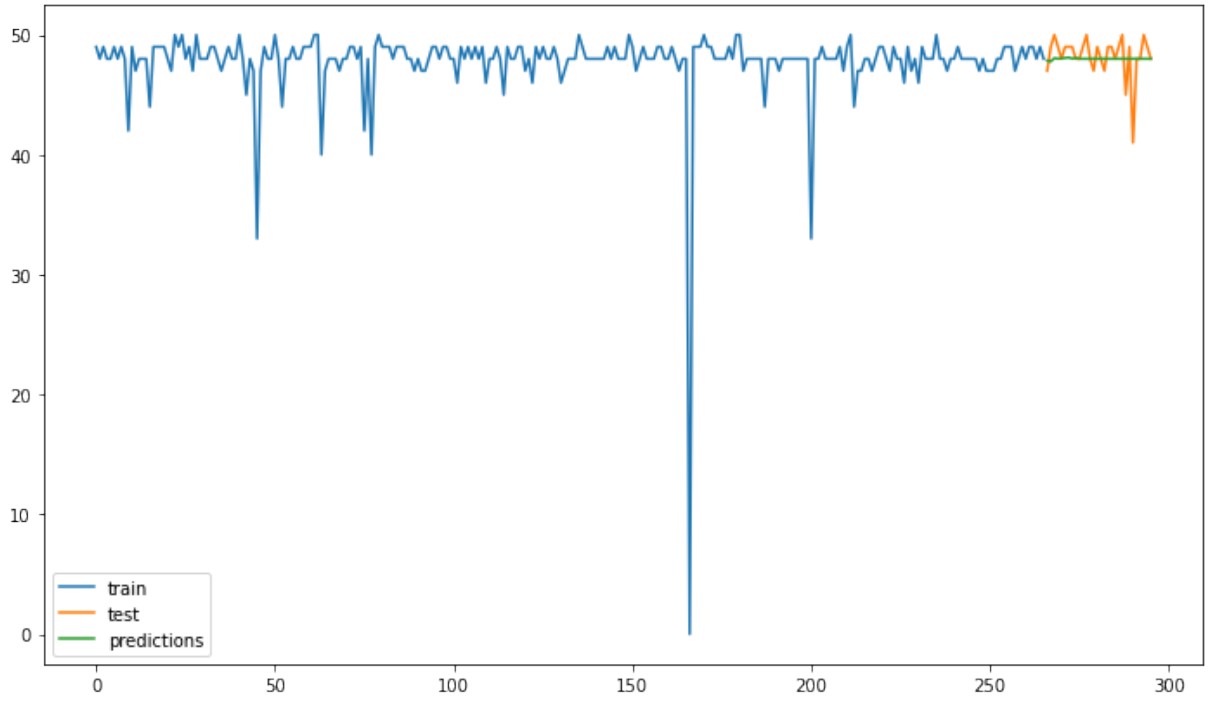


Figure 24. Prediction of Attributable for I 600 dataset with random forest regression model with grid search. The performance of the model was: Test error (MSE): 2.9538063879668854.

Some automated prediction and custom-made models follow a prediction strategy in which one prediction is based on the previous one. So, an alternative is to train the model for each step we need to anticipate. This strategy is the immediate prediction of multiple steps and is more accurate than the retrospective as it requires training of multiple models however in some scenarios it can and gives better results. In contrast to the automatic prediction models with automatic regression where we had to indicate the number of steps. For this we must use a linear model with a lasso value for regression. [51] These models require the standardization of forecasts, so it is combined with a Standard Scaler using Pipeline. [52] This model shows Test error (MSE) 3.5926376838689946

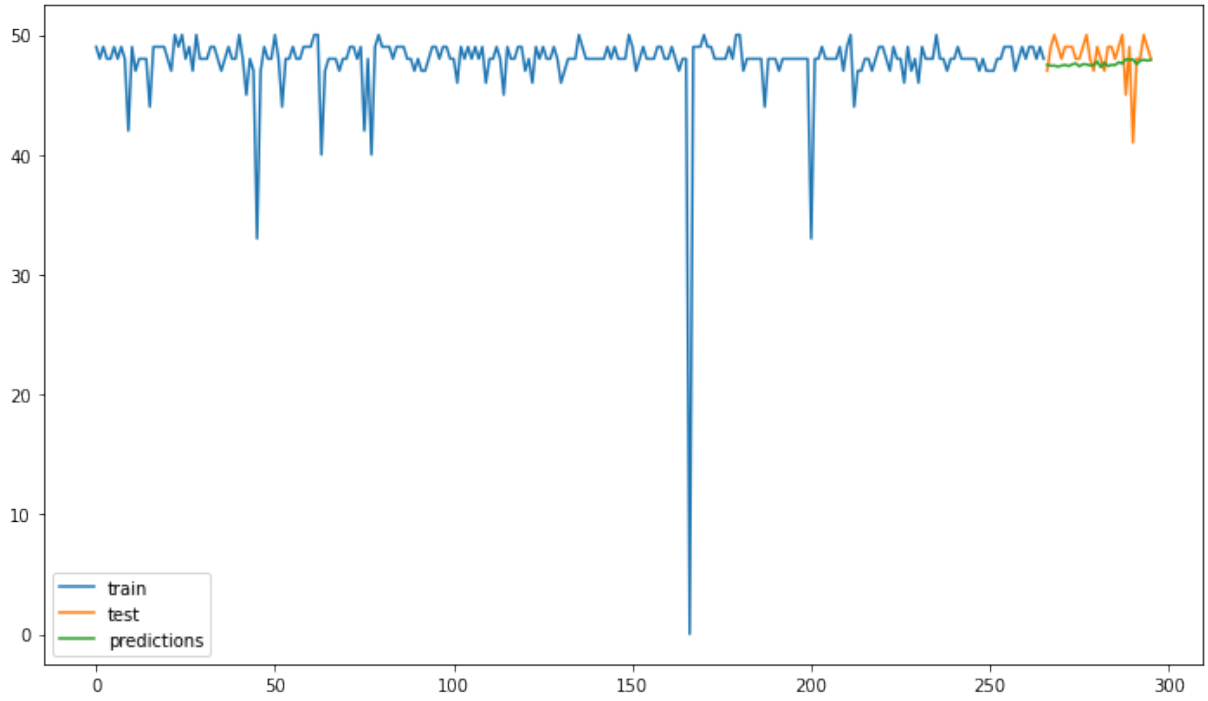


Figure 25. Prediction of Attributable for I 600 dataset with auto regressor model with Lasso Penalty. The performance of the model was: Test error (MSE) 3.5926376838689946.

For the final and in order to find the prediction error we train a simple linear regression model with the 30 last ALCOA values and the result was, test error (MSE): 3.395375850595643. This result can be seen in the figure below.

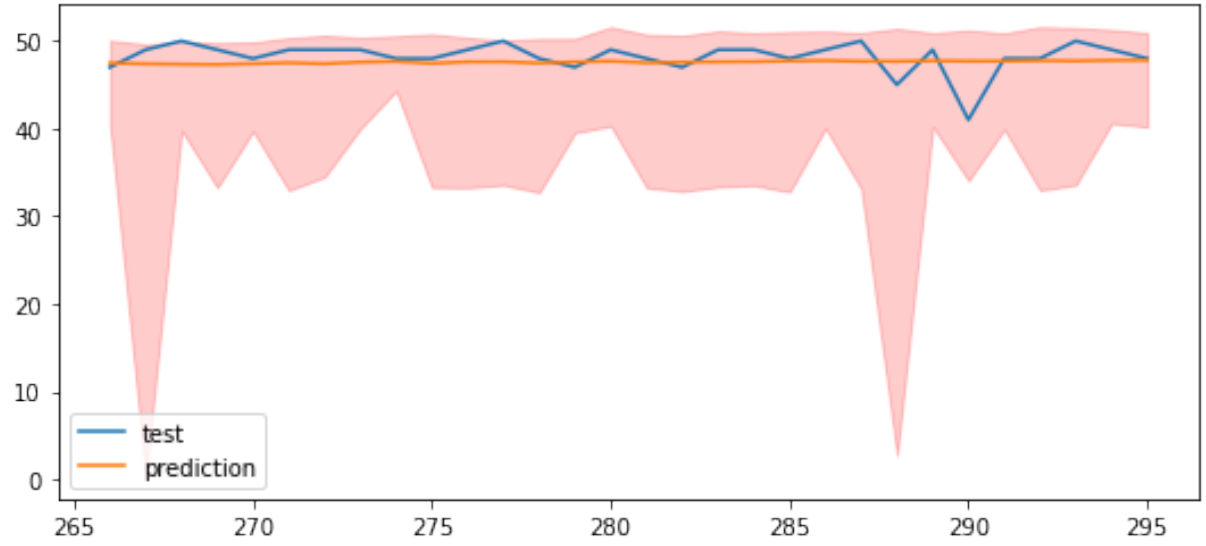


Figure 26. Prediction of Attributable for I 600 dataset with a linear regression model. The figure shows the last 30 batches which used for evaluate the prediction. The performance of the model was: Test error (MSE): 3.395375850595643.

The table below shows that the best model for I 600 dataset and Attributable principle is the Random Forest model with Test error (MSE): 2.83422. The Random Forest with best Hyperparameters Test error (MSE): 2.9538063879668854 is the second-best model. The Linear Regression model has a Test error (MSE): 3.395375850595643 and it is in the third place. The worst performance model is Lasso Alpha Penalty model with test error (MSE) 3.5926376838689946.

Dataset	ALCOA	Model	Performance
I600	Attributable	Random Forest	Test error (MSE): 2.83422
I600	Attributable	Random Forest with best Hyperparameters	Test error (MSE): 2.9538063879668854
I600	Attributable	Lasso Alpha Penalty	Test error (MSE) 3.5926376838689946
I600	Attributable	Linear Regression	Test error (MSE): 3.395375850595643

Table 7. Performance comparison of the four regression models used for prediction of Attributable for I 600 dataset.

5.2 Attributable principle and I 1000 dataset

For the I 1000 dataset and for the attributable the Bi-LSTM Network the performance of MAE is 1.5814 and VMAE is 0.8381. For the GRU Network the MAE is 2.0532 and the and VMAE is 0.8437 and for the LSTM Network the MAE is 1.6865 and for the LSTM Network the VMAE is 0.8565.

Network	MAE	VMAE
Bi-LSTM	**1.5814**	0.8381
GRU	2.0532	0.8437
LSTM	1.6865	0.8565

Table 8. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 1000 dataset for with no scaling.

The hyperparameters was loss: mean_squared_error, the optimizer set as: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-3 for the Bi-LSTM and learning rate set to: 1e-4 for GRU and LSTM Network too, lstm_units: 200, epochs: 15 for the Bi-LSTM and 20 epochs for GRU and LSTM Networks, the batch_size: 8, es_patience : 0.5 for Bi-LSTM and 0.4 for GRU and LSTM Networks too.

HYPERPARAMETERS I 1000 dataset with no scaling								
NETWORK	LOSS	OPTIMIZER	DROPOUT	LR	LSTM UNITS	EPOCHS	BATCH SIZE	ES PATIENCE
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	15	8	0.5
GRU	MSE	ADAM	0,2	1e-4	200	20	8	0.4
LSTM	MSE	ADAM	0,2	1e-4	200	20	8	0.4

Table 9. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 1000 dataset with no scaling.

For the I 1000 dataset with no scaling in our data the best performance was with Bi-LSTM with MAE 1.5814 the LSTM network has the second performance with 1.6865 and GRU Network with 2.0532 was third. The two results 1.5878 for I 600 dataset and 1.5814 for I 1000 dataset are very close and with similar hyperparameter tuning.

For the I 1000 dataset the Bi-LSTM network in the loss and val/ loss curve shows the fewest losses between train and validation. The similarity with I 600 dataset above is very big. This also applies to the other two networks, GRU and LSTM. In I 1000 dataset the GRU and LSTM shows better performance because the VMAE in these two networks is better as we can see in two graphs below.

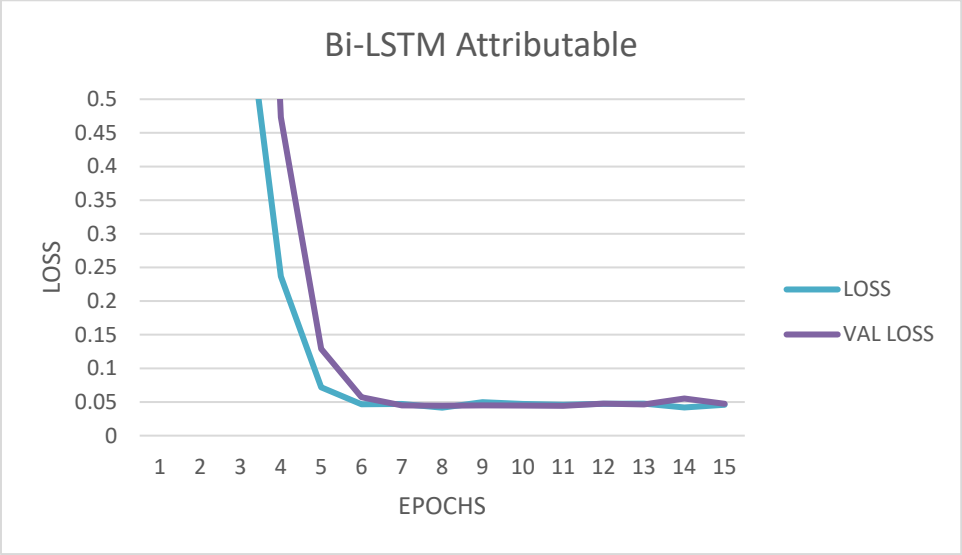


Figure 27. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 1000 dataset with no scaling.

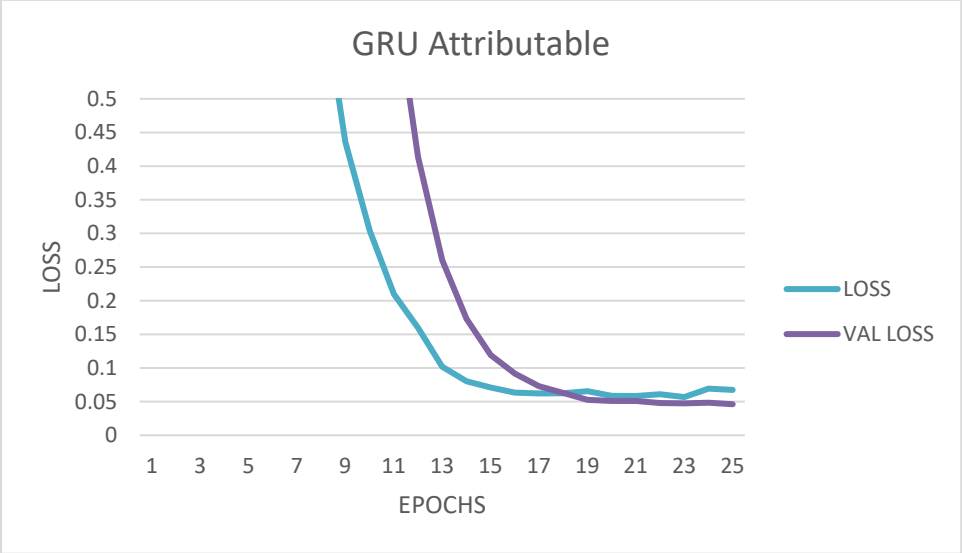


Figure 28. Loss/ Val loss curve of GRU network for Attributable and I 1000 dataset with no scaling.

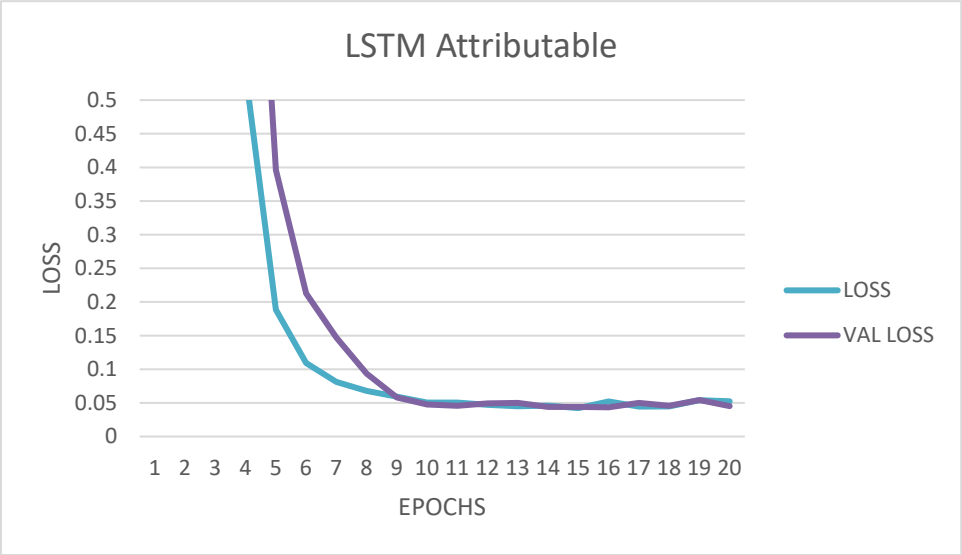


Figure 29. Loss/ Val loss curve of GRU network for Attributable and I 1000 dataset with no scaling.

5.2.1 Minmax Scaler

In order to take better performance to our neural networks we applied Minmax Scaler algorithm to our data. So, for the attributable and the I 1000 dataset with Minmax Scaler the Bi-LSTM Network the MAE is 1.5866 and the VMAE for the Bi-LSTM 0.8429. The GRU Network the MAE is 1.7525 and the VMAE is 0.8345 with Minmax Scaler and for the LSTM Network

the MAE is 1.4858 and the VMAE is 0.8705 with the same algorithm applied. The MAE 1.4858 of LSTM network with minmax scaling is the best performance of Attributable and I 1000 dataset.

Network	**MAE**	**VMAE**
Bi-LSTM	1.5866	0.8429
GRU	1.7525	0.8345
LSTM	**1.4858**	0.8705

Table 10. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 1000 dataset for with Minmax Scaler algorithm applied.

The hyperparameters for each network in Minmax Scaler was loss: mean_squared_error, the optimizer: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-4 for GRU and LSTM Network and for Bi-LSTM the learning rate is 1e-3, lstm_units set to: 200, epochs: 15 for Bi-LSTM and 20 epochs for GRU and LSTM, the batch_size: 8, es_patience : 0.5.

HYPERPARAMETERS with Mixmax Scaler								
NETWORK	**LOSS**	**OPTIMIZER**	**DROPOUT**	**LR**	**LSTM UNITS**	**EPOCHS**	**BATCH SIZE**	**ES PATIENCE**
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	15	8	0.5
GRU	MSE	ADAM	0,2	1e-4	200	20	8	0.4
LSTM	MSE	ADAM	0,2	1e-4	200	20	8	0.4

Table 11. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 1000 dataset with Minmax Scaler algorithm applied.

With Minmax Scaler applied on the network, we did not see any improvement in our results as in the I 600 dataset. The best result was the single LSTM network with 1.4858 MAE and Bi- LSTM 1.5866.

The loss/ val loss curves come the results with Minmax Scaliling where the LSTM network got the best results along with the Bi-LSTM network. As we see, these two networks are trained in only 4-5 epochs.

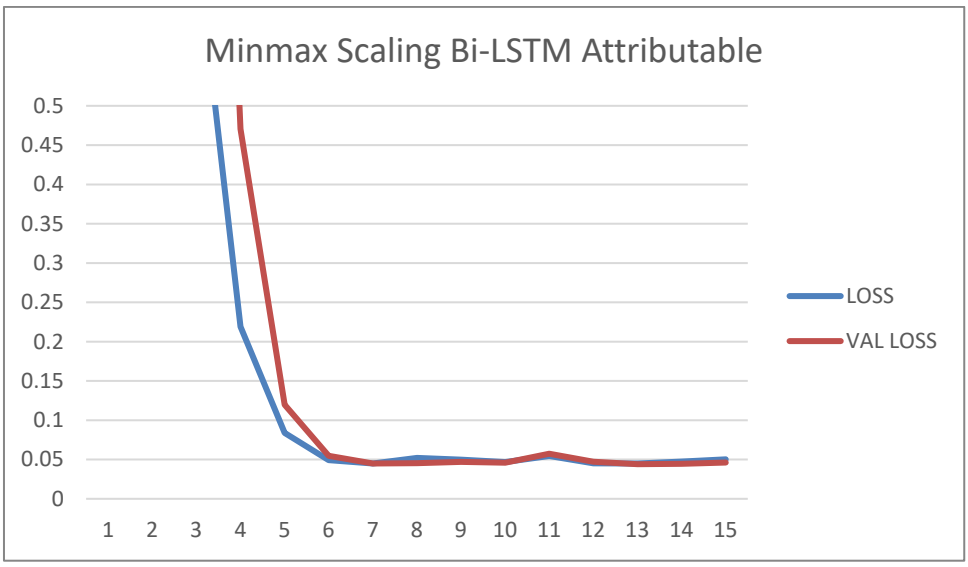


Figure 30. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 1000 dataset with Minmax Scaling.

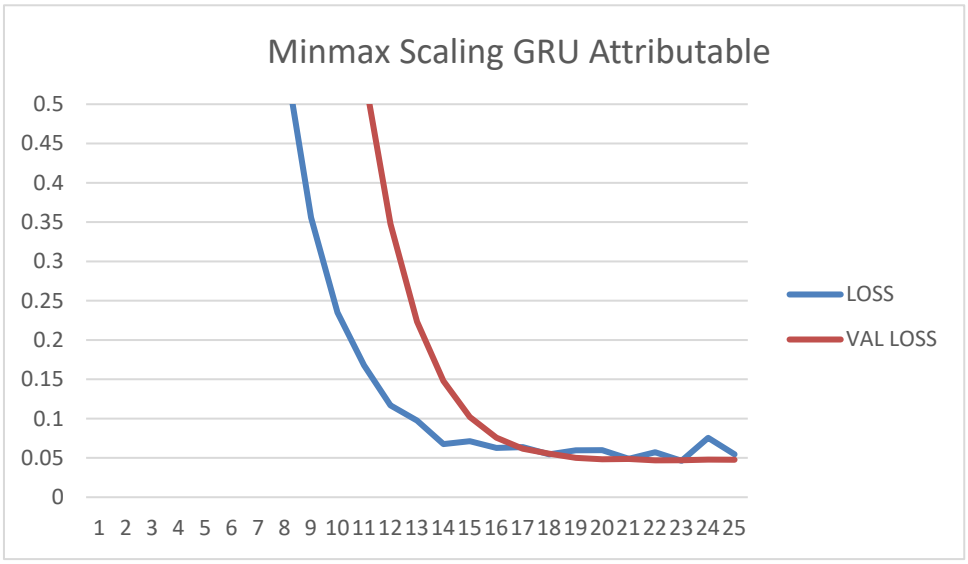


Figure 31. Loss/ Val loss curve of GRU network for Attributable and I 1000 dataset with Minmax Scaling.

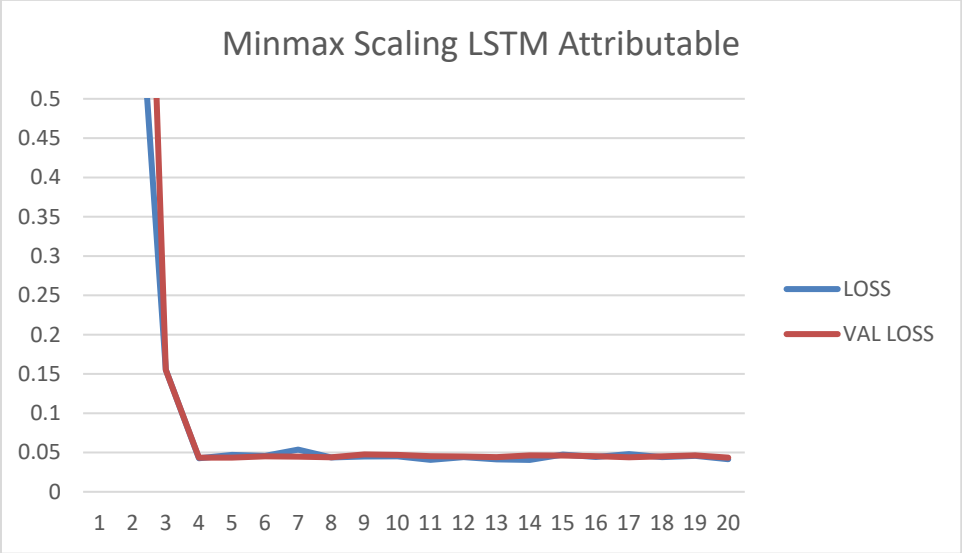


Figure 32. Loss/ Val loss curve of LSTM network for Attributable and I 1000 dataset with Minmax Scaling.

5.2.2 Standard Scaler

With Standard Scaler applied in our dataset the Bi-LSTM network has a performance 1.5718 in MAE and 0.8400 in VMAE. The GRU Network has MAE 1.7616 and VMAE 0.8345 with Standard Scaler applied and the LSTM Network the MAE is 1.4905 and the VMAE is 0.8506 with the same algorithm applied.

Network	**MAE**	**VMAE**
Bi-LSTM	1.5718	0.8400
GRU	1.7616	0.8345
LSTM	**1.4905**	0.8506

Table 12. Bi-LSTM, GRU, LSTM networks performance comparison for Attributable and I 1000 dataset for with Standard Scaler algorithm applied.

The hyperparameters for the networks in Standard Scaler was loss: mean_squared_error, the optimizer set as: adam, the dropout: 0.2, the optimal learning rate set to: 1e-4 for GRU and LSTM and 1e-3 for Bi-LSTM, lstm_units: 200, epochs:15 for Bi-LSTM and 20 epochs for GRU and LSTM Networks, the batch_size: 8, es_patience : 0.5.

HYPERPARAMETERS with StandardScaler

NETWORK	LOSS	OPTIMIZER	DROPOUT	LR	LSTM UNITS	EPOCHS	BATCH SIZE	ES PATIENCE
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	15	8	0.5
GRU	MSE	ADAM	0,2	1e-4	200	20	8	0.4
LSTM	MSE	ADAM	0,2	1e-4	200	20	8	0.4

Table 13. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Attributable and I 1000 dataset with Standard Scaler algorithm applied.

With Standard Scaler applied on the network too, we did not see any significance improvement in our results. The best result was the single LSTM network as with Minmax Scaler with 1.4905 MAE and Bi- LSTM 1.5718.

As in the other scaling, the loss val/ loss curves completely confirm the result for the Standard Scaling. The LSTM network shows better results along with the Bi-LSTM and they only need 4 epochs to train what this has to do with system optimization.

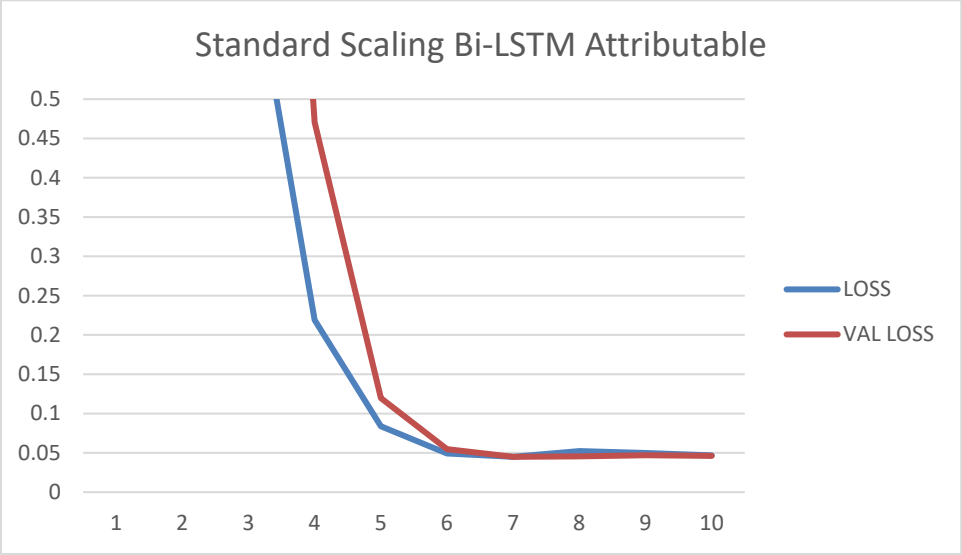


Figure 33. Loss/ Val loss curve of Bi-LSTM network for Attributable and I 1000 dataset with Standard Scaling.

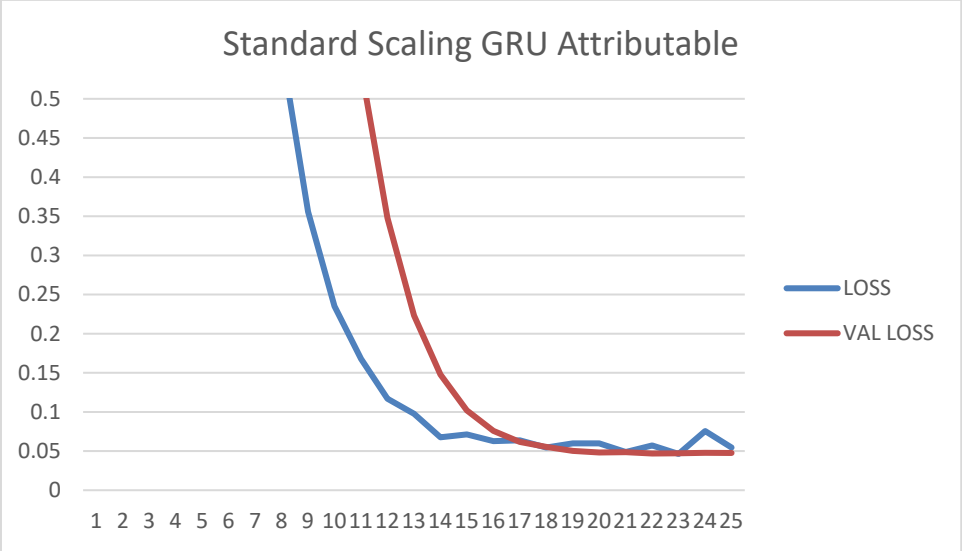


Figure 34. Loss/ Val loss curve of GRU network for Attributable and I 1000 dataset with Standard Scaling.

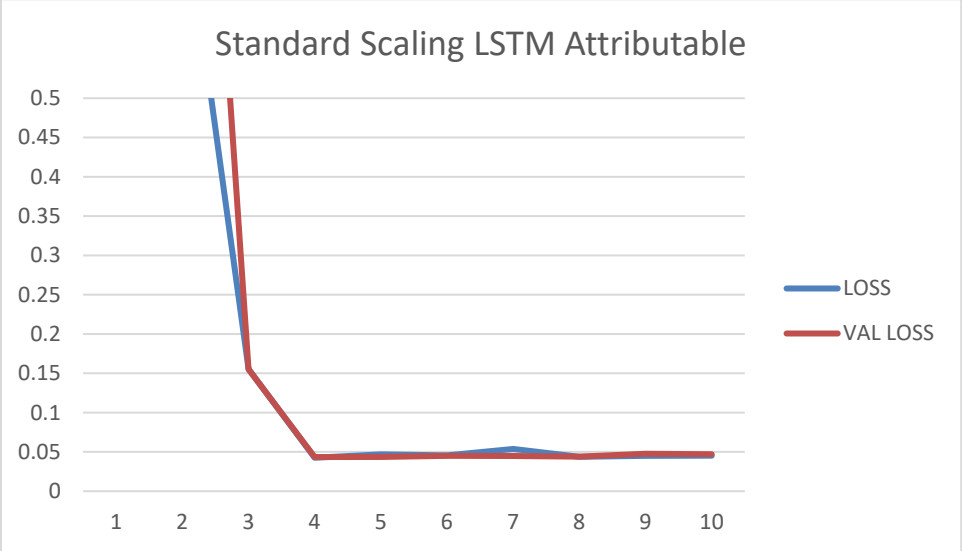


Figure 35. Loss/ Val loss curve of LSTM network for Attributable and I 1000 dataset with Standard scaling.

5.2.3 Prediction from previous Attributable prices I 1000 dataset

As in the attributable of the I 600 the first prediction model of the attributable from the ALCOA for I 1000 dataset is the random forest regression model for the last 30 attributable values of ALCOA. The result is Test error (MSE): 1.3849400000000012.

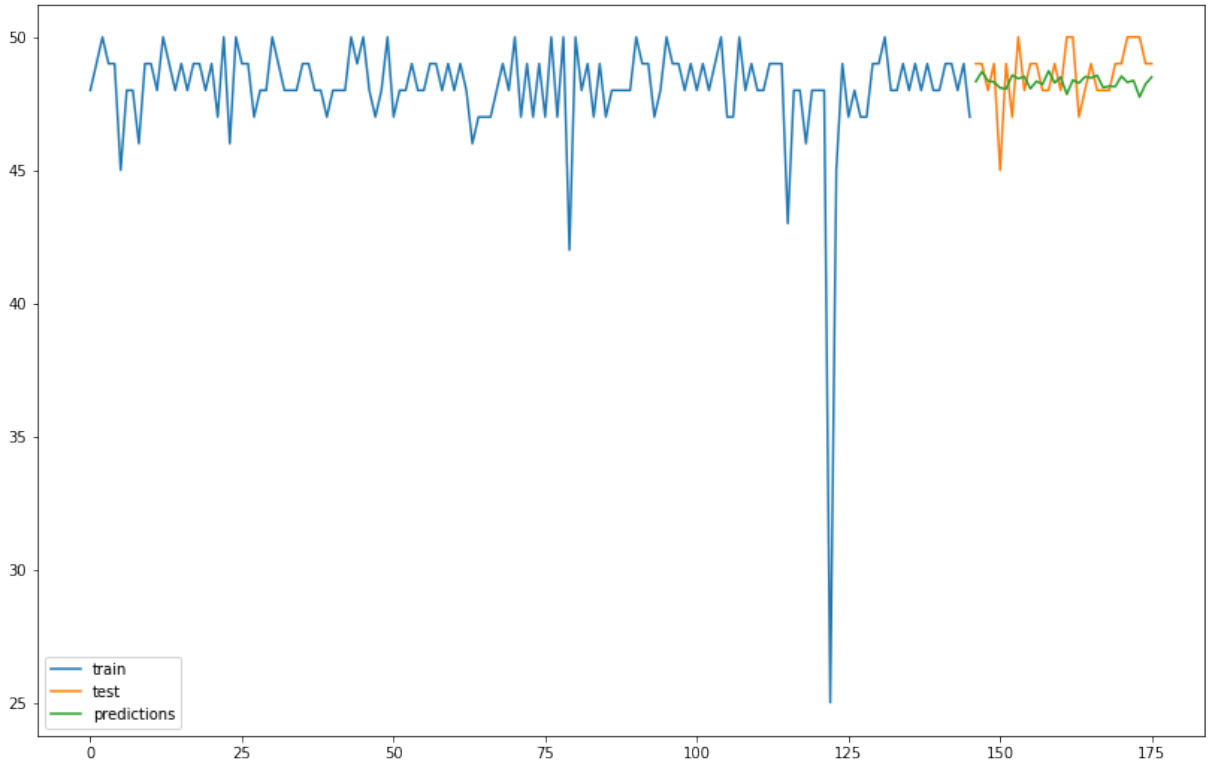


Figure 36. Prediction of Attributable for I 1000 dataset with random forest regression model. The performance of the model was: Test error (MSE): 1.3849400000000012.

Then to improve the model we looked to see which of these default parameters is best. So, we used a library that provides the grid search forecaster function like I 600 dataset that it compares the results obtained with each model configuration. The result is: (MSE) 1.7344624646202926

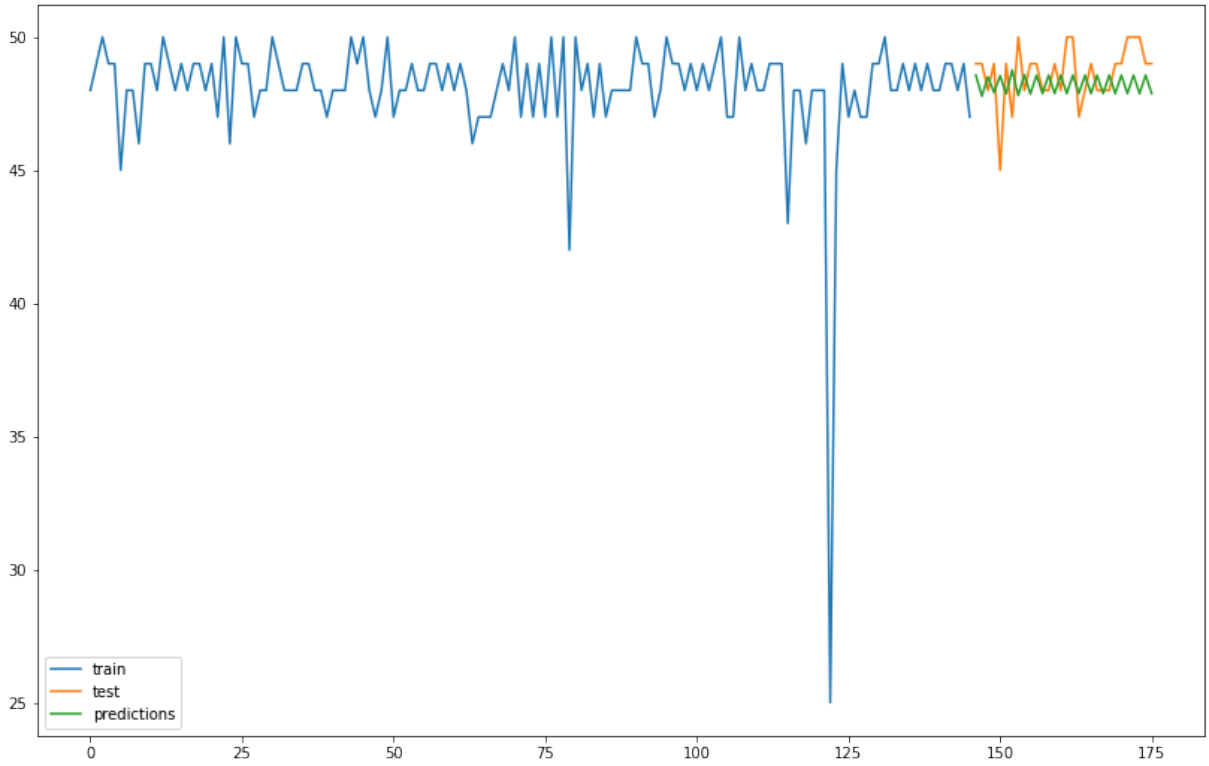


Figure 37. Prediction of Attributable for I 600 dataset with random forest regression model with grid search. The performance of the model was: Test error (MSE) 1.7344624646202926.

And for final like the method, we used in I 600 dataset we applied the linear model with Lasso penalty like a regressor and the result was Test error (MSE) 1.6038020049755768.

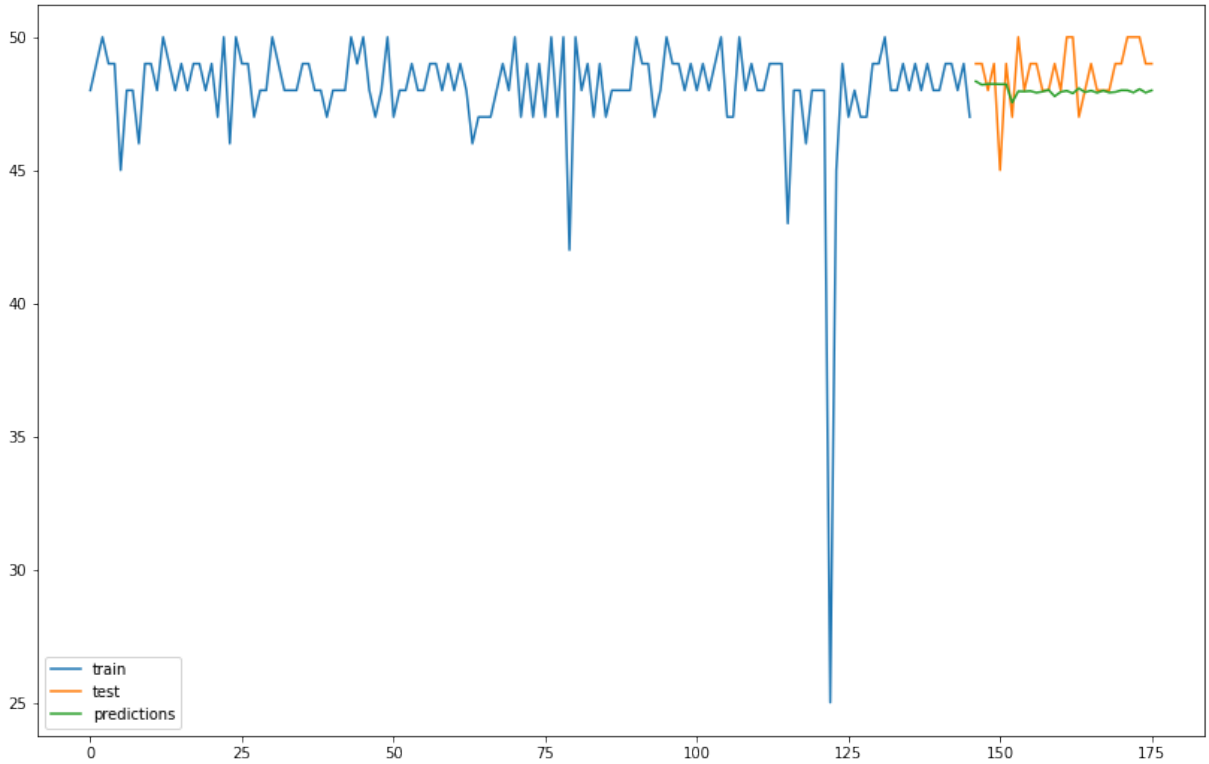


Figure 38. Prediction of Attributable for I 600 dataset with auto regressor model with Lasso Penalty. The performance of the model was: Test error (MSE) 1.6038020049755768.

In order to find the prediction error when using an autoregressor method like linear regression. We train a model and the last 30 prices for the Attributable of the ALCOA data used for prediction. The test error (MSE): 1.5267208846444105

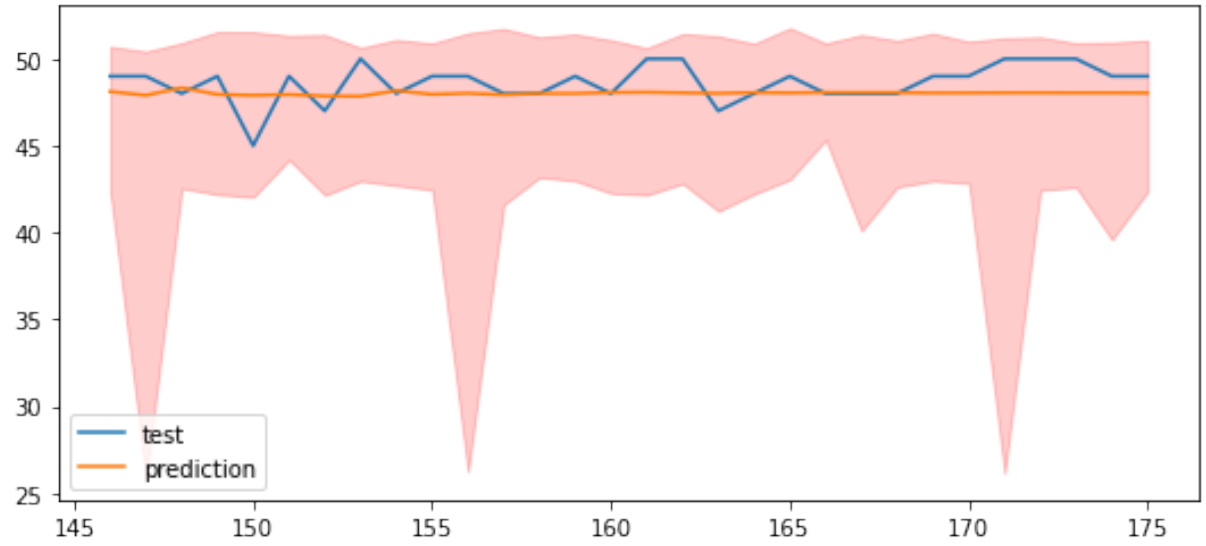


Figure 39. Prediction of Attributable for I 600 dataset with a linear regression model. The figure shows the last 30 batches which used for evaluate the prediction. The performance of the model was: The test error (MSE): 1.5267208846444105.

As we can see in the table below the best model for Attributable and I 1000 dataset, is the Random Forest with (MSE): 1.3849400000000012. The random forest model worked very well but also failed to predict the future ALCOA values. But this is an interesting observation that simple machine learning models can be used to predict ALCOA values without resorting to more complex neural networks, and that this possibility should also be considered.

The second-best model is the Linear Regression model with Test error (MSE): 1.5267208846444105. The third best model is the autoregreesor model with Lasso Alpha Penalty with (MSE) 1.6038020049755768. The worst model is Random Forest with best Hyperparameters with Test error (MSE): 1.7344624646202926.

Dataset	ALCOA	Model	Performance
I1000	Attributable	Random Forest	Test error (mse): 1.3849400000000012
I1000	Attributable	Random Forest with best Hyperparameters	Test error (mse): 1.7344624646202926
I1000	Attributable	Lasso Alpha Penalty	Test error (mse) 1.6038020049755768
I1000	Attributable	Linear Regression	Test error (mse): 1.5267208846444105

Table 14. Performance comparison of the four regression models used for prediction of Attributable for I 1000 dataset.

5.3 Contemporaneous principle and I 600 Dataset

For the same reason and to determine whether the artificial intelligence models used above can predict whether the data was collected at the time it was created, this means that the data is Contemporaneous from the ALCOA acronym. This should be done in such a way that their continuity is maintained and the data can be displayed whenever required.

Using the I 600 dataset and for the Contemporaneous, the Bi-LSTM Network the MAE is 3.2269 and VMAE is 3.2292. This price 3,229 is the best performance of all networks for I 600 dataset. For the GRU Network the MAE is 3.7810 and the and VMAE is 3.2096 and for the LSTM Network the MAE is 3.9712 and for the LSTM Network the VMAE is 3.3155.

Network	MAE	VMAE
Bi-LSTM	**3.2269**	3.2292
GRU	3.7810	3.2096
LSTM	3.9712	3.3155

Table 15. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 600 dataset for with no scaling.

The hyperparameters was loss: mean_squared_error, the optimizer: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-3 for the Bi-LSTM for GRU and LSTM Network, lstm_units: 200, epochs: 20, batch_size: 4 for Bi-LSTM and GRU and for LSTM batch size set to 8, es_patience : 0.5 for Bi-LSTM and 0.4 for GRU and LSTM.

HYPERPARAMETERS								
NETWORK	**LOSS**	**OPTIMIZER**	**DROPOUT**	**LR**	**LSTM UNITS**	**EPOCHS**	**BATCH SIZE**	**ES PATIENCE**
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	20	4	0.5
GRU	MSE	ADAM	0,2	1e-3	200	20	4	0.4
LSTM	MSE	ADAM	0,2	1e-3	200	20	8	0.4

Table 16. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 600 dataset with no scaling.

As we can see to the graphs below, the best results were obtained by the Bi-LSTM like the attributable network with MAE 3.2269 the second was GRU Network with MAE 3.7810 and the third network was the simple LSTM with 3.9712 of MAE.

As we saw from the MAE above, it is expected that the models here will be presented with a larger underfit. This can also be seen from the loss val/ loss curves. The Bi- LSTM network shows smaller underfit than to two others. The network wants only 4-5 epochs to train than 9-11 epochs to GRU and 15-20 the single LSTM.

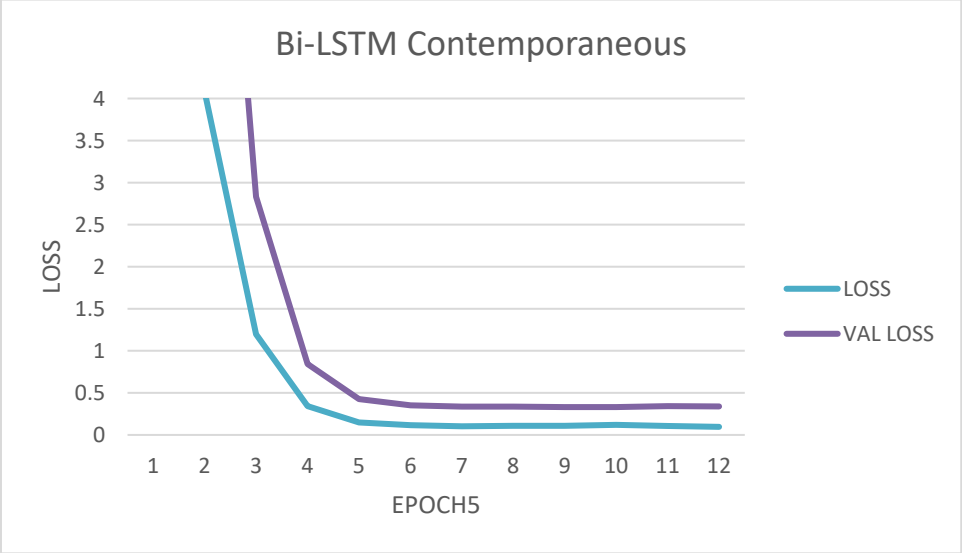


Figure 40. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 600 dataset with no scaling.

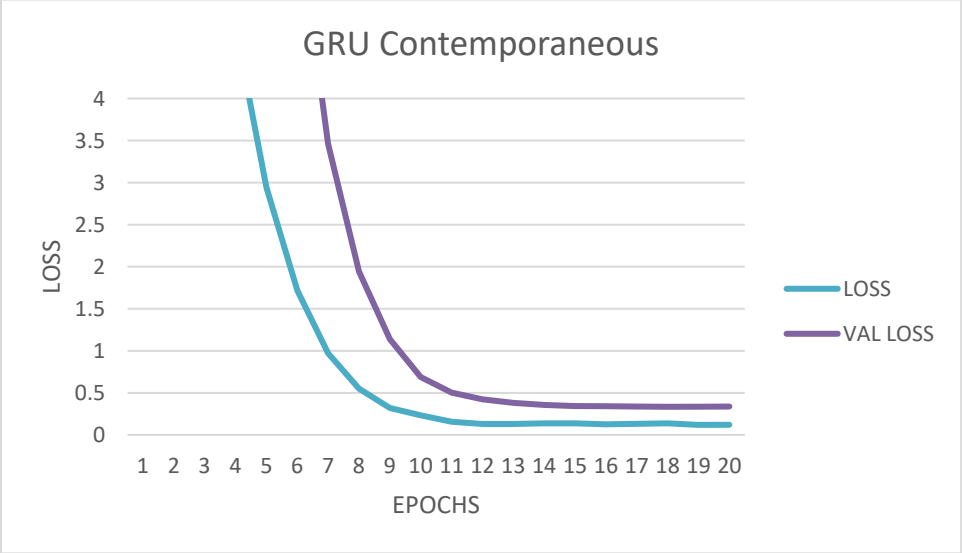


Figure 41. Loss/ Val loss curve of GRU network for Contemporaneous and I 600 dataset with no scaling.

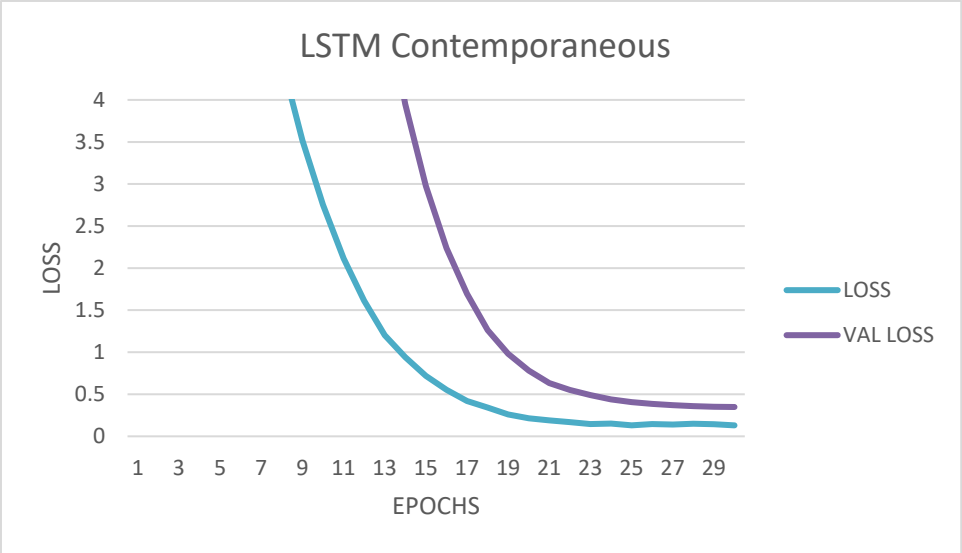


Figure 42. Loss/ Val loss curve of LSTM network for Contemporaneous and I 600 dataset with no scaling.

5.3.1 Minmax Scaler

With Minmax Scaler applied in I 600 dataset for the Contemporaneous the Bi-LSTM has a performance 3,5263 in MAE and 3,2226 in VMAE. The GRU Network has MAE 4,0007 and VMAE 3,1866 with Standard Scaler applied and the LSTM Network the MAE is 3,8919 and the VMAE is 3,2131with the same algorithm applied.

Network	**MAE**	**VMAE**
Bi-LSTM	**3.5263**	3.2226
GRU	4.0007	3.1866
LSTM	3.8919	3.2131

Table 17. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 600 dataset for with Minmax Scaler.

The hyperparameters was loss: mean_squared_error, the optimizer: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-3 for the Bi-LSTM for GRU and LSTM Network, lstm_units: 200, epochs: 20 for Bi-LSTM and GRU and 40 epochs for LSTM the the batch_size: 4 for Bi-LSTM and GRU and for LSTM batch size set to 8, es_patience: 0.5 for Bi-LSTM and 0.4 for GRU and LSTM.

HYPERPARAMETERS with Mixmax Scaler								
NETWORK	**LOSS**	**OPTIMIZER**	**DROPOUT**	**LR**	**LSTM UNITS**	**EPOCHS**	**BATCH SIZE**	**ES PATIENCE**
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	20	4	0.5
GRU	MSE	ADAM	0,2	1e-3	200	20	4	0.4
LSTM	MSE	ADAM	0,2	1e-3	200	40	8	0.4

Table 18. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 600 dataset with Minmax Scaler.

In order to improve the performance of our networks we applied Mixmax Scaler in our data too. The best performance was Bi-LSTM network with 3.5263 with second the LSTM network with MAE 3.8919.

The Minmax Scaling only helps the three networks to train faster but the underfit is the same. Again, the Bi- LSTM Network has the smaller underfit.

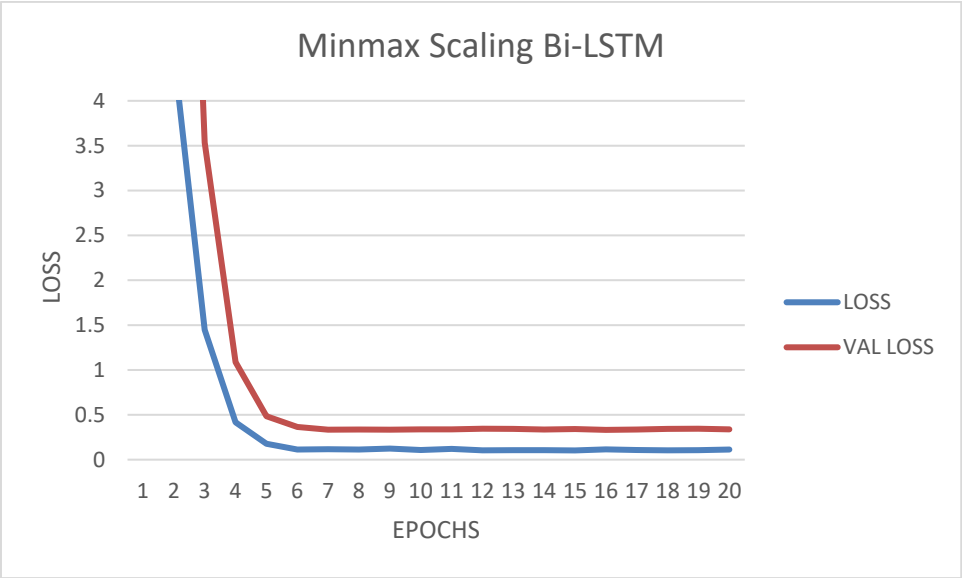


Figure 43. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 600 dataset with Minmax Scaling.

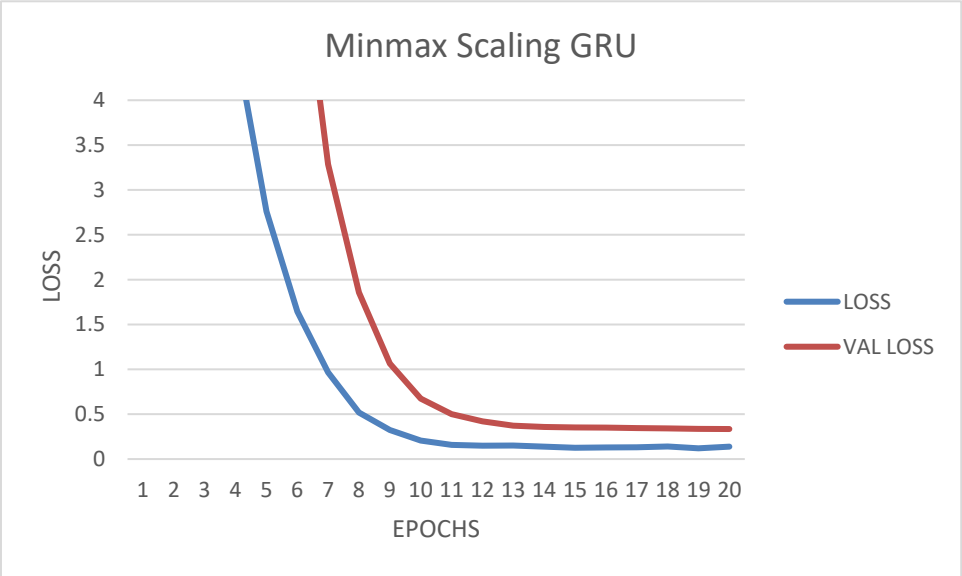


Figure 44. Loss/ Val loss curve of GRU network for Contemporaneous and I 600 dataset with Minmax Scaling.

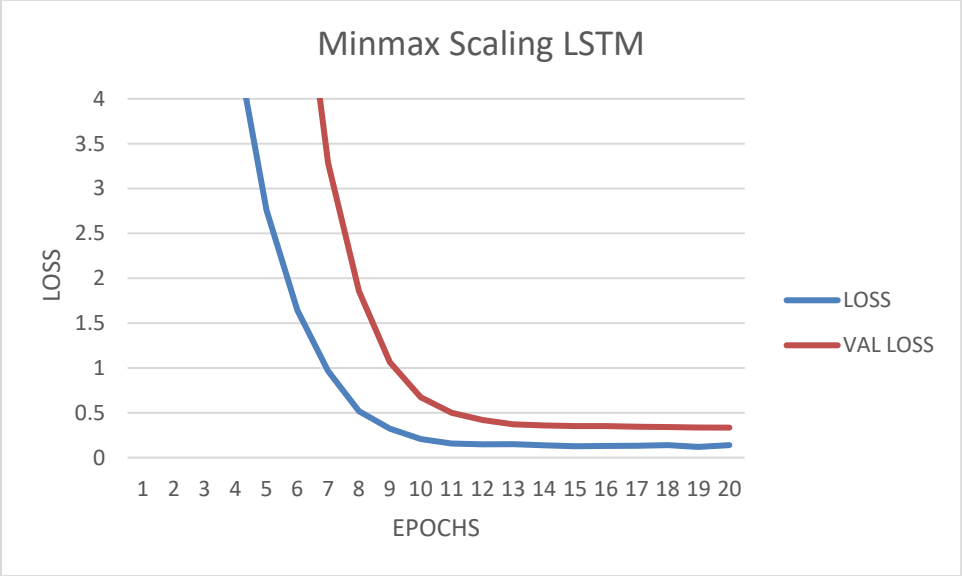


Figure 45. Loss/ Val loss curve of LSTM network for Contemporaneous and I 600 dataset with Minmax Scaling.

5.3.2 Standard Scaler

With Standard Scaler applied in our dataset the Bi-LSTM has a performance 3.4792 in MAE and 3.1656 in VMAE. The GRU Network has MAE 3.7858 and VMAE 3.1861 with Standard Scaler applied and the LSTM Network the MAE is 3.999 and the VMAE is 3.2258 with the same algorithm applied.

Network	**MAE**	**VMAE**
Bi-LSTM	**3,4792**	3,1656
GRU	3,7858	3,1861
LSTM	3,9999	3,2258

Table 19. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 600 dataset for with Standard scaler.

The hyperparameters was loss: mean_squared_error, the optimizer as: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-3, lstm_units: 200, epochs: 20 for Bi-LSTM and GRU and 40 epochs for LSTM the the batch_size: 4 for Bi-LSTM and GRU and for LSTM batch size set to 8, es_patience : 0.5 for Bi-LSTM and 0.4 for GRU and LSTM.

HYPERPARAMETERS with Standard Scaler								
NETWORK	**LOSS**	**OPTIMIZER**	**DROPOUT**	**LR**	**LSTM UNITS**	**EPOCHS**	**BATCH SIZE**	**ES PATIENCE**
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	20	4	0.5
GRU	MSE	ADAM	0,2	1e-3	200	20	4	0.4
LSTM	MSE	ADAM	0,2	1e-3	200	40	8	0.4

Table 20. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 600 dataset with standard Scaler.

With Standard Scaler applied the MAE for Bi-LSTM is 3.4792 which is the best performance like Minmax scaler. Like above the Standard Scaler algorithm has no significance performance improvement.

As we see from above the Minmax Scaling helps our networks to train faster. The underfit is the same and again the Bi-LSTM shows better performance than the other two networks as we can see from our loss and val/ loss curves.

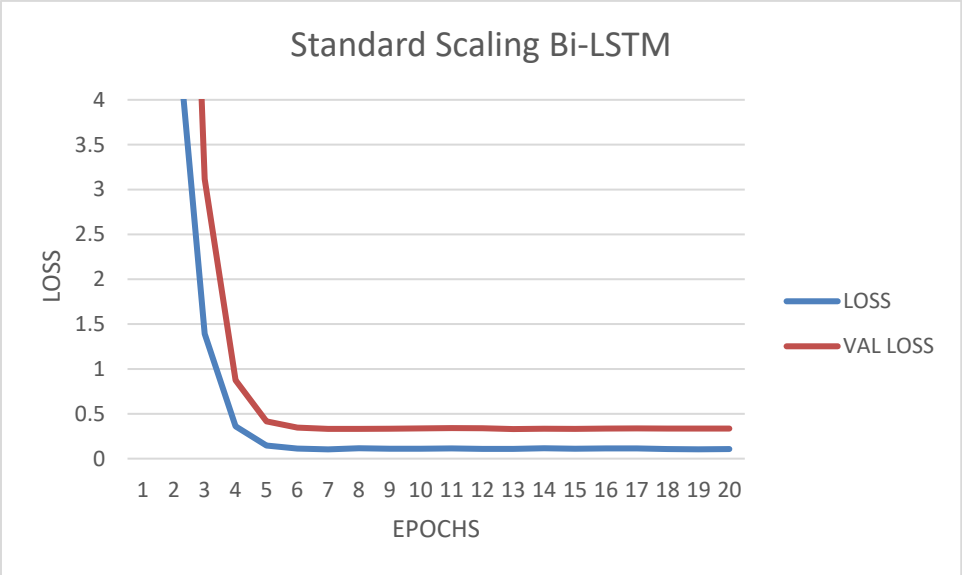


Figure 46. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 600 dataset with Standard Scaler.

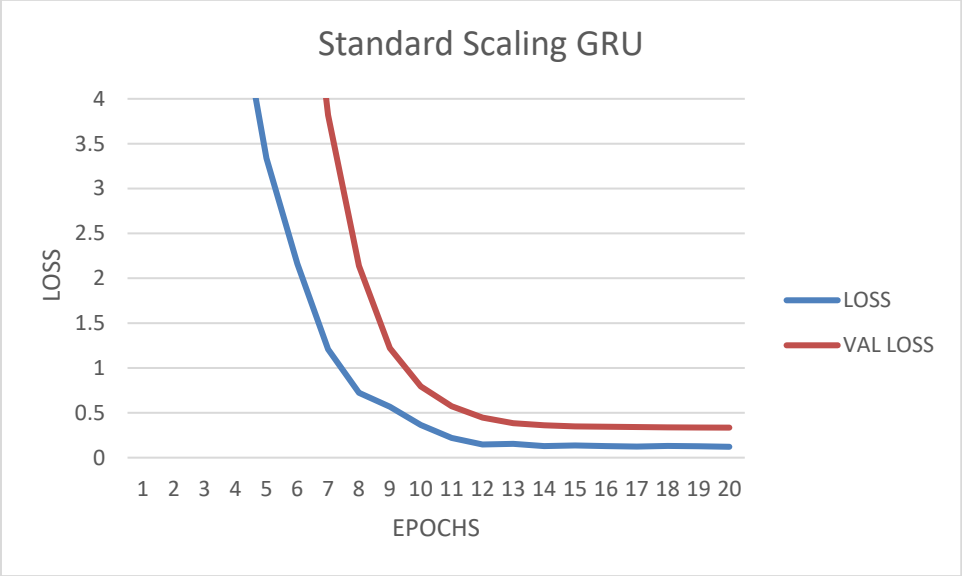


Figure 47. Loss/ Val loss curve of GRU network for Contemporaneous and I 600 dataset with Standard Scaler.

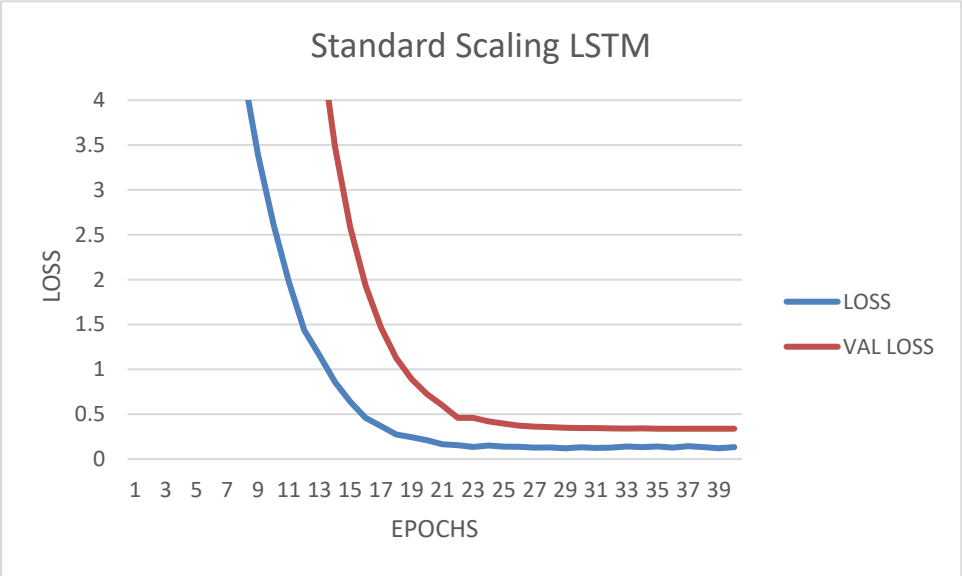


Figure 48. Loss/ Val loss curve of LSTM network for Contemporaneous and I 600 dataset with Standard Scaler.

5.3.3 Prediction from previous Contemporaneous prices I 600 dataset.

And for Contemporaneous principle we used the same algorithms and models to predict the future Contemporaneous from ALCOA prices. Our first model is a random forest regression.

This model takes the last 30 prices of Contemporaneous ALCOA prices. This model has a Test error (MSE): 45.466243333333324. The prediction curve is shown below.

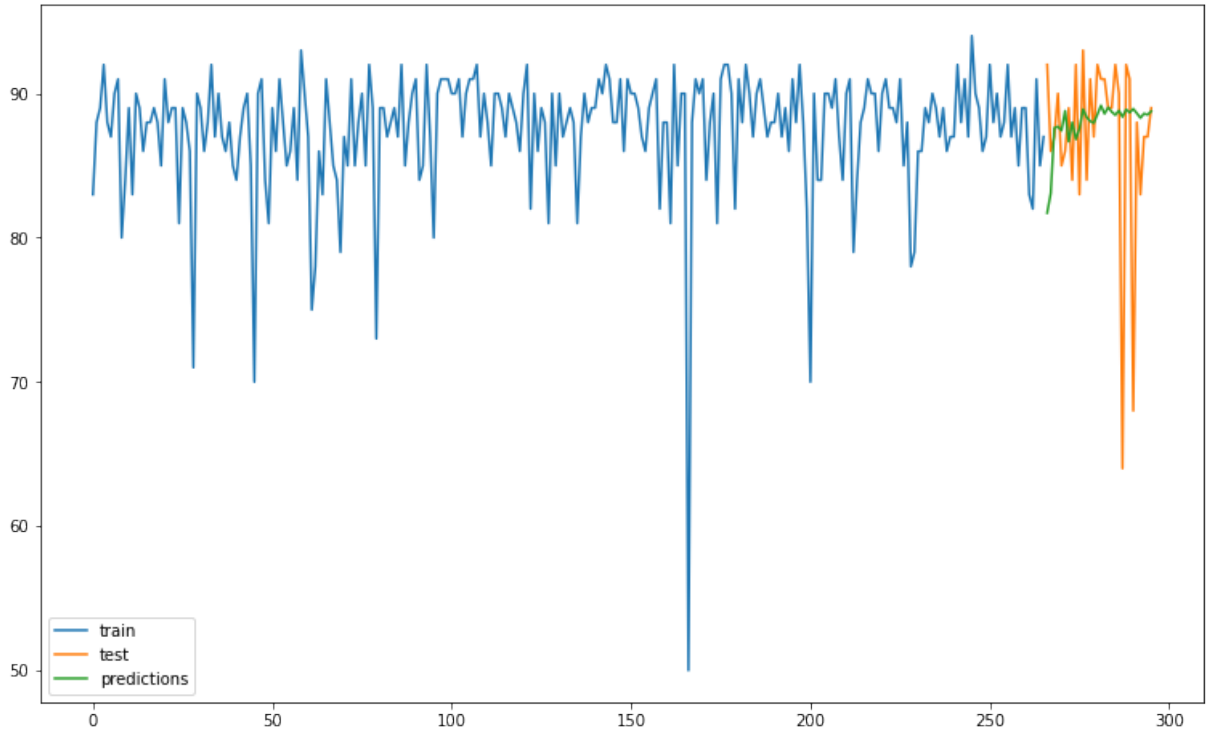


Figure 49. Prediction of Contemporaneous for I 600 dataset with random forest regression model. The performance of the model was: Test error (MSE): 45.466243333333324.

The second model is a forecaster with the best hypermeters in order to improve the model and the MSE and implemented a random forest model like the previous Attributable data. The test error (MSE) of this model is: 40.6529824518856.

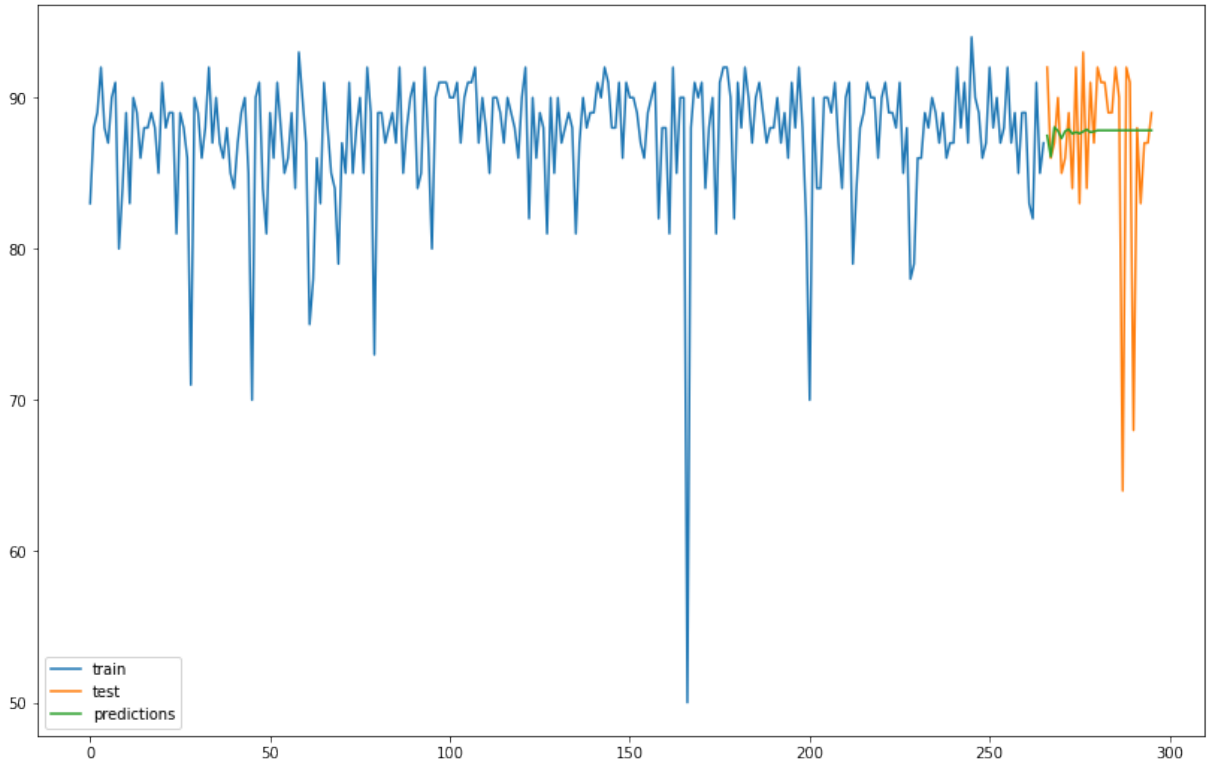


Figure 50. Prediction of Contemporaneous for I 600 dataset with random forest regression model with grid search. The performance of the model was: (MSE) of this model is: 40.6529824518856.

The third model is the auto regressor with lasso penalty is used as a regression. The model performance is Test error (MSE) 40.59479708636836. We can see this performance in the graph below.

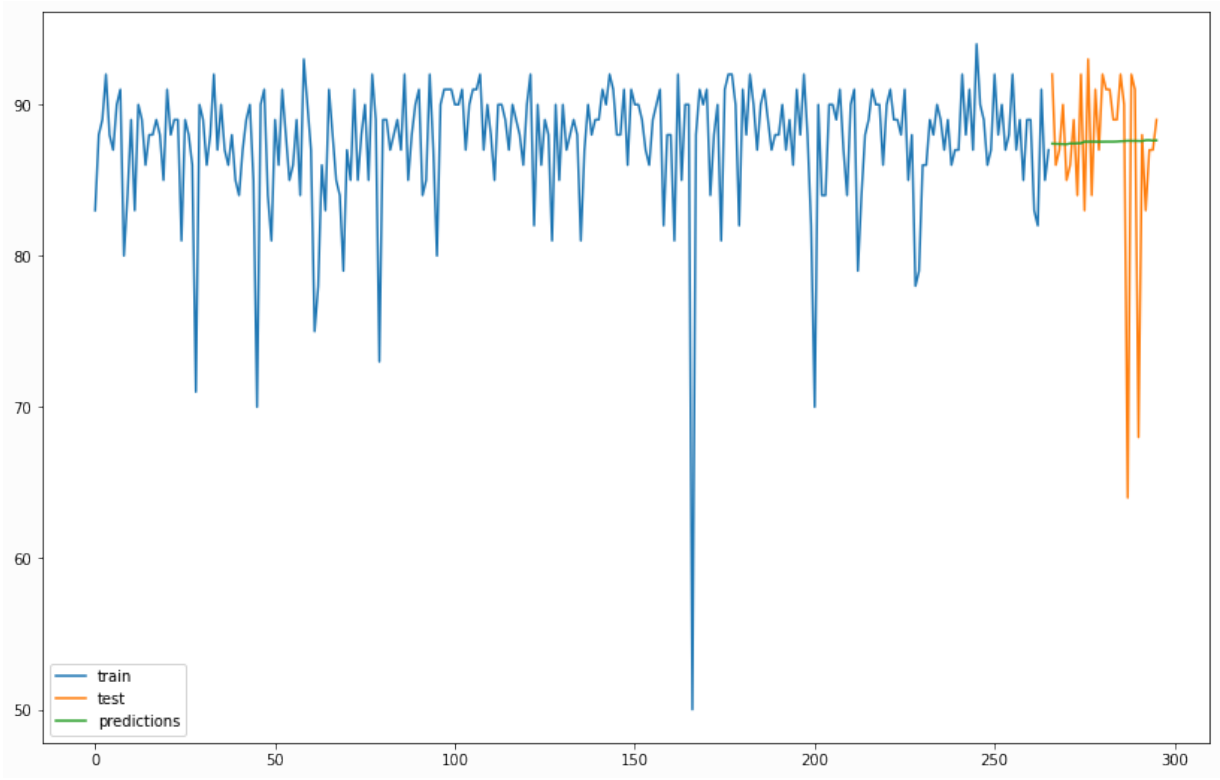


Figure 51. Prediction of Contemporaneous for I 600 dataset with auto regressor model with Lasso Penalty. The performance of the model was: Test error (MSE) 40.59479708636836.

In order to calculate the network prediction error, we trained a regression model with linear regression. This model has a Test error (MSE): 40.05476184483597. The performance of this model can be seen below.

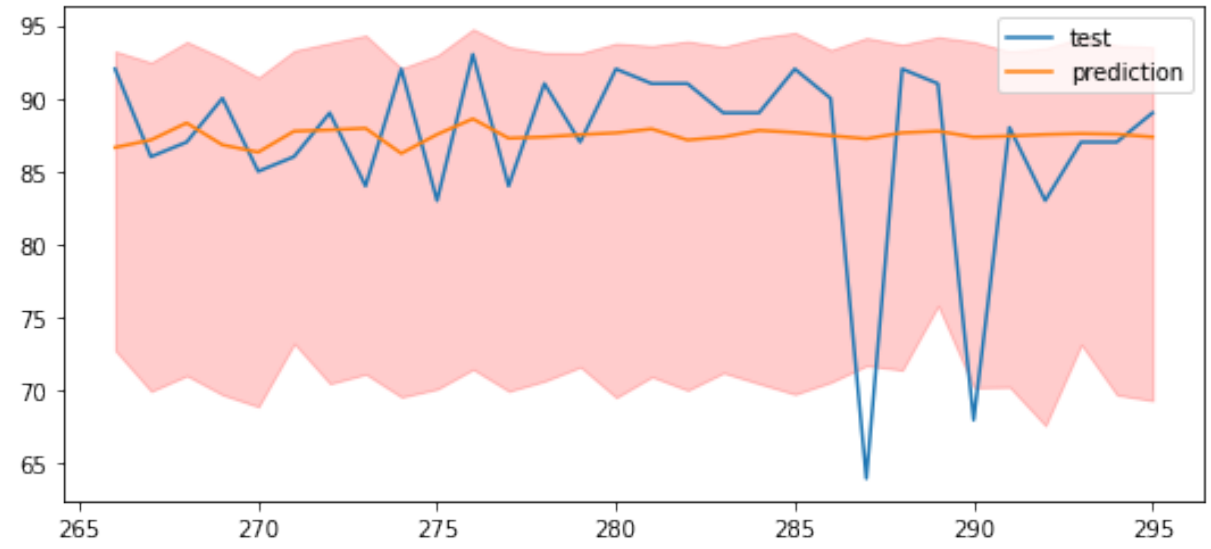


Figure 52. Prediction of Contemporaneous for I 600 dataset with a linear regression model. The figure shows the last 30 batches which used for evaluate the prediction. The performance of the model was: The test error (MSE): 40.05476184483597.

The table below shows the best performance the Linear Regression model with test error 40.05476184483597. This price is close to Random Forest with best Hyperparameters and the model with lasso penalty.

Dataset	ALCOA	Model	Performance
I600	Contemporaneous	Random Forest	Test error (MSE): 45.466243333333324
I600	Contemporaneous	Random Forest with best Hyperparameters	Test error (MSE): 40.6529824518856
I600	Contemporaneous	Lasso Alpha Penalty	Test error (MSE): 40.5947970866836
I600	Contemporaneous	Linear Regression	Test error (MSE): 40.05476184483597

Table 21. Performance comparison of the four regression models used for prediction of Contemporaneous for I 600 dataset.

5.4 Contemporaneous principle and I 1000 Dataset

Using the I 1000 dataset we will try to predict and see if the data is contemporaneous. The Bi-LSTM network has a MAE 4.5098 and VMAE 4.5926 the GRU network MAE is 4.8771 and VMAE 4.5934 and LSTM performance in MAE is 5.7342 and VMAE 5.7881.

Network	MAE	VMAE
Bi-LSTM	**4.5098**	4.5926
GRU	4.8771	4.5934
LSTM	5.7342	5.7881

Table 22. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 1000 dataset for with no scaling.

The hyperparameters was loss: mean_squared_error, the optimizer: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-3 for the Bi-LSTM for GRU and LSTM Network, lstm_units: 200, epochs: 40, the batch_size: 8 for Bi-LSTM and for GRU, LSTM batch size set to 4, es_patience : 0.5 for Bi-LSTM and 0.4 for GRU and LSTM.

HYPERPARAMETERS								
NETWORK	LOSS	OPTIMIZER	DROPOUT	LR	LSTM UNITS	EPOCHS	BATCH SIZE	ES PATIENCE
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	40	8	0.5
GRU	MSE	ADAM	0,2	1e-3	200	40	4	0.4

LSTM	MSE	ADAM	0,2	1e-3	200	40	4	0.4

Table 23. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 1000 dataset with no scaling.

For the I 1000 dataset and the best performance was by Bi-LSTM network with 4.5098 MAE. The GRU Network has 4.8771 MAE and LSTM Network has 5.7342 MAE.

The contemporaneous for I 1000 dataset has a big underfit in loss/ val loss curves. The best performance is for Bi-LSTM network. Bi-LSTM and GRU networks need 11-15 epochs to train the LSTM need around 25 epochs to train, with better performance the Bi-LSTM network.

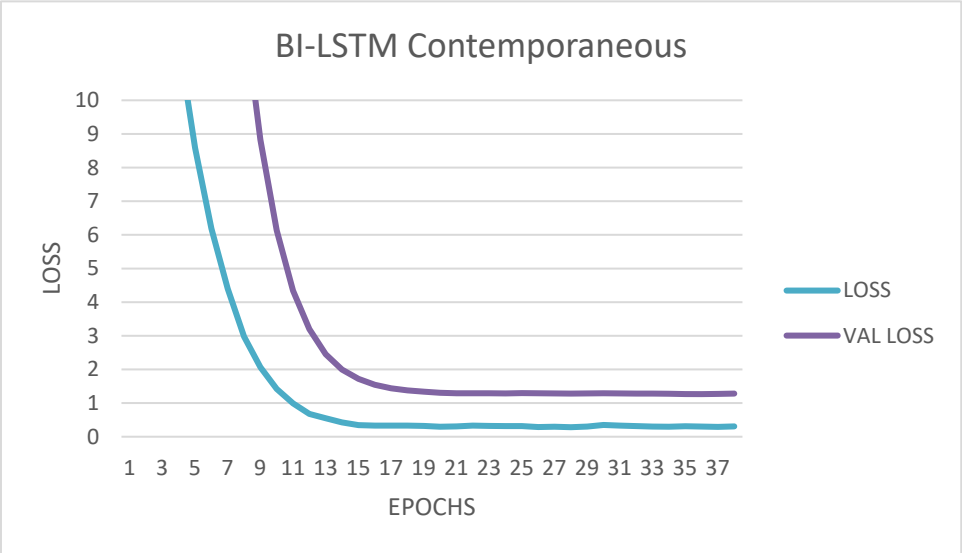


Figure 53. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 1000 dataset with no Scaling.

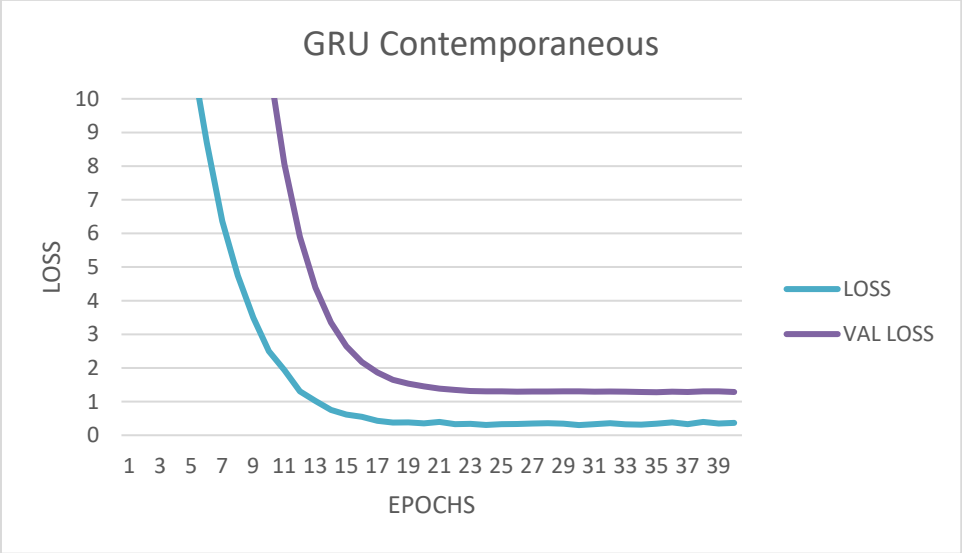


Figure 54. Loss/ Val loss curve of GRU network for Contemporaneous and I 1000 dataset with no scaling.

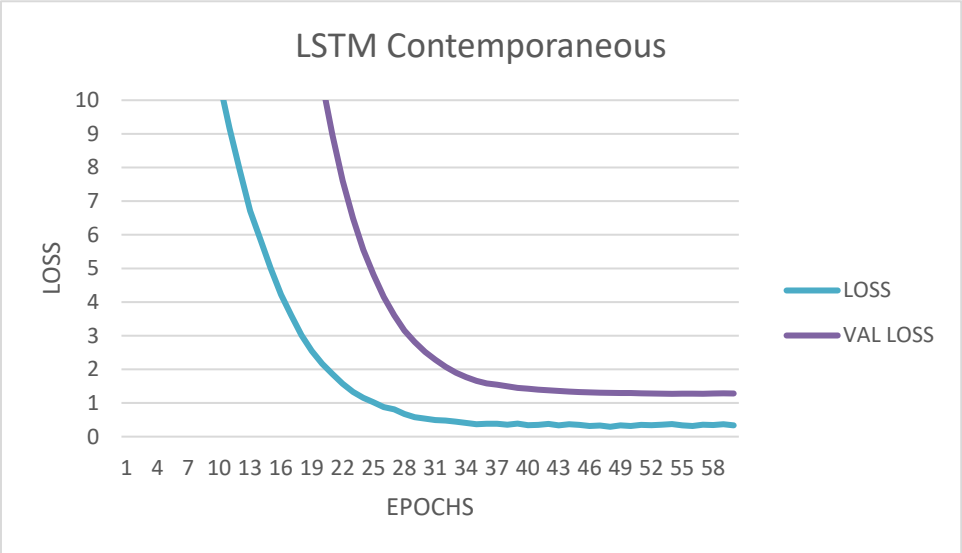


Figure 55. Loss/ Val loss curve of LSTM network for Contemporaneous and I 1000 dataset with no scaling.

5.4.1 Minmax Scaler

With Minmax Scaler applied in I 1000 dataset for the Contemporaneous the Bi-LSTM has a performance 4.4756 in MAE and in 4.5612 VMAE. The GRU Network has MAE 4.5399 and VMAE 4.5984 with Standard Scaler applied and the LSTM Network the MAE is 4.8464 and the VMAE is 4.6192 with the same algorithm applied.

Network	MAE	VMAE
Bi-LSTM	**4.4756**	4.5612
GRU	4.5399	4.5984
LSTM	4.8464	4.6192

Table 24. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 1000 dataset for with Minmax Scaler.

The hyperparameters was loss: mean_squared_error, the optimizer: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-3 for the Bi-LSTM for GRU and LSTM Network, lstm_units: 200, epochs: 20 for Bi-LSTM and GRU and 40 epochs for LSTM the the batch_size: 4 for Bi-LSTM and GRU and for LSTM batch size set to 8, es_patience : 0.5 for Bi-LSTM and 0.4 for GRU and LSTM.

HYPERPARAMETERS with Mixmax Scaler								
NETWORK	LOSS	OPTIMIZER	DROPOUT	LR	LSTM UNITS	EPOCHS	BATCH SIZE	ES PATIENCE
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	20	4	0.5
GRU	MSE	ADAM	0,2	1e-3	200	20	8	0.4
LSTM	MSE	ADAM	0,2	1e-3	200	40	8	0.4

Table 25. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 1000 dataset with Minmax Scaler.

With Minmax Scaler applied the performance has no significant improvement with MAE 4.4756 in Bi- LSTM network and GRU Network 4.5399 MAE. The Minmax scaler it did not help our network to train faster at all, nor did it reduce the underfit. Again, the Bi- LSTM network shows better performance than the other two networks but the underfit is big again too.

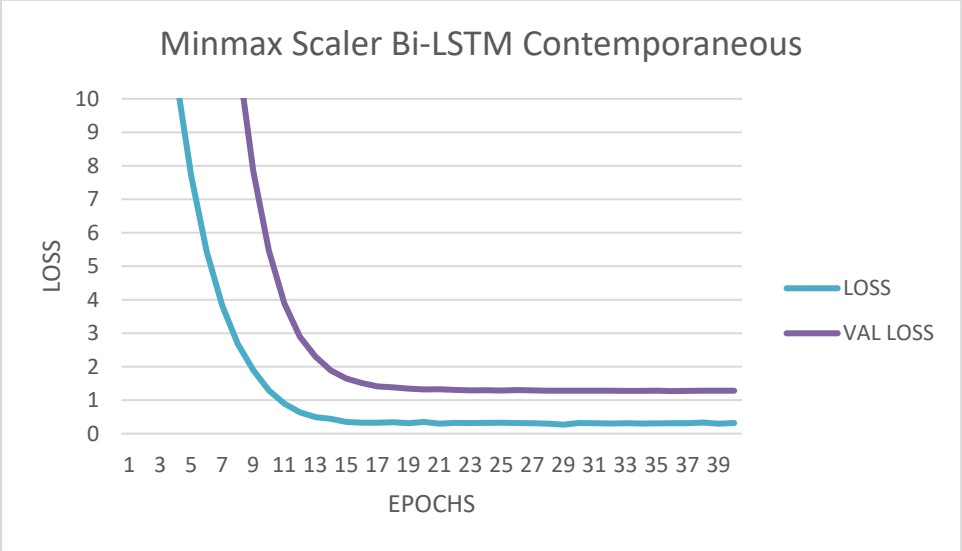


Figure 56. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 1000 dataset with Minmax Scaler.

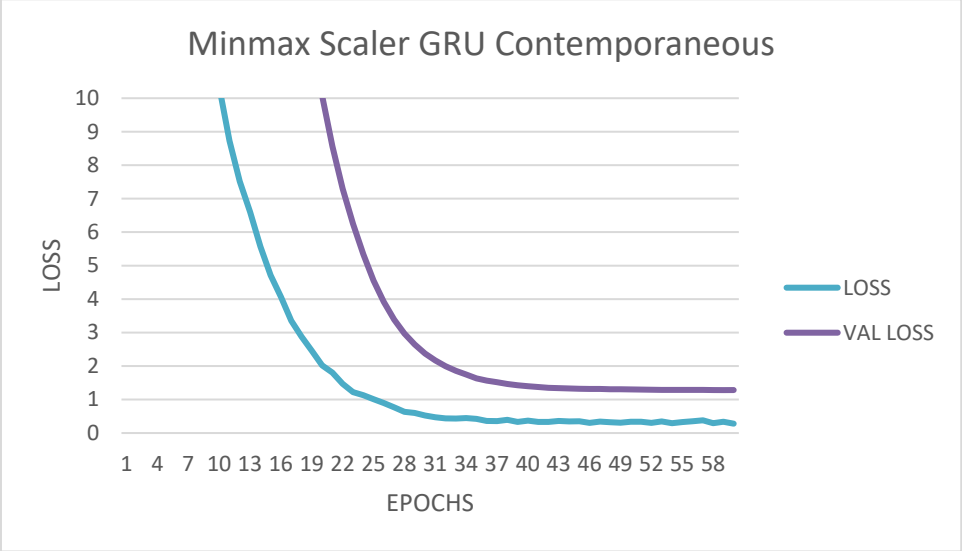


Figure 57. Loss/ Val loss curve of GRU network for Contemporaneous and I 1000 dataset with Minmax Scaler.

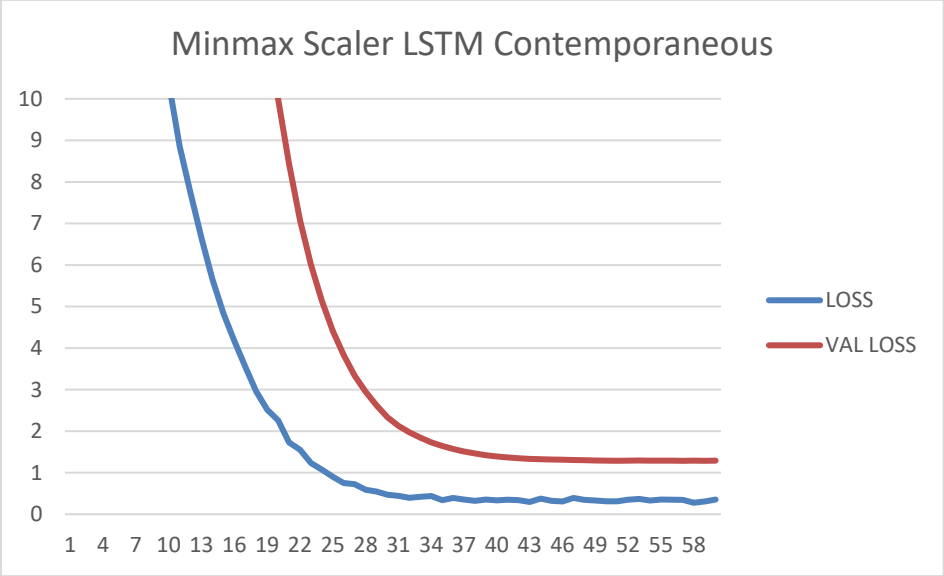


Figure 58. Loss/ Val loss curve of LSTM network for Contemporaneous and I 1000 dataset with Minmax Scaler.

5.4.2 Standard Scaler

With Standard Scaler applied in our I 1000 dataset the Bi-LSTM has a performance 4.4318 in MAE and 4.5008 VMAE. The GRU Network has MAE 5.0021 and VMAE 5.3025 with Standard Scaler applied and the LSTM Network the MAE 4.5948 is and the VMAE is 4.6140 with the same algorithm applied. The performance of Bi-LSTM with 4,4318 MAE, of this network is the best performance of all networks for Contemporaneous and I 1000 dataset.

Network	MAE	VMAE
Bi-LSTM	**4.4318**	4.5008
GRU	5.0021	5.3025
LSTM	4.5948	4.6140

Table 26. Bi-LSTM, GRU, LSTM networks performance comparison for Contemporaneous and I 1000 dataset for with Standard Scaler.

The hyperparameters was loss: mean_squared_error, the optimizer as: adam, the dropout set to: 0.2, the optimal learning rate set to: 1e-3 for Bi -LSTM and 1e-4 for GRU and LSTM, lstm_units set to: 200, epochs: 40 for Bi-LSTM and 60 epochs GRU and LSTM the the batch_size: 8 for Bi-LSTM, GRU and for LSTM the batch size set to 4, es_patience: 0.5 for Bi-LSTM and 0.4 for GRU and LSTM.

HYPERPARAMETERS with Standard Scaler								
NETWORK	**LOSS**	**OPTIMIZER**	**DROPOUT**	**LR**	**LSTM UNITS**	**EPOCHS**	**BATCH SIZE**	**ES PATIENCE**
Bi-LSTM	MSE	ADAM	0,2	1e-3	200	40	8	0.5
GRU	MSE	ADAM	0,2	1e-4	200	60	4	0.4
LSTM	MSE	ADAM	0,2	1e-4	200	60	4	0.4

Table 27. Bi-LSTM, GRU, LSTM networks optimal hyperparameter tuning for Contemporaneous and I 1000 dataset with Standard scaler.

With Standard Scaler algorithm applied the results in MAE in Bi- LSTM was close to Minmax Scaler algorithm with 4.4318 in MAE the LSTM has MAE 4.5948 and for GRU network the MAE is 5.0021.

The same goes for Standard Scaling. It did not help to reduce the underfit, nor should the models run and learn faster. The GRU model went even worse so he needed about 60 seasons to train. And here the Bi-LSTM model shows better behavior than the two others. It takes less time and resources to train and display the smallest underfit. As shown in the curves loss and val/ loss below.

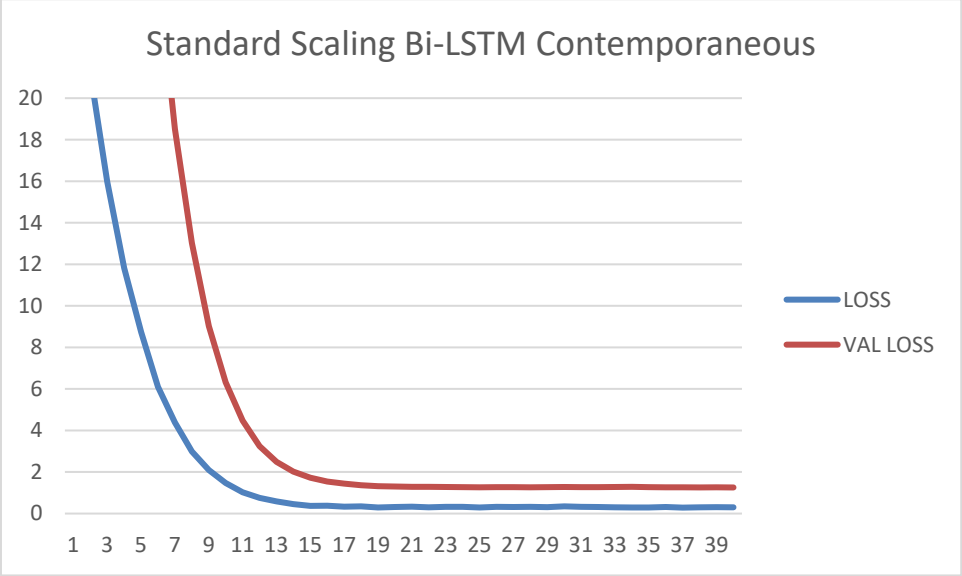


Figure 59. Loss/ Val loss curve of Bi-LSTM network for Contemporaneous and I 1000 dataset with Standard Scaler.

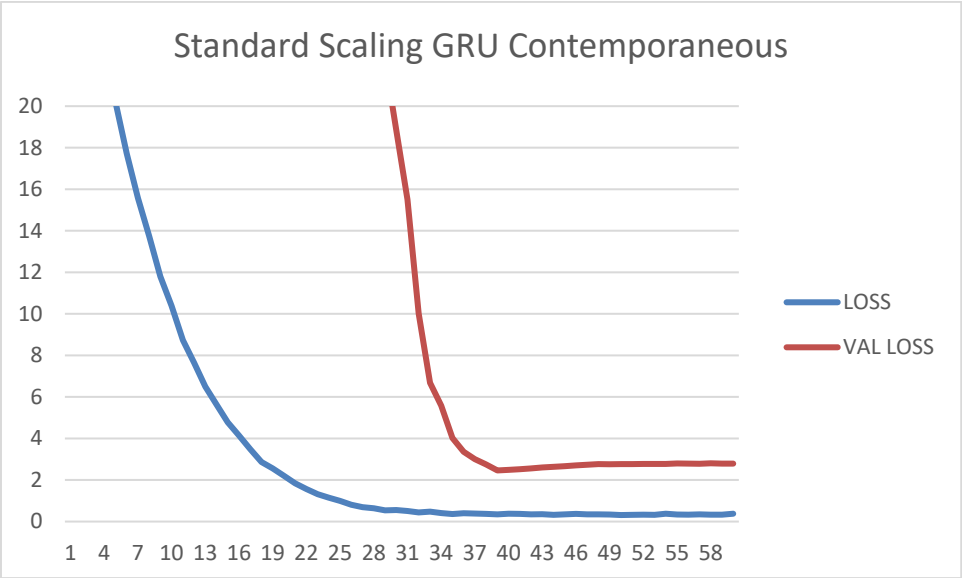


Figure 60. Loss/ Val loss curve of GRU network for Contemporaneous and I 1000 dataset with Standard Scaler.

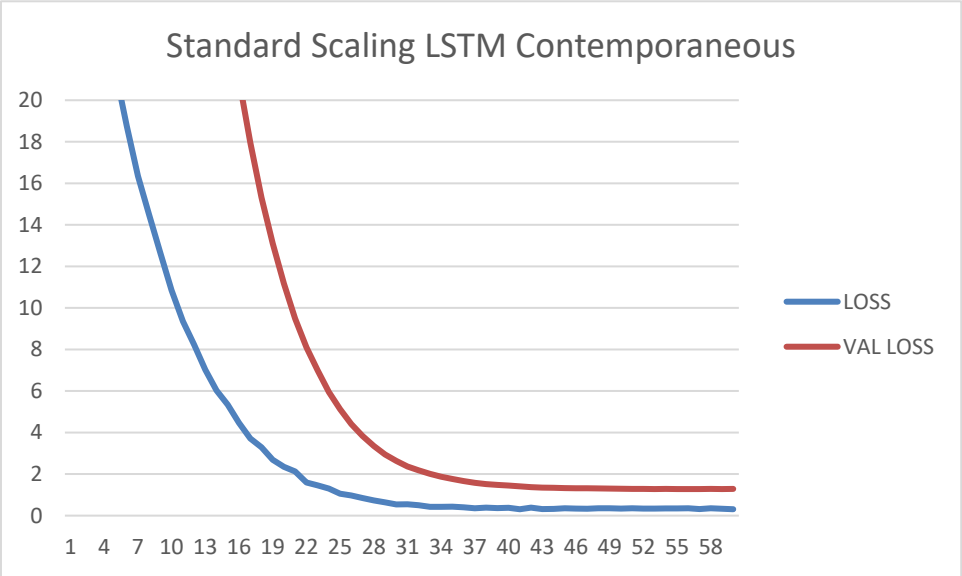


Figure 61. Loss/ Val loss curve of LSTM network for Contemporaneous and I 1000 dataset with Standard Scaler.

5.4.3 Prediction from previous Contemporaneous prices I 1000 dataset.

In order to predict the future prices of contemporaneous from the previous 30 contemporaneous prices of ALCOA, we trained a random forest model as we did for the I 600

dataset. The random forest regressor has a total Test error (MSE): 77.08506000000001. This can see to the chart below.

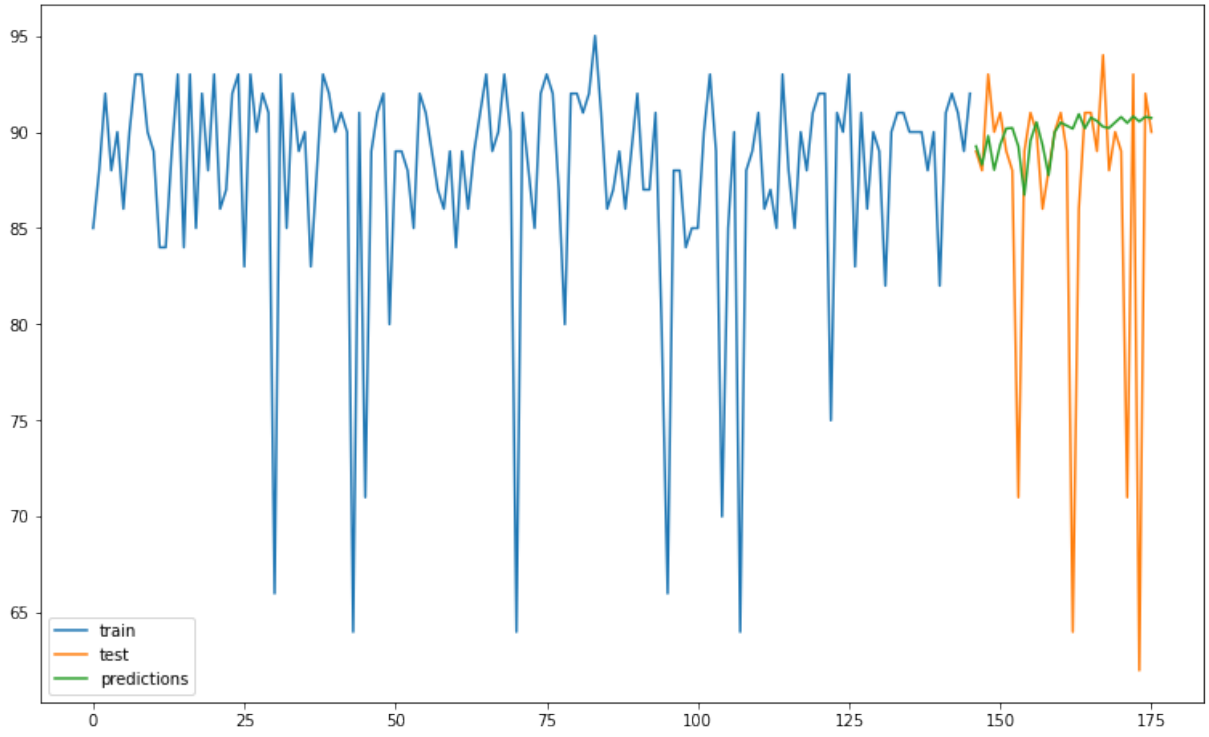


Figure 62. Prediction of Contemporaneous for I 1000 dataset with random forest regression model. The performance of the model was: Test error (MSE): 77.08506000000001.

The next model as we have seen is a random forest model with the best hyperparameters. The performance of this model is Test error (MSE): 70.67621716463478. As we can see this model failed to predict the ALCOA too.

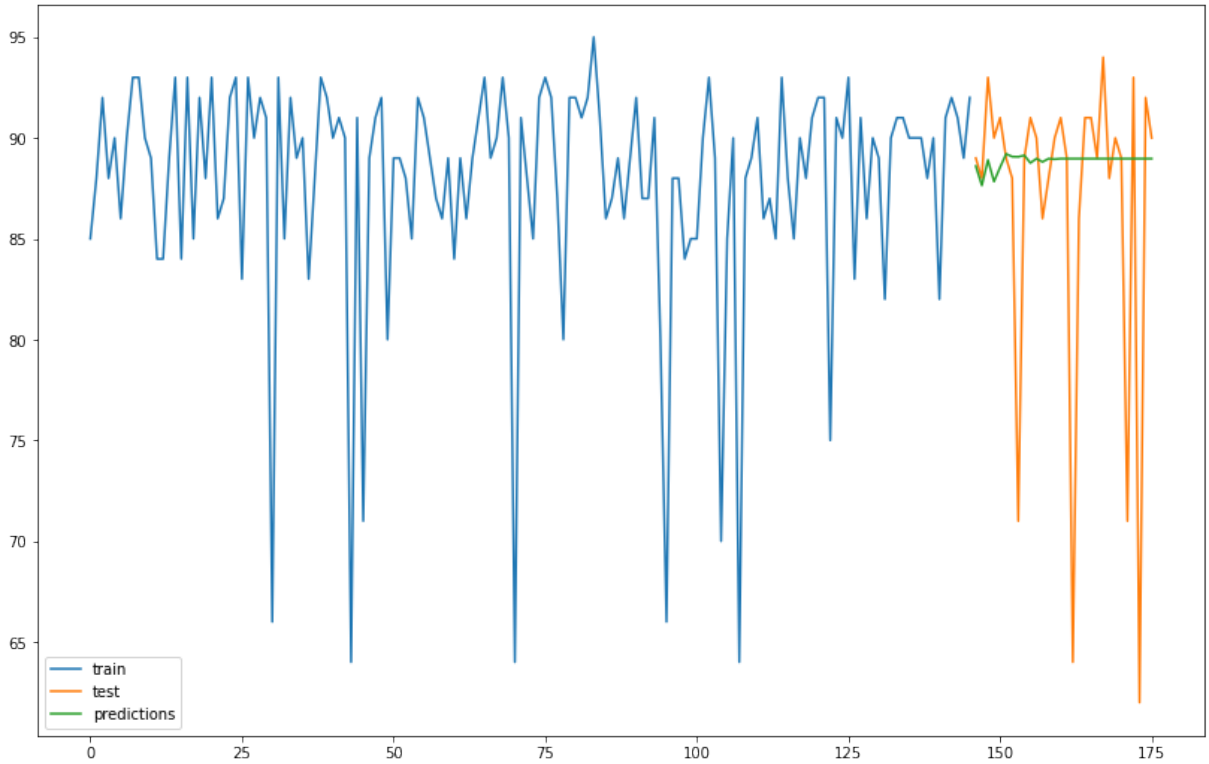


Figure 63. Prediction of Contemporaneous for I 1000 dataset with random forest regression model with grid search. The performance of the model was: Test error (MSE): 70.67621716463478.

The next model is the autoregressor model with lasso penalty which is used as a regression model. Having a Standard Scaler and one pipeline. The model performance is Test error (MSE) 65.75462106717687.

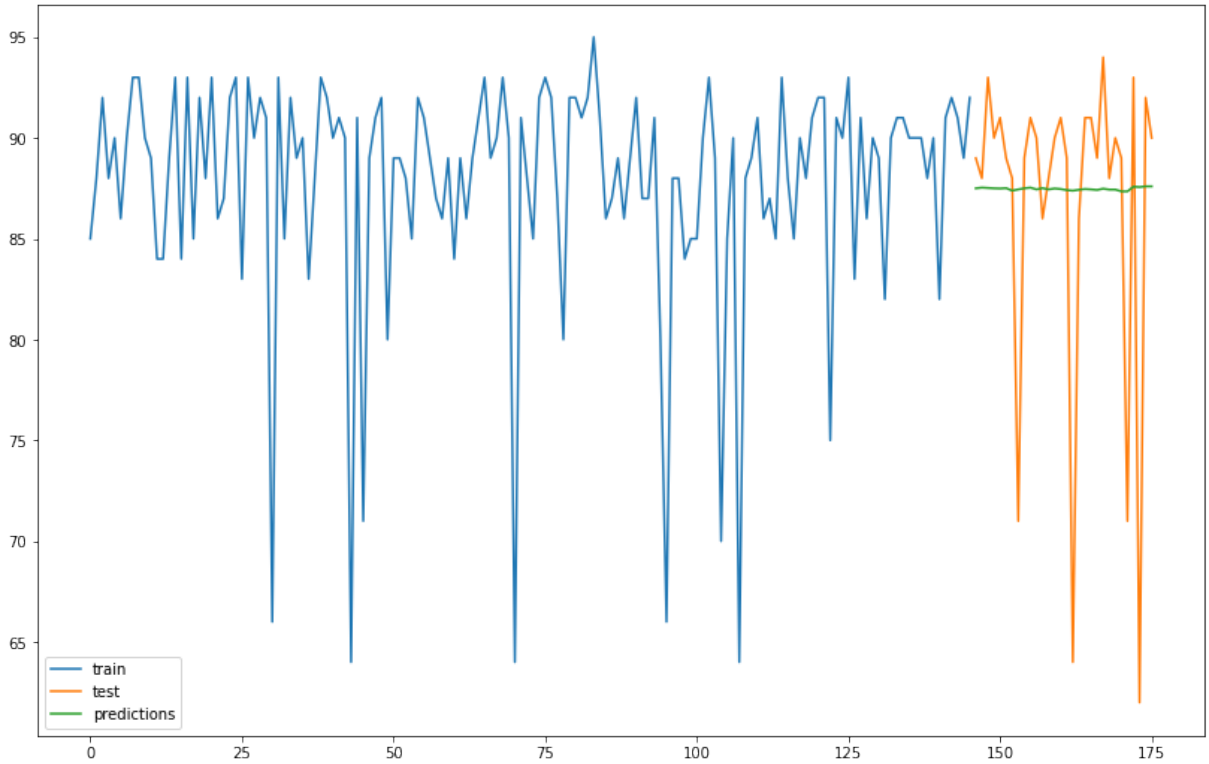


Figure 64. Prediction of Contemporaneous for I 1000 dataset with auto regressor model with Lasso Penalty. The performance of the model was: Test error (MSE) 65.75462106717687.

The last is the forecaster with linear regression in order to find the prediction error. The test error (MSE): 60.135593776713236 and the plot is seen in the chart below.

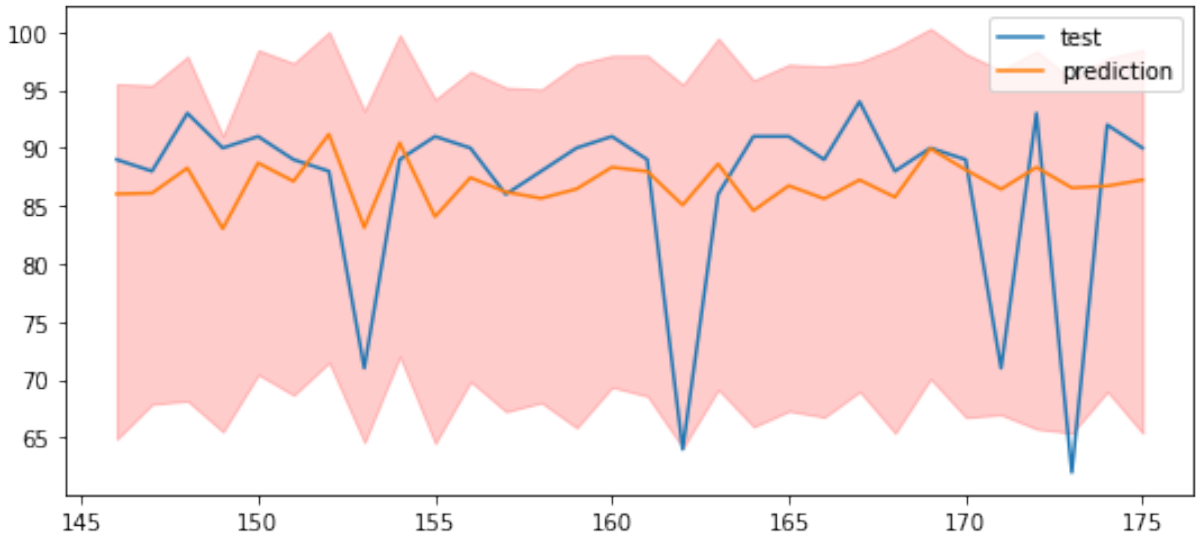


Figure 65. Prediction of Contemporaneous for I 1000 dataset with a linear regression model. The figure shows the last 30 batches which used for evaluate the prediction. The performance of the model was: Test error (MSE): 60.135593776713236.

As we can see in the table below the best performance is the Linear Regression model with the test error 60.135593776713236.

Dataset	ALCOA	Model	Performance
I 1000	Contemporaneous	Random Forest	Test error (MSE): 77.08506000000001
I 1000	Contemporaneous	Random Forest with best Hyperparameters	Test error (MSE): 70.67621716463478
I 1000	Contemporaneous	Lasso Alpha Penalty	Test error (MSE) 65.75462106717687
I 1000	Contemporaneous	Linear Regression	Test error (MSE): 60.135593776713236

Table 28. Performance comparison of the four regression models used for prediction of Contemporaneous for I 1000 dataset.

6 Conclusion

Certainly, the prices of Attributable and Contemporaneous can in no case be used to predict future ALCOA in a pharmaceutical industry. The Attributable has a much better performance than the Contemporaneous, but in no case can this be used as a prediction model of Attributable in a pharmaceutical industry. The possibility of data interference is limited to impossible. We could not add new data in any way and try to keep the data in its original form because such a thing would undermine the principle of ALCOA that his data that the data should be original and accurate.

The best result for Attributable principle for I 600 dataset is the Bi-LSTM Network with MAE 1.5878 and without any scaling. For I 1000 dataset the best performance result was the LSTM network with MAE 1.4858 with Minmax Scaling. As we see in dataset description for the I 600 dataset the standard deviation is 3,444936 for the prices and for the I 1000 dataset the standard deviation of the prices is 2,139274. The MAE performance of neural networks is better but not significance better to use them in prediction of this ALCOA principle.

For Contemporaneous principle the standard deviation of I 600 dataset is 4,615831 and for I 1000 dataset the standard deviation is 6,316490. The best performance is for I 1000 dataset is 4.4318 and achieved with Bi-LSTM network with Standard Scaling. The best performance for I 600 dataset is the 3.2269 and this this was achieved with Bi-LSTM with no scaling. The same as Attributable the performance of neural network is better than standard deviation of dataset prices but under no circumstances can these values be used to predict ALCOA prices since they contain a large percentage of error.

Another interesting observation is that for the I 1000 dataset and Attributable principle the Random Forest Model has Test error (MSE): 1.3849400000000012 which is a very good result. This should push us to use simple machine learning algorithms in the future and not to ignore

them because they might be able to perform better than the most complex deep learning models for certain datasets and for certain ALCOA.

As we saw our data was quite imbalanced and this seems to have affected the results of the artificial intelligence models we used. Also, the small range of data values from the time series seems to have affected the performance of the models used. This happens both production lines the I 600 and I 1000. Certainly, the designers of the system must give a new escalation to the data or a new way of recording so that they are not so imbalanced. Also, the designers of the system must give a new escalation to the data or a new way of recording so that they are not so invalid. In the short-sighted e.g., and for the I 1000 production line the lowest price is 62 and the maximum is 95. This could change with a scaling from 0 to 100 or from 0-10. ALCOA scales must be taken into account by the manufacturers of the system as they in the future should be ready to be introduced into machine learning models or artificial intelligence. Even from where we have a lot of imbalanced data such as the contemporaneous and accurate, we can use here binary classification. This would help us to use other artificial intelligence and machine learning systems with less and simpler data.

Also, the optimal hyperparameter tuning used did not help enough to make the algorithms learn better. The results were very close and usually whichever model produced the best results from the beginning did not change even after the optimal hyperparameter tuning. It generally did not help the performance of the system at all and this should be taken into account in the future where more and more data will be added.

Minmax Scaling and Standard Scaling were also done to help some algorithms learn faster. Scaling the data makes it easier for our model to learn and understand the problem. In the case of our neural networks, an independent variable with a value spread can lead to a large loss in training and testing and cause an unstable learning process. As we have seen this worked in part. The algorithms learned best by reducing the seasons to about 5-10 like we see in GRU at Contemporaneous in which seem to learn quickly when we apply Standard Scaling while the GRU continue to learn even after 60 epochs. This certainly offers a trump card in terms of system performance. Not so much in the level of results but in the possibility of learning if in the future other such experiments are done with much more data that will have to be collected.

After all this was done, an attempt was made to predict the future ALCOA from the previous ones, which, as we saw, did not work. None of the four models used performed properly to make a reliable prediction. The forecasting models could not predict future prices of ALCOA at all and in fact in the case of the contemporaneous there are chaotic values of deviation of the order of Test error (MSE): 77.085 for the I 1000 dataset.

It is the nature of the pharmaceutical industry such that even a forecast model with MSE 1.58 like the Bi-LSTM model of Attributable like we have achieved in I 600 when the standard deviation for our data is 3.34 is quite high value and especially when it comes for ALCOA.

7 Future Research

As we have seen, the prediction of ALCOA Attributable and Contemporaneous is not possible with these three neural networks presented in this dissertation. Definitely a suggestion in the future is to use other learning networks or a combination of these. Also, simple machine learning algorithms can enter the equation. Maybe they can perform better in some datasets and certain ALCOA's. Surely another research should focus on other remaining ALCOA like Legible, the Original and Accurate with the same neural network that used in this thesis, in order to predict the future ALCOA's.

This research has many limitations. Definitely a limitation is the available literature. When there is no bibliography there is nothing to compare or rely on to continue. There is not much research on ALCOA and artificial intelligence models. Also, a limitation is the small volume of data since few pharmaceutical companies have such a system installed for research purposes only. Also, in terms of data we saw that it is very imbalanced. Certainly no one was ready to enter data for processing into intelligent and machine learning algorithms. Also, our data was data from sensors during the production line. What would happen if our data, for example, were photos or text? There we could use other models of artificial intelligence and machine learning with better results.

Of course, this graduate is, as we said, something new. An attempt to predict ALCOA's from a pharmaceutical derivative line. This may give the impetus for other pharmaceutical industries to incorporate this system that ensures the timeliness of the data and by extension the quality of the medicine. It can also encourage other researchers since they will have more data at their disposal to deal with the prediction of the specific ALCOA's as well as the other three that were not mentioned here.

The data generated as we said, during the production process is really a lot. A few of them are important though. The torch source industry analytics could certainly help in a future investigation. It could focus, for example, on points that have been shown to have the most errors or no errors at all. Where the data follows a sequence. Certainly, with our data such a thing could not be done because the sample was already very small. Certainly, a lot needs to be done at the level of sample preparation before they are introduced into the models to be tested. The acronym ALCOA has a bottom-up philosophy, from the base to the top or from the top to the base or 360 model. Perhaps the data should not have focused only on the production process but on the production of the raw material before it was imported to the factory. Its completion can be with the delivery of the drug to the final consumer. There would be other data and certainly more.

In the light of the importance, we have placed on our data, there is a serious issue of possible selection bias when sampling our sample. Our sampling strategy explicitly included a sampling approach from two single production lines, we believed that some significant development was taking place and some conclusions could be drawn for our research, samples that have no significance. These two production lines I 1000 and I 600 we do not even know if they produce the same drug. If they do not produce the same drug, other physicochemical parameters have, for example, the production of a cream and other physicochemical parameters have the production of a pill. so we have different alarms. In the case of the production of a pill the temperature can be ambient temperature so the values in the temperature sensors do not exist

while in the case of the cream the temperature is higher and the temperature sensors are recorded normally. This was evident, for example, with some values from some sensors which were missing.

Perhaps the only case where the prediction can be used for these ALCOA's and especially for the Attributable should be investigated whether these models could be used as early prediction models. A statement if it is to become an anomaly during a production so that we can intervene 20% -30% faster and take corrective action. This can be important because a corrective action can save a drug from destroying the entire batch. Something like this should not be considered negligible. Many drugs cost thousands of euros especially when they are on the derivative. And the fact that we will be able to intervene before something happens means great savings for the company that produces it.

Certainly, the ALCOA field has a lot to offer. As we have seen in the literature, artificial intelligence will be everywhere in a pharmaceutical industry. Quality will also be everywhere in terms of the pharmaceutical industry, so the marriage of one philosophy with another will surely take place somewhere with very good results. If the ALCOA acronym is finally adopted by the pharmaceutical industry and these pharmaceutical companies come together, they can work as nodes that will feed artificial intelligence algorithms and will be able to draw conclusions in general or in particular.

With the help of artificial intelligence, we can move from supporting only business goals to models that ensure the safety, quality and effectiveness of the pharmaceutical product. It would be a very good start of a very good definition of data governance in pharmaceutical industries. Effective data management with the help of artificial intelligence will provide added value, including increasing consistency and confidence in decision-making and improving data security. It will maximize the potential of the data while minimizing or eliminating the failure of the result and the repetition of a task or a procedure. The concepts and definitions of data governance in public authority guidance generally have a rather limited scope in terms of data integrity and do not cover the broader concept of data quality. This with the help of artificial intelligence can change and breathe new life into this field.

Bibliography – References – Online sources

8 Bibliography

[1] N. S. A. C. F. T. X. Y. L. L. Kopcha, "Industry 4.0 for pharmaceutical manufacturing: Preparing for the smart," *International Journal of Pharmaceutics,* vol. Volume 602, 2021.

[2] A. H. R. P. S. a. R. S. Mohd Javaid, "Artificial Intelligence Applications for Industry 4.0: A Literature-Based Study," *Journal of Industrial Integration and Management,* vol. 07, no. 1, pp. 83-111, 2022.

[3] J. L. R. L. Y. Z. a. W. Z. Sheela Kolluri, "Machine Learning and Artificial Intelligence in Pharmaceutical Research," *The AAPS Journal,* 2022.

[4] Moderna, "Important Safety Information & EUA," Moderna, 2022. [Online]. Available: https://eua.modernatx.com/covid19vaccine-eua/providers/storage-handling. [Accessed 23 6 2022].

[5] N. B. P. E. R. L. T. L. D. R. N. Schneider, "Artifcial intelligence in chemistry and drug design," *Journal of Computer-Aided Molecular Design,* pp. 709-715, 2020.

[6] T. Z. W. J. S. S. Yue Li, "Materials discovery and design using machine learning," *Journal of Materiomics,* vol. 3, no. 3, pp. 159-177, 2017.

[7] G. S. S. S. D. K. K. K. a. R. K. T. Debleena Paul, "Artificial intelligence in drug discovery and development," *Drug Discovery Today,* no. 1, p. 80–93., 2021.

[8] S. D. S. S. J. C. M. B. &. M. J. A. Suresh Dara, "Machine Learning in Drug Discovery: A Review," *Artificial Intelligence Review,* no. 1, p. 1947–1999, 2021.

[9] https://clinicaltrials.gov/, "Clinical Trial For SARS-CoV-2 Vaccine (COVID-19)," Us National Library of Medicine, [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT04582344. [Accessed 23 6 2022].

[10] J. W. M. S. E. C. S. C.-T. N. G. J. C. E. D.-C. B. K. H. G. a. J. S. Y. Arash Keshavarzi Arshadi, "Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development," *frontiers in Artificial Inteligence,* 2020.

[11] M. H. D. A. D. R. R. S. M. F. L. K. K. G. A. K. R. S. P. L. A. L. Ajay Vikram Singh, "Artificial Intelligence and Machine Learning in Computational Nanotoxicology: Unlocking and Empowering Nanomedicine," *Advanced Healthcare Materials,* vol. 9, no. 17, 2020.

[12] C. f. T. Innovation, "Protecting privacy in an AI-driven world," Brookings, 10 2 2020. [Online]. Available: https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/. [Accessed 23 7 2022].

[13] IBM, "What is artificial intelligence in medicine?," IBM, [Online]. Available: https://www.ibm.com/topics/artificial-intelligence-medicine?gclid=Cj0KCQjwntCVBhDdARIsAMEwACk6mYgmcrAEIcIFmnw9MCAnG6au-9aSxq4WI6gVtP3FoX9slgQCrS8aAufKEALw_wcB&gclsrc=aw.ds. [Accessed 23 6 2022].

[14] M.I.T., "M.I.T, News Artificial intelligence yields new antibiotic," M.I.T, 20 2 2020. [Online]. Available: https://news.mit.edu/2020/artificial-intelligence-identifies-new-antibiotic-0220. [Accessed 23 6 2022].

[15] A. H. R. P. S. S. R. R. S. Mohd Javaid, "Significance of sensors for industry 4.0: Roles, capabilities, and applications," *Sensors International,* vol. 2, 2021.

[16] E. M. Agency, "Guidance on good manufacturing practice and good distribution practice: Questions and answers," https://www.ema.europa.eu/, 2022. [Online]. Available: https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-manufacturing-practice/guidance-good-manufacturing-practice-good-distribution-practice-questions-answers. [Accessed 22 6 2022].

[17] E. M. Agencey, "Good manufacturing practice," European Medicines Agencey, 2022. [Online]. Available: https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-manufacturing-practice. [Accessed 23 6 2022].

[18] FDA, "Facts About the Current Good Manufacturing Practices (CGMPs)," FDA, 6 1 2021. [Online]. Available: https://www.fda.gov/drugs/pharmaceutical-quality-resources/facts-about-current-good-manufacturing-practices-cgmps. [Accessed 23 6 2022].

[19] "Counting the cost of failure in drug development," https://www.pharmaceutical-technology.com/, 17 6 2017. [Online]. Available: https://www.pharmaceutical-technology.com/analysis/featurecounting-the-cost-of-failure-in-drug-development-5813046/. [Accessed 22 6 2020].

[20] COPADATA, "6 ways to increase productivity and quality in pharmaceutical manufacturing," COPADATA, [Online]. Available: https://www.copadata.com/en/industries/pharmaceutical/life-sciences-pharmaceutical-insights/six-ways-to-increase-quality-and-productivity-in-pharmaceutical-manufacturing/. [Accessed 23 6 2022].

[21] M. M. K. Z. I. V. W. Uthayasankar Sivarajah, "Critical analysis of Big Data challenges and analytical methods," *Journal of Business Research,* vol. 70, pp. 263-286, 2017.

[22] P. Toni Manzano Gilad Langer, "Getting Ready for Pharma 4.0™," ISPE, 10 2015. [Online]. Available: https://ispe.org/pharmaceutical-engineering/september-october-2018/getting-ready-pharma-40tm. [Accessed 23 6 2022].

[23] VIMACHEM, "A PRACTICAL GUIDE TO PHARMA 4.0 REALIZATION," https://www.vimachem.com/, [Online]. Available: https://www.vimachem.com/resources/pharma-4-0/. [Accessed 23 6 2022].

[24] "Big Data in the pharmaceutical industry: benefits and applications," doxee.com, 7 4 2022. [Online]. Available: https://www.doxee.com/blog/technology/big-data-in-pharmaceutical-sector/. [Accessed 23 6 2022].

[25] TechTarget, "What is cloud backup and how does it work?," TechTarget, 8 2020. [Online]. Available: https://www.techtarget.com/searchdatabackup/definition/cloud-backup. [Accessed 23 6 2022].

[26] A. J. Analyst(s): Douglas Laney, *100 Data and Analytics Predictions Through 2021,* https://mscdss.ds.unipi.gr/wp-content/uploads/2018/02/100_data_and_analytics_predictions_through_2021.pdf, 2017.

[27] C. S. r. M. Craig Stedman, "https://www.techtarget.com/," Techtarget, [Online]. Available: https://www.techtarget.com/searchcio/definition/data-collection. [Accessed 23 6 2022].

[28] M. K. Pratt, "How big data collection works: Process, challenges, techniques," TechTarget, 7 2 2022. [Online]. Available: https://www.techtarget.com/searchdatamanagement/feature/Big-data-collection-processes-challenges-and-best-practices. [Accessed 23 6 2022].

[29] WHO, "WHO good practices for pharmaceutical," [Online]. Available: https://www.who.int/docs/default-source/medicines/norms-and-standards/guidelines/quality-control/trs957-annex1-goodpractices-harmaceuticalqualitycontrol-laboratories.pdf?sfvrsn=ca0c211c_0. [Accessed 23 6 2022].

[30] A. Zeneca, "Data Science & Artificial Intelligence: Unlocking new science insights," Astra Zeneca, [Online]. Available: https://www.astrazeneca.com/r-d/data-science-and-ai.html. [Accessed 26 6 2022].

[31] C. Kathleen Walch, "Big data vs. machine learning: How they differ and relate," 27 4 2021. [Online]. Available: https://www.techtarget.com/searchbusinessanalytics/tip/Big-data-vs-machine-learning-How-they-differ-and-relate. [Accessed 23 7 2022].

[32] "Blockchain Changes the Pharmaceutical Industry," intellectsoft, 27 5 2021. [Online]. Available: https://www.intellectsoft.net/blog/blockchain-in-the-pharmaceutical-industry/. [Accessed 23 6 2022].

[33] IBM, "IaaS versus PaaS versus SaaS," IBM, [Online]. Available: https://www.ibm.com/cloud/learn/iaas-paas-saas. [Accessed 23 6 2022].

[34] G. H. S. H. McFalone-Shaw, "Digitalization in pharmaceutical industry: What to focus on under the," *International Journal of Pharmaceutics: X,* vol. 4, 12/2022.

[35] Forbes, "What Is Bitcoin And How Does It Work?," Forbes, [Online]. Available: https://www.forbes.com/advisor/investing/cryptocurrency/what-is-bitcoin/. [Accessed 23 6 2022].

[36] M. G. &. C. V. Meet Kumari, "Blockchain in Pharmaceutical Sector," *Applications of Blockchain in Healthcare ,* vol. 83, pp. 199-220, 2020.

[37] Techtarget, "Top 11 cloud security challenges and how to combat them," Techtarget, [Online]. Available: https://www.techtarget.com/searchsecurity/tip/Top-11-cloud-security-challenges-and-how-to-combat-them. [Accessed 23 6 2022].

[38] P. José Rodríguez-Pérez, Data Integrity and Compliance, A Primer for Medical Product Manufacturers, Milwaukee, Wisconsin: ASQ Quality Press, 2019.

[39] M. Inc, *GAMP® 5: A Risk-based Approach to Compliant GxP Computerized Systems,* MasterControl Inc, 2019.

[40] D. A. (.). P. Kevin C Martin, "GAMP 5 quality risk management approach," *Pharmaceutical Engineering,* 5 2008.

[41] B. ParkWoolf, "Chapter 7 - Machine Learning," in *Building Intelligent Interactive Tutors*, 2009, pp. 221, 223-297.

[42] Z. Y. H. F. T. a. M. D. Frank Emmert-Streib, "An Introductory Review of Deep Learning for Prediction Models With Big Data," *Frontiers in Artificial Intelligence publishes,* 2020.

[43] D. Kalita, "A Brief Overview of Recurrent Neural Networks (RNN)," Analytics Vidhya, 11 03 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/03/a-brief-overview-of-recurrent-neural-networks-rnn/. [Accessed 24 6 2022].

[44] E. S. N. K. S. ManiSarathy, "Chapter 9 - Using deep learning to diagnose preignition in turbocharged spark-ignited engines," in *Artificial Intelligence and Data Driven Optimization of Internal Combustion Engines*, 2022, pp. 213-257.

[45] N. Arbel, "How LSTM networks solve the problem of vanishing gradients A simple, straightforward mathematical explanation," medium.datadriveninvestor.com, 21 12 2018. [Online]. Available: https://medium.datadriveninvestor.com/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577. [Accessed 24 6 2022].

[46] "Bidirectional LSTM," Meta AI, 23 6 2022. [Online]. Available: https://paperswithcode.com/method/bilstm. [Accessed 23 6 2022].

[47] S. Kostadinov, "Understanding GRU Networks," https://towardsdatascience.com/, 16 12 2017. [Online]. Available: https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be. [Accessed 23 6 2022].

[48] F. Chollet, Deep Learning with Python, MANNING, 2017.

[49] J. Brownlee, "How to Use StandardScaler and MinMaxScaler Transforms in Python," https://machinelearningmastery.com, 10 6 2020. [Online]. Available: https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/. [Accessed 24 6 2022].

[50] J. E. O. Joaquín Amat Rodrigo, "Skforecast: time series forecasting with Python and Scikit-learn," https://www.cienciadedatos.net, 1 2 2021. [Online]. Available: https://www.cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html. [Accessed 26 6 2022].

[51] G. L. Team, "A Complete understanding of LASSO Regression," Great Learning, 26 12 2021. [Online]. Available: https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/. [Accessed 2022 12 24].

[52] J. E. O. Joaquín Amat Rodrigo, "Skforecast: time series forecasting with Python and Scikit-learn," https://www.cienciadedatos.net/, 1 2 2021. [Online]. Available: https://www.cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html. [Accessed 24 6 2022].

[53] E. D. P. Supervisor, "EUROPEAN DATA PROTECTION SUPERVISOR," 16 3 2018. [Online]. Available: https://edps.europa.eu/sites/edp/files/publication/18-03-16_cloud_computing_guidelines_en.pdf. [Accessed 23 6 2022].

[54] https://www.pqegroup.com/, "SPuMoNI, the European project about Big Data and process modelling for smart industry," pqegroup.com, [Online]. Available: https://www.pqegroup.com/blog/2019/12/spumoni-the-european-project-about-big-data-and-process-modelling-for-smart-industry/. [Accessed 23 6 2022].

[55] CHIST-ERA, "SPuMONI - Smart Pharmaceutical MaNufacturIng," CHIST-ERA, [Online]. Available: https://www.chistera.eu/projects/spumoni. [Accessed 23 6 2022].

[56] S. J. Schniepp, "ALCOA+ and Data Integrity," https://www.pharmtech.com/, 2 10 2019. [Online]. Available: https://www.pharmtech.com/view/alcoa-and-data-integrity. [Accessed 23 7 2022].

[57] FDA, *Data Integrity for the FDA Regulated Industry,* Quality Systema Compialance, 2019.

[58] Á. R. ·. H. A. ·. P. R. ·. C. ·. I. O. ·. M. Editors, Trends and Innovations in Information Systems Technologies, Conference proceedings, 2020.

[59] L. Piccioli, "SPuMoNI, the European project about Big Data and process modelling for smart industry," PQE, [Online]. Available: https://www.pqegroup.com/blog/2019/12/spumoni-the-european-project-about-big-data-and-process-modelling-for-smart-industry/. [Accessed 23 6 2022].

[60] c. A. E. S. H. J. M. M. A. C. A. V. C. A. E. T. E. FátimaLeala, "Smart Pharmaceutical Manufacturing: Ensuring End-to-End Traceability and Data Integrity in Medicine Production☆," *Big Data Research,* vol. 24, 2021.

Appendix A

Code A I 600 Bi-LSTM GRU LSTM Models.

Appendix B

Code B I 1000 Bi-LSTM GRU LSTM Models.

Appendix C

Code C I 600 Machine Learning Models.

Appendix D

Code D I 1000 Machine Learning Models.