AA20 - Real time Prediction of Evaporator Circuit Performance Using ML Based Soft Sensors

Abhishek Seth[1], Bharathesh Kumar[2], Chandrakala Kari[3], Vilas Tathavadkar[4],
Prince Siddhartha[5], Ameet Chaure[6], Babji Pusarla[7], Rajesh Khuntia[8], Kausikisaran Misra[9]
1. Scientist (ABSTC)
2. Scientist (ABSTC)
3. Lead Scientist (ABSTC)
4. Chief Technology Officer Hindalco & Joint President
Aditya Birla Science & Technology Company Pvt. Ltd. Taloja MIDC Panvel, Maharashtra, India.
5. Assistant Manager (Technical)
6. Assistant General Manager (Technical)
7. Manager (Operations)
8. Assistant General Manager (Operations)
9. Vice president (Operations)
Utkal Alumina International Limited, Rayagada, Odisha, India
Corresponding author: chandrakala.kari@adityabirla.com

Abstract

In alumina refineries, evaporation is the critical process used to concentrate spent liquor to achieve desired caustic concentration for the digestors and to maintain the process water balance. The evaporation circuit contributes to ~ 20–25 % of the total energy requirement; this is third highest energy consuming process after calcination and digestion. The energy or heat source is live steam which demands are as high as 40–50 % of total refinery steam requirement. Therefore, this process needs special attention to control the steam input and maintain consistent performance of the evaporator steam economy. To optimize and control the evaporator performance, soft sensors have been developed to predict quality parameters such as the concentration of strong evaporated liquor (SEL), and steam economy. This will help to minimize the variation in evaporator steam economy while maintaining consistent product quality. An advanced machine learning (ML) algorithm i.e., random forest, is used to develop predictive models based on historian data base for both SEL concentration & steam economy (SE). These models are optimized with hyper parameter tuning to achieve higher accuracy. The model accuracy for the prediction of SEL concentration is obtained in the acceptable range of ± 2–3 g/L. The minute-wise prediction dashboard has been developed for the area concerned to take prompt decisions to achieve consistently high steam economy which is utmost important towards saving energy. The predictions were validated using real-time DCS (distributed control system) plant data on a minute-wise basis. Through a web-based GUI (graphical user interface) platform, real time predictions from the model and actual plant measurements were pulled, displayed, and compared to showcase the critical parameters for the prediction accuracy.

Keywords: Soft sensor, Machine learning model, Spent liquor, Strong evaporator liquor, Steam economy.

1. Introduction

The evaporator is the second highest energy intensive section in the liquor circuit of the Bayer process. It plays a key role to concentrate spent liquor caustic to the desired level; this liquor is circulated to the digester to minimize the loss of caustic and maintain the plant water balance. Considering the energy efficiency, a multi-effect falling evaporator having steam economy ranging 3–4 is widely used in alumina refineries. Obtaining reduced steam consumption while maintaining the quality of product liquor i.e. SEL (Strong Evaporated Liquor) is crucial for improving the efficiency of the evaporator circuit.

Many researchers have made efforts to model the industrial-scale evaporator; one approach follows an empirical correlation using residence time distribution to model the falling film evaporator [1]. However, due to the complexity involved in the process because of simultaneous evaporation and condensation, the empirical correlation may not provide an appropriate description of the process. The dynamic behavior of multi-effect evaporators was investigated by disturbing the input parameters as a different approach [2]. A generalized mathematical model, which could be applicable to any number of effects was also attempted. This lumped and distributed mathematical models were developed for four effects falling film evaporator [3]. The results indicated that the distributed model had better prediction than the lumped model. Thereafter, combined model of lumped and distributed was used to predict the outlet liquor concentration for the evaporator circuit [4]. The simulated values of concentration were close to the measured values. From the perspective of energy consumption point, it was investigated for multi-effect evaporators with and without mechanical vapor recompression (MVR) [5]. The evaporator integrated with MVR consumed one-third of the energy compared with the conventional one. Although MVR technology seems promising, it requires further evaluation to be implemented in the alumina refineries.

Despite having the advantages with mathematical modelling, the online prediction of desired parameters through advanced data analytics resolves the issues related to offline analysis. Since process data is a direct reflection of the performance of the evaporator, the development of an online predictive model through soft sensing method seemed to be a reliable approach for accurate representation.

The technological development in the alumina industry has aided in developing approaches and algorithms to predict certain physical or quality parameters through soft sensors. A stepwise linear regression approach was used to predict the strength of alumina crystal conglomerates from alumina refinery [6]**.** Another approach, where-in comparison between different methods such as artificial neural network and multiple linear regression method was made for prediction of the alumina recovery efficiency. The amount of red mud produced, red mud settling rate and bound-soda losses in the red mud were studied [7]. The Radial Basis Function under the neural network function turned out to be the best-fitted approach for the considered data set. A predictive model using Support Vector Machine (SVM) was developed for predicting the leaching rate [8]**.** Leaching rate of alumina in the digestion process obtained through laboratory analysis delays the course of action. This causes less control of the process creating problems such as a low leaching rate and excess energy. Distributed support vector machine-based soft sensor to predict the quality of the digested slurry online was developed [9]. Based on expert knowledge and the mechanism of the blending and digestion process, a hybrid expert control system for supervisory control of the blending process was developed to optimize the raw material proportioning.

Limited attempts have been reported on developing soft sensors in the evaporator section with industrial validation. This present work is aimed to fill this gap by finding optimum independent parameters which can open opportunities for steam reduction. The development of a soft sensor would be beneficial to predict SEL (g/L) to take proactive actions to maintain desired SEL (g/L) and achieve consistent steam consumption.

2. Technical Approach

2.1 Overview of evaporator circuit

The circuit consists of a falling film multi-effect evaporator (feed backward) with the storage flash vessels attached, whose vapors are transported from one effect to another, under vacuum. There are 6 effects in falling film multi-effect evaporator. The first effect evaporator uses steam, which exchanges heat between steam and spent liquor in tubes. In storage flash tanks, liquor flashes and vapors are transported to another effect through demister. Demister removes entrained liquid carried away with vapors thereby improving heat transfer.

The evaporators in series are shown in Figure 1, where feed SPL splits into 2 streams (A & B), one is fed to 1st feed flash drum (FFD) & another to 4th effect preheater (PH). SPL passes in series from 4th to 1st effect and then it enters multiple product flash drum, where it flashes. Another stream enters feed flash drums, then into 2 effects and finally thick liquor from the product flash drum (PFD) & 5th effect is mixed to give SEL.

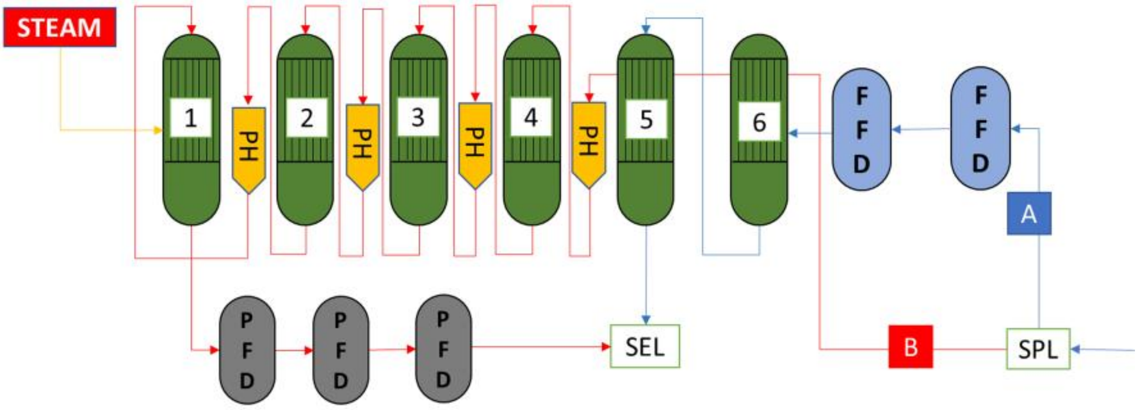


Figure 1. Schematic of falling film evaporator

Steam condensate is flashed, overflow vapor is sent to effect and underflow to the power plant. Similarly, the vapor condensate of the rest of the effects are flashed, and finally, all the process condensate is mixed and circulated for different application like hydrate washing etc.

2.2 Methodology

Minute-wise DCS process data and concentrations and densities of SPL and SEL are provided, which are measured in the laboratory once per shift (8 hourly). Instantaneous data is taken from the DCS readings and mapped with the lab data for the same time interval, considering residence time, which provides us with timestamp data. Treating null values, removing duplicate rows, converting features to appropriate datatype & removing outliers, to clean the dataset was carried out [10].

This methodology being supervised machine learning, correlation matrix is obtained which shows the relationship between the independent and dependent parameter, to choose best-related features with output & to drop all irrelevant features. This decision is taken by clubbing this result with process domain knowledge.

To predict SEL (g/L) and steam economy, a random forest machine learning algorithm was developed after tuning hyperparameters. Random forest forms various decision trees which brings randomness in samples as well as features [10, 11]. Total data is divided into training and test datasets. With the help of the training dataset it develops rules and regulations between known input and output, then it is validated by test data. The random forest uses method of bagging, where it randomly selects samples and features to create decision tree to arrive at the class label (prediction of SEL and steam economy). The whole dataset has been divided into training and testing dataset, 80 - 20 % split. Multiple decision trees are built, which take features and samples randomly. Every node poses different conditions on dataset, thereby making perfect split. Gini index is used to ensure perfect split and maximize information gain. Finally, each decision tree predicts class labels, whose mean is taken to report the predicted value. The overall development of the predictive model used as a soft sensor for real-time predictions of SEL (g/L) and SE from the evaporator is depicted in Figure 2.

Currently, the SPL (g/L) is found to vary between a range of 15 g/L. SEL majorly depends on SPL concentration, temperature, steam flow rate and vacuum pressure maintained. There are several hyperparameters viz., N_estimator, Max_features, Max_depth, Min_samples_leaf, Min_samples_split, Oob score, Bootstrap, N_jobs, specific to the random forest are considered to optimally train the model to arrive at prediction with high accuracy.

Where,

- N_estimator = Maximum number of decision trees.
- Max_features = Maximum features to be taken for building decision tree.
- Max_depth = Maximum depth of the tree.
- Min_samples_leaf = Minimum number of samples in leaf node.
- Min_samples_split = Minimum number of samples in terminal node to split.
- Oob score = fraction of data to test/validate model.
- Bootstrap = whether samples to be chosen by the method of bagging.
- N_jobs = Number of processors for parallel computing.

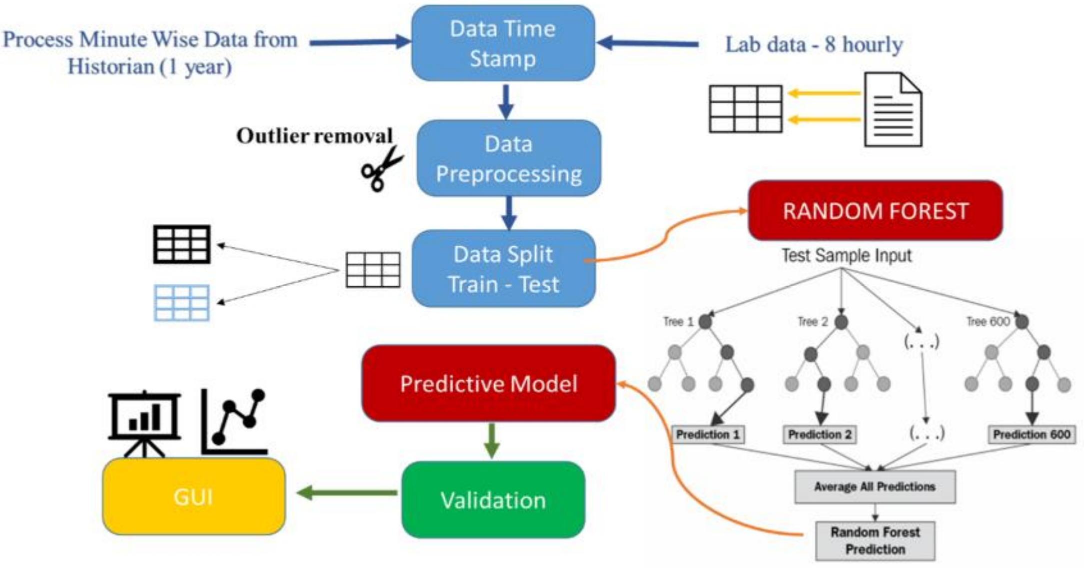


Figure 2. Schematic of development of predictive model

To know the best fit of the hyperparameter for the used training dataset, grid search or randomized grid search cross-validation method is applied. It is provided with a set of combination values for all the hyperparameters mentioned above and the best values which fits train_x and train_y well, are chosen for the model.

3. Results and Discussions

3.1 Exploratory Data analysis

Univariate, multivariate and time series data analysis are carried out for the process parameters using Tableau and Excel. Several insights are obtained which are validating the actual scenario.

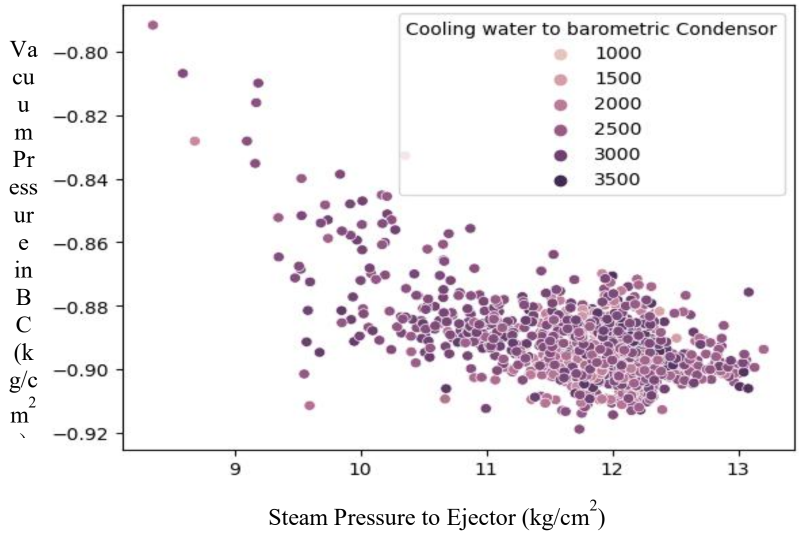


Figure 3. Dependency of vacuum in BC on steam pressure with respect to cooling water (m^3/h).

To maintain vacuum in the barometric condenser, it is steam pressure and cooling water flow rate responsible (refer to Figure 3). MP Steam pressure is crucial for vacuum in the barometric condenser (BC).

Cooling water temperature plays a great role. Low temperature of cooling water would be favorable for condensing the 6th effect vapors. The higher cooling water flow rate results in creating higher pressure drop. The higher steam flow rate results in the higher chest pressure in 1st effect (refer to Figure 4).

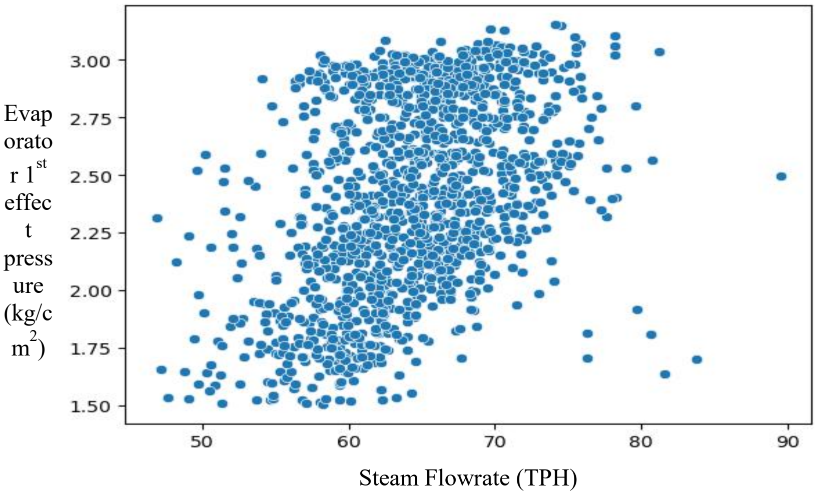


Figure 4. Dependency of steam flowrate on 1st effect pressure.

Steam entering 1st effect should be at its saturated temperature. If steam enters at its superheated temperature, both sensible and latent heat need to be extracted during heat transfer. At first, heat is extracted from sensible heat and its continuous flow increases chest pressure of the effect.

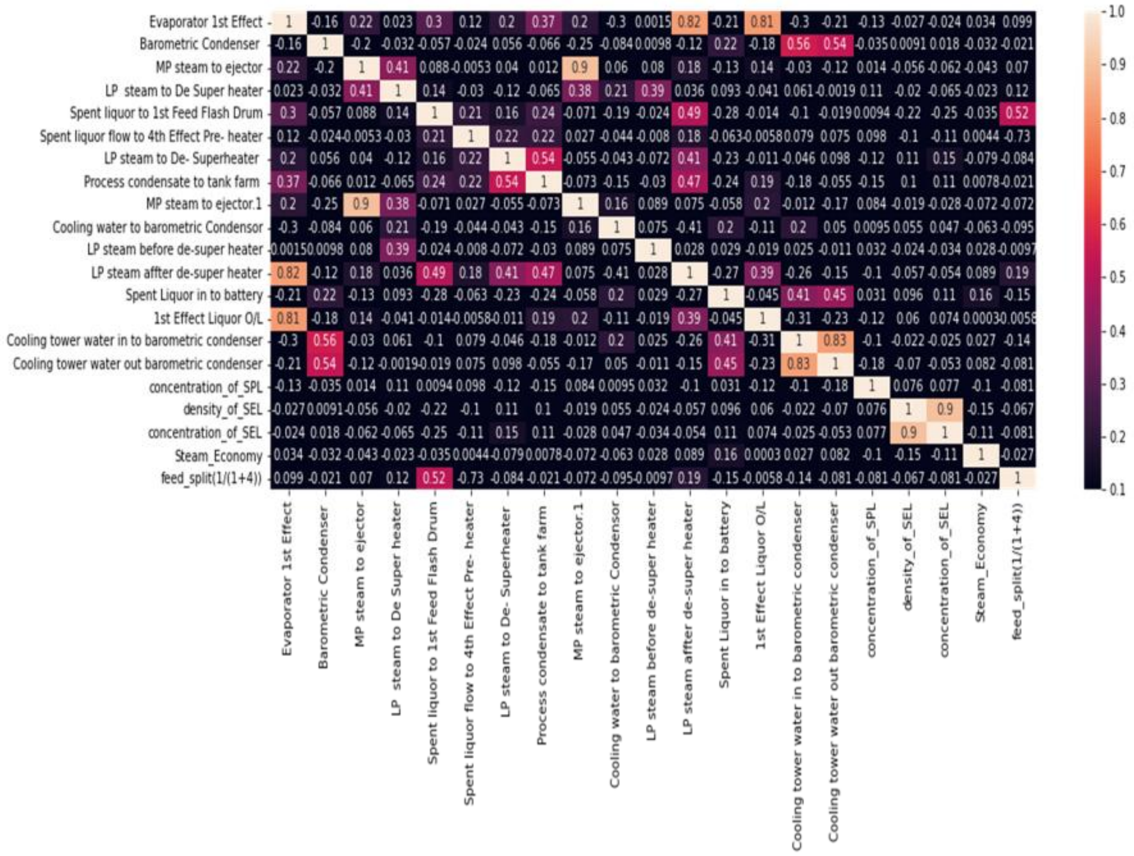
Splitting feed in different ratios could save steam, preferably higher towards 1st FFD by taking advantage of flashing under vacuum. However, a trade-off needs to be explored to maintain the temperature of SEL. The low temperature of process condensate indicates efficient evaporation.

3.2. Predictive Model Development

To develop the machine learning model, python 3.8.1 is used involving various steps as discussed in the technical approach. The plant time-stamped data along with the measurements of the quality variables for a year has been utilized. All the variables have been scaled to a range of '0' to '1' based on their minimum and maximum value. Scaling is performed to bring all the variables into one uniformity to have a mean equal to 0 and standard deviation of 1. This aids in removing unnecessary biasness. In Python, Standard Scaler is used on the data set from the sklearn preprocessing module.

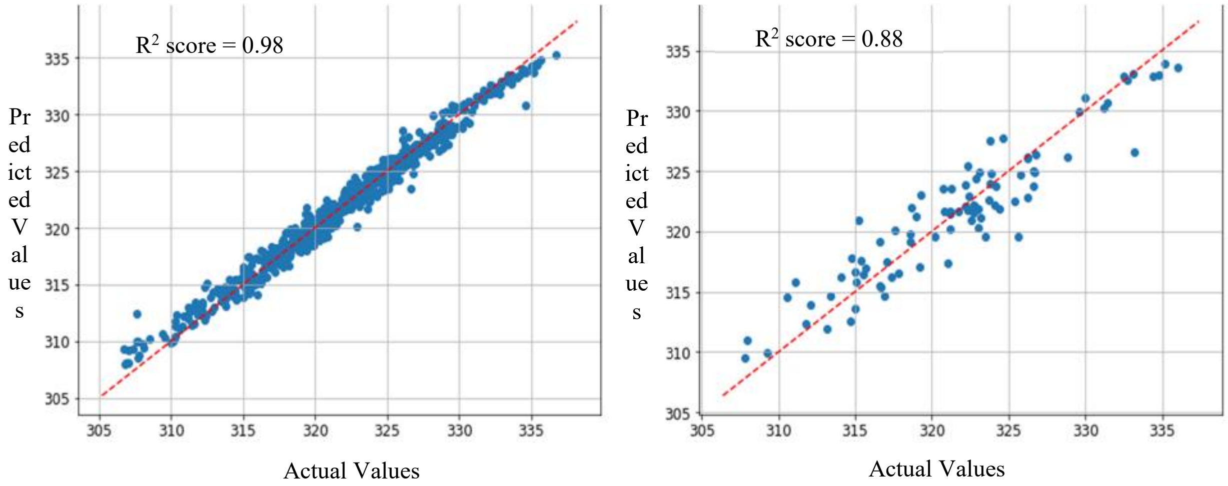
A correlation matrix (refer to Figure 5) using a heat map generated from the Seaborn library of Python for all the variables critical in the quality parameters has been plotted. Correlation matrixes explain the degree to which independent variables influence the dependent variables. Dependent variables are SEL (g/L) and SE whereas all input variables are independent. In addition to the correlation matrix, process knowledge is applied while finalizing the variables for predictive model development for SEL concentration and steam economy. The entire data is processed by eliminating outliers and scaling to the desired range for each variable is performed.

Due to the high non-linearity in the data, a robust machine learning algorithm 'Random Forest' has been considered for model development. After pre-processing the raw data by eliminating outliers and scaling to the desired range, 900 data points have been obtained out of which 720 have been utilized for training and 180 for model validation purposes. The hyperparameters have been tuned using a randomized grid search algorithm with 3-fold cross-validation and the optimum values of the tuned parameters are obtained. The hyperparameter tuning is a crucial step for good accuracy of the model. Obtained hyperparameters are used to develop the model, which achieves 88 % accuracy of SEL (g/L) and 97 % accuracy for the steam economy (refer to Figures 6 - 7). The model is validated for 6 months, it predicts SEL concentration with 86 % of the time with an error less than ± 3 g/L (laboratory error is 2-3 g/L).



	Evaporator 1st Effect	Barometric Condenser	MP steam to ejector	LP steam to De Super heater	Spent liquor to 1st Feed Flash Drum	Spent liquor flow to 4th Effect Pre- heater	LP steam to De- Superheater	Process condensate to tank farm	MP steam to ejector.1	Cooling water to barometric Condensor	LP steam before de-super heater	LP steam affter de-super heater	Spent Liquor in to battery	1st Effect Liquor O/L	Cooling tower water in to barometric condenser	Cooling tower water out barometric condenser	concentration_of_SPL	density_of_SEL	concentration_of_SEL	Steam_Economy	feed_split(1/(1+4))
Evaporator 1st Effect	1	-0.16	0.22	0.023	0.3	0.12	0.2	0.37	0.2	-0.3	0.0015	0.82	-0.21	0.81	-0.3	-0.21	-0.13	-0.027	-0.024	0.034	0.099
Barometric Condenser	-0.16	1	-0.2	-0.032	-0.057	-0.024	0.056	-0.066	-0.25	-0.084	0.0098	-0.12	0.22	-0.18	0.56	0.54	-0.035	0.0091	0.018	-0.032	-0.021
MP steam to ejector	0.22	-0.2	1	0.41	0.088	-0.0053	0.04	0.012	0.9	0.06	0.08	0.18	-0.13	0.14	-0.03	-0.12	0.014	-0.056	-0.062	-0.043	0.07
LP steam to De Super heater	0.023	-0.032	0.41	1	0.14	-0.03	-0.12	-0.065	0.38	0.21	0.39	0.036	0.093	-0.041	0.061	-0.0019	0.11	-0.02	-0.065	-0.023	0.12
Spent liquor to 1st Feed Flash Drum	0.3	-0.057	0.088	0.14	1	0.21	0.16	0.24	-0.071	-0.19	-0.024	0.49	-0.28	-0.014	-0.1	-0.019	0.0094	-0.22	-0.25	-0.035	0.52
Spent liquor flow to 4th Effect Pre- heater	0.12	-0.024	-0.0053	-0.03	0.21	1	0.22	0.22	0.027	-0.044	-0.008	0.18	-0.063	-0.0058	0.079	0.075	0.098	-0.1	-0.11	0.0044	-0.73
LP steam to De- Superheater	0.2	0.056	0.04	-0.12	0.16	0.22	1	0.54	-0.055	-0.043	-0.072	0.41	-0.23	-0.011	-0.046	0.098	-0.12	0.11	0.15	-0.079	-0.084
Process condensate to tank farm	0.37	-0.066	0.012	-0.065	0.24	0.22	0.54	1	-0.073	-0.15	-0.03	0.47	-0.24	0.19	-0.18	-0.055	-0.15	0.1	0.11	0.0078	-0.021
MP steam to ejector.1	0.2	-0.25	0.9	0.38	-0.071	0.027	-0.055	-0.073	1	0.16	0.089	0.075	-0.058	0.2	-0.012	-0.17	0.084	-0.019	-0.028	-0.072	-0.072
Cooling water to barometric Condensor	-0.3	-0.084	0.06	0.21	-0.19	-0.044	-0.043	-0.15	0.16	1	0.075	-0.41	0.2	-0.11	0.2	0.05	0.0095	0.055	0.047	-0.063	-0.095
LP steam before de-super heater	0.0015	0.0098	0.08	0.39	-0.024	-0.008	-0.072	-0.03	0.089	0.075	1	0.028	0.029	-0.019	0.025	-0.011	0.032	-0.024	-0.034	0.028	-0.0097
LP steam affter de-super heater	0.82	-0.12	0.18	0.036	0.49	0.18	0.41	0.47	0.075	-0.41	0.028	1	-0.27	0.39	-0.26	-0.15	-0.1	-0.057	-0.054	0.089	0.19
Spent Liquor in to battery	-0.21	0.22	-0.13	0.093	-0.28	-0.063	-0.23	-0.24	-0.058	0.2	0.029	-0.27	1	-0.045	0.41	0.45	0.031	0.096	0.11	0.16	-0.15
1st Effect Liquor O/L	0.81	-0.18	0.14	-0.041	-0.014	-0.0058	-0.011	0.19	0.2	-0.11	-0.019	0.39	-0.045	1	-0.31	-0.23	-0.12	0.06	0.074	0.0003	-0.0058
Cooling tower water in to barometric condenser	-0.3	0.56	-0.03	0.061	-0.1	0.079	-0.046	-0.18	-0.012	0.2	0.025	-0.26	0.41	-0.31	1	0.83	-0.1	-0.022	-0.025	0.027	-0.14
Cooling tower water out barometric condenser	-0.21	0.54	-0.12	-0.0019	-0.019	0.075	0.098	-0.055	-0.17	0.05	-0.011	-0.15	0.45	-0.23	0.83	1	-0.18	-0.07	-0.053	0.082	-0.081
concentration_of_SPL	-0.13	-0.035	0.014	0.11	0.0094	0.098	-0.12	-0.15	0.084	0.0095	0.032	-0.1	0.031	-0.12	-0.1	-0.18	1	0.076	0.077	-0.1	-0.081
density_of_SEL	-0.027	0.0091	-0.056	-0.02	-0.22	-0.1	0.11	0.1	-0.019	0.055	-0.024	-0.057	0.096	0.06	-0.022	-0.07	0.076	1	0.9	-0.15	-0.067
concentration_of_SEL	-0.024	0.018	-0.062	-0.065	-0.25	-0.11	0.15	0.11	-0.028	0.047	-0.034	-0.054	0.11	0.074	-0.025	-0.053	0.077	0.9	1	-0.11	-0.081
Steam_Economy	0.034	-0.032	-0.043	-0.023	-0.035	0.0044	-0.079	0.0078	-0.072	-0.063	0.028	0.089	0.16	0.0003	0.027	0.082	-0.1	-0.15	-0.11	1	-0.027
feed_split(1/(1+4))	0.099	-0.021	0.07	0.12	0.52	-0.73	-0.084	-0.021	-0.072	-0.095	-0.0097	0.19	-0.15	-0.0058	-0.14	-0.081	-0.081	-0.067	-0.081	-0.027	1

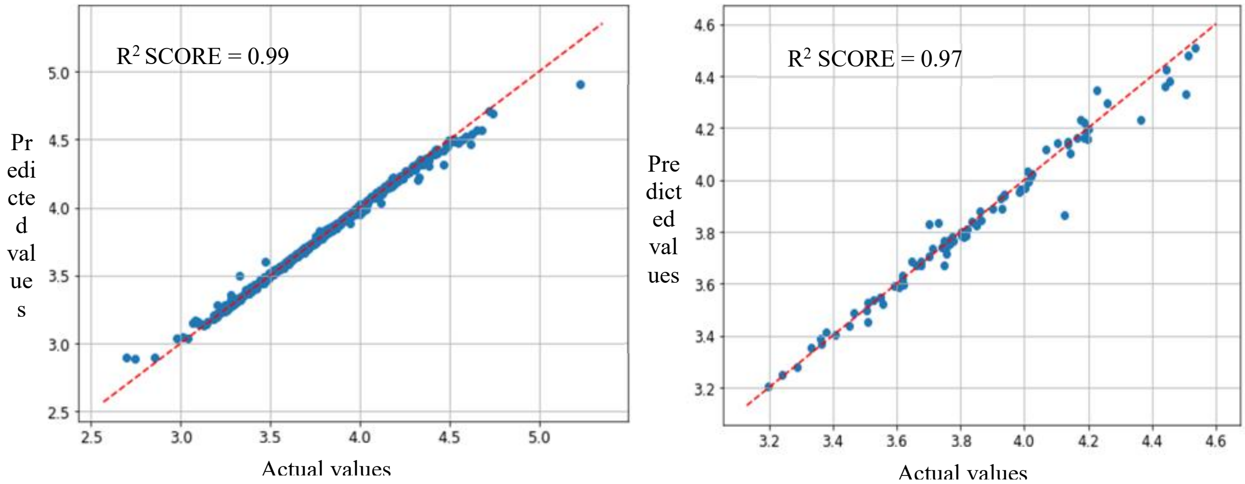
Figure 5. Heat map for the process variables



(a) Prediction on training data of SEL

(b) Prediction on test data of SEL

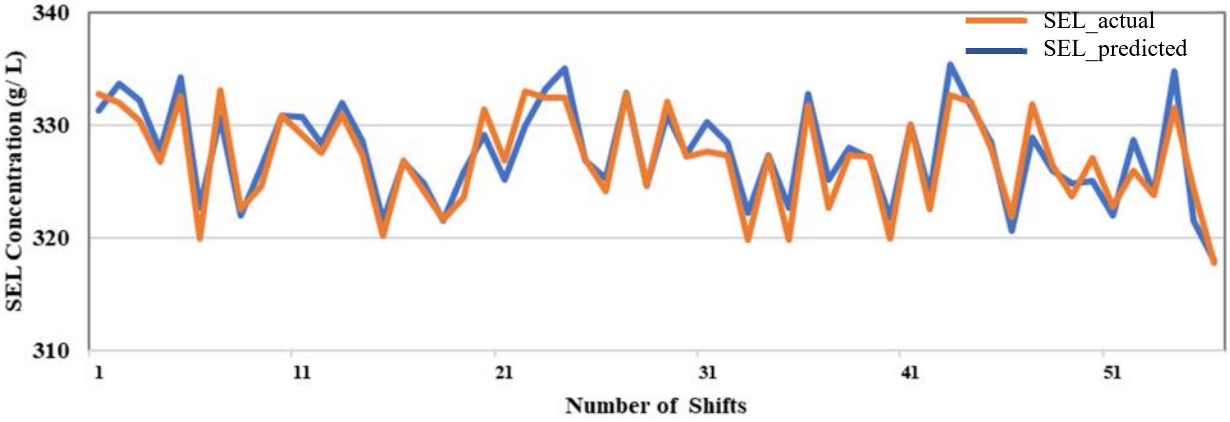
Figure 6. Training and test data for SEL concentration (g/L).



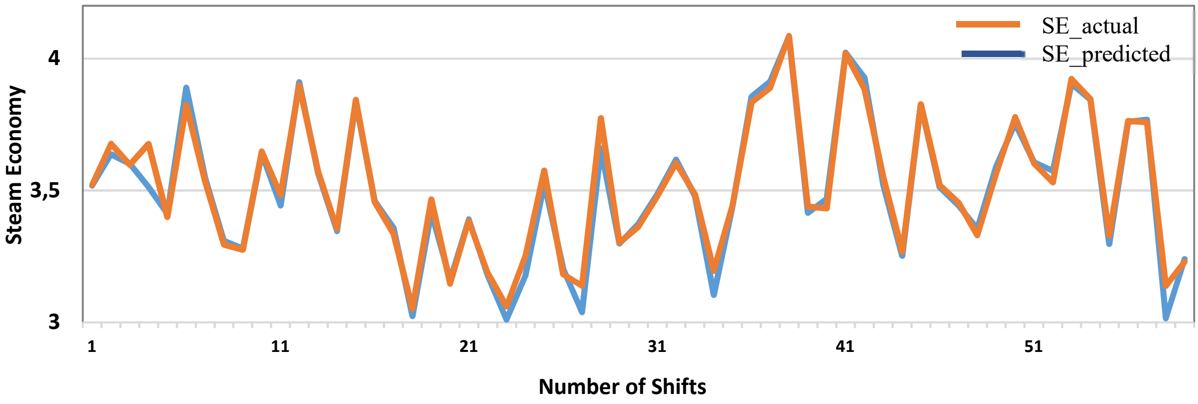
(a) Prediction on training data of SE

(b) Prediction on test data of SE

Figure 7. Training and test data for the steam economy.



(a) SEL Concentration



(b) Steam Economy

Figure 8. Validation for unseen data SEL Concentration & Steam Economy.

Thus, a randomized grid search runs iteratively in such a way that the accuracy is intact. Using the finalized hyperparameters, the model is trained and tested using the data after the train-test split as discussed in the technical approach.

For the rescaling inverse Laplace transform is applied in Python. Further, several metrics i.e., mean absolute error, root mean square error etc., available in Python are used. Further, new data which is not used in training and testing is utilized for validation of the predictive model. On the same regressor, predictions are made for the unseen data i.e., for the approx. 400 shifts, prediction for a month are as follows (refer to Figure 8).

The sample dashboard designed for GUI development (refer to Figure 9) graphic user interface/dashboard is developed to display the results of the predicted SEL concentration and steam economy as needed (minute-wise/hourly). In addition, it also involves the display of important variables /features which play a key role in the model prediction in the order of their respective ranking. On top of that, past trends of each variable along with predictions could be visualized. Reports could also be generated for the benefit of the operating personnel in making proactive decisions to tune operational parameters.

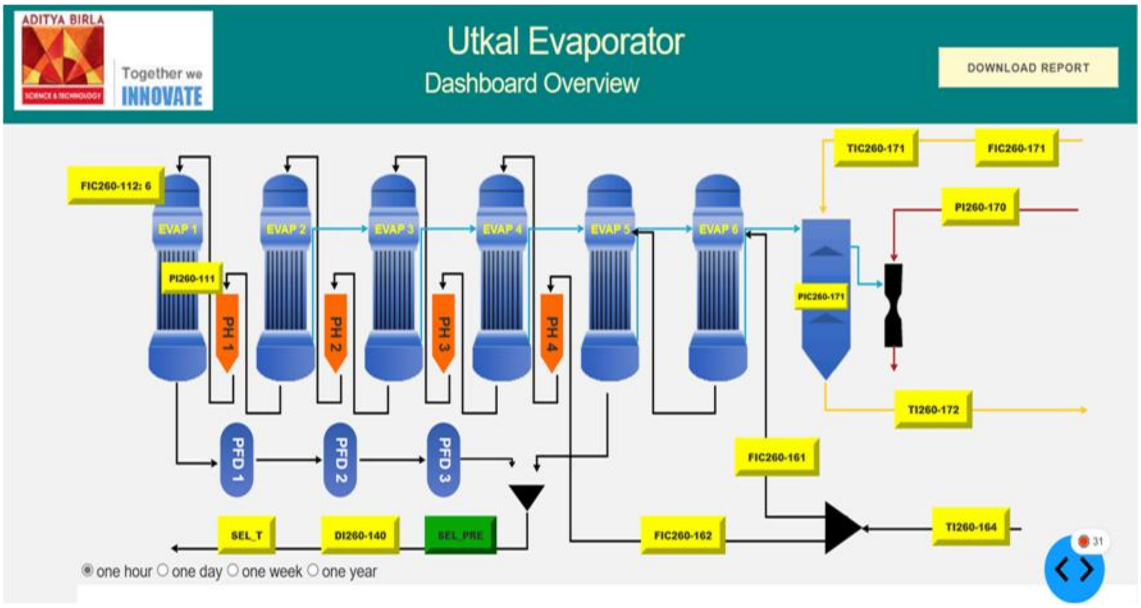


Figure 9. GUI for evaporator circuit integrated with predictive model

4. Conclusions

The current work focused on carrying the following technical methodologies viz., A) Data Analysis: Identification of critical variables affecting the process, B) Predictive model development based on data analytics techniques for product quality. After testing various models, the Random Forest Machine Learning algorithm is finalized due to its robustness, giving satisfactory prediction results. The developed Machine learning model is capable to predict the SEL concentration with an accuracy of 88 % and steam economy with 97 %. This model is further integrated with the web-based platform where predicted vs actual SEL concentration are displayed along with other critical process parameters. This platform is useful for operators to control the process in a better way by taking prompt corrective actions instead of depending on laboratory analysis of the SEL concentration.

5. Acknowledgment

The authors would like to extend our sincere gratitude to the management of Utkal Alumina Refinery, particularly the technical & operational teams for their invaluable support, cooperation, and constructive discussions throughout this research work. We are truly grateful to the management of ABSTCPL, especially Mr. Dilip Gaur and Dr. Aspi Patel, for their continuous guidance and unwavering support.

6. References

1. A.C.P. Silveira et al., Flow regime assessment in falling film evaporators using residence time distribution functions, J. Food Eng. 2015, 160, 65–76.
2. Deepak Kumar, Vivek Kumar, and V.P. Singh, Modeling, and dynamic simulation of mixed feed multi-effect evaporators in paper industry, Applied Mathematical Modelling, 2013, 37, 384–397.
3. F. M. Bojnourd, M.A. Fanaei, and H. Zohreie, Mathematical modelling, and dynamic simulation of multi-effect falling-film evaporator for milk powder production, Math. Comput. Model. Dyn. Syst. 2015, 21, 336–358.
4. Wang Xiaoli et al., Modeling and simulation of an industrial falling film evaporator for alumina production, Chemical Engineering Research and Design, 2020, 154, 303–315.
5. Yechun Zhang et al., Modelling of a milk powder falling film evaporator for predicting process trends and comparison of energy consumption, Journal of Food Engineering, 2018, 225, 26-33.
6. V. Cregan, W. T. Lee, and L. Clune, A soft sensor for the Bayer process, J. Math. Ind., 2017, Vol. 7, No. 1, 1-6.
7. Mostafa Mahmoudian et al., Comparing the Capability of Various Models for Predicting the Bayer Process Parameters, Journal of Advanced Materials and Processing, 2018, 6, 71-86.
8. Yongfang Xie et al., A new prediction model based on the leaching rate kinetics in the alumina digestion process, Hydrometallurgy, 2016, 164, 7-14.
9. Yong Gang Li et al., Soft sensor and expert control for blending and digestion process in alumina metallurgical industry, Journal of Process Control, 2013, 23, 1012-1021.
10. J. Zhu et al., Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data, Annu. Rev. Control, 2018, vol. 46, pp. 107-133.
11. P. Kadlec, B. Gabrys, and S. Strandt, Data-driven soft sensors in the process industry, Comput. Chem. Eng., 2009, Vol. 33, 759-814.