Grenze International Journal of Engineering and Technology, Special Issue

Optimization of a Process System in Nuclear Power Plant- A Data Mining Approach

S.Narasimhan[1] and Dr. Rajendran[2]
[1]Bharatiya nabhikiya Vidyut Nigam Limited(BHAVINI), Kalpakkam,India
Email: snsimhan@igcar.gov.in
[2] VELS Institute of Science,Technology and Advanced Studies(VISTAS),Chennai,India
Email: director.ece@velsuniv.ac.in

***Abstract*—Nuclear Power generation is a capital intensive process requiring continual efforts to reduce cost burdens to stay competitive in the market with many stipulations from statuary, legal and regulatory agencies. This necessitates more emphasis towards safety, efficiency, availability and reliability during operation. Operating the power plant processes inefficiently can result in a low capacity factor of the plant with high significant costs. It will also accelerate the ageing process of equipment and systems. Hence, monitoring the condition of the process, estimating critical parameters based on the variations seen in other parameters and their optimal control is a vital part of today's plant operation and control schemes.**

Data science has emerged as a major technology player in implementing concepts like Industry 4.0 and Internet of things. Basic principles of mechanics and process systems remained the same but the efficiency, productivity and reliability were improved immensely with this. Data mining concepts were widely accepted in all the areas since addition of an intelligence factor can optimize any system independent of industry, Data Mining has found promising scope in process optimization and performance improvements in nuclear power plants.

This paper deals with application of data mining techniques for a representative process in a nuclear power plant to identify various steady state operating regimes, detect anomalous operating conditions, identify key performance indicators (KPIs) so that steady state estimation of KPIs can be carried out subsequently. This will prompt the operation personnel to adhere to the optimum region of operation and estimate the process conditions for enhanced performance. In addition, this study will also enhance the scope for implementation of predictive maintenance concepts and hence an improved reliability.

***Index Terms*— Association Rule Mining, Data Mining, Data modelling, Data Transformation, Inter Correlation, k-Means Clustering, Principal Component Analysis.**

I. Introduction

Nuclear Power Plants are capital intensive and require continual efforts to reduce cost burdens to stay competitive in the market. They should operate with high availability factors, better fuel efficiency and better thermodynamic efficiency to minimize the cost of power generation. The process systems are designed and engineered to cover all the modes of operation without compromising on the nuclear safety aspects and

Grenze ID: 01.GIJET.6.2.1

ensuring high availability. As all nuclear power plants are operated as a base load generating stations, the optimization of the process is necessary to have consistent high availability factor, efficiency and reduced failures not compromising on quality, technical specifications, safety, environment, legal and regulatory requirements. The nuclear power plant is operated from a central control room where the operators by the virtue of their training, skill and experience, monitors the process conditions, interpret the observations and take corrective actions. Whereas many computerized tools and models are being developed to assist the operator in decision making using the large amount of data being measured and stored in the computers and servers of the plant data acquisition systems. These models will help the operator to understand the process behaviors and dynamics, predict the values and take corrective action to improve the performance. One such data mining tool is demonstrated in this study for a representative nuclear power plant process taking full advantage of readily available plant data to identify key performance indicators (KPI) and their optimum values.

The study involves performance optimization analysis in the steady state operating conditions. It primarily focusses on identification of various operating regimes, identification of associated and correlated variables and their values in each of the operating regimes. The operating regimes are identified using principal component analysis and k-means clustering. The time series data is converted to a transaction data type and the associations are identified and pruned with thresholds of significance to identify the key parameters of the process. This study further widens the scope for estimating the values of these parameters so that meaningful comparisons can be made with respect to process optimization.

II. Literature Review

Many publications are dealing with process industries performance monitoring using sensor data analysis and development of models using various statistical, mathematical and data modelling techniques.

The need for the process companies in the field of Process Supervision, real-time production monitoring, process tuning are well described in ref[1] considering a bio diesel process for the study. Ref [2] uses principal component analysis and partial least square models to demonstrate a viable solution. Ref [1, 5, 6, 7, and 8] deals with various industrial process monitoring methodologies like time series analysis, partial least squares, neural networks and soft sensing. Ref [9] deals with performance monitoring of redundant process sensors for process optimization.

Ref [2] also brings out the use of distributed control systems (DCS) in power plants that has become standard and facilitating effective plant operations. However, during a transient/upset conditions, with large number of alarms flooding the screen of the operator, the task of monitoring the plant operation and performance becomes increasingly difficult. The use of Data mining techniques facilitates to develop models that provide an efficient tool for the operator to assist him in performance monitoring from the data obtained from DCS historian server. These models can form the basis for detecting deviations in the performance from trained behavior.

III. Data Mining Steps Followed In This Study

The broad steps that are followed in this study for optimization of the process using data mining techniques are as follows:

1. Identify a representative dataset— select the period corresponding to steady operations of the plant.
2. Pre-process the dataset by missing value processing, inter correlation analysis and feature reduction.
3. Use a clustering model i.e. K-Means clustering to separate the data samples into clusters.
4. Identify out of bound clusters /Anomalous operations.
5. Identify key performance indicator (KPI) variables using association rule mining techniques and its metrics.
6. Identify the optimum values of KPIs so that process can be optimized to run at the efficient configuration.

A. Data Collection

The process data on the bottom mechanical seal oil cooling circuit of the main coolant pump in a nuclear power plant was taken for this study. This circuit consists of monitoring different parameters using continuous sensors and switching sensors. For the subject study, to understand the performance of the process, only analogue variables are taken for analysis. The process consists of parameters monitored by analogue sensors as listed in Table.1.

The subject Nuclear Power plant has a distributed digital Control System where all the plant parameters are scanned and logged in the historian. The data on the past measurements of the process variables under study for a period of one month was collected from the server for the analysis at a sampling rate of 5 seconds. The total data tuples taken for the study for the above period is 414730.

B. Data Pre-processing

The collected data was inspected for a possible missing values, irrational values etc. Missing values in the context of a nuclear power plant have various causes. The most common causes are the failure of the field sensor or removal of the sensor for maintenance and calibration etc.

TABLE I. PROCESS PARAMETERS AND RANGES OF MEASUREMENTS

Sl.No	Parameter ID	Parameter Description	Range of Measurement
1	TTRol20_801	Oil inlet temperature to bottom mechanical seal	0 to 600^o C
2	TTRol20_804A	Oil outlet temperature from bottom mechanical seal	0 to 600^o C
3	TTRol20_804B	Oil outlet temperature from bottom mechanical seal	0 to 600^o C
4	TTRol20_804C	Oil outlet temperature from bottom mechanical seal	0 to 600^o C
5	TTRol20_808	Return oil temperature after cooling by blowers	0 to 600^o C
6	PTRol20_801	Oil Filter differential pressure	0-7 kg/cm^2
7	PTRol20_803	Oil inlet pressure to bottom mechanical seal	0-7 kg/cm2
8	LTRol20_801	Bottom Circuit Oil tank Level	0-500 mm
9	FTRol20_801	Oil outlet flow from mechanical seal	0-50 m3/hr.

Missing data can also result due to loss of communication between the sensors and the data acquisition system, problems in assessing the server database, issues in historians, mismatch in connectivity protocols etc.

There are different methods to deal with such a situation. A commonly applied practice is to fill the missing values with the mean value of the observations so far. Another method is to skip the data tuple consisting of the missing values, i.e., case deletion. In this study, as steady state analysis was performed, the data tuples that had missing values were deleted from the data set for analysis.

The data now is unlabeled with approximately 4 lakh samples having 18 feature sets which correspond to the analogue measurements of temperatures, pressure, level and flow at various locations of the process system

C. Inter Correlation Analysis

As can be seen from the Table-1 above, redundant sensors (A, B, C) are provided for the measurement of oil outlet temperature from bottom mechanical seal in the lube oil system. Some of the process measurements are interdependent. Further the range of measurements for each type of process measurements are also different. Hence to identify possible reduction in dimensions, the inter correlation analysis was done. The correlation matrix calculated for the variables of bottom oil circuit are as shown in Table II.

As can be seen from the matrix, most of the variables are inter correlated. This calls for reduction of the dimensions using principle component analysis after standardization.

D. Data Transformation

The process instrumentation values as collected by the data acquisition system may have different ranges depending on the parameter measured, the units of measurement and the behavior of the process. This characteristic will make identification of the system behavior difficult in a data analysis process since variables with larger range of values may dominate over variables with smaller ones. Hence, to bring all the variables to a common standard for processing, we may perform data scaling. The most common scaling method is the z-score normalizations as defined by [4]

$$\text{Z-score normalization: } x' = (x_i - \mu)/\delta$$

Where x_i is the unscaled measurement, x' is the scaled version of the variable, μ is the mean value of measurement and δ is the standard deviation of the variable. All variables are normalized with a mean of 0 and standard deviation of 1 with the z-score.

IV. PRINCIPAL COMPONENT ANALYSIS

Subsequent to normalization, Principal Component Analysis (PCA) is applied on the data set. The essence of PCA is that it transforms a set of possibly correlated features into a transformed set of uncorrelated features.

Hence, PCA sets the data up for effective clustering, while also affording the possibility of reducing the number of features. This is because the PCA coordinate axes are aligned with directions of maximum variance in the data. Therefore, a reduced set of PCA coordinates could potentially account for the variance of a much larger number of the original set of features. [10]

While applying PCA analysis for the dataset under study, the Eigen values were obtained for each of the principal components so arrived as in Table III.

TABLE II. INTER-CORRELATION ANALYSIS OF PROCESS PARAMETERS

	TTRol20_801	TTRol20_804A	TTRol20_804B	TTRol20_804C	TTRol20_808	FTRol20_801	LTRol20_801	PTRol20_801	PTRol20_803
TTRol20_801		0.91	0.91	0.92	0.91	0.9	0.02	0.22	0.89
TTRol20_804A	0.91		1	1	0.99	0.84	0.02	0.25	0.82
TTRol20_804B	0.91	1		1	0.98	0.84	0.02	0.24	0.82
TTRol20_804C	0.92	1	1		0.99	0.84	0.03	0.27	0.82
TTRol20_808	0.91	0.99	0.98	0.99		0.84	0.08	0.27	0.82
FTRol20_801	0.9	0.84	0.84	0.84	0.84		0.01	0.24	0.99
LTRol20_801	0.02	0.02	0.02	0.03	0.08	0.01		0.2	0.01
PTRol20_801	0.22	0.25	0.24	0.27	0.27	0.24	0.2		0.24
PTRol20_803	0.89	0.82	0.82	0.82	0.82	0.99	0.01	0.24	

TABLE III. EIGEN VALUES FOR PCA COMPONENTS

Principal Components	1	2	3	4	5	6	7	8	9
Eigen Values	6.51719	1.15623	0.77383	0.43928	0.08706	0.01720	0.00861	0.00031	0.00025
%values	72.4132	12.8470	8.59815	4.88097	0.96735	0.19112	0.09576	0.00348	0.00283

TABLE IV. PATTERN MATRIX FOR PCA COMPONENTS

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
TTRol20-801	0.96	-0.05	0.04	0.07	0.27	0.01	0	0	0
TTRol20-804A	0.98	-0.04	0.03	-0.02	-0.04	0.03	0	0.01	0.01
TTRol20-804B	0.98	-0.04	0.04	-0.2	-0.05	0.05	0	0	-0.01
TTRol20-804C	0.98	-0.04	0.05	-0.19	-0.04	0.03	0	-0.01	0.01
TTRol20-808	0.97	0.02	0.05	-0.19	-0.02	-0.11	0	0	0
FTRol20-801	0.93	-0.05	-0.02	0.36	-0.06	0	-0.07	0	0
LTRol20-801	0.04	0.83	0.56	0.04	0	0.01	0	0	0
PTRol20-801	0.3	0.68	-0.68	-0.02	0.01	0	0	0	0
PTRol20-803	0.91	-0.04	-0.01	0.39	-0.06	0	0.06	0	0

The standardized loadings (pattern matrix) for each of the PCA component based on the correlation with respect to the dimensions considered are listed in Table IV. With the above data following inferences can be made from PCA analysis:

1. 95% of the variances in the data are explained with the first three PCA components viz. PC1, PC2 and PC3.
2. PC1 explains the variances of all temperature data, inlet pressure and the outlet flow data. This can also be seen by high correlation as observed in the table for the corresponding signals.
3. PC2 and PC3 explains the variances of oil tank level and the oil strainer differential pressure.
4. PC4 explains the variations in inlet pressure and the outlet flow data.

As 95% of the total variances of the data are explained within PC1, 2 and 3, these dimensions are taken into consideration for cluster analysis.

V. CLUSTER ANALYSIS OF THE SAMPLES

A. K-means Clustering

The data was split in to 3 clusters by using K-means clustering technique. Random centroids were generated and the distance are calculated. Centroids were moved by taking the mean within the cluster. After many iterations, centroids are fixed for maximum inter cluster distance and minimum intra cluster distance. With this 2 major clusters and 1 minor cluster were obtained. The outcome of this clustering model applied to the data set of our study is depicted in Table V.

TABLE V. CLUSTER VALUES FOR PCA COMPONENTS

Parameter		Cluster-1	Cluster-2	Cluster-3
Cluster Centres	PC1	0.08529056	-3.94198043	0.12816901
	PC2	-0.03678369	-4.48500434	0.27447676
	PC3	1.0315072	-0.8070725	-0.9486174
Cluster Size in terms of number of data sets		197900	10975	205855
Custer size in %		47.71779	2.64630	49.63591

B. Data Samples in the Non-Majority (Anomalous) Cluster

The density plot of variables in each of the clusters are plotted in the Figure.1 below:

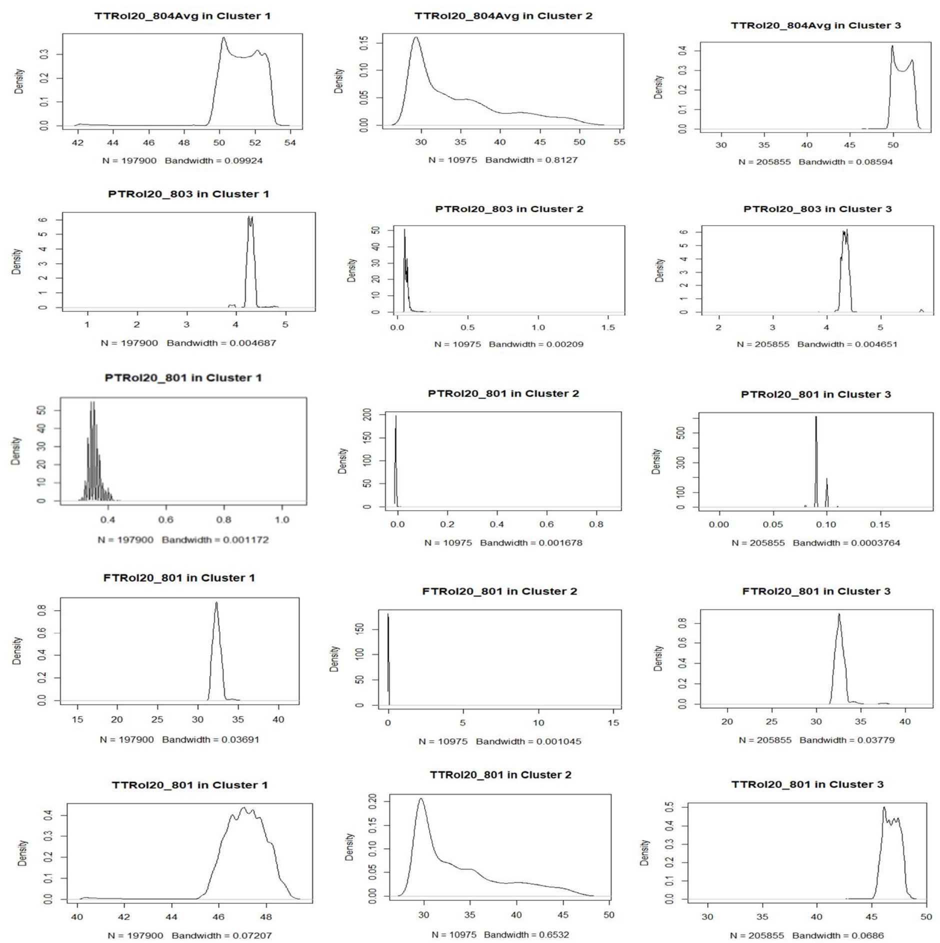

Fig. 1. Density plots of variables in each clusters

The density plot of the variables in each of the clusters indicate the following:

1. The process exhibits two stable operation regimes (cluster-1 and 3) with the bandwidth spread of oil outlet temperature, oil inlet pressure and oil outlet flow exhibiting similar values. However the spread of values of oil filter differential pressure value indicates a unimodal spread in cluster-1 and a bimodal spread in cluster-3. On analyzing the process operation, it was found that the filter bed was changed in between during the operation.

2. Since the spread in the density plots of the variables of cluster-2 exhibit a large variation compared to cluster-1 and 3, it can be inferred that, unless some anomalous process conditions were present, these values are unlikely to occur entirely by chance or due to erroneous measurements. Also, remarkably, the sample measurements are made consecutively in time. This provides additional substantiation for our inference that these are anomalous measurements and not simply outliers. It would be extremely unlikely for outliers of such significance to occur in large consecutive measurements. It is statistically sound.
3. On reviewing the operation log, it is concluded that Cluster-2 is an anomalous operation where the main coolant pump tripped and the lube oil system was started afterwards to normalize the cooling pump. Hence cluster-2 is not considered for further study.

VI. ASSOCIATION RULE MINING

Association rule mining is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness. [4]

Association rule mining is a procedure which aims to observe frequently occurring patterns, correlations, or associations from datasets found in various kinds of databases such as relational databases, transactional databases, and other forms of repositories.

Let $I = \{i_1, i_2, i_3, ... i_n\}$ be a set of n binary attributes called *items*.

Let $D = \{t_1, t_2, t_3, ... t_m\}$ be a set of m transactions called *database*

Each transaction in D has a unique transaction ID and contains a subset of the items in I .

A rule is defined as an implication of the form:

$$X \Rightarrow Y \text{ Where } X, Y \subseteq I .$$

Every rule is composed by two different sets of items, also known as *item sets* X and Y , where X is called *antecedent* or left-hand-side (LHS) and Y *consequent* or right-hand-side (RHS). An antecedent is something that is found in data, and a consequent is an item that is found in combination with the antecedent.

In order to select interesting rules from the set of all possible rules, constraints on various measures of significance and interest are used. The best-known constraints are minimum thresholds on support, confidence and Lift.

Support: This says how frequent an item set is, as measured by the proportion of transactions in which an item set appears.

Let X, Y be item sets, $X \Rightarrow Y$ an association rule and T a set of transactions of a given database. The support of X with respect to T is defined as the proportion of transactions t in the dataset which contains the item set X .

Confidence: This says how likely item Y is present when item X is present. This is measured by the proportion of transactions with item X, in which item Y also appears. Confidence is an indication of how often the rule is found to be true.

The confidence value of a rule $X \Rightarrow Y$, with respect to a set of transactions T , is the proportion of the transactions that contains X which also contains Y.

Lift. This says how likely item Y is present when item X is present, while controlling for how frequent the item Y is. The Lift can also be defined as the ratio of observed support value for the rule to the expected value if X and Y were statistically independent.

If the rule had a lift of 1, it would imply that the probability of occurrence of the antecedent and that of the consequent are independent of each other. When two events are independent of each other, no rule can be drawn involving those two events.

If the lift is > 1, the lift value indicates the degree to which those two occurrences are dependent on one another, and makes those rules potentially useful for predicting the consequent in future data sets.

If the lift is < 1 implies that the items are substitute to each other. This means that presence of one item has negative effect on presence of other item and vice versa.

So, in a given transactional data with multiple items, Association Rule Mining primarily tries to find the rules that govern how or why such items are often found together.

Association Rule Mining is sometimes referred to as “Market Basket Analysis”, as it was the first application area of association mining. The aim is to discover associations of items occurring together more often than being expected from a randomly sampled data tuples.

Most machine learning algorithms work with numeric datasets and hence tend to be mathematical. However, association rule mining is suitable for non-numeric, categorical data.
Thus by converting the data sets to transactional data type by transforming the numerical variables to categorical values with identified classes, it is possible to identify the key performance indicators (KPI) of the process under study by identifying frequent occurrence of the patterns in the values and associations between the variables by applying association rule mining techniques.

A. Data Transformation for Association Rule Mining

As can be seen from the descriptive study of the data, the various ranges of values observed in general are
Temperature Sensors 20^0 C to 60^0 C
Pressure Sensors -0.5 kg/cm^2 to 7.0 kg/cm^2
Level Sensor 420mm to 450 mm
Flow Sensors 0.5 m^3/hr. to 45 m^3/hr.
Based on the above, the variable ranges has been split into 25 bins of equal width for the purpose of defining classes for each variable. Each of the bin corresponds to a class for that variable. The various bin ranges and classes are tabulated in Table VI.
With this classification a new data frame is created with the class values for each of the variables. Thus the time series data recorded for the process has been converted to a transactional data base with multiple classes for each of the variable for each of the clusters to implement association rules.

B. Association Rule Mining for cluster-1

The data frame of cluster-1 now consists of transactions as item matrix in sparse format with 197900 rows (i.e., elements/item sets/transactions) and 116 columns (items). The frequency of occurrence of items in the data frame of cluster-1 are shown in Figure.2. Considering a minimum frequency of occurrence of 50%, the most frequent patterns are listed in Table VII.

TABLE VI. CLASS LABEL VALUES FOR PARAMETERS

variable	Class Label Values												
	Card1	Card2	Card3	Card4	Card5	Card6	Card7	Card8	Card9	Card10	Card11	Card12	Card13
Flow	-0.5 – 1.32	1.32 - 3.13	3.13 - 4.96	4.96 - 6.78	6.78 - 8.60	8.60 - 10.42	10.42- 12.24	12.24- 14.06	14.06- 15.88	15.88- 17.70	17.70 - 19.52	19.52 - 21.34	21.34 - 23.15
Pressure	-0.5 - -0.2	-0.2- 0.1	0.1- 0.4	0.4- 0.7	0.7- 1.0	1.0- 1.3	1.3- 1.6	1.6- 1.9	1.9- 2.2	2.2- 2.5	2.5- 2.8	2.8- 3.1	3.1- 3.4
Temperature	20- 21.6	21.6- 23.2	23.2- 24.8	24.8- 26.4	26.4- 28.0	28.0- 29.6	29.6- 31.2	31.2- 32.8	32.8- 34.4	34.4- 36.0	36.0- 37.6	37.6- 39.2	39.2- 40.8
Level	420.0- 421.2	421.2- 422.4	422.4- 423.6	423.6- 424.8	424.8- 426.0	426.0- 427.2	427.2- 428.4	428.4- 429.6	429.6- 430.8	430.8- 432.0	432.0 - 433.2	433.2 - 434.4	434.4 - 435.6

variable	Class Label Values											
	Card14	Card15	Card16	Card17	Card18	Card19	Card20	Card21	Card22	Card23	Card24	Card25
Flow	23.16- 24.98	24.98- 26.80	26.80- 28.62	28.62- 30.44	30.44- 32.26	32.26- 34.08	34.08- 35.90	35.90- 37.72	37.72- 39.54	39.54- 41.36	41.36 - 43.18	43.18 - 45.00
Pressure	3.4- 3.7	3.7- 4.0	4.0- 4.3	4.3- 4.6	4.6- 4.9	4.9- 5.2	5.2- 5.5	5.5- 5.8	5.8- 6.1	6.1- 6.4	6.4- 6.7	6.7- 7.0
Temperature	40.8- 42.4	42.4- 44.0	44.0- 45.6	45.6- 47.2	47.2- 48.8	48.8- 50.4	50.4- 52.0	52.0- 53.6	53.6- 55.2	55.2- 56.8	56.8- 58.4	58.4- 60.0
Level	435.6- 436.8	436.8- 438.0	438.0- 439.2	439.2- 440.4	440.4- 441.6	441.6- 442.8	442.8- 444.0	444.0- 445.2	445.2- 446.4	446.4- 447.6	447.6 - 448.8	448.8 - 450

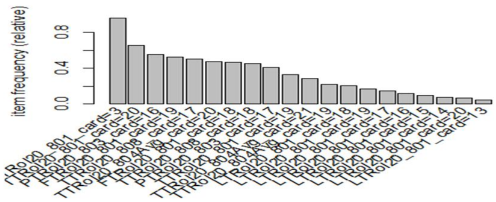

Fig. 2. Frequency plots of variables in Cluster-1

TABLE VII. FREQUENT ITEMS IN CLUSTER-1

Sl.No	Card	Count
1	PTRol20_801 Card-3	190048
2	TTRol20_801 Card-20	130048
3	PTRol20_803 card-16	110214
4	FTRol20_801 Card-19	103228
5	TTRol20_808 Card-17	99390

Further to identify the key performance indicators, association rule mining is applied to the data with the support and confidence threshold at 0.4. i.e. the absolute minimum support count to be 79160. Applying apriori algorithm to identify the frequent item sets, with the set threshold value, 67 items are identified out of 197900 transactions. 23 association rules could be generated. For the lift values being greater than 1, the valid rules that can be taken for assessment are tabulated in Table.8. However the results indicate a weak association as the lift values are only marginally higher than 1.To ascertain the admissibility of association, the correlation values for each of the item sets in a rule are verified as per Table IX. The table indicates very weak correlation and hence these rules cannot be admitted as valid.

C. Association Rule Mining for cluster-3:

The data frame of cluster-3 now consists of transactions as item matrix in sparse format with 205855 rows (i.e., elements/item sets/transactions) and 116 columns (items). The frequency of occurrence of items in the data frame of cluster-3 are shown in the figure.3 below. Considering a minimum frequency of occurrence of 50%, the most frequent patterns are tabulated in Table-10. Further to identify the key performance indicators, association rule mining is applied to the data with the support and confidence threshold at 0.4. I.e. the absolute minimum support count to be 82342.Applying apriori algorithm to identify the frequent item sets, with the set threshold value, 93 items are identified out of 20585 transactions. 47 association rules could be generated. For the lift values greater than 1, the valid rules that can be taken for assessment are as in Table XI.

TABLE VIII. ASSOCIATION RULES FOR CLUSTER-1 WHERE LIFT IS >1

Sl.No	Rule	Support	Confidence	Lift	Count
1	{TTRol20_808_card=18}=> {PTRol20_801_card=3}	0.4524861	0.9987842	1.0400	89547
2	{PTRol20_801_card=3} => {TTRol20_808_card=18}	0.4524861	0.4711810	1.0400	89547
3	{TTRol20_804Avg_card=20} => {PTRol20_801_card=3}	0.4670541	0.9915254	1.0325	92430
4	{PTRol20_801_card=3} => {TTRol20_804Avg_card=20}	0.4670541	0.4863508	1.0325	92430
5	{PTRol20_803_card=16} => {PTRol20_801_card=3}	0.5520313	0.9912262	1.0321	109247
6	{PTRol20_801_card=3} => {PTRol20_803_card=16}	0.5520313	0.5748390	1.0321	109247
7	{FTRol20_801_card=18} => {PTRol20_801_card=3}	0.4621425	0.9839484	1.0246	91458
8	{PTRol20_801_card=3} => {FTRol20_801_card=18}	0.4621425	0.4812363	1.0246	91458
9	{TTRol20_801_card=20}=> {PTRol20_801_card=3}	0.6401617	0.9741634	1.0144	126688
10	{PTRol20_801_card=3}=> {TTRol20_801_card=20}	0.6401617	0.6666105	1.0144	126688

TABLE IX. INTER CORRELATION VALUES OF ITEM SETS IN ASSOCIATION RULES OF CLUSTER-1

Sl.No	Rule	Item sets	Correlation value
1	{TTRol20_808_card=18} => {PTRol20_801_card=3}	TTRol20_808, PTRol20_801	0.27
2	{PTRol20_801_card=3} => {TTRol20_808_card=18}		
3	{TTRol20_804Avg_card=20} =>{PTRol20_801_card=3}	TTRol20_804Avg, PTRol20_801	0.25
4	{PTRol20_801_card=3} =>{TTRol20_804Avg_card=20}		
5	{PTRol20_803_card=16} => {PTRol20_801_card=3}	PTRol20_803, PTRol20_801	0.24
6	{PTRol20_801_card=3} => {PTRol20_803_card=16}		
7	{FTRol20_801_card=18} => {PTRol20_801_card=3}	FTRol20_801, PTRol20_801	0.24
8	{PTRol20_801_card=3} => {FTRol20_801_card=18}		
9	{TTRol20_801_card=20} => {PTRol20_801_card=3}	TTRol20_801, PTRol20_801	0.22
10	{PTRol20_801_card=3} => {TTRol20_801_card=20}		

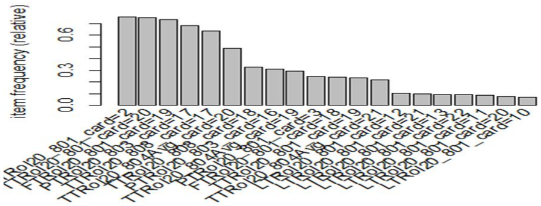

Fig. 3. Frequency plots of variables in Cluster-3

TABLE X. FREQUENT ITEMS IN CLUSTER-3

Sl.No	Card	Count
1	PTRol20_801 Card-2	155217
2	TTRol20_801 Card-20	153715
3	FTRol20_801 card-19	151142
4	PTRol20_803 Card-17	139981
5	TTRol20_808 Card-17	131008

TABLE XI. ASSOCIATION RULES FOR CLUSTER-3 WHERE LIFT IS >1

Sl.No	Rule	Support	Confdence	Lift	Count
1	{TTRol20_808_card=17, FTRol20_801_card=19} => {PTRol20_803_card=17	0.4731923	0.9057511	1.331991	97409
2	{PTRol20_803_card=17, FTRol20_801_card=19} => {TTRol20_808_card=17}	0.4731923	0.8223014	1.292096	97409
3	{TTRol20_801_card=20, PTRol20_803_card=17} => {TTRol20_808_card=17}	0.4031187	0.8034468	1.262469	82984
4	{TTRol20_808_card=17, PTRol20_801_card=2} => {PTRol20_803_card=17}	0.4174978	0.8575363	1.261086	85944
5	{TTRol20_808_card=17} => {PTRol20_803_card=17}	0.5412548	0.8504824	1.250713	111420
6	{PTRol20_803_card=17} => {TTRol20_808_card=17}	0.5412548	0.7959652	1.250713	111420
7	{TTRol20_801_card=20, TTRol20_808_card=17} => {PTRol20_803_card=17}	0.4031187	0.8488457	1.248306	82984
8	{PTRol20_801_card=2, PTRol20_803_card=17} => {TTRol20_808_card=17}	0.4174978	0.7936760	1.247116	85944
9	{TTRol20_808_card=17, PTRol20_803_card=17} => {FTRol20_801_card=19}	0.4731923	0.8742506	1.190727	97409
10	{PTRol20_801_card=2, FTRol20_801_card=19} => {PTRol20_803_card=17}	0.4426465	0.7901715	1.162020	91121
11	{TTRol20_801_card=20, PTRol20_803_card=17} => {FTRol20_801_card=19}	0.4264895	0.8500266	1.157734	87795
12	{PTRol20_803_card=17} => {FTRol20_801_card=19	0.5754487	0.8462506	1.152591	118459
13	{FTRol20_801_card=19} => {PTRol20_803_card=17}	0.5754487	0.7837596	1.152591	118459
14	{PTRol20_801_card=2, PTRol20_803_card=17} => {FTRol20_801_card=19}	0.4426465	0.8414846	1.146100	91121
15	{TTRol20_801_card=20, FTRol20_801_card=19} => {PTRol20_803_card=17}	0.4264895	0.7769538	1.142582	87795
16	{TTRol20_808_card=17} => {FTRol20_801_card=19}	0.5224308	0.8209041	1.118069	107545
17	{FTRol20_801_card=19} => {TTRol20_808_card=17}	0.5224308	0.7115494	1.118069	107545
18	{TTRol20_801_card=20, PTRol20_803_card=17} => {PTRol20_801_card=2	0.4035365	0.8042794	1.066668	83070
19	{TTRol20_801_card=20, FTRol20_801_card=19} => {PTRol20_801_card=2}	0.4386826	0.7991664	1.059886	90305
20	{PTRol20_801_card=2, FTRol20_801_card=19} => {TTRol20_801_card=20	0.4386826	0.7830954	1.048721	90305

TABLE XII. INTER CORRELATION VALUES OF ITEM SETS IN ASSOCIATION RULES FOR CLUSTER-3

Sl.No	Rule	Item sets	Correlation value
1	{TTRol20_808_card=17, FTRol20_801_card=19} => {PTRol20_803_card=17}	TTRol20_808, FTRol20_801	0.84
2	{PTRol20_803_card=17, FTRol20_801_card=19} => {TTRol20_808_card=17}	FTRol20_801, PTRol20_803	0.99
3	{TTRol20_808_card=17, PTRol20_803_card=17} => {FTRol20_801_card=19}	PTRol20_803, TTRol20_808	0.82
4	{TTRol20_801_card=20, PTRol20_803_card=17} => {TTRol20_808_card=17}	TTRol20_801, PTRol20_803	0.89
5	{TTRol20_801_card=20, TTRol20_808_card=17} => {PTRol20_803_card=17}	PTRol20_803, TTRol20_808	0.82
6	{TTRol20_808_card=17, PTRol20_801_card=2} => {PTRol20_803_card=17}	TTRol20_808, TTRol20_801	0.91
7	{PTRol20_801_card=2, PTRol20_803_card=17} => {TTRol20_808_card=17}	PTRol20_801, PTRol20_803 PTRol20_803, TTRol20_808 TTRol20_808, PTRol20_801	0.24 0.82 0.27
8	{TTRol20_808_card=17} => {PTRol20_803_card=17}	TTRol20_808, PTRol20_803	0.24
9	{PTRol20_803_card=17} => {TTRol20_808_card=17}		
10	{PTRol20_801_card=2, PTRol20_803_card=17} => {FTRol20_801_card=19}	PTRol20_801, PTRol20_803	0.24
11	{PTRol20_801_card=2, FTRol20_801_card=19} => {PTRol20_803_card=17}	PTRol20_803, FTRol20_801 FTRol20_801, PTRol20_801	0.99 0.24
12	{TTRol20_801_card=20,PTRol20_803_card=17} => {FTRol20_801_card=19}	TTRol20_801, PTRol20_803	0.89
13	{TTRol20_801_card=20,FTRol20_801_card=19} => {PTRol20_803_card=17}	PTRol20_803, FTRol20_801 FTRol20_801, TTRol20_801	0.99 0.9
14	{PTRol20_803_card=17} => {FTRol20_801_card=19}	PTRol20_803, FTRol20_801	0.99
15	{FTRol20_801_card=19} => {PTRol20_803_card=17}		
16	{TTRol20_808_card=17} => {FTRol20_801_card=19}	TTRol20_808, FTRol20_801	0.84
17	{FTRol20_801_card=19} => {TTRol20_808_card=17}		
18	{TTRol20_801_card=20, PTRol20_803_card=17} => {PTRol20_801_card=2}	TTRol20_801, PTRol20_803 PTRol20_803, PTRol20_801 PTRol20_801, TTRol20_801	0.89 0.24 0.22
19	{TTRol20_801_card=20,FTRol20_801_card=19} => {PTRol20_801_card=2}	TTRol20_801, FTRol20_801	0.9
20	{PTRol20_801_card=2, FTRol20_801_card=19} => {TTRol20_801_card=20}	FTRol20_801, PTRol20_801 PTRol20_801, TTRol20_801	0.24 0.22

TABLE XIII. PARAMETERS AND THEIR VALUES IN ASSOCIATION RULES FOR CLUSTER-3

Sl.No	Parameter Tag ID & Card	Parameter Description	Associated value range
1	PTRol20_803 Card-17	Oil inlet pressure to bottom mechanical seal	4.3-4.6 kg/cm^2
2	TTRol20_801 Card-20	Oil inlet temperature to bottom mechanical seal	50.4-52.0 ^{0} C
3	FTRol20_801 Card-19	Oil outlet flow from mechanical seal	32.26-34.08 m^3/hr
4	TTRol20_808 card-17	Return oil temperature after cooling by blowers	45.6-47.72 ^{0} C

To ascertain the admissibility of association, the correlation values for each of the item sets in a rule as identified during intercorrelation analysis are assessed in Table.12.

The table indicates very weak correlation for the rules 7, 8,9,10,11,18,19 and 20. Hence these association rules cannot be admitted as valid. For the remaining rules the variables and their values that are involved are indicated in Table.13.

These variables shows high association among themselves for a particular range of values. Any change in the value of one of the variable affects the others and hence can be considered as important parameters for the process that can be controlled for optimised performance.

It is also to be noted that the values of these parameters are same as that of cluster-1 except for Oil inlet pressure to bottom mechanical seal. In cluster-1 the inlet pressure value was in the class 4.0-4.3 kg/cm^2 as against 4.3-4.6kg/cm^2 in Cluster-3. However no valid association rule was admissible in cluster-1.

This may indicate that the process system is not tuned to deliver a steady and a stable performance unless the oil inlet pressure is higher than 4.3kg/cm^2.

VII. CONCLUSION

The results of the study indicated that there was a stable operating regime of the process with the combinations of the redundant supply pumps and oil coolers as indicated by cluster-3. Further there was a distinct anomalous condition indicated by cluster-2 where the process was disturbed due to tripping of the centrifugal pump and subsequent settling of the process through the control actions initiated by the operators.

In the identified stable regime, applying association rule mining techniques, the critical parameters that shows high association among themselves for a particular range of values could be identified. Any change in the value of one of the parameter affects the other parameters and hence can be considered as important parameters for the process that can be controlled for optimised performance. **Thus these variables can be considered as Key performance Indicators (KPI)** and their associated values as the parametric values of the optimised process under study.

This study has brought out the use of association rule mining technique which is predominantly used in management applications for the analysis of power plant process optimisation under steady state operating regime. By identifying the KPIs the model will provide the scope for the plant owners to compare the performances and develop strategies to improve efficiency.

FURTHER SCOPE OF WORK

The association rule mining model deployed in this work had categorised the continuous data in the full range of measurement to 25 class variables. The accuracy of the work can be further examined by increasing the class labels and categorising the data based on the observed range of measurement. Further, it is also possible to extend the study to estimate the values of key performance indicators by developing a model based on artificial neural networks, Bayesian estimation etc.

By deploying the model, the values of one KPI can be estimated when other feature values are tweaked. The process can be operated based on the estimated value so that the process runs in optimum mode with sufficient control and safety margins.

ACKNOWLEDGMENT

I am grateful to all who have supported me in every manner for this study and would especially like to thank Dr. Kallol Roy, CMD, BHAVINI, Shri V. Rajan Babu, Director (Technical) and Shri.Vijith Venugopal. I am greatly indebted for the permission given to me to take up the study in the process system and to have an access to the data for the purpose of study.

REFERENCES

[1] Ana Sofia Ramos Brásio, "Industrial process monitoring methodologies", *University of Coimbra, PhD thesis in Chemical Engineering*, 2015.

[2] D. Flynn, J. Ritchie and M. Cregan, "Data Mining Techniques Applied To Power Plant Performance Monitoring", *The Queen's University of Belfast, N. Ireland*, 16th Triennial World Congress, Prague, Czech Republic.

[3] Kadlec.P, Gabrys.B., &Strandt.S,"Data-driven Soft Sensors in the Process Industry", *Computers and Chemical Engineering*,2008.

[4] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques,"

[5] Juha Juselius, "Advanced condition monitoring methods in thermal power plants", *Lappeenranta University of Technology , LUT School of Energy Systems*, , Master's thesis, 2018

[6] Li J., Niu C., Liu J., Zhang L. "Research and Application of Data Mining in Power Plant Process Control and Optimization," in: Yeung D.S., Liu ZQ., Wang XZ., Yan H. (Eds) Advances in Machine Learning and Cybernetics. *Lecture Notes in Computer Science*, vol 3930. Springer, Berlin, Heidelberg, 2006.

[7] Tony Ogilvie, B W Hogg, "Use of data mining techniques in the performance monitoring and optimization of a thermal power plant", *IEEE Colloquium on Knowledge Discovery and Data Mining*, 1998.

[8] Zhiqiang Ge 1, (Senior Member, IEEE), Zhihuan Song, Steven X. Ding, and Biao Huang, (Senior Member, IEEE), "Data Mining and Analytics in the Process Industry: The Role of Machine Learning", *IEEE Access*, 2017.

[9] Narasimhan, S., & Rajendran. "Application of data mining techniques for sensor drift analysis to optimize nuclear power plant performance." *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 3087-3095,2019,doi:10.35940/ijitee.A9139.119119

[10] Jolliffe, IT., "Principal Component Analysis",*Springer*, 2002.

[11] Dote.Y., Ovaska.S.J."Industrial applications of soft computing: A review". In: *Proceedings of the IEEE*. Vol. 89. pp. 1243–1265, pp. 1243–1265, 2001.