Prediction of Biogas Production from Anaerobic Treatment of Distillery Wastewater Using R Programming

[1]Sachita Jadhav, [2]Sonali Nagane, [3]Yashoda Bhanvase, [4]Sunil Khare
[1 2 3 4] *Department of Computer Science & Engineering*
[1 2 3 4] *SKN Sinhgad College of Engineering, Pandharpur*

ABSTRACT

In the field of development of computer algorithm for transforming data into an intelligent action probably known as machine learning. It is used to generate the sense from complex data. Data analysis is the process to make the sense from raw data. With the help of some statistical techniques of machine learning and data mining predictive analytics make the prediction about unknown future event. R is one of the latest programming language support for statistical analysis. It provides many tools to build a prototype which benefit for predictive analytics. This paper focus on Anaerobic treatment of distillery waste water which generate Biogas production. Based on the organic load, pH, Temp and influent COD, Biogas production, Alkalinity will be predicted. It has been found that accuracy rate of the prediction is about 80%.

Keywords

ML, R Programming, Predictive Analysis, Linear Regression

1. INTRODUCTION

Predictive Analytics is the branch of advanced analytics which is used to make prediction about unknown future events. Predictive analytics uses many techniques from data mining statistics modelling, machine learning and artificial intelligent to analyze current data to make prediction about future. It uses number of data mining, predictive modelling and analytical techniques to bring together the management, information technology and modelling business process to make prediction about future. The patterns found in historical and transactional data can be used to identify risk and opportunity for future. Predictive analytics models capture relationships among many factors to access risks with a particular set of conditions to assign a score or weightage.

By successfully applying predictive analytics. The business effectively interprets big data for their benefits.

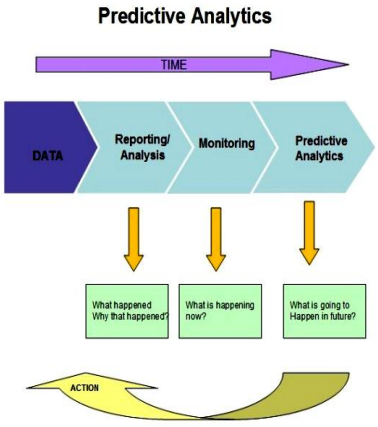

Fig 1. Predictive Analytics

Predictive analytics Process

Step 1: Define project:

Define the project outcome deliverables, scoping of effort, business objectives, identify the data sets which are going to be used.

Step 2: Data Collection:

Data Mining for predictive analysis prepares data from multiple source for analysis. The

provides a complete view of the customer interactions.

Step 3: Data Analysis:

Data analysis is the process of inspecting, clearing, transforming and modelling data with the objective of discovering useful information, arriving at conclusions.

Step 4: Statistics:

Statistical analytics enables to validates the assumptions, hypotheses and test them with using standard statistical models.

Step 5: Modelling:

Predictive modelling provides the ability to automatically create accurate predictive models about future. There are also options to choose the best solution with multi-model evaluation.

Step 6: Deployment:

Predictive model deployment provides the option to deploy the analytics results in to the everyday decision making process to get results, reports, and output but automating the decision based on the modelling.

Step 7: Model Monitoring:

Models are managed and monitored to review the model performance to ensure that it is providing the results expected.

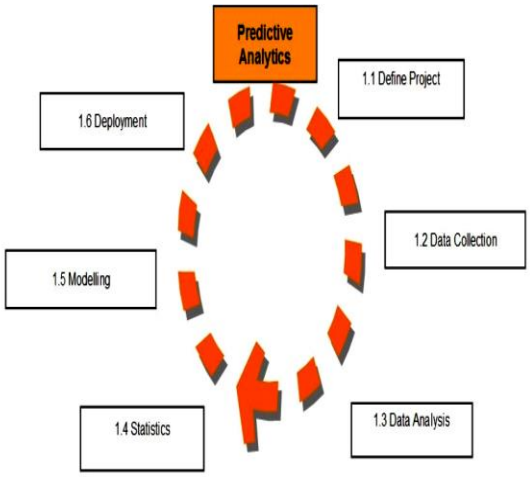

Fig 2. Predictive analytics Process

The R environment

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

The term "environment" is intended to characterize it as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.

IMPLEMENTATION AND RESULTS

To make the prediction of Biogas production from the distillery wastewater treatment plant following experimental laboratory data set has been considered.

Table 1.1

Influent.C	OLR.COD.	HLR.m3..d	HRTDays	pH	Temp0C	MLSS.mg.L	VFA.mg.L	Alkalinity.mg.L	Biogas.m3.d
128	5000	39.06	282	6.9	31	31400	493	39536	1100
130	10000	76.92	143	7	31	31920	905	39353	2300
116	15000	129	85	6.9	32	33680	1066.243	38808	3600
116	20000	172.4	64	7.2	33	34900	1275	36512	5000
137	25000	182.4	60	7.4	33	35200	1974	36232	6500
130	30000	230.8	48	7.3	34	35490	2016	36108	7900
122	35000	286.9	38	7.5	35	36380	2057	35394	9500
116	40000	344.9	32	7.6	34	38500	2262.876	34884	10700
122	45000	368.8	30	7.6	35	39400	2262	34760	12400
122	50000	409.8	27	7.8	36	39480	2386	34320	14200
119	55000	462.2	24	7.6	37	39940	2509	32704	15700
119	60000	504.2	22	7.5	38	40200	3233	30700	17300
115	65000	565.1	19	7.7	38	40200	3459.509	28044	18800
119	70000	588.2	19	7.8	38	41380	3390	27400	19900
119	75000	630.3	17	7.8	38	41380	4032	25320	21100

Based on the environmentally condition such as influent COD, OLR COD, HLR, HRT, pH, temp, MLSS we are predicting value of VFA, Alkalinity and Biogas production.

Prediction: Here first choose the csv input file. This function predicts the future value by analyzing past and present value. It uses many techniques from data mining for collection of data,

machine learning for finding relation between dependent/target data and independent data/predictor data. To find the correlation between data regression algorithm is used. Regression is concerned with specifying the relationship between a single numeric dependent variable and one or more dependent variables. Regression analysis is commonly used for modeling complex relationship among data elements, estimating the impact of a treatment on an outcome, and extrapolating into future. It also uses statistical approach to predict future data. Lm() function finds relation between selected dependent and independent variable.

```
r1 <- lm(Biogas ~ OLR.COD.kg..day , data=dataset13)

a <- coef(r1)[1]

a1 <- coef(r1)[2]

df$Biogas[i]= a + a1* df$OLR.COD.kg..day[i]
```

Display Graph: Graph is the diagram which shows the interrelation between one or more variables. Data is displayed as collection of points, each having the value of one variable determining the position on horizontal axis other on vertical axis. and determining position on vertical axis. This function plots the graph of selected dependent and independent variable and download the graph in the png format.

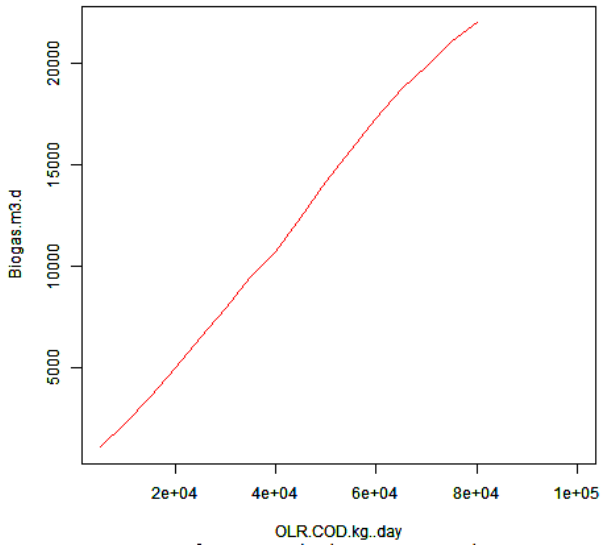

Fig 3. OLR COD kg. day Vs BIOGAS.m3.d

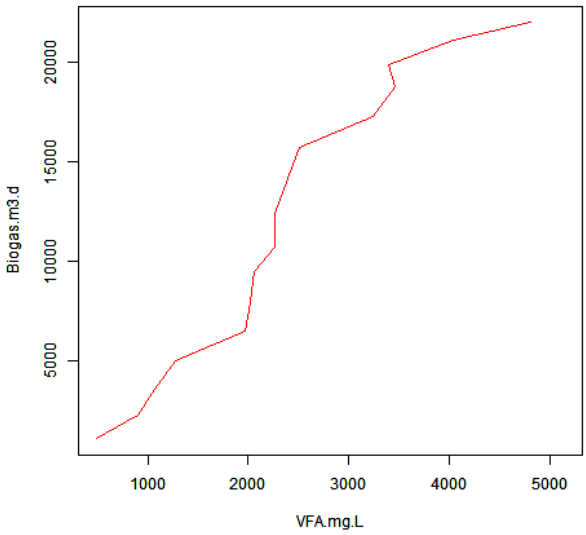

Fig. 4. VFA.mg.l vs BIOGAS.mg.d

CONCLUSION

In this analysis, the original value of Biogas production is 21100. after using this prediction analysis, the value becomes 19900. It has been found that accuracy rate of the prediction is about 90.57%. If value of OLR COD increases the value of Biogas production also increases.

REFERENCES

[1] Machine Learning with R by Brett Lantz.

[2] Rpubs Missing value imputation using Linear Regression.

[3] http://en.proft.me/2015/12/24/types-machine-learning-algorithms/

[4] www.dummies.com

[5] https://machinelearningmastery.com/machine-learning-in-r-step-by-step/

[6] https://en.wikipedia.org/wiki/R_(programming_language)

[7] https://shiny.rstudio.com/