International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 01 | Jan 2022 www.irjet.net p-ISSN: 2395-0072

Prognosis of Biogas Production from Sewage Treatment Plant using Machine Learning

Sabitha S G[1], Rupashire P[2], Mathu Priya J[3*]

[1,2]Student, Department of Biotechnology, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamilnadu, India

[3]Assistant Professor, Department of Biotechnology, Bannari Amman Institute of Technology, Sathyamangalam Erode, Tamilnadu, India*

Abstract - *Biogas generation is a complex process and mathematical modelling is necessary for efficient plant management. Machine learning has emerged as a revolutionary approach for model creation that has the potential to determine and optimize biogas production. On using the data obtained from the Sewage Treatment Plant(STP), Bannari Amman Institute of Technology, Sathyamangalam, Erode, the relationship between biogas production and its system variables are to be determined using a machine learning algorithm. Two algorithms were applied, namely Multiple linear regression and Random forest algorithm to understand the significance of operational parameters influencing the biogas production. Results of the Random forest algorithm showed good accuracy of 97%. Kitchen food waste and Bathing waste were inputs for STP to biogas production. Data visualization has been done to infer the effect on the quantity of biogas production with varying pH, temperature, and Biological Oxygen Demand(BOD) through various graph inferences. These two algorithms assist in legitimate objective use and, as a result, ensure more environmentally friendly energy generation. This research has a lot of significance for biogas plant operators to improve their facility's performance by introducing machine learning into the analytical process.*

Key Words: Biogas, Sewage treatment plant, Machine learning, Random forest, Sustainable energy

1. INTRODUCTION

Anaerobic digestion is a process of breaking down biodegradable wastes like food and kitchen wastes with the help of microorganisms without oxygen intake. The process in return results in the production of biogas and bio fertilizers [1]. The sewage water altogether enters the equalization tank to equalize the parameters, which is then moved to the buffer tank where the pH is adjusted and the nutrient for the anaerobes is fed. From buffer tank, it enters the up-flow blanket anaerobic reactor tank where the anaerobic digestion takes place, the anaerobic microbes grow by intake of nutrients supplied and convert the sewage into renewable energy i.e., biogas such as methane and carbon-di-oxide by performing activated sludge process. Here, the process is of four major steps: Hydrolysis, acedogenesis, acetogenesis, and methanogenesis. In hydrolysis, the larger biomolecules like carbohydrates, fats, etc. are converted into smaller molecules like fatty acids, hydrogen, and acetate. Further on acetogenesis, an acidic environment is being created by acidogenic bacteria, where the molecules are converted to ammonia, carbon dioxide, and hydrogen. This process is followed by acetogenesis, formation of acetate and final products including acetic acid, hydrogen and carbon-di-oxide. In the process of methanogenesis, the carbon-di-oxide and hydrogen together get converted into methane and the acetic acid gets converted to methane and carbon-di-oxide. After all this process of anaerobic digestion, it moves to the aeration tank and clarifier.

The aeration tank reduces the level of BOD and the clarifier separates the sludge and the treated water. And further, the sludge moves to multi-grade filtration and UV treatment. The gas released is stored in the gas balloon. A mixture of gases t is released including methane, hydrogen sulfide, and carbon-di-oxide which are then converted into electricity [2]. Biogas is a mixture of gases like methane of about 55-75%, carbon dioxide of about 25-45%, nitrogen of about 0-5%, hydrogen of about 01%, hydrogen sulphide of about 0-1%, and oxygen of about 0-2%. This serves as a source of sustainable energy preventing the exploitation of conventional fuels [3]. This waste can be converted into eco-friendly bioenergy which involves the action of microorganism's percent in the environment and the feed that is given into the bioreactor that decomposes this waste into biogas. The sludge that is left over after the treatment can be used as a bio fertilizer for agricultural crops [4] due to the high organic content like nitrogen and potassium which is indeed a cheap way to a healthier ecosystem [5]. The interchanging of the top soils in monoculture less fertile soil with STP soil is thought to improve soil fertility and facilitate plant nutrient uptake [6]. Biogas production is influenced by a variety of factors, including operating conditions. pH and temperature are the operational conditions. The organic waste content, the concentration of substrate digested and hazardous chemicals, digestion inhibitors, volatile fatty acids, and ammonia are all affected by power shortages [7]. The anaerobic digestion is highly affected by the

microorganism (mostly methanogens) as it is sensitive to the various environmental conditions [8]. Optimization of these parameters can be achieved by implementing advanced technologies like artificial intelligence. Machine learning comes under artificial intelligence that is revolutionizing the world with its fantastic knowledge capacity. It has wide applications including speech recognition, spam detection, medical diagnosis, wastewater treatment, resource utilization improvements [9], etc. There are lots of algorithms that are developed to perform those specific kinds of functions and give us the desired output, which includes Linear Regression, Logistic Regression, Random Forest, Support Vector Machine, K-nearest neighbors, Decision Tree, etc. Three main categories of machine learning are supervised, unsupervised, and reinforcement learning classified based on the labeling of data. The tree-based pipeline optimization tool was one of the successful automated ML systems developed [10]. The biogas generation has been predicted previously using Artificial Neural Networks [11], data mining algorithms, adaptive neural fuzzy inference systems [12,13], Gradient boosting [14] and Genetic algorithm [15]. Random forest is a supervised algorithm that seems to work better on this large data of this specific Sewage Treatment Plant which can handle continuous variables (regression) as well as categorical variables(classification). This works as a combination of decision trees where based on the majority of the vote, the output is predicted. If machine learning models are integrated into project analytics systems, they can considerably simplify the decision-making process [16]. Approaching this method of automation and prediction for process parameters and outcomes in bioprocess has been favoring the industries recently with explicit technology.

2. METHODOLOGY

2.1. Data collection

The data incorporated in the model is obtained from the Sewage Treatment Plant setup in Bannari Amman Institute of Technology, Sathyamangalam, Erode. We took about 3-4 months (June - Sept)of data including parameters like total inlet flow, pH, temperature, biological oxygen demand, conductivity, mixed liquor suspended solids, and biogas units.

2.2. Data pre-processing

The curated data is then pre-processed. Here, the missing values and null values are checked for and adjusted accordingly. Any outliers detected are omitted and the necessary details are alone taken into consideration to the next level to prevent noise and error in the model.

2.3. Model building and evaluation metrics

Since the data is continuous, we tried to apply regression-based algorithms like multiple linear regression as it comprises two or more inputs [17]. But the model doesn't turn out well in prediction. The other algorithm we employed is Random Forest, is an ensemble learning, where a group of decision trees works along and based on the maximum number of votes, it gives the output. This algorithm is a type of supervised machine learning [18]. Deeply developed trees, in particular, tend to over fit the training sets, resulting in very little bias and large variance. A random forest is an approach of minimizing variation by averaging many decision trees that are trained on several regions within the same training set. This causes an increase in bias and loss in interpretation. But it improves the model performance in the end.

Random forest is similar to decision tree algorithms coming together as a whole in their efforts. By using the collective intelligence of multiple trees on enhancing the performance of a single random tree. Forests provide the impact of K-fold cross-validation, though they are not identical [19].

Bagging:

To train tree learners, the random forest algorithm uses the bootstrap aggregation(bagging) technique [20]. By repeatedly updating the training set, bagging takes a random sample, and then fits the trees to these samples: Bagging takes a random sample from a training set A = a_1..., a with answers B = b_1..., b_n by continually replacing the training set B then fit the trees to these samples: Call these A_x, B_x. Take n training instances from A, B, using replacement, for x= 1..., X. Train a classification or regression tree f_x on A_x, B_x. Next to training, summing the predictions from all the various regression trees on a' can be used to make predictions for unseen samples a':

$$\hat{f} = \frac{1}{X}\sum_{x=1}^{X} f_x(a')$$

(1)

A majority vote can be employed in the case of classification trees. This method improves the model's performance since it minimizes model variance without raising bias. While a single tree's predictions are highly sensitive to noise in its training set if the trees are not correlated, the average of numerous trees' predictions is not. Bootstrap sampling de-correlates the trees and presents them with various training sets, resulting in significantly correlated trees when several trees are

trained inside a single training set [21]. In addition, the standard deviation of the predictions from all of the separate regression trees on a' can be used to evaluate uncertainty:

$$\sigma = \sqrt{\frac{\sum_{x=1}^{X} \left(f_x(a') - \hat{f}\right)^2}{X-1}} \quad (2)$$

The number of samples or trees is indicated by the letter X in equation (3). A hundred to thousand trees are commonly utilized, depending on the size and type of training set. The mean prediction error on each training sample is called cross-validation. a_i can be used to find the ideal number of trees X by considering trees that did not have a_i in their bootstrap sample [22]. The training and test error decreases after a few trees have been fitted.

Bagging to the random forest:

The technique explained here is the original tree bagging algorithm. Another type that is used by random forests in bagging algorithms is a modification of a tree learning algorithm. During the learning process, each candidate is given a random subset of features. This technique is defined by the term "Feature bagging". Because in a conventional bootstrap sample, if one or a few features are exceptionally effective predictors of the response variable (target output), they will be selected in many of the X trees, such that it is correlated. Ho investigates how bagging and random subspace projection contribute to the accuracy attained in various settings [23]. For a classification problem with p features, the square root of p (rounded down) features are commonly used in each split. For regression problems, the inventors recommend using p/3 (rounded down) as the default, with a minimum node size of 5. The ideal settings for these characteristics will vary depending on the error or issue, it is considered as tuning parameters [25,26].

This model gave an accuracy of about 97% i.e., a regression coefficient of 0.97. This seemed to be effective and working well on this data.

2.4. Data visualization

These data are visualized through graphical representations like scatter plots and line charts. Based on the nutrient sources used like cow dung and food wastes, a comparison chart has been plotted on the biogas production to find out which is capable of higher production. The input parameters like total inlet flow, pH, temperature, BOD are scatter plotted with the amount of biogas produced to effectively optimize the process efficiently.

3. RESULTS AND DISCUSSION

Obtaining the results from the machine learning model that could be used to predict methane production numerically, the random forest model showed the best performance when applied to the STP data, with an R square value of 0.97. The results from the Multiple linear regression model showed less performance compared to the random forest model. Using the data obtained, we were able to find the connection between the operational parameters and the amount of biogas produced. Temperature control is a critical parameter for the development of anaerobes in the digester, this has a strong influence over the quantity of biogas production. From figure 1 we can see that the biogas produced was relatively high during 28° to 29°C . Another factor that influences the digestion process is the pH of the anaerobic digestion process. From figure 2 better yield of biogas was observed pH at a range of 4.5to 5.5. Biogas formation showed a significant difference between biogas formation from wastewater at pH 4.5 to 5.5 producing a higher amount of biogas than pH above 6. The amount of biodegradable material in a sample is measured by the biological oxygen demand (BOD). It is based on bacteria consuming biodegradable material and depleting oxygen. Municipal waste usually contains a relatively low concentration of BOD. Below figure 3 shows the amount of biogas produced concerning the BOD level. From the below figure 4, the orange line indicated the amount of biogas produced daily when the feed used was food waste and the blue line indicates the biogas produced when sugar press was used as feed. From this graph, we can show the significant difference between the biogas production relative to the feed given for microbes.

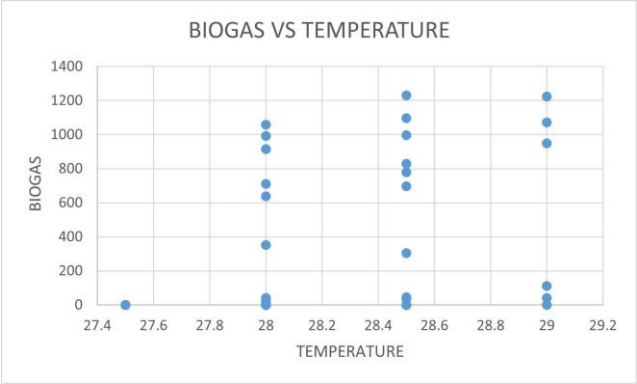

Fig -1: Biogas Vs Temperature

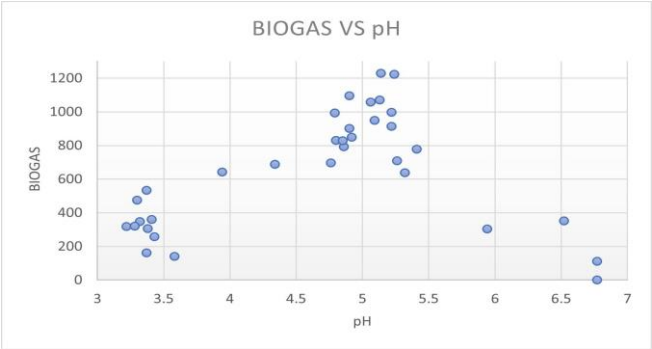

Fig -2: Biogas Vs pH

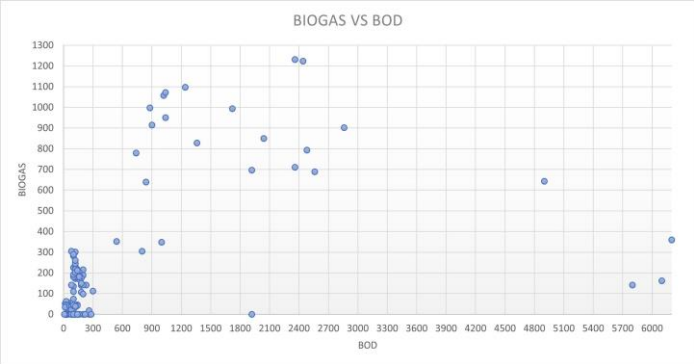

Fig -3: Biogas Vs BOD

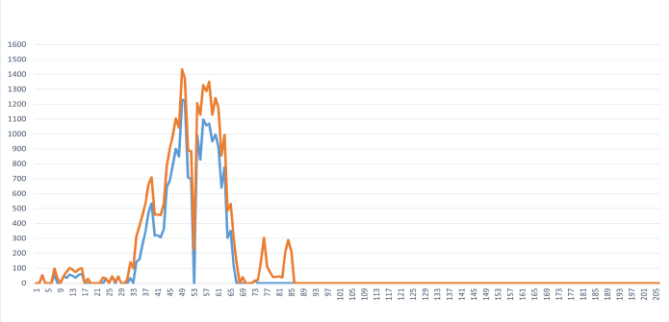

Fig -4: Comparison graph on the various inputs given as feed

4. CONCLUSIONS

An attempt has been made to predict the biogas production from STP plants digitally using advanced technologies like machine learning with production environmental variables as input parameters, the results showed good accuracy of 97 percent using the Random Forest algorithm. Randomly few data were collected and fed as an input to the developed model, with very little variance the model was able to predict the desired output. The graphical representation of the parameters like pH, Biological oxygen demand(BOD), and temperature affecting the biogas yield was extrapolated. It was found that at the temperature of 28°-29°C and a pH of about 4.5-5.5 the production of biogas was at maximum level. Also, the various feed sources like cow dung, food, and kitchen wastes were compared and the latter seems to be effective. This helps in the optimization of the bioprocess thus making it efficient. The biogas produced can be utilized for a variety of purposes, including cooking, lighting, heating, cooking and as a biofuel that can be pumped into the city's gas network [26]. Artificial intelligence is used to create prediction models between functional parameters and target outputs in order to discover the best combination of parameters [27]. The enhancement of biogas in terms of quantity and quality at every cycle of a biogas production process saves energy and efficiency at a wastewater treatment facility's current generation unit and sludge drying plant [28].

REFERENCES

[1] Hagos, K.; Zong, J.; Li, D.; Liu, C.; Lu, X. Anaerobic co-digestion process for biogas production: Progress, challenges, and perspectives. Renewable Sustainable Energy Rev. 2017, 76, 1485– 1496, DOI: 10.1016/j.rser.2016.11.184.

[2] R. Borja,2.55 - Biogas Production, Editor(s): Murray Moo-Young, Comprehensive Biotechnology (Second Edition), Academic Press,2011, Pages 785-798, ISBN 9780080885049, https://doi.org/10.1016/B978-0-08-088504-9.00126-4.

[3] Demirbas, A., Taylan, O., & Kaya, D. (2016). Biogas production from municipal sewage sludge (MSS). Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 38(20), 3027–3033.

[4] Sakiewicz, P.; Piotrowski, K.; Ober, J.; Karwot, J. Innovative artificial neural network approach for integrated biogas wastewater treatment system modeling: Effect of plant operating parameters on process intensification. Renewable Sustainable Energy Rev. 2020, 124, No. 109784

[5] Appels, L., Baeyens, J., Degreve, J., and Dewil, R. 2008. Principles and potential of the anaerobic digestion of waste-activated sludge. Prog. Energy Combust. Sci. 34:755–781

[6] Pandiyan, Balaganes & Mangottiri, Vasudevan & Saragur, Suneeth. (2019). Chemical Characterization and Environmental Implications of Recycled Sewage Sludge in the Proximity Soil of a treatment plant.

[7] Wang, Luguang; Long, Fei; Liao, Wei; Liu, Hong (2020). Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. Bioresource Technology, 298(), 122495–. DOI: 10.1016/j.biortech.2019.122495

[8] Yetilmezsoy, K.; Turkdogan, F. I.; Temizel, I.; Gunay, A. Development of Ann-Based Models to Predict Biogas and Methane Productions in Anaerobic Treatment of Molasses Wastewater. Int. J. Green Energy 2013, 10, 885–907

[9] Guo, Hao-nan; Wu, Shu-Biao; Tian, Ying-Jie; Zhang, Jun; Liu, Hongtao (2021). Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. Bioresource Technology, 319(), 124114.

[10] Yan Wang, Tyler Huntington, and Corinne D. Scown Tree-Based Automated Machine Learning to Predict Biogas Production for Anaerobic Co-digestion of Organic Waste ACS Sustainable Chemistry & Engineering 2021 9 (38), 12990-13000

[11] Rego, A. S.; Leiteb, S. A.; Leiteb, B. S.; Grilloc, A. V.; Santosa, B. F. Artificial Neural Network Modelling for Biogas Production in Bio digesters. Chem. Eng. Trans 2019, 74, 25–30

[12] A. Kusiak and X. Wei 2011.Prediction of methane production in wastewater treatment facility: a data-mining approach Ann. Op. Res., 216 (2011), pp. 71-81

[13] David P.B.T.B. StrikAlexander M. DomnanovichLoredana ZaniRudolf BraunPeter Holuba (2005). Prediction of trace compounds in biogas from anaerobic digestion using the

[14] De Clercq, D., Wen, Z., Fei, F., Caicedo, L., Yuan, K., Shang, R.: Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion. Sci. Total Environ. 712, 134574 (2020).

[15] H. AbuQdais, K. BaniHani, N. Shatnawi Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. Resour. Conserv. Recycle., 54 (2010), pp. 359-363

[16] Djavan De Clercq, Devansh Jalota, Ruoxi Shang, Kunyi Ni, Zhuxin Zhang, Areeb Khan, Zhongguo Wen, Luis Caicedo, Kai Yuan, "Machine learning-powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data", Journal of Cleaner Production, Volume 218,2019, Pages 390-399.

[17] Dandikas, V.; Heuwinkel, H.; Lichti, F.; Drewes, J. E.; Koch, K. Predicting methane yield by linear regression models: A validation study for grassland biomass. Bioresour. Technol. 2018, 265, 372–379.

[18] Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. Archived from the original (PDF) on 17 April 2016. Retrieved 5 June 2016.

[19] Genuer, R., Poggi, J.-M., Tuleau-Malot, C., Villa-Vialaneix, N., 2017. Random Forests for Big Data. Big Data Res. 9, 28–46.

[20] N. Bibi, I. Shah, A. Alsubie, S. Ali and S. A. Lone, "Electricity Spot Prices Forecasting Based on Ensemble Learning," in IEEE Access, vol. 9, pp. 150984-150992,2021, doi:10.1109/ACCESS. 2021.3126545.

[21] Y. Chen, M. Huang, C. Hu, Y. Zhu, F. Han and C. Miao, "A coarse-to-fine feature selection method for accurate detection of cerebral small vessel disease," 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 2609-2616, doi: 10.1109/IJCNN.2016.7727526.

[22] Ho, Tin Kam (2002). "A Data Complexity Analysis of Comparative Advantages of Decision Forest Constructors" (PDF). Pattern Analysis and Applications. **5** (2):102–112. doi:10.1007/s100440200009. S2CID 7415435

[23] Gareth James; Daniela Witten; Trevor Hastie; Robert Tibshirani (2013). An Introduction to Statistical Learning. Springer. pp. 316–321.

[24] Rocha, Anderson & Scheirer, Walter & Forstall, Christopher & Cavalcante, Thiago & Theophilo, Antonio & Shen, Bingyu & Carvalho, Ariadne & Stamatatos, Efstathios. (2016). Authorship Attribution for Social Media Forensics. IEEE Transactions on Information Forensics and Security. 12. 5. 10.1109/TIFS.2016.2603960.

[25] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2008). The Elements of Statistical Learning (2nd ed.). Springer. ISBN 0-387-95284-5.

[26] Laskri, Nabila & Hamdaoui, Oualid & Nedjah, Nawel. (2015). Experimental Factors Affecting the Production of Biogas during Anaerobic Digestion of Biodegradable Waste. International Journal of Environmental Science and Development. 6. 451-454. 10.7763/IJESD. 2015.V6.635

[27] Dong, Cuiying; Chen, Juan (2019). Optimization of process parameters for anaerobic fermentation of corn stalk based on least-squares supports vector machine. Bio resource Technology, 271(), 174–181. doi: 10.1016/j.biortech.2018.09.085

[28] Akbaş, H., Bilgen, B., Turhan, A.M.: An integrated prediction and optimization model of the biogas production system at a wastewater treatment facility. Biores. Technol. 196, 566–576 (2015).