Journal of Cleaner Production 218 (2019) 390–399

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data

Djavan De Clercq [a, b, 1], Devansh Jalota [e, 1], Ruoxi Shang [e], Kunyi Ni [d], Zhuxin Zhang [b], Areeb Khan [b, c], Zongguo Wen [a, *], Luis Caicedo [f], Kai Yuan [g]

[a] State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, 100084, China
[b] Department of Industrial Engineering and Operations Research, University of California, Berkeley, USA
[c] Department of Computer Science, University of California, Berkeley, USA
[d] Department of Chemistry, University of California, Berkeley, USA
[e] College of Engineering, University of California, Berkeley, USA
[f] EARTH University, Costa Rica
[g] University of Edinburgh, Edinburgh Centre for Robotics, UK

ARTICLE INFO

Article history:
Received 18 July 2018
Received in revised form
2 January 2019
Accepted 4 January 2019
Available online 21 January 2019

Keywords:
Biogas
Machine learning
China
Graphical user interface

ABSTRACT

The search for appropriate models for predictive analytics is currently a high priority to optimize anaerobic fermentation processes in industrial-scale biogas facilities; operational productivity could be enhanced if project operators used the latest tools in machine learning to inform decision-making. The objective of this study is to enhance biogas production in industrial facilities by designing a graphical user interface to machine learning models capable of predicting biogas output given a set of waste inputs. The methodology involved applying predictive algorithms to daily production data from two major Chinese biogas facilities in order to understand the most important inputs affecting biogas production. The machine learning models used included logistic regression, support vector machine, random forest, extreme gradient boosting, and k-nearest neighbors regression. The models were tuned and cross-validated for optimal accuracy. Our results showed that: (1) the KNN model had the highest model accuracy for the Hainan biogas facility, with an 87% accuracy on the test set; (2) municipal fecal residue, kitchen food waste, percolate, and chicken litter were inputs that maximized biogas production; (3) an online web-tool based on the machine learning models was developed to enhance the analytical capabilities of biogas project operators; (4) an online waste resource mapping tool was also developed for macro-level project location planning. This research has wide implications for biogas project operators seeking to enhance facility performance by incorporating machine learning into the analytical pipeline.

1. Introduction

1.1. The global biogas industry lacks control methodologies powered by machine learning algorithms

Anaerobic digestion (AD) is commonly acknowledged as an environmentally friendly way to convert organic waste into biogas (Naroznova et al., 2016; Thyberg and Tonjes, 2017). During the anaerobic digestion process, organic substrate is biologically degraded and converted into biogas, which can be used for various applications including electricity production or vehicle fuel. The AD process can be separated into four stages: hydrolysis, acidogenesis, acetogenesis and methanogenesis (Appels et al., 2011). The microbiology behind this four-stage process is highly complex, and controlling organic substrate input carefully is essential for maintaining process stability and optimal biogas output. Maintaining process stability becomes additionally complex during anaerobic co-digestion, which involves digesting various waste types simultaneously.

In the last 40 years, various control methodologies have been proposed to maintain optimal levels of organic substrate input. Such control methods must compromise between maximizing process stability, biogas output, economic yield, environmental performance, and more (Alcaraz-González et al., 2005; Barnett and

* Corresponding author.
E-mail address: wenzg@tsinghua.edu.cn (Z. Wen).
[1] Dual first-authorship.

https://doi.org/10.1016/j.jclepro.2019.01.031
0959-6526/

Andrews, 1992; Chynoweth et al., 1994; Holubar et al., 2003; Méndez-Acosta et al., 2007; Mu et al., 2008; Zaefferer et al., 2016).

However, (Gaida et al., 2017) state that "analyzing AD control strategies, it is astonishing how few full-scale applications are published in the literature, although the demand is high", and, "new robust and low-maintenance online measurement systems need to be developed", which are "financially feasible and robust". This means that automated process control strategies have not yet been successfully implemented and validated at full-scale; challenges include a slow uptake in advanced analytical solutions in the biogas industry, and an absence of precision instruments capable of process-monitoring across the four AD stages in real-time. In addition, many control strategies were developed based on laboratory-scale studies; these strategies rely on extensive equipment for process monitoring, which is often unavailable or costly to operate. Furthermore, the results of such lab-scale experiments often do not generalize to full-fledged industrial biogas conversion facilities.

For instance, (Matuszewska et al., 2016) found that optimal biogas production was achieved for substrate mixtures of swine manure/maize silage and whey/grass silage in the ratio of 6:4 respectively. However, the authors also revealed that such mixtures were highly improper for industrial applications, due to extended periods of process inhibition, where little or no biogas production occurred. Studies like these highlight the difficulties of maintaining optimal process control in industrial facilities based on lab-scale experimentation, and that the optimal design of biogas plants varies according to local conditions, i.e. feedstock supply and local infrastructure (Chen et al., 2017).

1.2. Despite rapid biogas sector growth in China and abroad, projects suffer from major operational inefficiencies

One region which is particularly beset by a lack of sophisticated control methods in the biogas sector is China. In China, anaerobic digestion technology use has grown rapidly in recent years (a detailed survey can be found Chen et al., 2017 and Chen and Liu, 2017). China had an estimated 100,000 modern biogas plants and 43 million residential-scale digesters in 2014, generating about 15 billion m^3 of biogas (equal to approximately 9 billion m^3 of biomethane) (Scarlat et al., 2018).

Clear policy support from the Chinese government will continue to strengthen the Chinese biogas industry. For instance, under the 12th Five-Year Plan, the Chinese government ratified the construction of 242 pilot-scale restaurant food waste treatment projects across 100 cities in China (De Clercq et al., 2017c). Many of these projects are still under construction or in the early phases of operation. In addition, the 2020 objectives of *The Medium-and-Long Term Development Plan for Renewable Energy* are to reach 80 million household biogas plants and 8000 large-scale biogas projects with an installed capacity of 3000 MW and annual biogas production of 50 billion m^3 (Scarlat et al., 2018). Furthermore, under China's 13th Five-Year Plan, the National Energy Administration set targets of 10 billion, 20 billion, and 40 billion cubic meters of annual biomethane output by 2020, 2025 and 2030 respectively ("biomethane" is a purified form of biogas) (De Clercq et al., 2017a).

Despite this vigorous policy support, the biogas industry in China is plagued by operational issues that impede production in biogas facilities. For example, Deng et al. (2017) found that many anaerobic sewage treatment plants had shut down due to process complexity, understaffing, poor management, and reduced biogas/electricity production. (De Clercq et al., 2016) found that a major anaerobic food waste treatment facility in Beijing suffered from low biogas production, weak process monitoring and feedback, inefficient biogas utilization, and excessive troubleshooting and downtime. Such operational bottlenecks have adverse impacts on the economic performance of biogas projects: (De Clercq et al., 2017c) conducted an economic performance evaluation of a Chinese biowaste pilot facility based on stochastic modeling of net present value and found that the pilot project suffered from weak economic performance.

China is not the only region with issues related to biogas facility performance. Similar problems have also been found in countries such as South Korea (Kim et al., 2012) and Brazil (De Clercq et al., 2017d).

1.3. Control methods powered by machine learning algorithms can enhance operational efficiency

Enhancing the efficiency of biogas plants is crucial for the economic sustainability of the biogas sector. Better efficiency can reduce operational costs related to biowaste, chemical, water, and electricity inputs; it can also reduce costs in biogas upgrading and transmission to end-use destinations. Given that tremendous government expenditure has been allocated to the proliferation of the industry (Gottfried et al., 2018), it is vital that efficiency is maximized so that public funding does not go to waste.

Given the aforementioned issues, this research addresses the following research question: *how can machine learning tools be applied to improving operational decision-making in biogas facilities?* This study collected primary data from industrial biogas facilities in China, and developed a web-based graphical user interface (GUI) powered by machine learning algorithms.

The remainder of the manuscript is structured as follows. Section 2 provides a literature review and highlights the novelty of this study. Section 3 elucidates the machine learning methods and programming tools used to develop the online GUI. Section 4 presents the results of the study and a discussion of the findings. Concluding remarks are provided in section 5.

2. Literature review

To the authors' knowledge, there is no prior literature that has developed an interactive software tool powered by machine learning algorithms for the biogas facilities (especially with a focus on Chinese biogas projects). According to Gaida et al. (2017), full-scale applications of control strategies for AD plants are extremely rare. Nevertheless, previous studies have developed techniques and models with the goal of promoting efficiency in biogas projects and other types of bio-energy projects. A brief review of the most relevant studies are summarized here, and their shortcomings with regard to industrial-scale generalizability and robust machine learning model development are highlighted.

2.1. Lab-scale tests lack generalizability to large-scale production

There has been a plethora of lab-scale experimental studies that investigate the effect of mixing various organic substrates on biogas production. However, these studies often do not capture the characteristics of large-scale anaerobic digestion in industrial projects. For example, some projects in China treat up to 1000 tons per day of mixed waste which can contain location-specific impurities. Such specifics are often hard to simulate in lab-scale experiments, and machine learning can help in this regard.

For instance, Matuszewska et al. (2016) investigated the biomethane potential (BMP) of co-digested swine/cattle manure, maize, grass silage, and acid whey. Corro et al. (2013) investigated the generation of biogas from coffee pulp and cow dung co-digestion in batch digestion. Cabbai et al. (2016) performed experimental co-digestion of sewage sludge and organic municipal

waste in a 3.4 m^3 Continuously Stirred Tank Reactor (CSTR), obtaining an increase of 192% in biogas production. Song and Zhang (2015) conducted lab-scale co-digestion of pretreated wheat straw with cattle manure, and analyzed the microbial community during the process. Miao et al. (2014) investigated the effect of co-digesting blue algae with swine manure on methane production in 500 mL reactors. Cabbai et al. (2013) conducted lab-scale biomethane potential tests on various types of organic municipal waste co-digested with sewage sludge. Zhang et al. (2013) investigated the effects on biogas production of co-digestion alkaline-pretreated banana stems with swine manure. Li et al. (2018) experimented with two-stage, high-solid anaerobic co-digestion of food waste, chicken manure, and grass waste in a 1000 mL bioreactor. Xing et al. (2014) evaluated the co-digestion performance of wheat straw and fruit/vegetable waste in a novel two-phase anaerobic reactor. Garfí et al. (2011) attempted to improve the anaerobic digestion performance of cow and guinea pig manure in low-cost unheated tubular digesters implemented at high altitudes, by comparing operating conditions and co-digesting both manures. Jacob and Banerjee (2016) modeled anaerobic co-digestion of potato waste and aquatic weeds with artificial neural networks coupled with genetic algorithms for biological processes. However, the results were based on experimental scale lab studies of two types of waste, rather than a full-fledged industrial co-digestion process. Sathish and Vivekanandan (2016) applied artificial neural networks to increasing biogas yield derived from agricultural waste, albeit this study was limited to an experimental set-up.

2.2. Modeling approaches

In addition to lab-scale studies, several studies have proposed physical mathematical models of the biological process in anaerobic digestion. However, many of these models are still informed by experimental AD set-ups, which impedes the generalizability of such models.

For instance, Fedailaine et al. (2015) modeled the biokinetics of anaerobic digestion based on several aspects such as microbial activity, substrate degradation, and methane production; however, eight simplifying assumptions were made in this model, which limits its application to industrial-scale reactors. Cook et al. (2017) developed a stability assessment tool for anaerobic co-digestion, which allowed for evaluation of the stability of real or simulated digesters. The study used seven stability indicators (including pH, alkalinity, free ammonia, etc.) which are commonly measured during digester operation. However, the model in the study was not informed by long-term operational data of a full-scale project, and did not make use of additional indicators.

Advanced simulation models such as the Anaerobic Digestion Model No. 1 (ADM1) and benchmark models for anaerobic wastewater treatment like the benchmark model no. 2 (BSM2) offer many possibilities for a comprehensive assessment of control strategies with different measurement variables and different degrees of uncertainty. However, such models were not explicitly designed to make use of industrial-scale production data to accurately forecast biogas production under unique engineering conditions, although they can be adapted for this case.

For instance, Grim et al. (2015) applied the Dynamic Biogas plant Model (DyBiM) and ADM1 to evaluate the technical requirements and economic implications of demand-oriented biogas production in a Swedish CHP biogas plant digesting cattle manure and sugar beet. In addition, Bensmann et al. (2016) used the ADM1-based operating scheme capable of detecting process disturbances in anaerobic digestion. Another popular model used for calculating the biogas potential of anaerobic digestion is Buswell's equation. This equation involves breaking down organic waste into its elemental components of carbon, hydrogen, oxygen, nitrogen, and sulfur. However, this method provides the theoretical biogas yield, and is not based on operational conditions in industrial-scale projects (De Clercq et al., 2016).

2.3. Traditional statistical methods

Other studies have sought to predict biogas production and the determinants of efficiency in anaerobic digestion processes based on traditional statistical methods. However, a key shortcoming of the models developed in these studies is that they do not incorporate the latest advances in machine learning to predict biogas output; instead, traditional statistical performance metrics such as R^2 and root-mean-square-error (RMSE) are reported as performance metrics.

Modern machine learning models, on the other hand, are evaluated on their ability to accurately predict previously unseen data. Datasets are separated into model training and model testing partitions for this reason (James et al., 2013), and out-of-sample evaluation metrics are preferred, since test-set measures reveal possible model overfitting. Examples of literature with took this traditional statistical approach include studies by De Clercq et al. (2017a, 2017e), which combined statistical tools (principal component analysis and multiple linear regression) with methods from operations research (data envelopment analysis) in order to model the determinants of efficiency in biogas projects, and found a number of inefficiencies such as decreasing returns to scale. (Terradas-Ill et al., 2014) developed a thermic model to predict biogas production in unheated fixed-dome digesters buried in the ground, although their model was not tested on a validation data set, and was not applicable to large-scale facilities. De Clercq et al. (2017b) used multi-criteria decision analysis to benchmark food waste and biowaste projects based on technical, economic, and environmental criteria, and made six major policy recommendations based on the results; however, this study did not apply generalizable modeling tools for project operators to enhance production based on waste inputs.

As demonstrated in the literature review, this study fills important gaps in the literature. We move beyond lab-scale experimentation of the effects of anaerobic co-digestion by applying machine learning models to industrial-scale production. Lab-scale studies often do not capture the operational realities in large-scale projects; moreover, conducting lab-scale experiments is costly and time-consuming. Project managers often desire real-time results with regards to the effect of different waste inputs on the anaerobic digestion process.

3. Methodology

This study required multiple steps to produce a functioning tool for biogas project operators. These steps included a comprehensive data collection effort, feature engineering/data transformation, model building and evaluation, and final development of a graphical user interface to the machine learning models deployed. An overview of the methodology is displayed in Fig. 1. The entire Python code behind these methods are provided online, to ensure full transparency and reproducibility of our results.

3.1. Raw data collection and preprocessing

Daily operational data was collected from two major anaerobic digestion facilities in the south of China. The first facility, “Hainan BioCNG”, is China's first major biowaste-to-biomethane vehicle fuel facility. The project has a biowaste input capacity of 750 tons/day, a daily maximum production of 30,000 m^3/d of biomethane vehicle

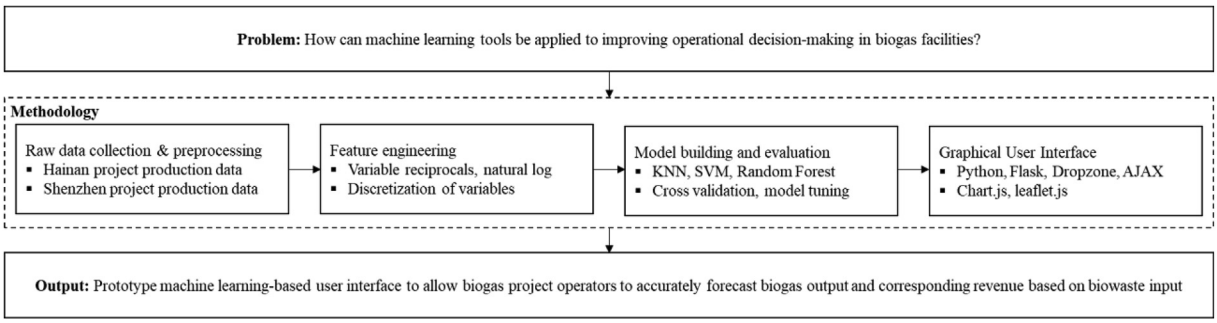

Fig. 1. Overview of methods.

fuel, and produces liquid/solid organic fertilizer. The facility treats a variety of agricultural (i.e. cow manure, straw, bagasse), municipal (food waste, human excrement), and industrial (fish processing waste, alcohol processing waste) biowastes. The second facility is Shenzhen's largest and most advanced food waste treatment project. It treats various types of municipal organic waste, including food waste, waste oil, and fruit/vegetable waste, and will soon be capable of treating between 600 tons/day to 800 tons/day.

For the Hainan biomethane project, daily production data was collected from May 2014 to February 2017. For the Shenzhen project, data was collected from December 2015 until February 2017. Regarding pre-processing, the various production data were dispersed across several spreadsheets corresponding to monthly production; these datasets were merged into a single data frame for each project. In addition, values with undesirable characters or spaces were removed. Extreme outliers were also discarded.

3.2. Feature engineering

After collecting and pre-processing the raw data, additional data transformation was conducted in order to enhance the variables for model training on data from both of the surveyed projects. For instance, in addition to the raw values for each variable, the reciprocals, natural logarithms, and squared values of these amounts were also taken. In addition, variables irrelevant to the biogas output, such as project water consumption, or electricity usage, were dropped. A time lag between biowaste input and biogas output was also introduced to reflect the time taken to convert biowaste into biogas under mesophilic conditions.

In addition to this feature engineering, basic transformations of the target variable (biogas output) were also introduced. Biogas production was "binned" into several buckets corresponding to the production amount: "low", "medium", and "high", which correspond approximately to 0 to 5033 m^3, 5034 m^3 to 10,067 m^3, and 10,068 m^3 to 15,100 m^3, since the values of biogas production is divided into 3 equal-width bins ranging from the minimum(0) of biogas production to the maximum(15100) of biogas production. Hence, "Low", "medium", and "high" are given labels 0, 1, and 2, which replace the numeric values. The continuous, numeric values for biogas production were also left intact. Machine learning models were then trained for classification of biogas outputs (i.e. predicting which bin biogas output would fall in) and regression (i.e. predicting a precise numeric value for biogas output).

Data cleaning and feature transformation was performed using Python's "pandas" package. Additional information on logic behind the feature engineering and its relationship to model performance is provided in the supplementary information. In addition, a full list of the original and transformed variables is provided in the supplementary information.

3.3. Model building and evaluation

3.3.1. Model selection

As stated in section 3.2. The study applied both classification and regression model to predict biogas output. Following the data cleaning and transformation, the data was partitioned into a training set and a test set. For the Hainan and Shenzhen data, the most recent 20% and 15% of the data was retained for the test data.

In order to explore which models could most accurately predict biogas production based on the model input features, a number of machine learning algorithms were trained and tested on the data. These included support vector machine (SVM) random forest, logistic regression, extreme gradient boosting (XGBoost) and k-nearest neighbors (kNN). For a full mathematical discussion of these methods, see the seminal texts on statistical learning, (Hastie et al., 2009; James et al., 2013); a brief introduction of the algorithms is presented here.

K-nearest neighbors (kNN): kNN is a non-parametric prediction algorithm which searches the k most similar features in a historical database to predict future values. The model structure is simple and has high computational efficiency (Wang et al., 2015). Given a positive integer K and a test observation x_0, KNN classifiers identify K points in the training data that are closest to x_0, represented by N_0. Then, it estimates conditional probability for class j as the fraction of points N_0 whose response values equal j (James et al., 2013).

$$\Pr(Y = j|X = x_0) = \frac{1}{K}\sum_{i \in N_0} I(y_i = j) \tag{1}$$

Support vector classifier (SVC): The fundamental idea behind support vector classification is to find a hyperplane in the feature space that divides different categories of observations to the largest separation (Shen et al., 2017). The support vector classifier is the solution to the optimization problem.

$$\max_{\beta 0,\beta 1,\ldots,\beta p,\varepsilon 1,\ldots,\varepsilon n} M$$

$$s.t.\ \sum_{j=1}^{p} \beta_j^2 = 1$$

$$yi(\beta 0 + \beta 1xi1 + \beta 2xi2 + \ldots + \beta pxip) \geq M(1 - \varepsilon i)$$

$$\varepsilon_i \geq 0,\ \sum_{i=1}^{n} \varepsilon_i \leq C \tag{2}$$

Where C is a nonnegative tuning parameter M is the width of the

margin, $,\varepsilon 1,\ldots,\varepsilon n$ are slack variables which allow for observations to be on the wrong side of the hyperplane. Solving this problem allows us to classify a test observation x^* based on the sign of $f(x^*) = \beta_0 + \beta_1 x_1^* + \ldots + \beta_p x_p^*$.

Random forests: Random forests are powerful non-parametric statistical methods that permit regression problems as well as two-class or multi-class classification problems. From a practical perspective, random forests are widely used due to their high performance and low amount of parameters to tune (Genuer et al., 2017).

Boosting: Unlike bagging algorithms, which fit base models in parallel, gradient boosting techniques combine a series of weaker base learners (generally regression trees) into a stronger one. The boosting method fits additive base learners in order to minimize the loss function provided; the process continues until the loss function reduction becomes limited (Xia et al., 2017). Scikit-learn's XGboost object was used in this research. The algorithm for boosting is given by first setting $\widehat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set. Then, for $b = 1, 2, \ldots, B$, the following steps are repeated (James et al., 2013):

(a) Fit a tree $\widehat{f}^b$ with d splits ($d+1$ terminal nodes) to the training data (X, r).
(b) Update $\widehat{f}$ by adding in a shrunken version of the new tree: $\widehat{f}(x) \leftarrow \widehat{f}(x) + \lambda \widehat{f}^b(x)$.
(c) Update the residuals: $ri \leftarrow ri - \lambda \widehat{f}^b(xi)$.
(a) Output the boosted model: $\widehat{f}(x) = \sum_{b=1}^{B} \lambda \widehat{f}^b(x)$

where B is the number of trees, λ is the shrinkage parameter, and d the number of splits in each tree.

Logistic regression: Logistic regression is a commonly used classification method to model the probability that a certain Y-value belongs to a particular category (Yang and Loog, 2018). A binary target variable is coded as a 1 or 0, where the aim is to model the relationship between the probability that a response Y is equal to 1 and a set of covariates. The ratio of the probability $Y = 1$ to the probability of $Y = 0$ is referred to as the log odds (Ekström et al., 2018). In this research, multiple logistic regression was used to classify response variables with more than two classes. Logistic regression uses the logistic function to fit the model (Huang et al., 2017):

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X+\ldots+\beta_p X_p}} \tag{3}$$

where the maximum likelihood method is used to estimate the model coefficients $\beta_0, \beta_1, \ldots, \beta_p$.

3.3.2. Feature importance

One of the outputs from running the algorithms was feature importance. For application of the random forest algorithm, for instance, feature importance was computed using the mean decrease in Gini index method. The Gini index is defined by

$$G = \sum_{k=1}^{K} \widehat{p}_{mk}(1 - \widehat{p}_{mk}) \tag{4}$$

where $\widehat{p}_{mk}$ represents the proportion of training observations in the mth region that are from the kth class. The Gini index measures variance across the K classes, and is a measure of node purity in tree-based methods (James et al., 2013).

3.3.3. Resampling and model tuning

The models were trained using the 10-fold cross-validation resampling method and rigorous hyper-parameter tuning; using Python's scikit-learn, gridsearch CV was performed with combinations of hyper-parameters for the various models. For instance, XGBoost models were tuned across parameters such as "number of estimators" and "max depth". For additional details, see the full online code sample referred to in the supplementary information.

3.3.4. Machine learning models and evaluation metrics

LogisticRegression and perceptron models are imported from sklearn.linear_model package, RandomForestClassifier model is imported from sklearn.ensemble package, KNeighborsRegressor model is imported from sklearn.neighbors package, XGBClassifier model is imported from xgboost package, and svm model is imported from sklearn. Those model functions are used directly with the input and output training data, and the test scores are generated by score method inherited in the model functions.

Classification model performance was judged according to the accuracy metric, which computes the percentage of instances where predicted labels are equal to the actual labels. Regression model performance evaluation was based on the out-of-sample R^2 value.

3.4. Graphical user interface development

This study developed a graphical user interface to the trained models highlighted in 3.3. The backend technology was based on Flask, a backend web application framework written in Python. This allowed data processing tasks to be sent to a front-end user interface for rendering. Dropzone and AJAX were used to accept CSV data and transfer it to the backend in Flask, where Python's pandas package was used to process the CSV data and feed it to the algorithms. The user interface front-end was based on a Bootstrap template with the gentalella theme by colorlib. Charting was based on the chart.js javascript library. Lastly, a mapping tool was based on leaflet.js, which renders maps and overlaying data online. The baseline map was derived from "openstreetmap", which is an open-source mapping tool.

The entire code behind the GUI development is provided online; see the supplementary information for more information. Open-sourcing the code allows for reproducibility and for biogas researchers to extend the interface according to their research/practical needs.

4. Results and discussion

4.1. Machine learning results on the Hainan data

Out of the model tested, the kNN model performed the best on the Hainan dataset, with an R^2 of 0.86 on the training data and 0.85 on the test data. The training and test R^2 value for the five other models that were tested (logistic regression, SVM, perceptron, random forest, and XGBoost), were considerably lower. The low accuracies of the Perceptron model can be explained by the fact that it is a fairly simple linear model that would not map any non-linearities present in the data. Moreover, due to the sporadic changes in input and output values in the biogas facility on a daily basis, the perceptron model is unable to capture such fluctuations.

Given that the kNN model performed the best on the Hainan dataset, the hyperparameters of the model were tuned to determine the number of nearest neighbors that provide the best testing accuracy (R^2). It can be seen that (see supplementary information) as the number of nearest neighbors increased from 1 to 10, the training accuracy reduced and the testing accuracy increased, reaching a peak at 7 nearest neighbors. Using one nearest neighbor ends up overfitting the model, as it results in a very complex model

that ends up capturing the noise in the training data, making the model less generalizable to the test set. A model with one nearest neighbor implies that bias is low, meaning that the model will be very close to the testing data, and that the variance is high, since optimizing on the nearest point means that the probability of the model to capture noise in the data is very high. As the number of nearest neighbors is increased, the model reduces in complexity and after about 7 nearest neighbors, the model may even be underfitting the data, as can be seen by the reduction in model accuracies on both the training as well as the test set. The model is clearly not overfitting the training data with 7 nearest neighbors, as the difference between the training and the test accuracies is very low, indicating that the model is far more generalizable. Fig. 2 provides a feature importance plot that highlights the variables with the strongest effect on the model.

Possible reasons for the kNN model's high accuracy include: (1) there was a sufficient amount of data for model training; (2) None of the independent variables (different waste factors) have disproportionately large variance to confer to the model; (3) kNN is a non-parametric method, meaning that it does not assume an explicit form for $f(x)$ when fitting the data, providing a more flexible approach.

Both SVM and random forest models had very high accuracies on the training set, with accuracies of 0.95 and 0.97 respectively; however the training accuracies were far lower indicating that the models were clearly overfitting the training dataset and were thus less generalizable. As the accuracy of the test set is near 0.5, we have that the both these models are only being able to “randomly guess” whether the output on a given day is high, medium or low. Even after tuning the hyper parameters of the random forest and SVM models, it was obtained that the accuracies of the KNN model on the testing set were far superior to that of the random forest or SVM models. The primary reason for not observing much of an increase in the model accuracies despite hyperparameter tuning was that the data set was too small for any hyperparameter tuning to have much of an impact on the testing accuracy.

4.2. Machine learning on the Shenzhen data

Out of the models tested on the Shenzhen data, the best-performing model was found to be XGBoost, which had an accuracy of 66%. Compared to the Hainan model performance results, the relatively lower prediction accuracy was most likely due to the limited data availability. There were less observations available for model training in Shenzhen project. In addition, the Shenzhen project had less features (originally 14, before feature engineering) than the Hainan project (20, before feature engineering). Model accuracies tend to increase with larger sets of observations and features.

According to (Liew et al., 2014), despite the unique advantages of biofuels, high production costs are preventing them from achieving mass commercialization.

4.3. Graphical user interface

Since the kNN model achieved strong predictive performance on the Hainan project's data, the model was used as the primary model for the GUI for making biogas output predictions. Based on this model, the GUI allows the biogas plant operator to specify the inputs to the model at any point in time, and based on the kNN algorithm, the GUI returns the expected biogas output.

The input to the GUI is designed to be practical for most plant operators, so it accepts a CSV file with the different time points as rows, as well as all available waste input types as columns. This data is then sent to the model in our backend, processed, and returns expected production data as well as various relevant statistics. In addition, the bioCNG output is translated into monetary value using a predetermined conversion rate, which also provides the facility operator with earnings estimates (Fig. 3).

On the GUI, an initial line graph is rendered to show expected revenue for each time point, followed by a second multiline graph to visualize the respective amounts of waste input provided by the user. This is augmented with a bar chart showing the input composition, with the 6 largest inputs plotted for comparison. Below that, basic statistics are provided detailing the model's accuracy, total expected revenue, and total input, followed by a tabulated version of the input data for reference (Fig. 2).

The second component of the user interface (Fig. 4) includes a tool that would be useful for investors and potential operators to identify opportunities for facility construction in China's many regions based on the waste resource availability for each of these regions. Based on 2016 data for restaurant numbers and agricultural production, a waste resource map was generated and rendered in the GUI. The choropleth map distinguishes each region by potential value, as determined by the model created in the first part of this study. Hovering over each region provides more detailed information, including the actual measured waste and production, as well as potential revenue earnings for a bioenergy facility in the area (see Figs. 5 and 6).

4.4. Discussion

A number of challenges remain for the development of novel control strategies in full-scale anaerobic digestion plants (Gaida et al., 2017). These include (1) financially feasible and low-maintenance online-measurement systems for crucial process variables; (2) a high number of different substrates; and (3) successful full-scale application examples. The authors advocate

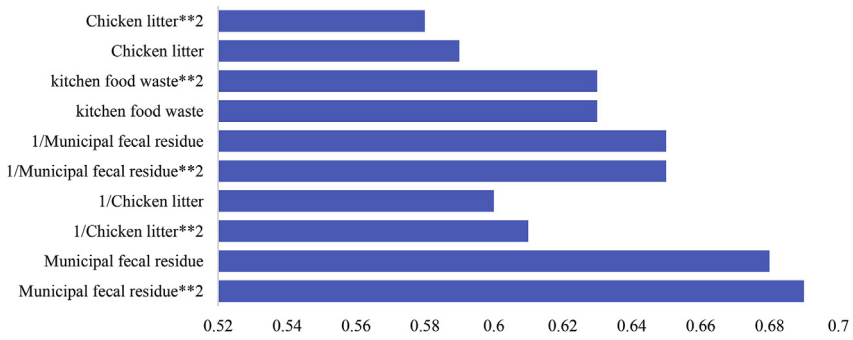

Fig. 2. Feature importance plot for the kNN model trained on the Hainan data.

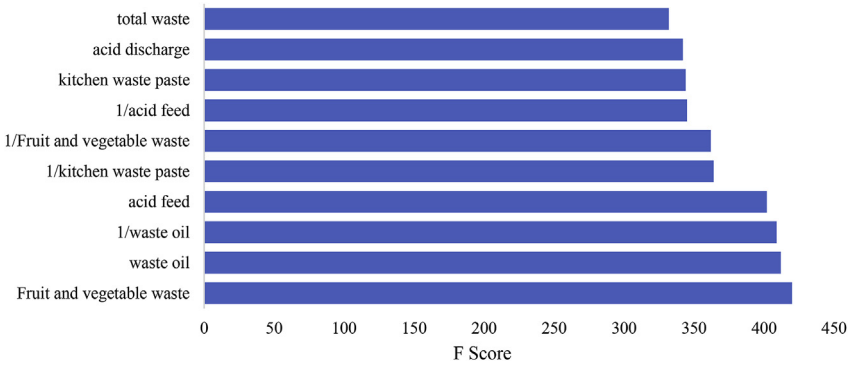

Fig. 3. Feature importance plot for the XGBoost model trained on the Hainan data.

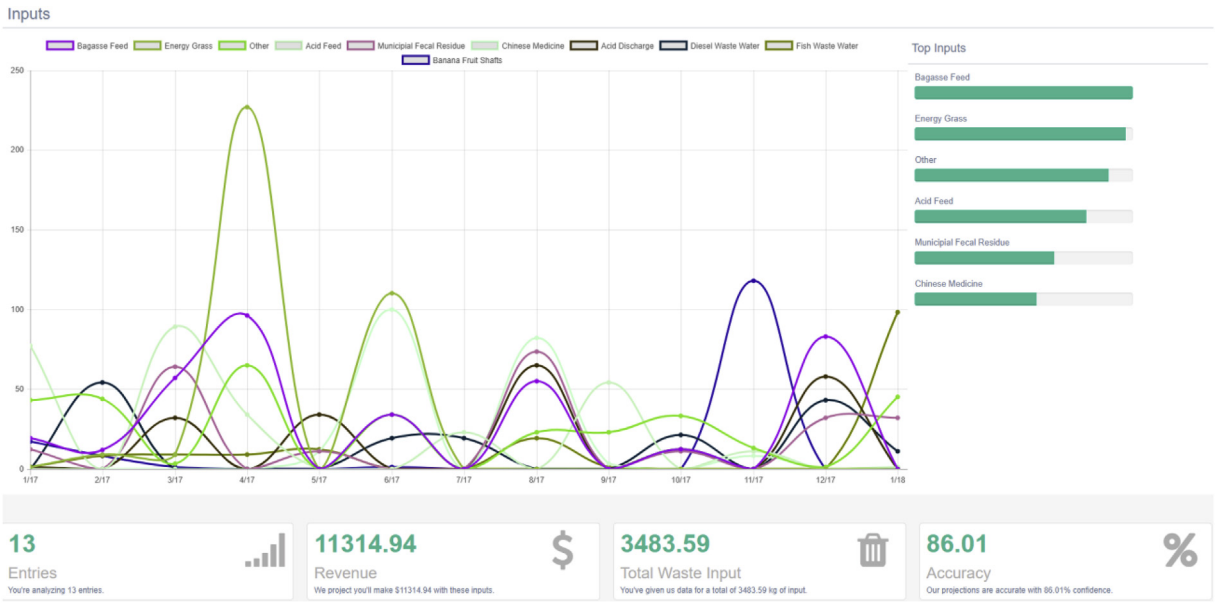

Fig. 4. Visualization of the waste inputs.

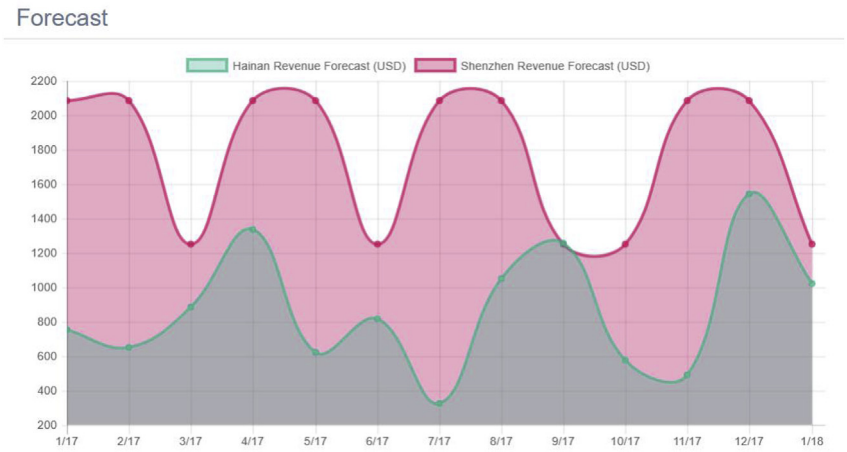

Fig. 5. Estimated biogas output and corresponding revenue forecast. Note: GUI displaying the differing time series profile of revenue predictions for Hainan and Shenzhen biogas facilities as a result of differing quantity and types of waste inputs into both these plants.

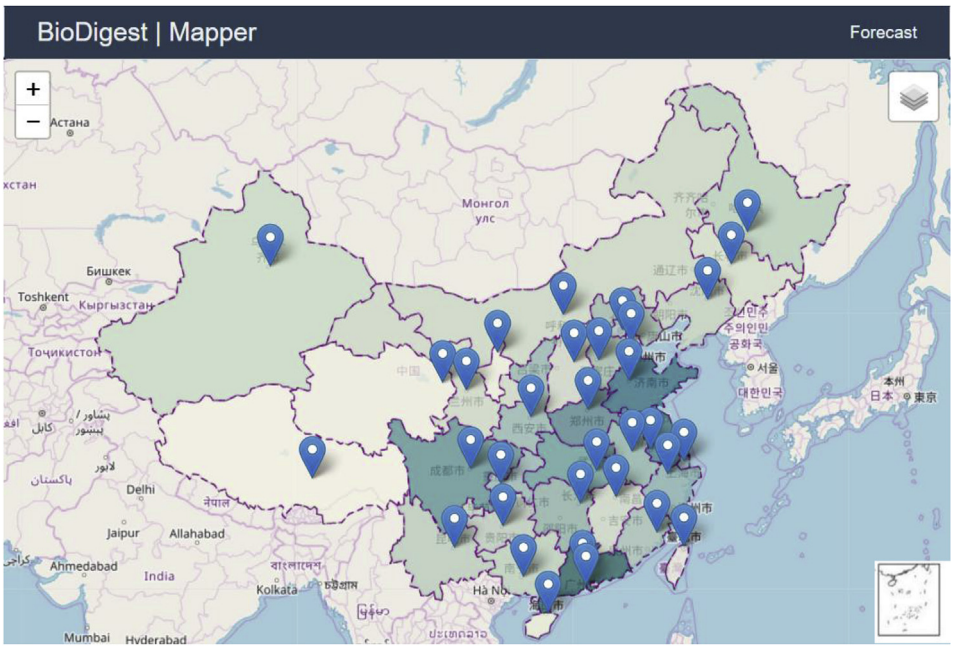

Fig. 6. Biowaste resource availability mapper. Note: Choropleth map of China displaying biogas production potential of each province using data on the information of waste production from each province obtained from China's annual statistical yearbook. Waste production data is converted to biogas production potential using the output from the KNN model.

applying machine learning methods to estimate key process variables as well as substrate composition and its effect on methane potential.

These challenges are especially pertinent in the Chinese context, since Chinese government policy has mainly focused on increasing the quantity of biogas facilities, increasing annual biogas output, expanding government funding, and encouraging private sector funding. Such financial policies offer little incentives for biogas project operators to maintain high operational efficiency after the initial construction phase. Insufficient emphasis has been placed on enhancing project efficiency by incentivizing the biogas sector to adopt cutting-edge tools in machine learning and artificial intelligence. Since biogas investment is constrained, it is crucial that investment into the sector is directed at highly efficient facilities (Gottfried et al., 2018).

Robust data analytics tools should be propagated not only in the biogas process, but also through the entire waste management chain. There have been increasing efforts to this end. For instance, one study (Niska and Serkkola, 2018) proposed novel data analytics for creating waste generation profiles based on container-specific weighing data in Helsinki, Finland. They used self-organizing maps (SOM) and the k-means algorithm for creating a set of waste generation type profiles. Another example of waste value chain analytics is a study (Wen et al., 2018) which designed, implemented, and evaluated a sensor-based Internet of Things (IoT) network technology for improving the management of restaurant waste in Suzhou City, China. Data analytics frameworks such as these, when combined with machine learning tools in final waste treatment methods such as anaerobic digestion, can make for vastly more efficient waste management systems. Given that machine learning tools have rapidly permeated other fields – for instance one study (Idowu et al., 2016) used support vector machine and neural networks to forecast heat load in district heating systems – waste management systems should make these tools a core component of the analytics pipeline.

Although Chinese biogas plants were used as a case study, there is compelling evidence that the results of this study are applicable to biogas projects around the world in both developing and developed regions. For instance, Mustafa et al. (2016) investigated a biogas plant in Norway, and found that there were significant opportunities for operational optimization that could lead in enhanced methane and electricity production (electricity output was just 35% of theoretical potential). In addition significant operational challenges related to process efficiency have been found in many biogas plants in South Korea (Kim et al., 2012). Machine learning-based analytics could alleviate some of these operational issues.

Despite the potential of machine learning to solve these challenges, there are nevertheless a number of limitations to its application that should be resolved in future research.

First, since machine learning algorithms are fueled by historical data, prediction error may increase when new types of organic waste enter a biogas facility. In this case, the machine learning algorithms may need to be supplemented with physical-chemical models while being trained on the new data. In addition, software tools which can incorporate constant retraining and algorithm optimization based on newly available data would make the tool more robust to frequent input fluctuation.

Secondly, the technological composition, input wastes, and output products in different biogas facilities can vary tremendously. This implies difficulty in building a one-size-fits all machine learning software tool that can be applied to solving operational bottlenecks across many different projects. Future research should focus on identifying the most common operational challenges faced by multiple projects around the world in order to continue the development of an open-source, general-purpose machine learning GUI software tool.

Thirdly, the uptake of such tools should be promoted by governments, biogas project investors, and industry associations. Many biogas projects around the world depend on public funding to subsidize operation, and it is imperative that money allocated towards biogas facilities is used efficiently. Together with lean project

management, well-informed technology choice, and well-defined incentive policy mechanisms, general purpose machine learning tools can contribute to achieving enhanced efficiency.

5. Conclusion

This study trained a number of machine learning models capable of predicting biogas output in industrial scale facilities based on fluctuating values of biowaste input. Moreover, a user interface to these predictive models was developed to facilitate forecasting for biogas project operators. The research is a departure from previous literature, which has tended to focus on lab-scale experimentation and traditional biogas models such ADM1 for evaluating the effect of biowaste input on biogas production.

We trained a number of models on industrial production data from industrial biowaste-to-biogas projects in Hainan and Shenzhen. The models trained included logistic regression, k-nearest neighbors, random forest, perceptron, and random forest. The kNN model was found to have the best performance, with an out-of-sample R^2 value 0f 0.87. In addition, a graphical user interface to the underlying machine learning models was developed in order to allow for model-based prediction of biogas output given a set of biowaste inputs. A prototype of this user interface tool is available online, and the code behind it was open-sourced to allow biogas researchers to extend upon it.

As mentioned in the discussion. Machine learning models are not panacea to the operational troubles found in industrial scale facilities. However, such models, if incorporated into the analytics systems of projects, can greatly facilitate the decision-making process.

Acknowledgements

The authors gratefully acknowledge the financial support from the "Thirteenth Five-Year" National Key R&D Program of China (2018YFC1903002) and General Programs of the National Natural Science Foundation of China (71774099, 71825006). The responsibility for any error rests solely with the authors. The contents of this paper reflect the views of the authors and do not necessarily indicate acceptance by the sponsors. We would also like to thank Ikhlaq Sidhu and Alexander Fred Ojala for their outstanding instruction in data science at UC Berkeley.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2019.01.031.

References

Alcaraz-González, V., Harmand, J., Rapaport, A., Steyer, J.P., González-Álvarez, V., Pelayo-Ortiz, C., 2005. Robust interval-based regulation for anaerobic digestion processes. Water Sci. Technol. 52 (1–2), 449–456.

Appels, L., Assche, A. Van, Willems, K., Degrève, J., Impe, J. Van, Dewil, R., 2011. Peracetic acid oxidation as an alternative pre-treatment for the anaerobic digestion of waste activated sludge. Bioresour. Technol. 102, 4124–4130. https://doi.org/https://doi.org/10.1016/j.biortech.2010.12.070.

Barnett, M.W., Andrews, J.F., 1992. Expert system for anaerobic-digestion-process operation. J. Environ. Eng. (United States) 118, 949–963. https://doi.org/10.1061/(ASCE)0733-9372(1992)118:6(949).

Bensmann, A., Hanke-Rauschenbach, R., Heyer, R., Kohrs, F., Benndorf, D., Kausmann, R., Plöchl, M., Heiermann, M., Reichl, U., Sundmacher, K., 2016. Diagnostic concept for dynamically operated biogas production plants. Renew. Energy 96, 479–489. https://doi.org/https://doi.org/10.1016/j.renene.2016.04.098.

Cabbai, V., Ballico, M., Aneggi, E., Goi, D., 2013. BMP tests of source selected OFMSW to evaluate anaerobic codigestion with sewage sludge. Waste Manag. 33, 1626–1632. https://doi.org/https://doi.org/10.1016/j.wasman.2013.03.020.

Cabbai, V., De Bortoli, N., Goi, D., 2016. Pilot plant experience on anaerobic codigestion of source selected OFMSW and sewage sludge. Waste Manag. 49, 47–54. https://doi.org/https://doi.org/10.1016/j.wasman.2015.12.014.

Chen, L., Cong, R.-G., Shu, B., Mi, Z.-F., 2017. A sustainable biogas model in China: The case study of Beijing Deqingyuan biogas project. Renew. Sustain. Energy Rev. 78, 773–779. https://doi.org/https://doi.org/10.1016/j.rser.2017.05.027.

Chen, Q., Liu, T., 2017. Biogas system in rural China: Upgrading from decentralized to centralized? Renew. Sustain. Energy Rev. 78, 933–944. https://doi.org/https://doi.org/10.1016/j.rser.2017.04.113.

Chynoweth, D.P., Svoronos, S.A., Lyberatos, G., Harman, J.L., Pullammanappallil, P., Owens, J.M., Peck, M.J., 1994. Real-time expert system control of anaerobic digestion. Water Sci. Technol. 30, 21–29.

Cook, S.M., Skerlos, S.J., Raskin, L., Love, N.G., 2017. A stability assessment tool for anaerobic codigestion. Water Res. 112, 19–28. https://doi.org/https://doi.org/10.1016/j.watres.2017.01.027.

Corro, G., Paniagua, L., Pal, U., Bañuelos, F., Rosas, M., 2013. Generation of biogas from coffee-pulp and cow-dung co-digestion: Infrared studies of post-combustion emissions. Energy Convers. Manag. 74, 471–481. https://doi.org/https://doi.org/10.1016/j.enconman.2013.07.017.

De Clercq, D., Wen, Z., Caicedo, L., Cao, X., Fan, F., Xu, R., 2017a. Application of DEA and statistical inference to model the determinants of biomethane production efficiency: A case study in south China. Appl. Energy 205, 1231–1243. https://doi.org/https://doi.org/10.1016/j.apenergy.2017.08.111.

De Clercq, D., Wen, Z., Fan, F., 2017b. Performance evaluation of restaurant food waste and biowaste to biogas pilot projects in China and implications for national policy. J. Environ. Manag. 189, 115–124. https://doi.org/https://doi.org/10.1016/j.jenvman.2016.12.030.

De Clercq, D., Wen, Z., Fan, F., Caicedo, L., 2016. Biomethane production potential from restaurant food waste in megacities and project level-bottlenecks: A case study in Beijing. Renew. Sustain. Energy Rev. 59, 1676–1685. https://doi.org/https://doi.org/10.1016/j.rser.2015.12.323.

De Clercq, D., Wen, Z., Fei, F., 2017c. Economic performance evaluation of bio-waste treatment technology at the facility level. Resour. Conserv. Recycl. 116, 178–184. https://doi.org/https://doi.org/10.1016/j.resconrec.2016.09.031.

De Clercq, D., Wen, Z., Gottfried, O., Schmidt, F., Fei, F., 2017d. A review of global strategies promoting the conversion of food waste to bioenergy via anaerobic digestion. Renew. Sustain. Energy Rev. 79, 204–221. https://doi.org/https://doi.org/10.1016/j.rser.2017.05.047.

De Clercq, D., Wen, Z., Lu, X., Caicedo, L., Cao, X., Fan, F., 2017e. Determinants of efficiency in an industrial-scale anaerobic digestion food waste-to-biogas project in an Asian megacity based on data envelopment analysis and exploratory multivariate statistics. J. Clean. Prod. 168, 983–996. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.09.062.

Deng, L., Liu, Y., Zheng, D., Wang, L., Pu, X., Song, L., Wang, Z., Lei, Y., Chen, Z., Long, Y., 2017. Application and development of biogas technology for the treatment of waste in China. Renew. Sustain. Energy Rev. 70, 845–851. https://doi.org/https://doi.org/10.1016/j.rser.2016.11.265.

Ekström, M., Esseen, P.-A., Westerlund, B., Grafström, A., Jonsson, B.G., Ståhl, G., 2018. Logistic regression for clustered data from environmental monitoring programs. Ecol. Inf. 43, 165–173. https://doi.org/https://doi.org/10.1016/j.ecoinf.2017.10.006.

Fedailaine, M., Moussi, K., Khitous, M., Abada, S., Saber, M., Tirichine, N., 2015. Modeling of the anaerobic digestion of organic waste for biogas production. Procedia Comput. Sci. 52, 730–737. In: https://doi.org/https://doi.org/10.1016/j.procs.2015.05.086.

Gaida, D., Wolf, C., Bongards, M., 2017. Feed control of anaerobic digestion processes for renewable energy production: A review. Renew. Sustain. Energy Rev. 68, 869–875. https://doi.org/https://doi.org/10.1016/j.rser.2016.06.096.

Garfí, M., Ferrer-Martí, L., Perez, I., Flotats, X., Ferrer, I., 2011. Codigestion of cow and guinea pig manure in low-cost tubular digesters at high altitude. Ecol. Eng. 37, 2066–2070. https://doi.org/https://doi.org/10.1016/j.ecoleng.2011.08.018.

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., Villa-Vialaneix, N., 2017. Random forests for big data. Big Data Res. 9, 28–46. https://doi.org/https://doi.org/10.1016/j.bdr.2017.07.003.

Gottfried, O., De Clercq, D., Blair, E., Weng, X., Wang, C., 2018. SWOT-AHP-TOWS analysis of private investment behavior in the Chinese biogas sector. J. Clean. Prod. 184, 632–647. https://doi.org/https://doi.org/10.1016/j.jclepro.2018.02.173.

Grim, J., Nilsson, D., Hansson, P.-A., Nordberg, Å., 2015. Demand-orientated power production from biogas: modeling and simulations under swedish conditions. Energy Fuel. 29, 4066–4075. https://doi.org/10.1021/ef502778u.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, second ed. Springer-Verlag, New York, NY, USA https://doi.org/10.1007/978-0-387-84858-7.

Holubar, P., Zani, L., Hager, M., Fröschl, W., Radak, Z., Braun, R., 2003. Start-up and recovery of a biogas-reactor using a hierarchical neural network-based control tool. J. Chem. Technol. Biotechnol. 78, 847–854. https://doi.org/10.1002/jctb.854.

Huang, T., Li, B., Shen, D., Cao, J., Mao, B., 2017. Analysis of the grain loss in harvest based on logistic regression. Procedia Comput. Sci. 122, 698–705. In: https://doi.org/https://doi.org/10.1016/j.procs.2017.11.426.

Idowu, S., Saguna, S., Åhlund, C., Schelén, O., 2016. Applied machine learning: Forecasting heat load in district heating system. Energy Build. 133, 478–488. https://doi.org/https://doi.org/10.1016/j.enbuild.2016.09.068.

Jacob, S., Banerjee, R., 2016. Modeling and optimization of anaerobic codigestion of potato waste and aquatic weed by response surface methodology and artificial

neural network coupled genetic algorithm. Bioresour. Technol. 214, 386–395. https://doi.org/https://doi.org/10.1016/j.biortech.2016.04.068.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, first ed. Springer-Verlag, New York, NY, USA https://doi.org/10.1007/978-1-4614-7138-7.
Kim, Y.-S., Yoon, Y.-M., Kim, C.-H., Giersdorf, J., 2012. Status of biogas technologies and policies in South Korea. Renew. Sustain. Energy Rev. 16, 3430–3438. https://doi.org/https://doi.org/10.1016/j.rser.2012.02.075.
Li, W., Loh, K.-C., Zhang, J., Tong, Y.W., Dai, Y., 2018. Two-stage anaerobic digestion of food waste and horticultural waste in high-solid system. Appl. Energy 209, 400–408. https://doi.org/https://doi.org/10.1016/j.apenergy.2017.05.042.
Liew, W.H., Hassim, M.H., Ng, D.K.S., 2014. Review of evolution, technology and sustainability assessments of biofuel production. J. Clean. Prod. 71, 11–29. https://doi.org/https://doi.org/10.1016/j.jclepro.2014.01.006.
Matuszewska, A., Owczuk, M., Zamojska-Jaroszewicz, A., Jakubiak-Lasocka, J., Lasocki, J., Orliński, P., 2016. Evaluation of the biological methane potential of various feedstock for the production of biogas to supply agricultural tractors. Energy Convers. Manag. 125, 309–319. https://doi.org/https://doi.org/10.1016/j.enconman.2016.02.072.
Méndez-Acosta, H.O., Steyer, J.P., Femat, R., González-Alvarez, V., 2007. Robust nonlinear control of a pilot-scale anaerobic digester. Lect. Notes Contr. Inf. Sci. https://doi.org/10.1007/978-3-540-73188-7_6.
Miao, H., Wang, S., Zhao, M., Huang, Z., Ren, H., Yan, Q., Ruan, W., 2014. Codigestion of Taihu blue algae with swine manure for biogas production. Energy Convers. Manag. 77, 643–649. https://doi.org/https://doi.org/10.1016/j.enconman.2013.10.025.
Mu, S., Zeng, Y., Wu, P., 2008. Multivariable control of anaerobic reactor by using external recirculation and bypass ratio. J. Chem. Technol. Biotechnol. 83, 892–903. https://doi.org/10.1002/jctb.1888.
Mustafa, M.Y., Calay, R.K., Román, E., 2016. Biogas from organic waste - a case study. Procedia Eng. 146, 310–317. https://doi.org/https://doi.org/10.1016/j.proeng.2016.06.397.
Naroznova, I., Møller, J., Scheutz, C., 2016. Global warming potential of material fractions occurring in source-separated organic household waste treated by anaerobic digestion or incineration under different framework conditions. Waste Manag. 58, 397–407. https://doi.org/https://doi.org/10.1016/j.wasman.2016.08.020.
Niska, H., Serkkola, A., 2018. Data analytics approach to create waste generation profiles for waste management and collection. Waste Manag. https://doi.org/https://doi.org/10.1016/j.wasman.2018.04.033.
Sathish, S., Vivekanandan, S., 2016. Parametric optimization for floating drum anaerobic bio-digester using Response Surface Methodology and Artificial Neural Network. Alexandria Eng. J. 55, 3297–3307. https://doi.org/https://doi.org/10.1016/j.aej.2016.08.010.
Scarlat, N., Dallemand, J.-F., Fahl, F., 2018. Biogas: developments and perspectives in Europe. Renew. Energy. https://doi.org/https://doi.org/10.1016/j.renene.2018.03.006.
Shen, X., Niu, L., Qi, Z., Tian, Y., 2017. Support vector machine classifier with truncated pinball loss. Pattern Recogn. 68, 199–210. https://doi.org/https://doi.org/10.1016/j.patcog.2017.03.011.
Song, Z., Zhang, C., 2015. Anaerobic codigestion of pretreated wheat straw with cattle manure and analysis of the microbial community. Bioresour. Technol. 186, 128–135. https://doi.org/https://doi.org/10.1016/j.biortech.2015.03.028.
Terradas-lll, G., Pham, C.H., Triolo, J.M., Martí-Herrero, J., Sommer, S.G., 2014. Thermic model to predict biogas production in unheated fixed-dome digesters buried in the ground. Environ. Sci. Technol. 48, 3253–3262. https://doi.org/10.1021/es403215w.
Thyberg, K.L., Tonjes, D.J., 2017. The environmental impacts of alternative food waste treatment technologies in the U.S. J. Clean. Prod. 158, 101–108. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.04.169.
Wang, X., An, K., Tang, L., Chen, X., 2015. Short Term Prediction of Freeway Exiting Volume Based on SVM and KNN. Int. J. Transp. Sci. Technol. 4, 337–352. https://doi.org/https://doi.org/10.1260/2046-0430.4.3.337.
Wen, Z., Hu, S., De Clercq, D., Beck, M.B., Zhang, H., Zhang, H., Fei, F., Liu, J., 2018. Design, implementation, and evaluation of an Internet of Things (IoT) network system for restaurant food waste management. Waste Manag. 73, 26–38. https://doi.org/https://doi.org/10.1016/j.wasman.2017.11.054.
Xia, Y., Liu, C., Li, Y., Liu, N., 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. Expert Syst. Appl. 78, 225–241. https://doi.org/https://doi.org/10.1016/j.eswa.2017.02.017.
Xing, W., Chen, X., Zuo, J., Wang, C., Lin, J., Wang, K., 2014. A half-submerged integrated two-phase anaerobic reactor for agricultural solid waste codigestion. Biochem. Eng. J. 88, 19–25. https://doi.org/https://doi.org/10.1016/j.bej.2014.03.016.
Yang, Y., Loog, M., 2018. A benchmark and comparison of active learning for logistic regression. Pattern Recogn. 83, 401–415. https://doi.org/https://doi.org/10.1016/j.patcog.2018.06.004.
Zaefferer, M., Gaida, D., Bartz-Beielstein, T., 2016. Multi-fidelity modeling and optimization of biogas plants. Appl. Soft Comput. 48, 13–28. https://doi.org/https://doi.org/10.1016/j.asoc.2016.05.047.
Zhang, C., Li, J., Liu, C., Liu, X., Wang, J., Li, S., Fan, G., Zhang, L., 2013. Alkaline pretreatment for enhancement of biogas production from banana stem and swine manure by anaerobic codigestion. Bioresour. Technol. 149, 353–358. https://doi.org/https://doi.org/10.1016/j.biortech.2013.09.070.