

Forecasting Sensor Values in Waste-To-Fuel Plants: a Case Study.

Bor Breclj*
University of Ljubljana, Faculty of
Mathematics and Physics
Ljubljana, Slovenia
bor.brecelj@gmail.com

Beno Šircelj*
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
beno.sircelj@ijs.si

Jože M. Rožanec
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
joze.rozanec@ijs.si

Blaž Fortuna
Qlector d.o.o.
Ljubljana, Slovenia
blaz.fortuna@qlector.com

Dunja Mladenić
Jožef Stefan Institute
Ljubljana, Slovenia
dunja.mladenic@ijs.si

ABSTRACT

In this research, we develop machine learning models to predict future sensor readings of a waste-to-fuel plant, which would enable proactive control of the plant's operations. We developed models that predict sensor readings for 30 and 60 minutes into the future. The models were trained using historical data, and predictions were made based on sensor readings taken at a specific time. We compare three types of models: (a) a naïve prediction that considers only the last predicted value, (b) neural networks that make predictions based on past sensor data (we consider different time window sizes for making a prediction), and (c) a gradient boosted tree regressor created with a set of features that we developed. We developed and tested our models on a real-world use case at a waste-to-fuel plant in Canada. We found that approach (c) provided the best results, while approach (b) provided mixed results and was not able to outperform the naïve consistently.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Applied computing**;

KEYWORDS

Smart Manufacturing, Machine Learning, Feature Engineering

ACM Reference Format:

Bor Breclj, Beno Šircelj, Jože M. Rožanec, Blaž Fortuna, and Dunja Mladenić. 2022. Forecasting Sensor Values in Waste-To-Fuel Plants: a Case Study.. In *Ljubljana '22: Slovenian KDD Conference on Data Mining and Data Warehouses, October, 2022, Ljubljana, Slovenia*. ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

There is a wide range of applications of ML (machine learning). One of them is the modeling and control of chemical processes, such as the production of biodiesel. Introducing machine learning

*Both authors contributed equally to this research.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SiKDD '22, October, 2022, Ljubljana, Slovenia
© 2022 Copyright held by the owner/author(s).

to such processes can improve quality and yield and help engineers predict anomalies to control the factory better.

We modeled the JEMS waste-to-fuel plant, which produces high-quality diesel from organic waste. The plant has numerous sensors that measure temperature, and pressure, among other variables. It is operated by experts who must control the process. Since the chemical process is complex and, therefore, difficult to control, we built forecasting models that can predict future sensor readings based on historical data and the current state of the plant.

The model will be used to give plant operators additional information about the future state of the plant, which will allow them to make an informed decision about changing the plant's parameters and, therefore, adjust the process before it is too late.

2 RELATED WORK

Organic wastes in energy conversion technologies are an active area of research aimed at reducing dependence on fossil fuels, optimizing production costs, improving waste management, and controlling emissions. Biochemical, physiochemical, and thermochemical processes produce different biofuels, such as bio-methanation, bio-hydrogen, biodiesel, ethanol, syngas, and coal-like fuels, which are studied by Stephen et al. [8]. Work is also being done on optimization, such as catalyst selection, reactor design, pyrolysis temperature, and other important factors [5].

Many ML methods have been developed to address waste management and proper processing for biofuel production, focusing on energy demand and supply prediction [3]. Aghbashlo et al. [2] provided a systematic review of various applications of ML technology with a focus on ANN (Artificial Neural Network) in biodiesel research. They provided an overview of the use of ML in modeling, optimization, monitoring, and process control. Models that predict the conditions of the biofuel production process that have the highest yield were created by Kusumo et al. [6] and Abdelbasset et al. [1]. The models used in these studies were kernel-based extreme learning machines, ANN, and various ensemble models.

3 USE CASE

The JEMS waste-to-fuel plant produces synthetic diesel (SynDi) from any hydrocarbon-based waste, such as wood, biomass, paper, waste fuels and oils, plastics, textiles, rubber, and agricultural residues. The plant uses a chemical-catalytic de-polymerization process, the advantage of which is that the temperature is too low

to produce carcinogenic gasses. It operates continuously and produces about 150 liters of fuel per hour. Although it uses the latest software available and allows remote control, there is no anomaly detection, prediction, or optimization. As a result, there is a great need for better understanding, optimization, and decision-making, given data availability. The company plans to sell and install over 1.500 SynDi systems over the next ten years. In practice, this means many SynDi plants in different locations worldwide.

There are three main chambers in the pipeline, which are named B100, B200, and B300. The plant can be conceptually split into four stages

- (1) Feedstock inspecting and feeding;
- (2) Drying and mixing (chamber B100);
- (3) Processing (chamber B200);
- (4) Distilling (chamber B300).

Since there are no sensors in the feedstock inspecting and feeding stage, we focused on the later stages, each of which takes place in one of the main chambers.

In the drying and mixing stage (B100), the starting material is mixed with process oil, lime, and catalyst and is heated. During mixing, the material is broken down into smaller particles, and the water is evaporated. The primary chemical reaction occurs in the processing stage (B200). The material is fed to a turbine, and the reaction product evaporates through the diesel distillation column. If the diesel obtained is not of sufficient quality, it is redistilled in the second distillation stage (B300).

Currently, the plants are operated with highly skilled personnel and high costs for personnel training. Implementing automation, remote control, optimization, and interconnection among the plants would greatly facilitate their operation. Therefore, the main challenge to be solved by integrating AI is the self-control of the chemical process and the plant itself by minimizing the human resources required to operate the plants. Furthermore, operating many SynDi plants also means a significant challenge for ensuring remote control for troubleshooting, maintenance, and repair. AI integration aims to minimize the workforce required to operate the plants, minimize the resulting downtime due to human interaction, enable self-control and predictive maintenance of the SynDi plants, and achieve less downtime and higher production efficiency.

In modeling the waste-to-fuel processes, we decided to model each chamber separately. No model was developed for chamber B300 because it was not active during the period for which we obtained the data. As described above, a second distillation of the fuel is performed in chamber B300 only if the fuel in chamber B200 is not pure enough.

4 METHODOLOGY

4.1 Data analysis

The sensor measurements are from the experimental JEMS plant, which is located in Canada. The data consists of 154 sensors from January 2016 to January 2017. The measurements are taken at one-minute intervals and mostly measure temperature or pressure, but there are also sensors for motor current and valve position, among others. Since the data is from the prototype version of the waste-to-fuel plant, it contains many missing values. Our data set contained an average of 61.607 data points per sensor. We discarded

all sensors with less than 6.000 data points and kept only those that corresponded to chambers B100 and B200, giving us data from 39 sensors.

Analysis of the dataset we received revealed that many values were missing. In particular, we noted that there were day-long intervals with a tiny number of measurements. We also noticed that specific sensor values remained constant at low temperatures - a condition best described by the waste-to-fuel plant's inactivity. We, therefore, decided to remove such values. Because there were many ten-minute gaps, we decided to resample the data at fifteen-minute intervals, taking the last value of each interval and assuming that conditions had not changed in the short time since the last measurement - a reasonable assumption for sensor values. The resulting data set contained an average of 7.884 data points per sensor.

We divided the dataset into a train and a test dataset, split on October 31st 2016. The resulting train set included a total of 11.000 samples, and the test set included 3.000 samples.

4.2 Model training

In this research, we compare models that we develop using two different approaches. We first tried the neural network approach, in which the model makes predictions based only on sensor readings from the last five hours. Since the model did not perform better than the baseline, we began the second approach, developing features to describe the time series and capture its patterns. We used linear regression and gradient-boosted tree regressor. All the developed models were compared with the last-value model, which we used as a benchmark.

4.2.1 Neural network approach. We used the model developed for forecasting Tüpras' sensor values. Tüpras is an oil refinery, which is very similar to the JEMS use case. The model was used to forecast sensor values in different units of LPG production. Some of Tüpras' units are distillation columns, similar to JEMS' chamber B200. The model takes only past sensor values as input and predicts values for the future together with the prediction interval. More specifically, it predicts 10th, 50th and 90th percentile, which is the case in all our models that give prediction interval.

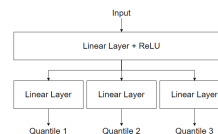


Figure 1: Architecture of the neural network model, which gives the prediction interval.

Figure 1 shows the architecture of the neural network. The model is a feedforward neural network with two layers. First, there is a linear layer with ReLU activation. The second layer has a separate linear layer for each quantile. The hidden dimension of the model is calculated from the number of features and the number of targets using the formula $\lfloor \frac{n_{features}}{2} \rfloor + n_{targets}$.

During training, we used the quantile loss function, which is defined as

$$\max \left\{ q \cdot (y_{\text{true}} - y_{\text{pred}}), (1 - q) \cdot (y_{\text{pred}} - y_{\text{true}}) \right\},$$

where q is the observed quantile (in our case, it can be 0.1, 0.5 or 0.9), y_{true} is the true target value and y_{pred} is the corresponding quantile of the prediction. In the case of $q = 0.5$, the loss is equal to the mean absolute error divided by two. When calculating the loss of 10th percentile ($q = 0.1$), a prediction that is greater than the true value is heavily penalized, while a prediction that is lower than the true value has a smaller loss and is therefore encouraged.

The model is implemented in the PyTorch library [7]. Since sensors measure different quantities, the values have to be scaled before learning. Here we used Min-Max scaler from the scikit-learn library, scaling all values between zero and one.

4.2.2 Feature engineering. The neural network model described above did not outperform the benchmark model. As a result, we decided to try another approach, where we developed features that better describe past sensor values and capture their patterns. One of the problems of the neural network model was that it had too many features. We decided to build a separate model for each sensor to tackle this problem. Each model uses only features calculated from the values of the sensor being predicted.

With the help of plant operators, we decided to consider at most five hours of data before the prediction point to issue a forecast. Since the latest data is usually more important in determining future sensor values, we created features on seven different time windows: 30, 45, 75, 120, 180, 240, and 300 minutes. For each time window, we computed the following features:

- average sensor value,
- fraction of peaks in the window,
- percentage change between first and last value in the time window,
- slope (coefficient of the least squares line through the points in the window),
- simple prediction (extension of the least squares line to the future),
- slope ratio (slope on the smaller window divided by the slope on the bigger window).

Besides features mentioned above, which depend on the window size, we also included features that were calculated only on the biggest time window (300 minutes):

- last value,
- maximal value,
- last value relative to the maximal value.

The features above attempt to capture different time series characteristics:

- *trend*: described by percentage change and slope;
- *growth pattern*: described by the fraction of peaks, which indicate whether the growth is steady or it has ups-and-downs. Furthermore, the slope indicates how aggressive such growth is;
- *expected value*: an approximation of the expected value is given through the average, last value, maximal value, and simple prediction.

Using developed features, we trained a linear regression model, and a gradient boosted tree regressor from the CatBoost library [4]. We used root mean squared error (RMSE) for the loss function.

5 RESULTS AND ANALYSIS

We built models for main chambers B100 and B200 with two forecasting horizons (30 and 60 minutes). Tables 1 and 2 show mean squared error (MSE) and mean absolute error (MAE) on chambers B100 and B200, respectively. There are three different neural network models (NN), which differ in the size of the window from which it gets the data.

	horizon = 30min		horizon = 60min	
	MSE	MAE	MSE	MAE
last-value model	21.0533	1.4320	50.6636	2.5128
NN, window = 5h	21.7525	1.6512	47.0545	2.5413
NN, window = 3h	19.7441	1.6109	45.3450	2.4127
NN, window = 2h	18.9717	1.6023	46.5047	2.5357
Linear regression	19.4264	1.4634	49.2268	2.5145
Catboost	16.9030	1.4478	38.3066	2.3164

Table 1: MSE and MAE on the test set of models when predicting for chamber B100.

	horizon = 30min		horizon = 60min	
	MSE	MAE	MSE	MAE
last-value model	52.3380	2.0577	124.9735	3.3768
NN, window = 5h	69.4678	3.8227	129.0330	4.9927
NN, window = 3h	57.9902	3.3601	121.1315	4.7431
NN, window = 2h	55.8769	3.1797	117.4154	4.7146
Linear regression	55.0218	3.2293	115.7457	4.5888
Catboost	49.3329	2.5305	109.5303	3.9745

Table 2: MSE and MAE on the test set of models when predicting for chamber B200.

From the tables 1 and 2 we can see that the five-hour window's neural network performed worse than the benchmark. The main reason for such poor results was too many features for the amount of data that we have. More precisely, the neural network model uses the values of all sensors in the chamber we are predicting. This means that there are six hundred features resulting in more than two hundred thousand trainable parameters for the model of chamber B200. We also have to consider that the neural network predicts future sensor values and prediction intervals. Therefore, there are too many features and target values for the amount of data that we have.

We included results of two more neural network models with three hours and two-hour windows since reduced window size results in a smaller number of features and trainable parameters. For example, the neural network model with a two-hour time window for chamber B200 had two hundred and forty features and almost fifty thousand trainable parameters. Neural network models with smaller window sizes performed better, which confirms that we had too many features.

The features that we developed using the second approach were used with two models, linear regression and the Catboost model. Comparing those two models, the Catboost model performed better because it can capture more than just linear relationships between the features and the target. The Catboost model also outperformed the neural networks, where one of the main differences is that the neural network uses all sensors from the chamber while the Catboost model uses only sensor values of the sensor which is being predicted. This results in forty-five features for the model that predicts one sensor, which solves the problem of too many features. In addition, the Catboost model produced better results than the benchmark when comparing the mean squared error (MSE). During the training, we used RMSE as a loss function, meaning that RMSE was minimized and, therefore, also MSE.

The tables show that although most models outperform the benchmark regarding MSE, almost all of them do not surpass the benchmark when considering MAE. When measuring MSE, predictions with strong spikes where such spikes do not take place are penalized more. Therefore, models with a competitive MSE are considered to rarely predict spikes when such spikes do not take place. This is a key feature for our use case, given that we are interested to understand whether an irregularity will take place or not. Therefore, the models give valuable information even though the average prediction is not entirely accurate.

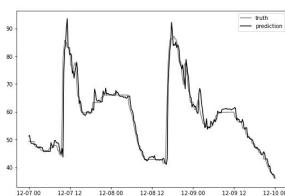


Figure 2: True value and prediction of the Catboost model for a temperature sensor in chamber B100.

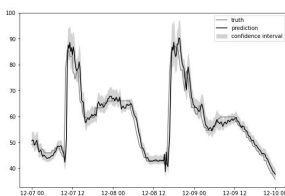


Figure 3: True value and prediction with a confidence interval of the neural network model with a two-hour window for a temperature sensor in chamber B100.

Figure 2 shows the Catboost model prediction on the test set together with the true values of the temperature sensor in chamber B100. The neural network model's prediction of the same sensor is presented in Figure 3. Since the neural network model also outputs prediction interval, it is shown in the abovementioned Figure.

From the plots, we can see that both models can closely predict future sensor values. In the case of the neural network model, the actual value is mainly inside the predicted confidence interval, except when there is a significant change in the sensor value.

However, there is no problem with models not being able to predict significant changes resulting from a manual change in plant setpoint parameters, which our data does not capture. Overall, we consider the best model was the Catboost model, given in all cases it outperformed the rest of the models when considering MSE, and also achieved the best MAE when predicting chamber B100 with a time horizon of 60 minutes.

6 CONCLUSION

We compared a set of models to predict sensor values for a waste-to-fuel plant: a neural network, linear regression, gradient-boosted tree regressor, and the last-value model. The last-value model was used as a benchmark. We developed three neural network models which were different in time window size. The neural network models were built based on the hypothesis that a simple neural network and raw sensor readings as features are enough to model the process. The results showed that this is not the case because the process is too complicated for the amount of data that we obtained. Lastly, we used feature engineering to develop features that better describe the time series. Features were used for learning linear regression, and the gradient boosted tree regressor, where the latter produced the best results in our case.

ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency and the European Union's Horizon 2020 program project FACTLOG under grant agreement number H2020-869951.

REFERENCES

- [1] Walid Kamal Abdelbasset, Safaa M Elkholi, Maria Jade Catalan Opulencia, Tazed-dinova Diana, Chia-Hung Su, May Alashwal, Mohammed Zwawi, Mohammed Algarni, Anas Abdelrahman, and Hoang Chinh Nguyen. 2022. Development of multiple machine-learning computational techniques for optimization of heterogeneous catalytic biodiesel production from waste vegetable oil. *Arabian Journal of Chemistry* 15, 6 (2022), 103843.
- [2] Mortaza Aghbashlo, Wanxi Peng, Meisam Tabatabaei, Soteris A Kalogirou, Salman Soltanian, Homa Hosseinzadeh-Bandbafha, Omid Mahian, and Su Shiung Lam. 2021. Machine learning technology in biodiesel research: A review. *Progress in Energy and Combustion Science* 85 (2021), 100904.
- [3] Hemal Chowdhury, Tamal Chowdhury, Pranta Barua, Salman Rahman, Nazia Hossain, and Anish Khan. 2021. *Biofuel production from food waste biomass and application of machine learning for process management*. 96–117. <https://doi.org/10.1016/B978-0-12-823139-5.00004-6>
- [4] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. 2018. CatBoost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363* (2018).
- [5] Bidhya Kunwar, HN Cheng, Sriram R Chandrashekar, and Brajendra K Sharma. 2016. Plastics to fuel: a review. *Renewable and Sustainable Energy Reviews* 54 (2016), 421–428.
- [6] F Kusumo, AS Silitonga, HH Masjuki, Hwai Chyuan Ong, J Siswantoro, and TMI Mahlia. 2017. Optimization of transesterification process for Ceiba pentandra oil: A comparative study between kernel-based extreme learning machine and artificial neural networks. *Energy* 134 (2017), 24–34.
- [7] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc., 8024–8035. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [8] Jilu Lizy Stephen and Balasubramanian Periyasamy. 2018. Innovative developments in biofuels production from organic waste materials: a review. *Fuel* 214 (2018), 623–633.