Monitoring and control of polymer production line based on machine learning

S Abdurakipov1,2
1Kutateladze Institute of Thermophysics SB RAS, Novosibirsk, Russia
2Novosibirsk State University, Novosibirsk, Russia

E-mail: s.s.abdurakipov@gmail.com

Abstract. The work is devoted to the development of an application for monitoring and controlling the state of equipment (extruder) for the petrochemical industry based on sensor readings using a machine learning model. The statistical relationships of the technological process parameters are analyzed, the most significant parameters influencing the occurrence of failures are determined using SHAP values. The hypotheses regarding the effectiveness of various machine learning algorithms in relation to the real problem of predicting the technical state of the extruder are tested. A gradient boosting model has been developed to predict the probability of extruder shutdown due to the formation of polypropylene agglomerates. The developed application allows interpreting the results of the model, which makes it possible to select the most important process parameters that have the greatest impact on the probability of extruder failure, and also proposing a prototype of an extruder monitoring system based on sensor readings using a machine learning model.

1. Introduction
Recently, in many industries, it is very important to optimize the cost of maintenance and repair. Traditional maintenance strategies (event-driven, routine maintenance, and condition-based maintenance) result in longer downtime on top of unnecessary downtime due to regular inspection. Thus, predictive maintenance, as a planning method based on predictive failures, has received increasing attention in industrial tasks in recent years to reduce costs and downtime. The decision-making and planning process of maintenance and repair activities can be improved by employing modern data analysis tools, which can be considered as a digital transformation course for any company with capital assets.

At present, the progress in machine learning algorithms allows developing process models based on failure-identifying data [1]. Using machine learning to determine patterns based on historical data (derived from IoT sensors) provides an additional approach to maintenance planning by analyzing datasets of individual characteristics of a machine, identifying abnormal situations, and providing predictable failures for individual parts [2-4]. In [5] an artificial neural network was developed for monitoring a coal furnace. The paper [6] provides a detailed analysis of various machine learning methods for solving the problem of predicting aircraft engine failures.

The considered technological process includes a granulation line, which is a hardware and software complex, where the main unit is an extruder. The extruder is a machine for continuous processing of polymer raw materials (granules, crushed agglomerate) into a homogeneous melt and shaping it by forcing through an extrusion head and a special calibrating device, whose cross-section corresponds to
the configuration of the finished product. The raw material is poured (manually or using a special loader) into the hopper of the extruder. From the hopper, bypassing the throat of the feed hopper, the raw material enters the feed zone of the auger, peroxide is added in advance, and then transported through the plasticizing cylinder. From squeezing, mixing, and contact with a heated cylinder and auger, the polymer raw material melts and turns into a homogeneous mass. The cylinder also has a slotted disc which is necessary for high melt homogenization. Next, the melt is pushed through a die (grate) and cut into granules (cassette with knives), which are lifted onto a special vibrating screen by an air flow, and poured into a special container for further distribution as a product [2]. Violation of the technological process can lead to both a deterioration in the quality of manufactured products, and to a complete shutdown with further disassembly, cleaning, and restart of the extruder, which entails costly downtime [7].

Timely detection of abnormal operating conditions of the extrusion line prevents degradation of both equipment and product quality. The implementation of the monitoring system based on the analysis of the behavior of the technological process parameters using various machine learning algorithms can potentially solve the problem of early detection of pre-failure situations. The schematic diagram of the polypropylene granules production line is shown in Figure 1.

This work aimed to develop a prototype of an application for monitoring and controlling the state of an extruder, based on sensor readings using a machine learning model.

![Diagram of polypropylene granule production line](image)

**Figure 1.** Schematic diagram of a polypropylene granule production line.

2. Data description

The data were provided by Sibur company as part of an open data science competition. Each instant measurement was a set of process parameters derived from sensors. All observations were collected in a dataset of more than four million lines. The measurement time step was 10 seconds. The dataset contained information about 43 process parameters for 15 months of historical observations:

- The current strength of electric motors;
- Rotation frequency of electric motors;
- Polypropylene temperature in different parts of the cylinder;
- External temperatures of the cylinder;
- The temperature of the slotted disk;
- Melt fluidity degree;
- Cooling water temperature;
- Position of knives;
- Consumption of polypropylene powder;
- Melt pressure before and after the sieves;
- Melt pressure before the die.

Target events were timestamps when the extruder stops due to the occurrence of polypropylene agglomeration. Time series of sensor readings and the facts of stops and degradation of product quality were combined into one dataset with dependent and target variables. The sensor readings were
compressed to 10 minutes sampling with data aggregation procedure. Mean, minimum, maximum, and standard deviation basic statistics were calculated from the original time series under the development of the baseline model. Target events were shifted by 30 minutes based on the required forecast horizon. The aggregation time window of 10 minutes was selected based on expert analysis and requirements for the monitoring system to predict a failure in 10-30 minutes. At that, 30 minutes was sufficient time for the process engineer to detect and prevent a failure by using control actions [3, 7]. Examples of time series of the most significant parameters are shown in Figure 2. The exploratory data analysis of the parameters shows data noise, sharp outliers, etc. These problems were solved at the stage of data preprocessing using Kalman filtering, time series clustering, etc.

![Figure 2. Time sequence of measurements of important process parameters.](image)

3. Statistical modeling
A typical binary classification problem is defined to determine the probability of extruder failure in the next 30 minutes from the moment of observation. The 30-minute horizon was chosen in consultation with a process engineer. This forecast-horizon allows timely detection and elimination of the problem before the extruder emergency shutdown.

To predict extruder failures, six different machine learning classifiers were tested [8, 9]:
- Logistic regression;
- K-nearest neighbors;
- Linear support vector machine (SVM);
- Random Forest;
- Gradient Boosting (XGBoost [12], CatBoost).

To solve the problem of class imbalance, the oversampling method SMOTE (Synthetic Minority Oversampling Technique) was applied [10]. The dataset was divided into train and test set in a ratio of 80 to 20%. For stratified cross-validation, StratifiedKFold for three folds (function from Sklearn library) was used to preserve the proportion of classes in each fold [9]. Hyperopt library [13] was used for the selection of the best hyperparameters for the model. The results of the quality metrics (Accuracy, Precision, Recall, F1 Score, ROC AUC [8]) for binary classification are shown in Table 1.
XGBoost Classifier was selected as the best model since the ROC AUC and F1 Score metrics were higher than other models. To calculate the quality metrics, the optimal probability cutoff was found.

| Model                        | ROC AUC  | Recall | Accuracy  | Precision | F1 Score |
|------------------------------|----------|--------|-----------|-----------|----------|
| Logistic Regression          | 0.971017 | 0.992192 | 0.915454 | 0.979307 | 0.995094 |
| Random Forest                | 0.947368 | 0.719526 | 0.020316 | 0.947368 | 0.394737 |
| KNN                          | 0.910341 | 0.988888 | 0.997110 | 0.907138 | 0.988204 |
| SVC Linear                   | 0.030457 | 0.171975 | 0.010000 | 0.028436 | 1.000000 |
| XGBoost Classifier           | 0.059016 | 0.278823 | 0.051202 | 0.057698 | 0.566038 |
| Catboost Classifier          | 0.480000 |        |           |           |          |

**Table 1.** Quality metrics of machine learning models.

The interpretation of the results of the XGBoost classifier model is carried out using the SHAP library, which allows interpreting the results of the model and visualizing the influence of each feature on the forecast of the model. SHAP values interpret the impact of having a certain value for a given feature in comparison to the prediction which could be made if that feature took some baseline value [11]. Rather than using a typical bar graph of the importance (histogram) of objects, one can use a scatter plot of the SHAP values density for each observation to determine how each observation affects the model output for individual cases in the validation dataset. Observations were sorted by the sum of the SHAP values across all samples. The result is shown in Figure 3. When examining the values of the variables, it can be seen that dispersion of loading level in V25001A, dispersion of polymer fluidity, minimum of powder consumption and pumping EX21401, and minimum pressure before shaft dominate the other variables, clearly standing out as the most important predictors of extruder failure. All hypotheses that were identified at the stage of interpreting the model results were tested with experts (process engineers) in this area.

![Figure 3](image.png)

**Figure 3.** Feature importance, visualized by the distribution of SHAP values for each significant parameter.

### 4. Conclusions
The present work aimed to develop a prototype of an extruder monitoring system based on sensor readings using a machine learning model.
The following tasks were completed to achieve the goal:
The current situation in the field of predictive maintenance has been analyzed;
Existing approaches to industrial maintenance have been analyzed;
Statistical relationships of parameters with the facts of extruder failures have been analyzed;
The most significant parameters influencing the occurrence of failures have been determined;
Hypotheses regarding the effectiveness of machine learning algorithms for predicting extruder failures have been tested;
Statistical model has been developed and an assessment of the model quality has been conducted;
An algorithm for interpreting the simulation results has been developed;
A prototype of an extruder monitoring system has been developed based on sensor readings using a machine learning model.

It is shown that the obtained gradient boosting model (XGBoost classifier [12]) has the best quality metrics among the considered models and can be used for predicting extruder failure. An algorithm for interpreting the model results based on SHAP values [11] has been developed. As a result, an application (in the form of an engineer's desktop, see Figure 4) for monitoring the technological process of polymer granules production was developed. The application allows monitoring a set of parameters, determining the probability of failure, and providing meaningful parameters that can affect the failure. To develop a prototype to a productive solution, it is necessary to increase the initial dataset (increase the number of failures) to improve the quality of the model. As part of an additional improvement of the prototype, it is planned to add more data to reduce the significance of synthetic examples obtained from SMOTE algorithm, add basic information about the equipment (technological diagrams, repair plans, passport data, main types of failures, their causes, and consequences, etc.), adjust data entry IoT in real time for integration into the developed application.

![Figure 4. Engineer's desktop for monitoring process parameters of extruder.](image)

**Acknowledgments**
This work was carried out within the framework of the state assignment for the IT SB RAS.
References

[1] Klimova E, Babkin I 2019 Scientific and technical statements of the St. Petersburg State Polytechnic University Economic sciences 3 132-141
[2] Rauwendaal C 2014 Polymer extrusion Carl Hanser Verlag GmbH Co KG
[3] Pong S, Lu C, Wang M, Chiu S 2005 Journal of applied polymer science 95 2 280-289
[4] Abdurakipov S, Tokarev M, Pervunin K, Dulin V 2019 Optoelectronics Instrumentation and Data Processing 55 2 205-211
[5] Abdurakipov S, Butakov E, Burdukov A, Kuznetsov A, Chernova G 2019 Combustion Explosion and Shock Waves 55 6 697-701
[6] Abdurakipov S, Butakov E 2020 Optoelectronics Instrumentation and Data Processing 56 6 586-597
[7] Robin A, Pimenov V 2004 4 46–51
[8] Mohri M, Rostamizadeh A, Talwalkar A 2018 Foundations of machine learning MIT press
[9] Shalev-Shwartz S, Ben-David S 2014 Cambridge university press
[10] Chawla N, Bowyer K, Hall L, Kegelmeyer W 2002 Journal of artificial intelligence research vol 16 no 3 pp 21-357
[11] Lundberg S, Lee S 2017 Proceedings of the 31st International Conference on Neural Information Processing Systems 4768-4777
[12] Chen T, Guestrin C 2016 Xgboost: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794
[13] Bergstra J, Yamins D, Cox D 2013 Proceedings of the 12th Python in Science Conferences 13 20