

РАЗВЕДОЧНЫЙ АНАЛИЗ ДАННЫХ, ПОЛУЧЕННЫХ НА ПРОМЫШЛЕННОМ ОБОРУДОВАНИИ (ФРЕЗЕРНЫЙ СТАНОК)

Самойлов В.Ю.¹, Губин Е.И.²

¹ НИ ТПУ, ИШИТР, 8К13, e-mail: vys27@tpu.ru

² НИ ТПУ, ИШИТР, доцент ОИТ, e-mail: gubine@tpu.ru

Введение

Прежде чем приступить к обработке данных, специалист должен получить общее представление об исследуемом датасете. В этом ему может помочь разведочный анализ данных [1], [2]. Он позволяет количественно оценить общий объем данных, определить число и типы переменных, найти и устранить повторяющиеся и не корректные данные, выделить целевую функцию, а также выдвинуть гипотезы о существующих закономерностях исследуемого процесса. В данной статье представлен пример разведочного анализа данных (подготовки данных), полученных на фрезерном станке. Для работы использовались синтетически сгенерированные данные, которые близки к реальным параметрам работы оборудования [3]. Исследование проводилось с использованием программы Microsoft Excel.

Основная часть

В исследуемом датасете содержится 10000 записей с 10 атрибутами. В результате рассмотрения значений полей исходных данных была составлена сводная таблица, характеризующая каждый атрибут.

Таблица 1

Сводная таблица атрибутов

Атрибут	Формат	Размерность	Имя
UDI	Числовой	–	уникальный идентификатор
Product ID	Текстовый	–	серийный номер
Type	Текстовый	–	уровень качества продукта (L – низкий, M – средний, H – высокий)
Air temperature	Числовой	градусы по Кельвину	температура окружающего воздуха
Process temperature	Числовой	градусы по Кельвину	температура технологического процесса
Rotational speed	Числовой	число радиан в минуту	скорость вращения фрезерного станка
Torque	Числовой	ньютон · метр	крутящий момент фрезерного станка
Tool wear	Числовой	минуты	износ инструмента
Target	Бинарный	–	наличие ошибки
Failure Type	Текстовый	–	тип ошибки

Исследуемый датасет не содержит повторяющихся строк, поврежденных и пропущенных данных и выбросов. Это объясняется тем, что датасет представляет собой набор синтетических данных, полученных методами случайного блуждания и нормализации.

Для атрибутов, представленных числовыми данными, были найдены минимальные, максимальные, медианные и средние значения. Из полученных данных следует, что средняя температура процесса на 3,3 % больше средней температуры окружающего воздуха и максимальное значение, которое проработал станок без замены фрезы, составило 253 минуты.

Далее было проведено исследование каждого параметра в отдельности. Параметр UDI представляет собой порядковый номер записи в списке и принимает значения от 1 до 10000. Он позволяет рассматривать процесс в хронологическом порядке.

Параметр Product ID содержит в себе серийный номер изделия и в первой букве несет информацию об уровне качества продукта (L – низкий, M – средний, H – высокий).

Параметр Type дублирует первый символ серийного номера и позволяет исследовать качество продукта в отрыве от остальных параметров. В результате построения диаграммы было обнаружено, что соотношение между продуктами низкого, среднего и высокого качества составило 60 %, 30 % и

10 % соответственно, что позволяет сделать вывод о том, что продукция низкого качества составляет больше половины от всей выпускаемой на станке.

Параметры Air temperature и Process temperature содержат информацию о температуре окружающей среды и фрезы. Их значения находятся в интервалах от 295,3 К до 304,5 К и от 305,7 К до 313,8 К соответственно.

Параметр Rotational speed содержит информацию о скорости вращения фрезы, атрибут Torque – о крутящем моменте. Вместе они описывают механические характеристики станка. Первый параметр меняется от 1168 до 2886 радиан в минуту, второй – от 3,8 до 76,6 ньютон на метр.

Параметр Tool wear характеризует время работы инструмента. Чем дольше работает фреза, тем больше становится ее степень износа. Атрибут меняется от 0 до 253 минут, среднее значение составляет 108 минут. При замене инструмента происходит сброс параметра до 0.

Параметр Target принимает два значения – 1, если произошел сбой машины, и 0, если сбоя в работе не было. Данный параметр является целевой функцией.

Параметр Failure Type характеризует режим сбоя станка и может принимать 6 значений: No Failure (управляющее устройство станка не ожидало возникновения ошибки), Power Failure (выход из диапазона допустимой мощности), Overstrain Failure (выход из строя из-за перенапряжения, то есть превышение допустимых значений для произведения времени работы инструмента и крутящего момента), Random Failures (случайная причина ошибки) и Heat Dissipation Failure (неправильное рассеивание тепла). Режим сбоя был зафиксирован в менее чем 3,5 % случаев, из них на Power Failure пришлось 27,3 %, на Tool Wear Failure – 12,93 %, на Overstrain Failure – 22,41 %, на Random Failures – 5,17 %, на Heat Dissipation Failure – 32,18 %. С одной стороны, в ряде случаев в режиме ожидания какого-либо сбоя (Power Failure, Heat Dissipation Failure и др.) выход из строя не происходил. С другой, несколько раз станок переставал работать в режиме No Failure. Это говорит о том, что между параметрами Target и Failure Type нет корреляции, поэтому их нельзя рассматривать как целевые функции одновременно.

Стоит уделить особое внимание значениям параметров, при которых произошел сбой, то есть те значения атрибутов, при которых целевая функция Target принимала значение 1. Наибольшее число сбоев произошло в режиме работы No Failure (98 %), то есть в тот момент, когда выход из строя не был ожидаем (рисунок 1).

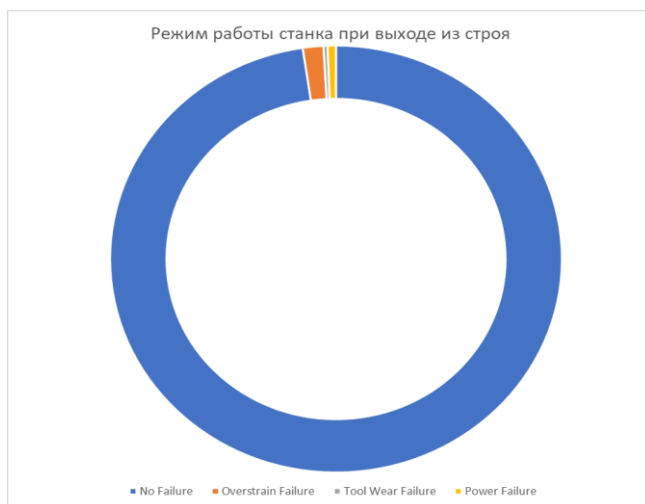


Рис. 1. Диаграмма распределения режимов работы станка при выходе из строя

Поломка станка происходила при обработке продукта высокого уровня качества в 14% случаев, среднего – 30%, низкого – 56%. При сравнении с указанным выше общим распределением можно отметить увеличение числа сбоев при обработке продукции более высокого качества. В результате рассмотрения других атрибутов были получены следующие сведения: наибольшее число сбоев происходило при температуре окружающего воздуха 298,5-298,6 К, температуре процесса в интервале от 309,6 до 309,9 К, времени работы станка от 0 до 33 минут, скорости вращения в интервале от 1375 до 1466 радиан в минуту и моменте вращения в диапазоне от 33,2 до 38,3 ньютон на метр.

Заключение

Таким образом, был проведен разведочный анализ и подготовка данных для дальнейших исследований. Были получены общие сведения о датасете, определены диапазоны изменения параметров, выбрана целевая функция, а также найдены некоторые закономерности в работе фрезерного станка.

Список использованных источников

1. Губин Е. И. Методика подготовки больших данных для прогнозного анализа. «Наука и бизнес: пути развития». Выпуск № 3(105). 2020, № 2020. – С. 33-35.
2. What is Exploratory Data Analysis? | by Prasad Patil. [Электронный ресурс]. – URL: <https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15> (дата обращения 15.02.2023).
3. Predictive Maintenance Dataset (AI4I 2020) | Kaggle. [Электронный ресурс]. – URL: <https://www.kaggle.com/datasets/stephanmatzka/predictive-maintenance-dataset-ai4i-2020> (дата обращения 27.10.2022).