

Анализ моделей машинного обучения для предсказания PassOrFail

Описание данных:

Количество наблюдений: 1256

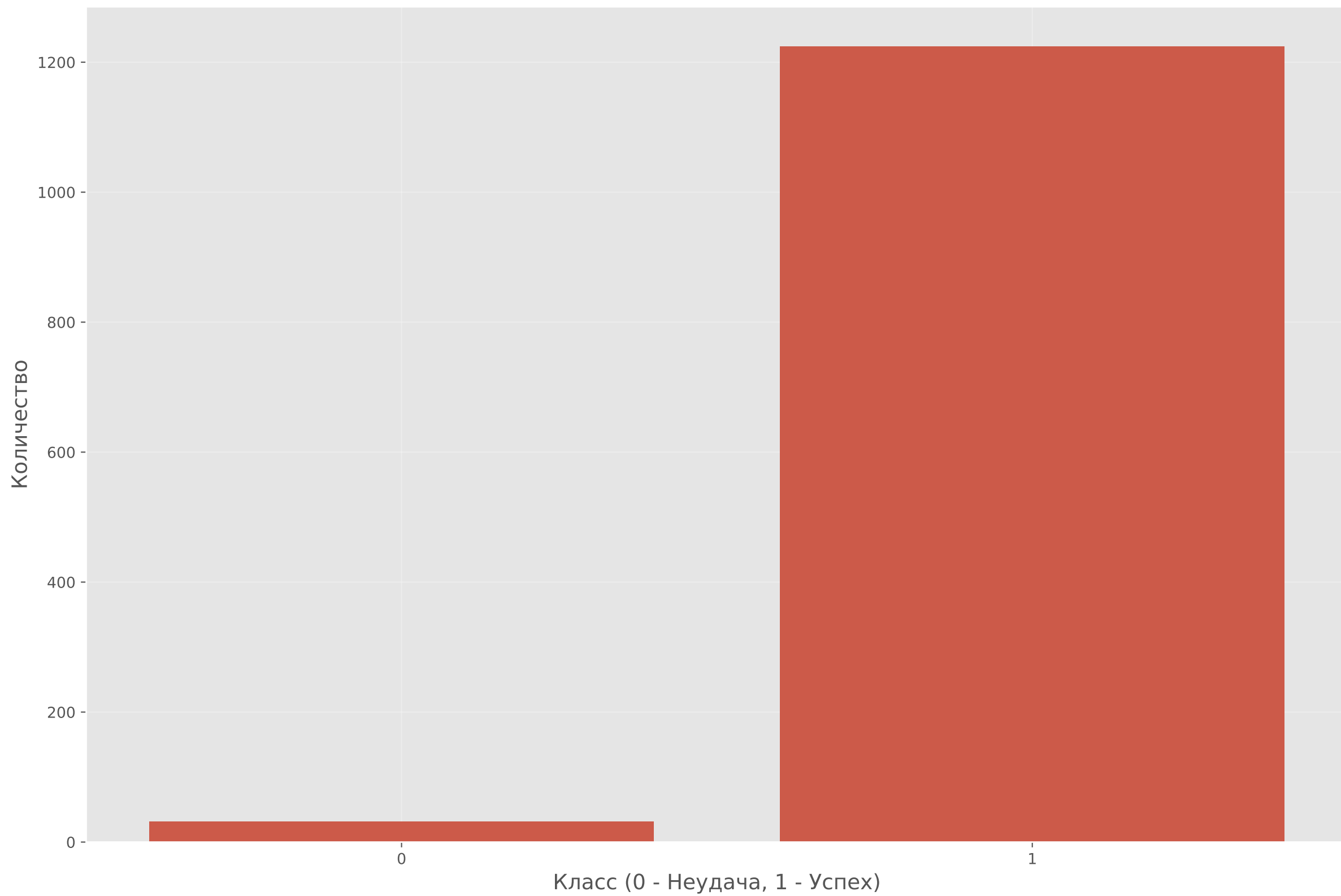
Количество признаков: 25

PassOrFail

1 0.974522

0 0.025478

Распределение классов в датасете



Предобработка данных:

1. Загрузка и фильтрация данных
2. Преобразование целевой переменной
3. Анализ типов данных
4. Удаление нечисловых колонок
5. Удаление колонок с нулевыми значениями
6. Удаление колонок с одной уникальной переменной
7. Удаление колонок с высокой долей пропущенных значений
8. Анализ корреляций
9. Удаление коррелирующих признаков
10. Обработка пропущенных значений
11. Обработка выбросов
12. Нормализация признаков

PART_FACT_SERIAL	int64
PassOrFail	int64
Injection_Time	float64
Filling_Time	float64
Plasticizing_Time	float64
Cycle_Time	float64
Clamp_Close_Time	float64
Cushion_Position	float64
Switch_Over_Position	float64
Plasticizing_Position	float64
Max_Injection_Speed	float64
Max_Screw_RPM	float64
Average_Screw_RPM	float64
Max_Injection_Pressure	float64
Max_Switch_Over_Pressure	float64
Max_Back_Pressure	float64
Average_Back_Pressure	float64
Barrel_Temperature_1	float64
Barrel_Temperature_2	float64
Barrel_Temperature_3	float64
Barrel_Temperature_4	float64
Barrel_Temperature_5	float64
Barrel_Temperature_6	float64
Hopper_Temperature	float64
Mold_Temperature_3	float64
Mold_Temperature_4	float64

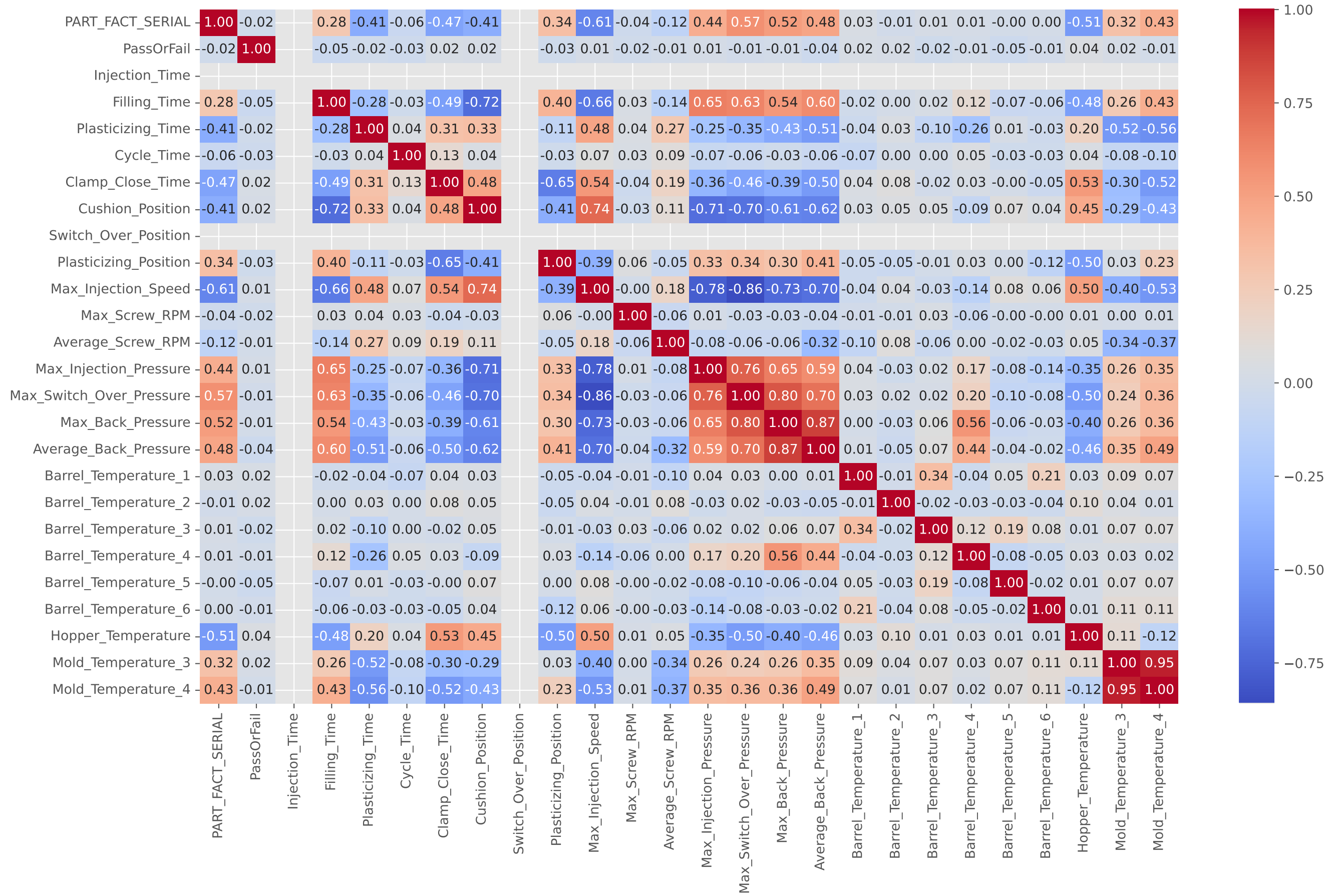
Анализ типов данных:

Количество пропущенных значений по признакам

Количество пропущенных значений

PART_FACT_SERIAL
PassOrFail
Injection_Time
Filling_Time
Plasticizing_Time
Cycle_Time
Clamp_Close_Time
Cushion_Position
Switch_Over_Position
Plasticizing_Position
Max_Injection_Speed
Max_Screw_RPM
Average_Screw_RPM
Max_Injection_Pressure
Max_Switch_Over_Pressure
Max_Back_Pressure
Average_Back_Pressure
Barrel_Temperature_1
Barrel_Temperature_2
Barrel_Temperature_3
Barrel_Temperature_4
Barrel_Temperature_5
Barrel_Temperature_6
Hopper_Temperature
Mold_Temperature_3
Mold_Temperature_4

Матрица корреляций

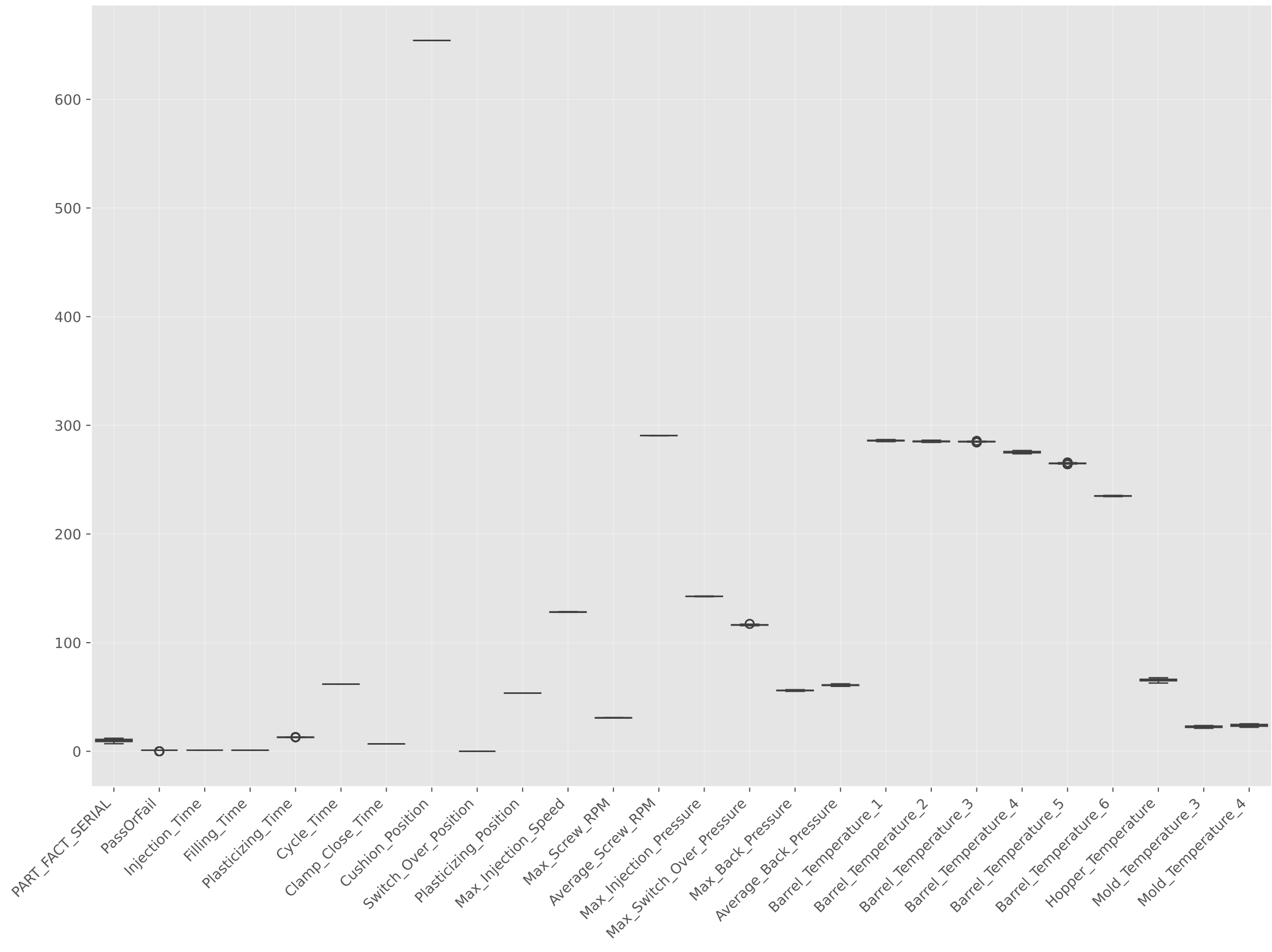


PART_FACT_SERIAL	0
PassOrFail	32
Injection_Time	0
Filling_Time	0
Plasticizing_Time	22
Cycle_Time	0
Clamp_Close_Time	0
Cushion_Position	0
Switch_Over_Position	0
Plasticizing_Position	0
Max_Injection_Speed	0
Max_Screw_RPM	0
Average_Screw_RPM	0
Max_Injection_Pressure	0
Max_Switch_Over_Pressure	3
Max_Back_Pressure	0
Average_Back_Pressure	0
Barrel_Temperature_1	0
Barrel_Temperature_2	0
Barrel_Temperature_3	169
Barrel_Temperature_4	0
Barrel_Temperature_5	50
Barrel_Temperature_6	0
Hopper_Temperature	0
Mold_Temperature_3	0
Mold_Temperature_4	0

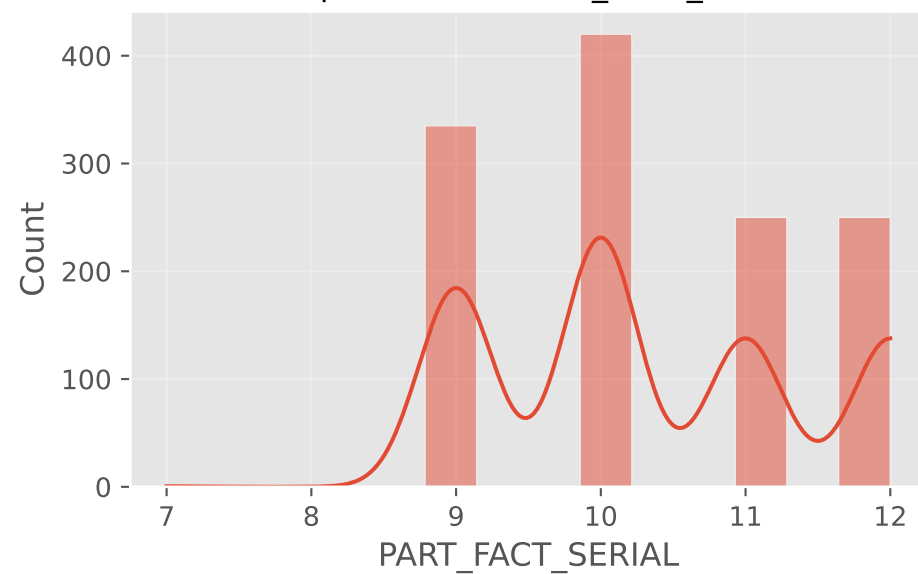
Анализ выбросов:

Количество выбросов по признакам:

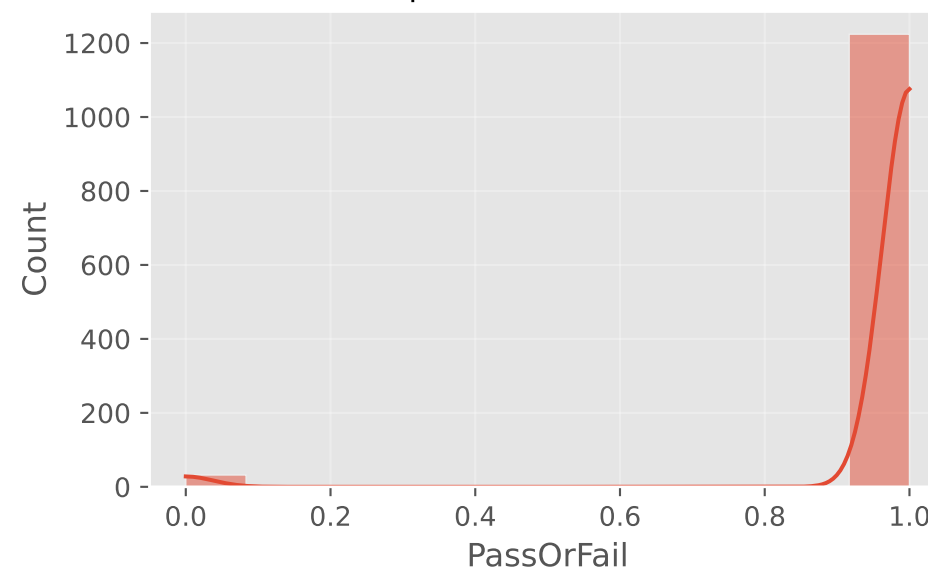
Boxplot признаков



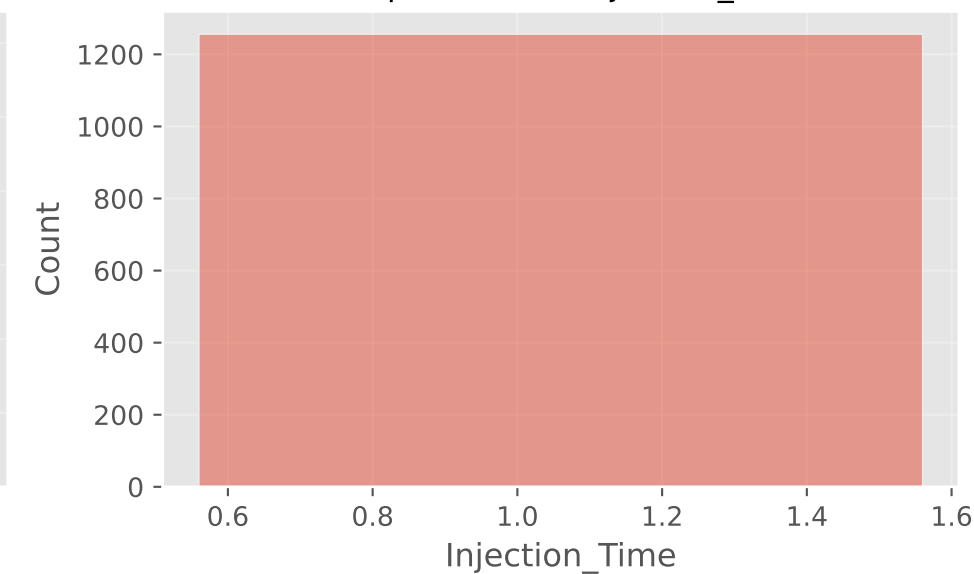
Распределение PART_FACT_SERIAL



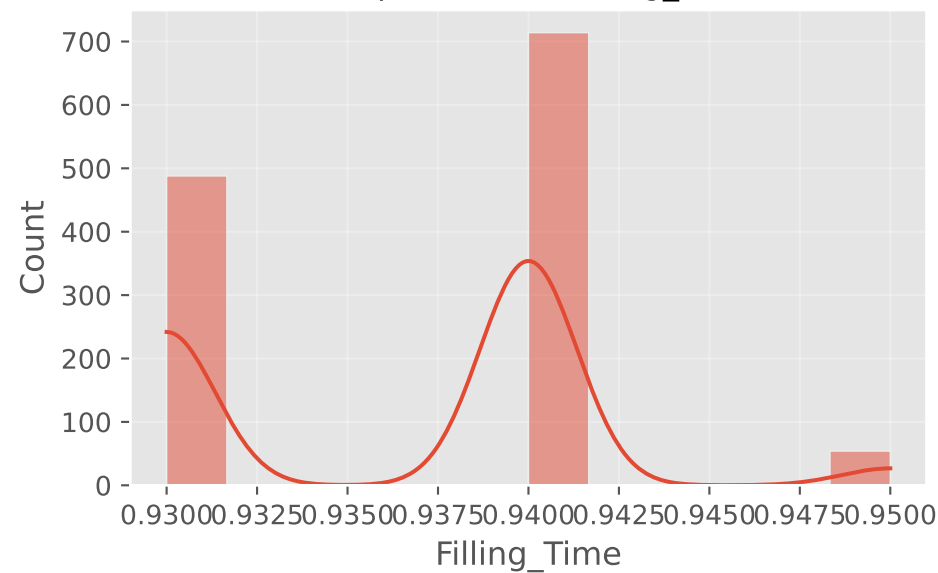
Распределение PassOrFail



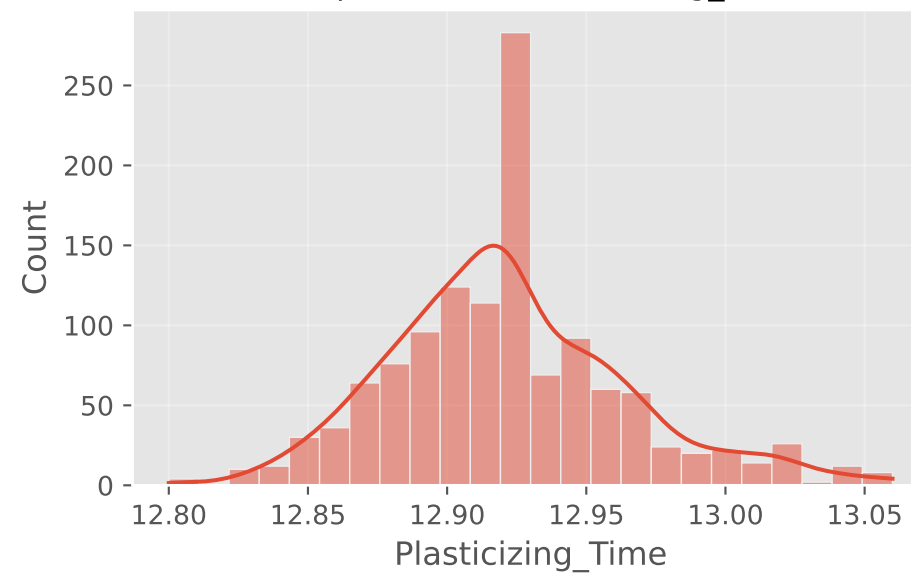
Распределение Injection_Time



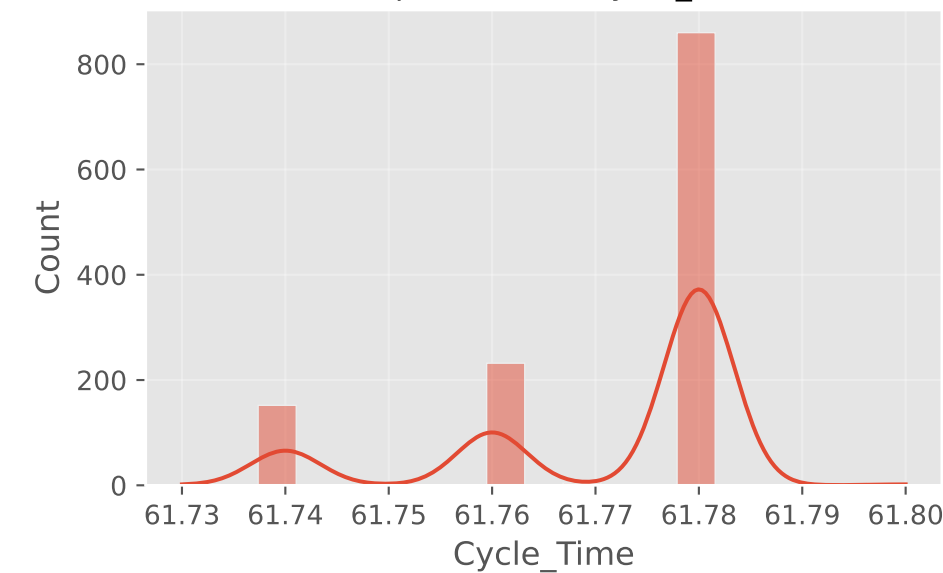
Распределение Filling_Time



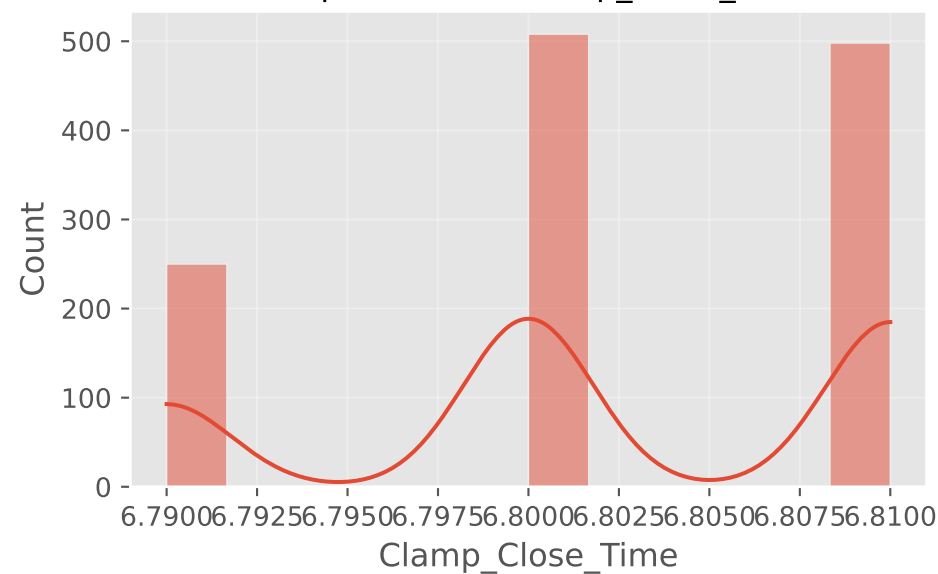
Распределение Plasticizing_Time



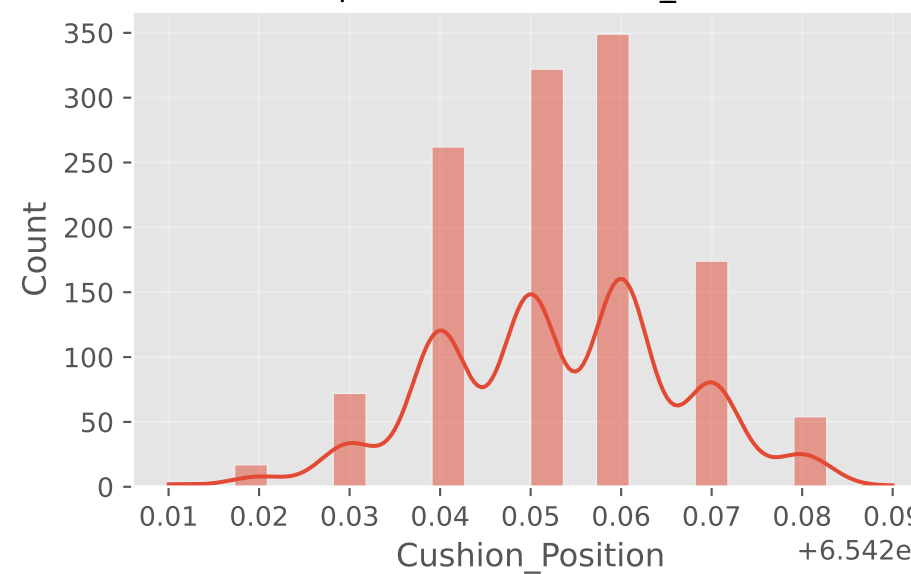
Распределение Cycle_Time



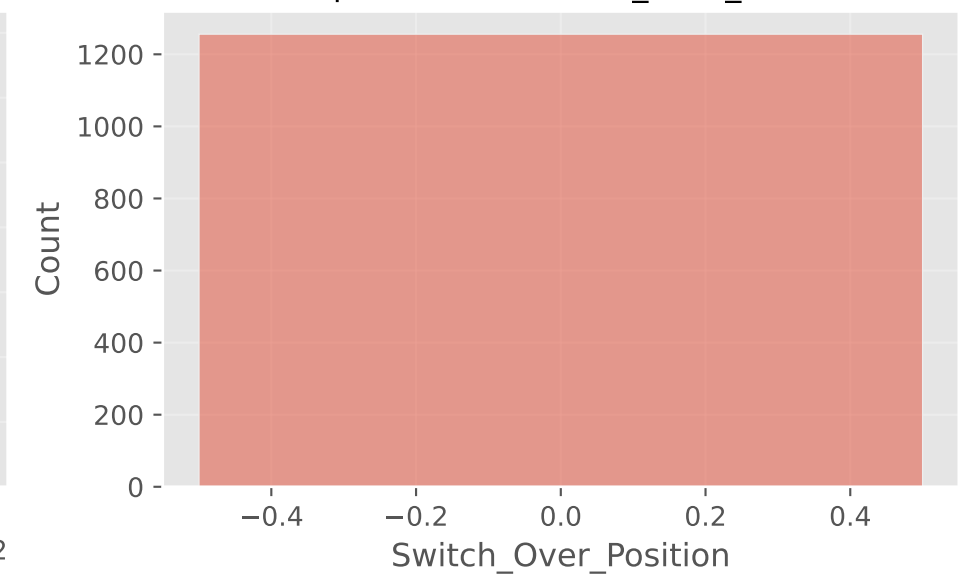
Распределение Clamp_Close_Time



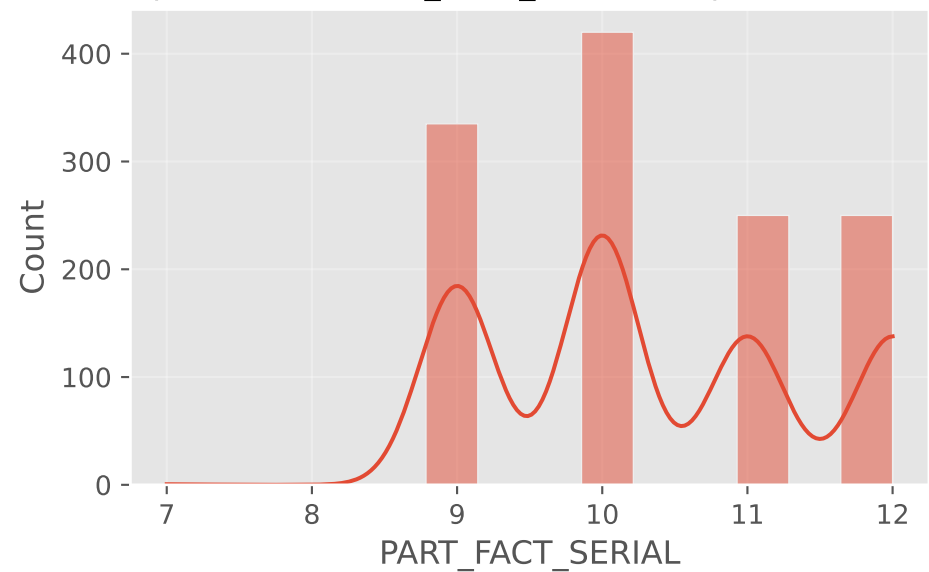
Распределение Cushion_Position



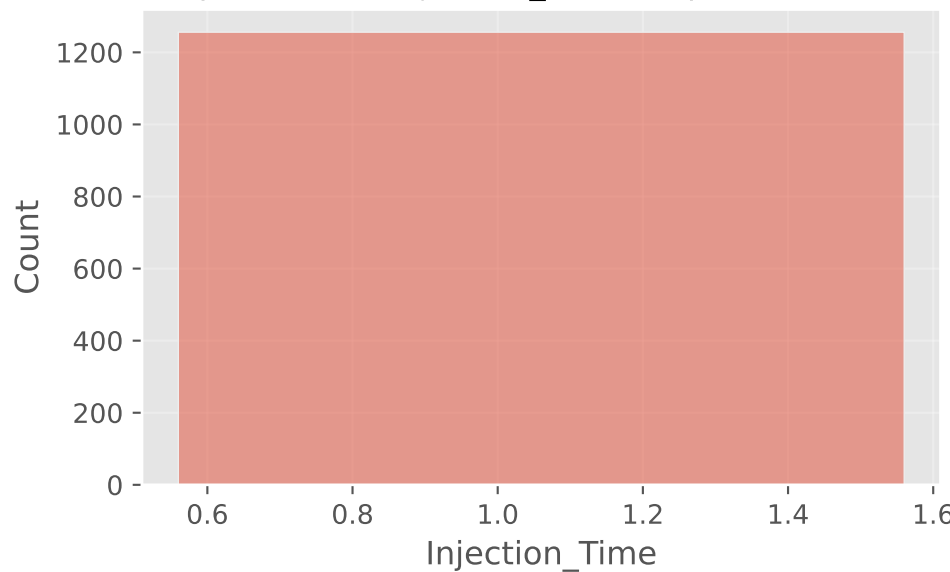
Распределение Switch_Over_Position



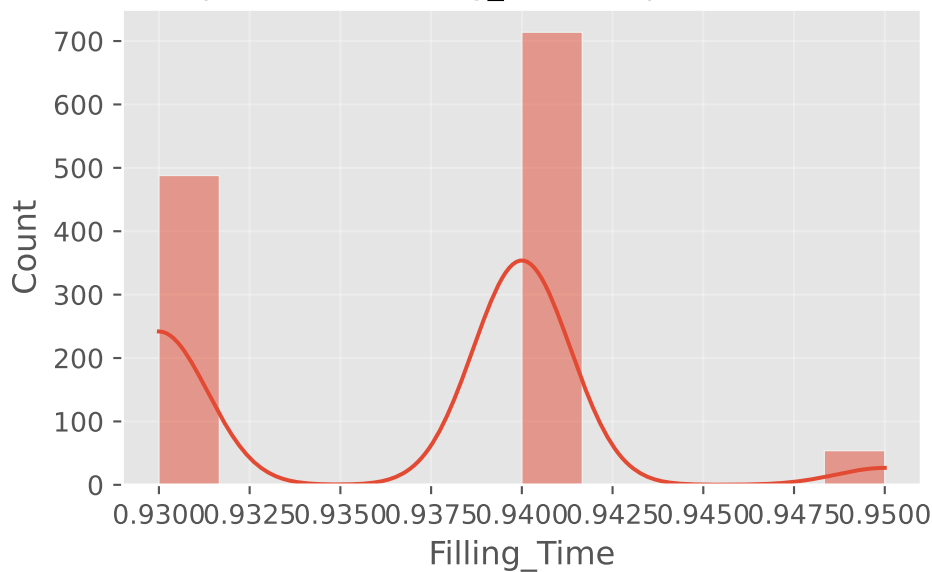
Распределение PART_FACT_SERIAL (нормализованное)



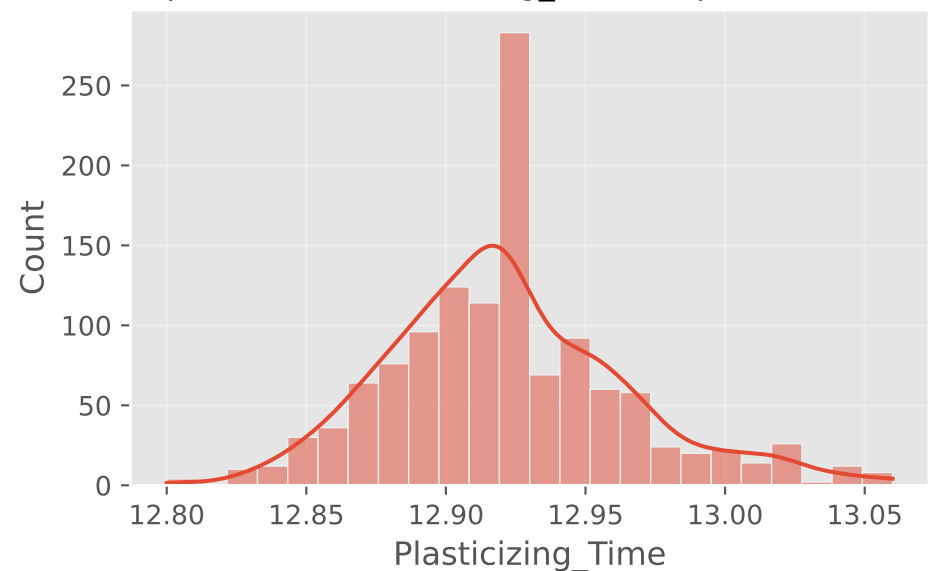
Распределение Injection_Time (нормализованное)



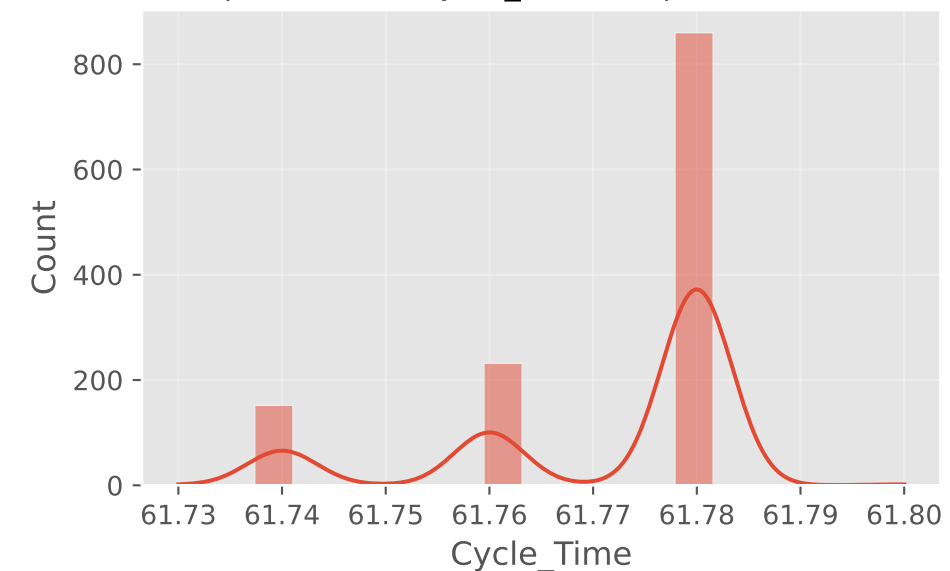
Распределение Filling_Time (нормализованное)



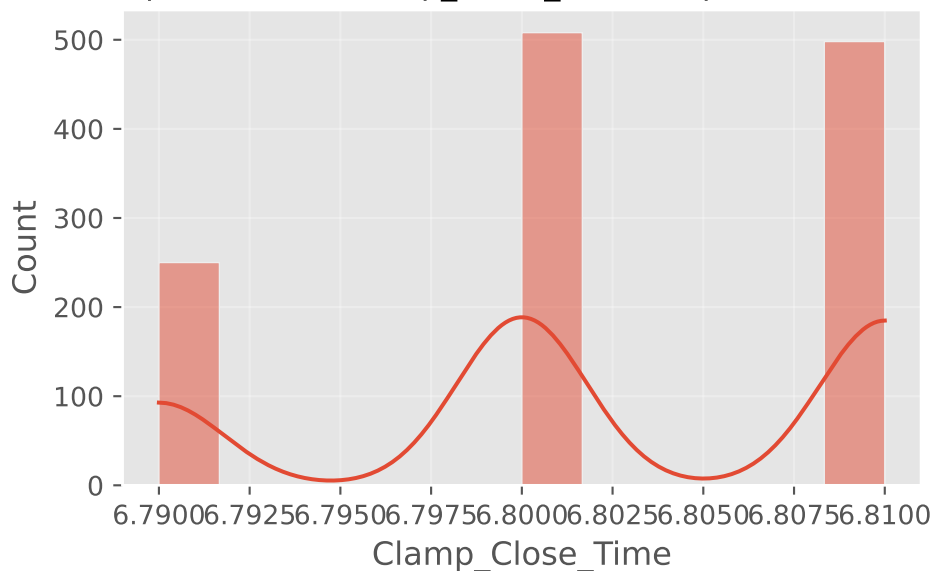
Распределение Plasticizing_Time (нормализованное)



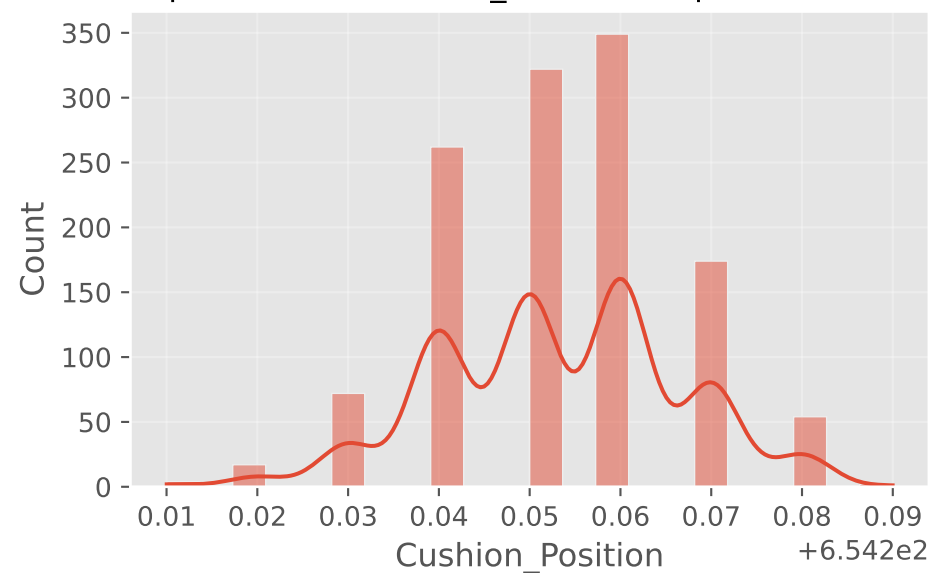
Распределение Cycle_Time (нормализованное)



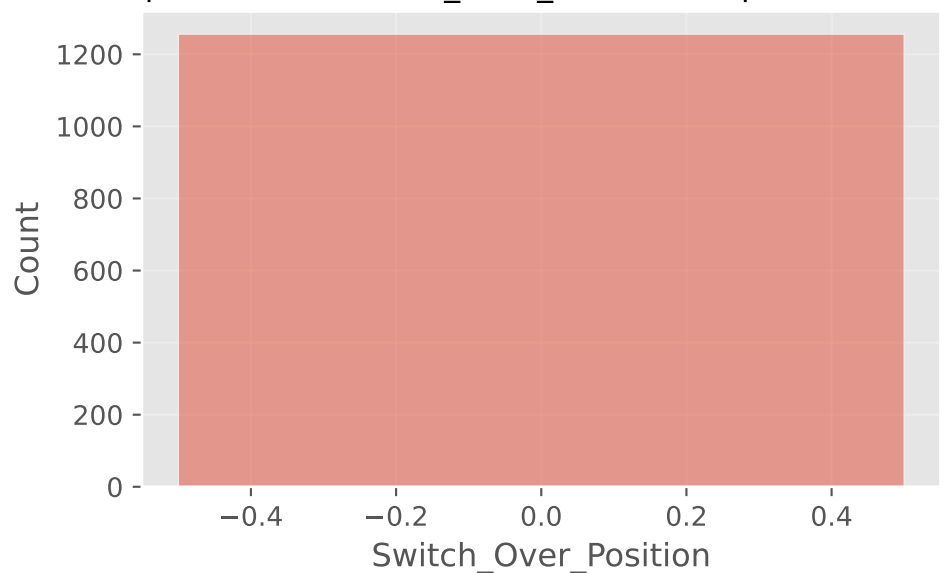
Распределение Clamp_Close_Time (нормализованное)



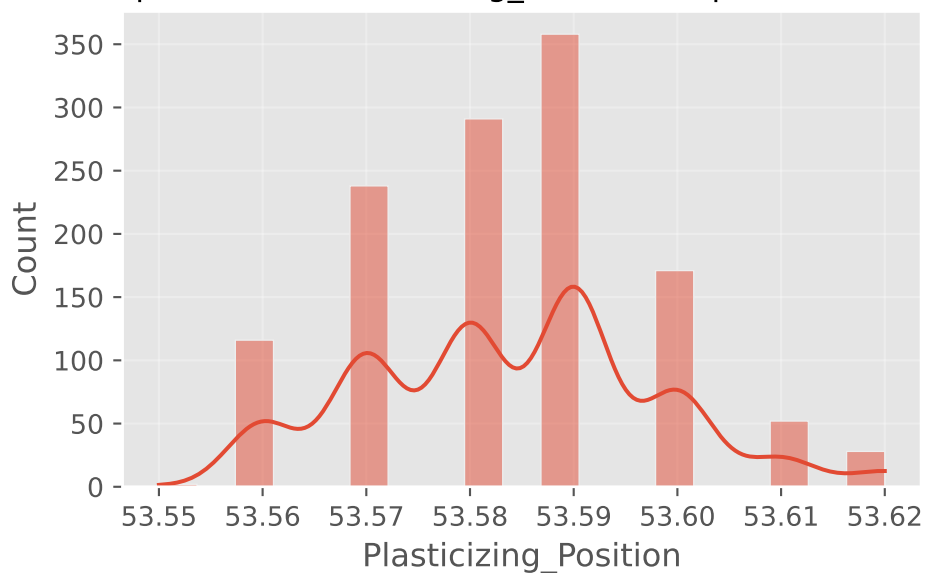
Распределение Cushion_Position (нормализованное)



Распределение Switch_Over_Position (нормализованное)



Распределение Plasticizing_Position (нормализованное)



PART_FACT_SERIAL
Injection_Time
Filling_Time
Plasticizing_Time
Cycle_Time
Clamp_Close_Time
Cushion_Position
Switch_Over_Position
Plasticizing_Position
Max_Injection_Speed
Max_Screw_RPM

Итоговый набор признаков:

Average_Screw_RPM
Max_Injection_Pressure
Max_Switch_Over_Pressure
Количество признаков до предобработки: 26
Max_Back_Pressure
Average_Back_Pressure
Barrel_Temperature_1
Barrel_Temperature_2
Количество признаков после предобработки: 25
Barrel_Temperature_3
Barrel_Temperature_4
Barrel_Temperature_5
Оставшиеся признаки:
Barrel_Temperature_6
Hopper_Temperature
Mold_Temperature_3
Mold_Temperature_4

Использованные модели:

1. Logistic Regression
2. Random Forest
3. Gradient Boosting
4. SVM
5. KNN
6. XGBoost
7. LightGBM
8. CatBoost
9. AdaBoost
10. Decision Tree
11. Extra Trees
12. Bagging

Методы балансировки классов:

1. Original (без балансировки)
2. SMOTE
3. Random UnderSampling
4. SMOTEENN
5. ADASYN
6. BorderlineSMOTE
7. Class Weight

ROC-кривые моделей:

ROC-кривая показывает соотношение между:

- True Positive Rate (TPR)
- False Positive Rate (FPR)

AUC-ROC показывает общую производительность модели

Результаты моделей:

Лучшая модель: Gradient Boosting

Метод балансировки: UnderSampling

AUC-ROC: 0.843

Precision: 1.000

Recall: 0.532

F1-score: 0.695

Важность признаков:

Анализ важности признаков показывает:

1. Какие признаки наиболее значимы для предсказания
2. Какие признаки можно исключить
3. Взаимосвязи между признаками

Итоговый пайплайн:

1. Предобработка данных
2. Нормализация признаков
3. Балансировка классов
4. Обучение модели
5. Предсказание

Спасибо за внимание!