|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Государственная корпорация по атомной энергии «Росатом» | | | | | | |
| Организация  УДК:  Инв. №: | | | | | | |
|  | | | | УТВЕРЖДАЮ | | |
|  | | | | |  | | --- | | Должность | | | |
|  | |  | |  | | Ф.И.О. |
|  | | | | «\_\_\_» \_\_\_\_\_\_\_\_\_ 2024 г. | | |
| Номер документа | | | | | | |
| ОТЧЕТ О проверке гипотезы: | | | | | | |
| бинарная классификация аномального режима работы насосного оборудования на данных о параметрах его работы | | | | | | |
|  | | | | | | |
| Договор № Курс ИИ | | | | | | |
|  | | | | | | |
|  | Руководитель проекта: | | Черкасова М.С. | | Ф.И.О.  2024 г. | | |
| « » | |
|  | |  | | |
|  |  | |  | | | | |
|  | Москва 2024 | | | | | | |

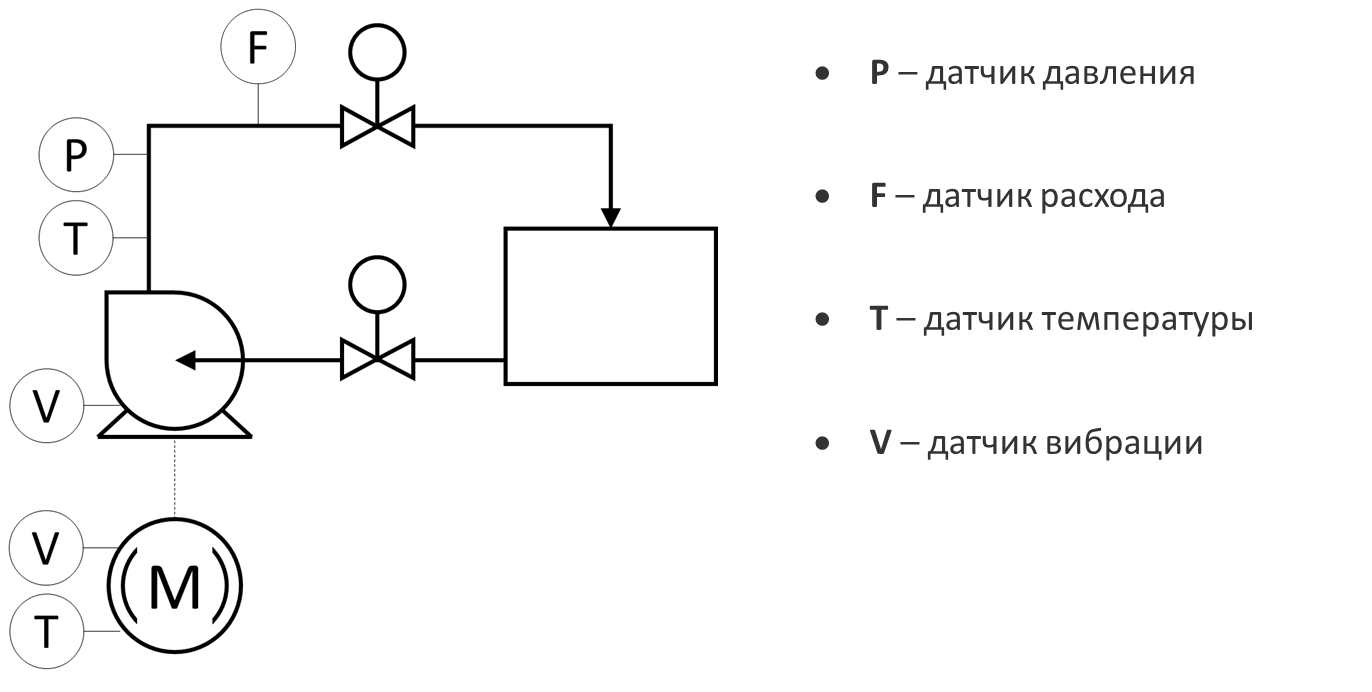
1. **Наименование работ**

Работы по проверке гипотезы «Бинарная классификация аномального режима работы насосного оборудования на данных о параметрах его работы».

1. **Описание процесса**

Технологический процесс остужения теплоносителя в конденсаторе заключается в прокачке хладагента по внутритрубному пространству конденсатора. Хладагент забирается из резервуара, прокачивается горизонтальным центробежным насосом через конденсатор и возвращается в резервуар.

Упрощенно схема процесса представлена на рисунке ниже:



1. **Гипотеза**

Бинарная классификация аномального режима работы насосного оборудования на данных о параметрах его работы лучше baseline.

Бинарная классификация – это задача классификации элементов множества в одну из двух групп (в данном случае – нормальный режим работы оборудования или аномальный режим работы оборудования).

1. **Критерий успешности**

Baseline побит (F1 > 0.79756).

1. **Метрики**

В качестве метрики используется F1-мера, представляющая собой гармоническое среднее между точностью и полнотой. Она особенно полезна в задачах с несбалансированными классами, где высокая точность может быть достигнута за счет низкой полноты и наоборот. Гармоническое среднее лучше отражает баланс между этими двумя метриками.

1. **Исходные данные**

Исходные данные представляют из себя многомерный временной ряд, собранный с датчиков технологической линии.

Перечень зарегистрированных параметров представлен в таблице ниже:

|  |  |  |
| --- | --- | --- |
| **Параметр** | **Описание** | **Единицы измерения** |
| Accelerometer1RMS | Виброускорение | - |
| Accelerometer2RMS | Виброускорение | - |
| Current | Сила тока питания электродвигателя | А |
| Pressure | Давление на выкиде | Bar |
| Temperature | Температура корпуса электродвигателя | oC |
| Thermocouple | Температура перекачиваемой среды (воды) | oC |
| Voltage | Напряжение питания электродвигателя | В |
| RateRMS | Расход перекачиваемой среды (воды) | л/мин |
| anomaly | Флаг аномалии (дискретный параметр) | - |
| changepoint | Флаг изменения состояния (дискретный параметр) | - |

Имеются значения указанных параметров за отдельные периоды в течение 2020 года.

Дискретизация параметров составляет для большинства данных – 1 секунда с периодическими пропусками.

1. **Модели**

В ходе выполнения работы были проанализированы различные модели машинного обучения и нейронные сети. На первом этапе перед проведением анализа была проведена подготовка данных:

* все данные с датчиков отмасштабированы с нулевым средним значением и стандартным отклонением, равным 1;
* сгенерированы признаки, имеющие физический смысл, а именно:
  + - мощность;
    - отношение расхода к мощности;
    - разница температур;
    - разница виброускорений (акселерометров);
* проведено сглаживание расхода при помощи скользящего среднего с размером скользящего окна, равным 10, и с нулевым минимальным количеством наблюдений в окне.

На втором этапе для проверки гипотезы использовался преподавательский подход с подбором моделей машинного обучения. Были рассмотрены модели без учителя, с учителем, а также многомерное (3D) прогнозирование. Обучение модели происходило на первых 400 данных из выборки.

Используемые модели, а также полученные значения метрики для каждой из них приведены в таблице ниже:

| **Метод** | **F1-мера** |
| --- | --- |
| Модели без учителя | |
| Изолирующий лес | 0.7037 |
| One class – SVM | 0.6751 |
| Модели с учителем | |
| Линейная регрессия | 0.4566 |
| Линейная регрессия (с регуляризацией) | 0.4271 |
| KNN | 0.5136 |
| Деревья решений | 0.5604 |
| SVR (Support Vector Machine for Regression) | 0.5999 |
| Случайный лес | 0.5989 |
| Бустинг | 0.5999 |
| Многомерное (3D) прогнозирование | |
| Линейная регрессия (с регуляризацией) | 0.3933 |
| Дерево решений | 0.6164 |
| SVR | 0.7625 |
| Случайный лес | 0.6784 |
| Бустинг | 0.7511 |

На основании приведенных в таблице результатов можно сделать вывод о том, что модель SVR в многомерном (3D) прогнозировании показала наилучшие результаты.

На третьем этапе для проверки гипотезы решено было использовать различные нейронные сети. Используемые нейронные сети, а также полученные значения метрики для каждой из них приведены в таблице ниже:

|  |  |
| --- | --- |
| **Нейронная сеть** | **F1** |
| Автоинкодер (LSTM AE) | 0.7469 |
| Рекуррентная нейронная сеть с долгой краткосрочной памятью (LSTM) | 0.6908 |
| Conv AE (одномерная сверточная нейронная сеть) | 0.8379 |
| AE | 0.6688 |

Таким образом, на основании приведенных в таблице результатов для дальнейшей проверки гипотезы была выбрана одномерная сверточная нейронная сеть Conv AE. Данная нейронная сеть является наиболее подходящей по метрике и по производительности в подобных типах задач потому, что показала наилучшие результаты по сравнению с LSTM и другими.

Далее для проверки гипотезы использовался преподавательский подход с ручным подбором гиперпараметров для выбранной нейронной сети Conv AE.

1. **Результаты**

Был получен результат F1 = 0.85374.

1. **Выводы**

На основании проделанной работы можно сделать вывод о том, что бинарная классификация аномального режима работы насосного оборудования на данных о его параметрах работы возможна с метрикой F1 = 0.85374.