



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное бюджетное образовательное учреждение высшего обра-  
зования  
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИ-  
ВЕРСИТЕТ»

Кафедра Информационных технологий и систем безопасности

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

(Бакалавр)

**На тему** «Сегментирование множества данных с учетом информации

воздействующих факторов для оценки состояния объектов морской техники»

**Исполнитель** Буравлев Даниил Александрович

(фамилия, имя, отчество)

**Руководитель** доктор технических наук

(ученая степень, ученое звание)

Сухопаров Михаил Евгеньевич

(фамилия, имя, отчество)

**«К защите допускаю»  
Заведующий кафедрой**

(подпись)

доктор технических наук, профессор

(ученая степень, ученое звание)

Бурлов Вячеслав Георгиевич

(фамилия, имя, отчество)

« » 2022 г.

Санкт-Петербург

2023

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра Информационные технологии и системы безопасности

«УТВЕРЖДАЮ»  
Заведующий кафедрой

\_\_\_\_\_  
(подпись) (фамилия, имя, отчество)  
«\_\_» \_\_\_\_\_ 20\_\_ года

ЗАДАНИЕ

на выпускную квалификационную работу

студенту Буравлеву Даниилу Александровичу

(фамилия, имя, отчество)

1. Тема Сегментирование множества данных воздействующих факторов для оценки состояния объектов морской техники

закреплена приказом ректора Университета от «\_\_» \_\_\_\_\_ 20\_\_ года, № \_\_\_\_\_.

2. Срок сдачи законченной работы «\_\_» \_\_\_\_\_ 20\_\_ года.

3. Исходные данные к работе

3. Нормативно-правовые документы

3.2 Нормативно-технические документы

...объекты исследования или их компоненты, условия их функционирования, т.п.

4. Основные вопросы, подлежащие разработке (краткое содержание ВКР)

- Изучение бортовых систем и их видов;
- Изучение сегментирования множества данных и его сравнение с другими методами машинного обучения
- Создание подхода на основе сегментирования;
- Оформление презентации и защита ВКР.

5. Перечень материалов, представляемых к защите:

- ВКР (пояснительная записка);
- отзыв руководителя;
- заключение о проверке работы в системе «Антиплагиат»
- иллюстративный материал (презентация)

6. Дата выдачи задания: «\_\_» \_\_\_\_\_ 20\_\_ года

Руководитель выпускной квалификационной работы

доктор технических наук Сухопаров Михаил Евгеньевич

(должность, ученая степень, ученое звание, фамилия, имя, отчество)

\_\_\_\_\_  
(подпись)

Задание принял к исполнению «\_\_» \_\_\_\_\_ 20\_\_ года

Студент Буравлев Даниил Александрович КВ-Б-19-1

(фамилия, имя, отчество, учебная группа)

\_\_\_\_\_  
(подпись)

(оборотная сторона)

## КАЛЕНДАРНЫЙ ПЛАН

| №<br>п/п | Наименование<br>этапов выпускной квалификационной<br>работы                | Срок выполнения<br>этапов выпускной<br>квалификационной<br>работы | Примечание |
|----------|--|---|------------|
| 1        | Обзор литературы   | С 01.11.2022<br>По 31.12.2022                                     |            |
| 2        | Изучение методов МО  | С 01.01.2022 по<br>20.01.2023                                     |            |
| 3        | Создание метода оценки с использова-<br>нием алгоритмов машинного обучения | С 1.02.2023 по<br>05.05.2023                                      |            |
| 4        | Оформление презентации и защита ВКР  | С 10.05.2023 по<br>29.05.2023                                     |            |

## РЕФЕРАТ

Выпускная квалификационная работа объёмом 63 страницы содержит 7 рисунков, 1 таблицу, 22 источник.

Объект исследования - устройства бортовых автоматизированных систем.

Предмет исследования – являются методы оценки состояния объектов морской техники.

Целью работы является разработка метода оценки состояния объектов морской техники с использованием моделей машинного обучения.

Задачи:

1. Анализ теоретической базы методов машинного обучения и их применения в задачах оценки состояния.

2. Изучение специфики работы устройств бортовых автоматизированных систем и анализ собранных данных.

3. разработка метода оценки состояния объектов морской техники с использованием моделей машинного обучения.

4. Проведение анализа данных с использованием выбранного метода и оценка его применимости.

5. Анализ полученных результатов

В выпускной квалификационной работе произведено исследование существующих методов и подходов к идентификации состояния устройств бортовых систем. В дальнейшем сегментирование множества данных было использовано для оценки состояния объектов морской техники.

## СОДЕРЖАНИЕ

|   |    |
|---|----|
| ВВЕДЕНИЕ.....   | 6  |
| ГЛАВА 1. ОЦЕНКА СОСТОЯНИЯ ОБЪЕКТОВ МОРСКОЙ ТЕХНИКИ....  | 10 |
| 1.1 Виды и классификация объектов морской техники .....   | 10 |
| 1.1.1 Навигационные системы .....   | 10 |
| 1.1.2 Энергетические системы .....  | 11 |
| 1.1.3 Системы связи и информационные системы .....  | 12 |
| 1.1.4 Системы безопасности и защиты .....   | 12 |
| 1.1.5 Автоматические системы управления .....   | 13 |
| 1.1.6 Электронные системы .....   | 14 |
| 1.2 Оценка состояния объектов морской техники.....  | 16 |
| 1.2.1 Оценка состояния объектов морской техники.....  | 16 |
| 1.2.2 Понятие надежности и ее модели.....   | 17 |
| 1.2.2.1 Расчет надежности судоводных бортовых систем.....   | 17 |
| 1.2.3 Виды состояний бортовых систем .....  | 18 |
| 1.2.4 Контроль качества оценки судоводных бортовых систем .....   | 20 |
| 1.3 Постановка задачи .....   | 21 |
| ГЛАВА 2 ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИИ О ВОЗДЕЙСТВУЮЩИХ<br>ФАКТОРАХ ДЛЯ ОЦЕНКИ СОСТОЯНИЯ ОБЪЕКТОВ МОРСКОЙ<br>ТЕХНИКИ.....               | 25 |
| 2.1 Применение методов машинного обучения для оценки состояния<br>бортовых систем.....  | 26 |
| 2.1.1 Области применения методов машинного обучения.....  | 26 |
| 2.1.2 Основные методы машинного обучения.....   | 30 |
| 2.1.3 Классификация методов машинного обучения для оценки состояния<br>объектов морской техники:.....                             | 34 |
| 2.2 Достоинства, недостатки и ограничения различных методов машинного<br>обучения.....  | 38 |
| 2.2.1 Вычислительные затраты и ресурсоемкость.....  | 38 |
| 2.2.2 Недостаток данных и избыточность данных .....   | 40 |
| 2.2.3 Свойства данных и их распределение .....  | 41 |
| 2.3 Сегментирования выборки данных для повышения показателей<br>полноты и точности оценки состояния объектов морской техники..... | 42 |

|   |    |
|---|----|
| 2.3.1 Существующие подходы.....   | 42 |
| ГЛАВА 3 СЕГМЕНТИРОВАНИЕ ВЫБОРКИ ДАННЫХ ПРИ ОЦЕНКЕ<br>СОСТОЯНИЯ ИБ БОРТОВЫХ СИСТЕМ .....           | 48 |
| 3.1 Предлагаемый подход .....   | 48 |
| 3.2 Анализ полученных экспериментальных результатов бортовые<br>телекоммуникационные системы..... | 51 |
| ЗАКЛЮЧЕНИЕ .....  | 58 |
| СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....  | 61 |

## ВВЕДЕНИЕ

В современном мире разработка и применение автоматизированных систем управления в различных сферах деятельности становится все более актуальной задачей. В частности, в морской области возникает потребность в создании средств мониторинга состояния бортовых систем судна, позволяющих выявлять сбои в процессе функционирования объектов морской техники и фиксировать возможные злонамеренные воздействия на систему. Большое количество элементов и узлов инфраструктуры, находящихся за пределами контролируемой зоны, определяет необходимость развития систем мониторинга состояния информационной безопасности. Непрерывный сбор информации о состоянии устройства, сетевого трафика при возникновении различных событий позволяет использовать модели и методы машинного обучения для решения ряда задач анализа и выявления воздействий, нарушающих работу систем.

Создание качественных выборок данных — одна из актуальных проблем машинного обучения. Появление доступных наборов данных, внедрение методов синтетической генерации позволяет создавать большие объемы множеств обучающих примеров, позволяющих оценивать состояние информационной безопасности (ИБ) различных элементов бортовых систем. Увеличение количества наблюдаемых объектов позволяет алгоритмам машинного обучения повысить качество вычислений характеристик для задач построения регрессий, классификаций и распознавания образов.

Для эффективного выявления проблем с ИБ внутри бортовых сетей объектов морской техники необходимо иметь точную информацию о состоянии каждого объекта, что в условиях ограниченных ресурсов бортовых систем невозможно без сегментирования этого множества данных.

В данной работе рассматривается проблема сегментирования множества данных с учетом информации воздействующих факторов для выявления проблем с обеспечением информационной безопасности объектов морской техники. Сегментация данных позволяет выделить группы объектов с похожими

характеристиками, что упрощает процесс анализа и принятия решений. В работе будут рассмотрены различные методы сегментации данных и их применение для выявления проблем связанных с ИБ объектов морской техники. Также будет проведен анализ воздействующих факторов на состояние систем мониторинга ИБ объектов морской техники и их учет при сегментации данных. Результаты данной работы могут быть использованы для повышения качества работы систем мониторинга информационной безопасности объектов морской техники.

Обоснование актуальности темы:

Тема «Сегментирование множества данных с учетом информации воздействующих факторов для оценки состояния объектов морской техники» является актуальной в современном мире, где судоходный транспорт играет важную роль в экономике и безопасности государства.

Оценка состояния систем мониторинга информационной безопасности объектов морской техники является важной задачей, которая позволяет предотвратить несанкционированное проникновение внутрь бортовых систем, что может спровоцировать дальнейшие отрицательные эффекты на работу всей внутренней системы судна. Для этого необходимо проводить мониторинг состояния ИБ объектов морской техники и анализировать полученные данные.

Однако, данные, получаемые при мониторинге, могут быть очень разнообразными и содержать множество факторов, влияющих на состояние объектов морской техники. Поэтому, для эффективного анализа данных необходимо проводить их сегментацию с учетом воздействующих факторов.

Сегментация данных позволяет разбить множество данных на группы схожих объектов, что упрощает анализ и позволяет выявить закономерности в данных. При этом, учет воздействующих факторов позволяет более точно проводить последующий мониторинг ИБ объектов морской техники и предотвратить возможные проблемы.



Таким образом, тема данной выпускной квалификационной работы является актуальной и важной для развития морской техники и повышения безопасности на море.

Целью данной выпускной квалификационной работы является разработка метода оценки состояния объектов морской техники с использованием моделей машинного обучения.

Объектом исследования являются устройства и узлы бортовых автоматизированных систем.

Предметом исследования являются методы оценки состояния объектов морской техники.

Для достижения поставленной в рамках выпускной квалификационной работы цели были решены следующие частные научно-технические задачи:

- проведен анализ возможности функционирования бортовых систем;
- обоснованы требования по надежности и безопасности функционирования этих систем;
- произведен обзор методов анализа работоспособности в ходе функционирования устройств и узлов бортовых систем;
- изучены существующие методы и модели машинного обучения, используемые для анализа и прогнозирования состояния бортовых систем судна;
- разработан метод сегментирования множества данных с учетом информации об объектах морской техники.
- произведена оценка эффективности предложенного решения.

Для решения поставленных задач в выпускной квалификационной работе использовались методы математического, теории вероятностей и математической статистики, теории классификации, теории графов, теории защиты информации, теории надежности, методы машинного обучения.

Для анализа и обработки данных были использованы современные языки программирования такие как Matlab и Python программные пакеты и библиотеки машинного обучения, TensorFlow, Keras, Pandas и другие.

Научная новизна работы заключается в разработке методов сегментирования множества данных с учетом информации об объектах морской техники. Данный метод позволяет улучшить точность и эффективность оценки состояния бортовых систем за счет более точного определения параметров систем и прогнозирования их состояния на основе адаптивной выборки.

Выпускная квалификационная работа состоит из введения, трех глав, заключения и списка использованной литературы. Содержание первой главы полностью посвящено теоретической части по теме объекта исследования. Во второй главе рассматривается предлагаемый метод решения поставленной проблемы. В третьей главе приводится, описание экспериментальной части и оценка применимости разработанного метода.

В процессе написания данной работы использовались научные и учебно–методические работы отечественных и зарубежных авторов, публикации в научных журналах и материалы периодической печати.

Практическая значимость работы заключается в возможности использования разработанного метода для создания дополнительных систем мониторинга состояния бортовых автоматизированных систем. Результаты исследования могут быть использованы морскими компаниями и судостроительными предприятиями для повышения надежности и безопасности судов.

Таким образом, предложенный подход имеет потенциал для применения в морской отрасли и может внести значительный вклад в область разработки морских информационных систем.

# ГЛАВА 1. ОЦЕНКА СОСТОЯНИЯ ОБЪЕКТОВ МОРСКОЙ ТЕХНИКИ

Первая глава посвящена изучению процессов оценки состояния объектов морской техники. В рамках данной главы будут рассмотрены виды и классификации бортовых систем, а также методы их оценки. В результате выполнения работы будет предложен метод сегментирования множества данных с учетом информации об объектах морской техники.

## 1.1 Виды и классификация объектов морской техники

Под понятием «объекты морской техники» далее будут рассматриваться судоходные бортовые системы. Судоходные бортовые системы представляют собой комплекс технических средств и оборудования, установленных на судне для обеспечения ее безопасности и эффективности во время плавания. Подробная классификация будет рассмотрена в следующих пунктах.

### 1.1.1 Навигационные системы

Навигационные системы являются одними из наиболее важных бортовых систем на судне, которые обеспечивают надежное определение местоположения и ориентацию судна в морском пространстве. Существует множество различных типов навигационных систем, включая системы глобальной навигации (GPS, GLONASS, Galileo), системы локальной навигации (ЛПСН), системы радионавигации (ДМБН, РЭБ), системы инерциальной навигации (ИНС) и многие другие [1].

Каждая из этих систем имеет свои особенности и применяется в различных условиях и задачах на судне. Например, системы глобальной навигации позволяют определять местоположение судна с высокой точностью в любых точках земного шара, а системы инерциальной навигации обеспечивают высокую точность определения местоположения и ориентации судна даже при отсутствии внешних ориентиров.

В условиях динамической оценки состояния навигационных систем на судне нам нужна возможность для быстрого перераспределения данных для их быстрой актуализации, и последующего применения в алгоритмах машинного обучения, что в свою очередь обеспечивает правильность и стабильность работы навигационных систем. Быстрая актуализация распределения водных данных также способствует более точному предсказанию неисправностей внутри судоводных бортовых систем.

### **1.1.2 Энергетические системы**

Энергетические системы включают в себя комплекс оборудования, необходимый для обеспечения энергией всех систем и устройств судна. Они играют важную роль в обеспечении энергетической безопасности судна, а также в работе его бортовых систем.

Современные энергетические системы на судах могут использовать различные источники энергии, такие как дизельные генераторы, газовые турбины, солнечные батареи, ветрогенераторы и др. Кроме того, могут быть установлены системы энергосбережения, например, системы рекуперации тепла и системы регулирования мощности генераторов.

Оценка состояния энергетических систем является важным элементом обеспечения надежной и безопасной работы судна. Для этого используются различные методы и средства мониторинга и контроля, такие как системы автоматического контроля энергопотребления, датчики параметров работы генераторов и турбин, системы контроля уровня топлива и масла, системы контроля температуры и давления в системах охлаждения.

Целью оценки состояния энергетических систем является обеспечение их надежной и безопасной работы во время эксплуатации судна. Это позволяет предотвратить возможные аварии, связанные с неисправностью энергетических систем, а также оптимизировать энергопотребление и снизить затраты на топливо.

### **1.1.3 Системы связи и информационные системы**

Неотъемлемой частью современного комплекса судоводных бортовых систем являются системы, которые обеспечивают передачу данных и команд между различными устройствами. Они включают в себя различные компоненты, такие как радиолокационное оборудование, радиосвязь, системы спутниковой связи и другие.

Радиолокационное оборудование используется для обнаружения и отслеживания объектов, а также для определения расстояний и направлений. Радиосвязь обеспечивает передачу голосовых и текстовых сообщений между различными членами экипажа, а также между кораблем и внешним миром. Системы спутниковой связи используются для передачи данных и команд на большие расстояния и в условиях отсутствия наземной связи.

Также важным компонентом являются информационные системы, которые обеспечивают сбор, обработку и передачу различных данных на борту корабля. Они могут включать в себя системы автоматической и ручной обработки данных, системы навигации и позиционирования, а также системы мониторинга и контроля за различными параметрами корабля.

Подобные системы могут быть использованы для управления бортовыми системами, обеспечения безопасности и защиты корабля, а также для сбора и передачи различных данных, необходимых для принятия решений экипажем.

### **1.1.4 Системы безопасности и защиты**

Системы безопасности и защиты являются неотъемлемой частью современных транспортных средств, в частности судоводного транспорта. Они обеспечивают безопасность перевозимых людей и грузов, а также защиты транспортных средств от угроз внешней и внутренней среды.

Системы безопасности и защиты могут включать в себя различные устройства и компоненты, такие как:

1. Системы контроля и управления движением транспортного средства, включая автоматические тормозные системы, системы управления ускорением и т.д.

2. Системы контроля за внутренней и внешней средой транспортного средства, включая системы контроля за температурой, освещением, качеством воздуха и т.д.

3. Системы обнаружения и предотвращения возгораний и аварийных ситуаций, включая системы пожарной безопасности, системы контроля за давлением в шинах и т.д.

4. Системы обеспечения личной безопасности пассажиров, включая подушки безопасности, ремни безопасности, системы автоматического оповещения о происшествиях и т.д.

В зависимости от типа транспортного средства и условий эксплуатации, могут использоваться различные системы безопасности и защиты. Например, для морских судов могут использоваться системы автоматической пожарной защиты, системы спасательных шлюпок и т.д., в то время как для автомобилей - системы контроля за давлением в шинах, системы стабилизации и т.д.

Оценка состояния систем безопасности и защиты транспортных средств является важной задачей, поскольку от их исправной работы зависит безопасность перевозимых людей и грузов. В рамках данной работы будут рассмотрены различные методы и подходы к оценке состояния систем безопасности и защиты на примере бортовых систем различных видов транспорта.

### **1.1.5 Автоматические системы управления**

Автоматизированные системы управления (АСУ) представляют собой комплекс технических средств и программных компонентов, предназначенных для автоматизации управления процессами в различных сферах, включая морскую промышленность. Они используются для контроля и управления различными системами судна, такими как система управления двигателями, система контроля и управления электрооборудованием, система управления

грузоподъемностью, система управления расходом топлива, система управления скоростью и направлением движения, система контроля загрузки, система навигации и другими.

Одним из главных преимуществ использования АСУ в морской промышленности является повышение эффективности и надежности управления процессами на судне. Автоматизированные системы управления позволяют сократить число ошибок, связанных с человеческим фактором, повысить скорость принятия решений и уменьшить время реакции на непредвиденные ситуации.

Существует несколько видов АСУ, в зависимости от сферы применения и уровня автоматизации процессов. Например, на малых судах могут использоваться относительно простые системы автоматизации, такие как системы автоматического управления движением, системы автоматического управления промышленными процессами и т.д. На больших судах, включая танкеры, контейнеровозы и другие, применяются более сложные системы автоматизации, включающие в себя датчики, контроллеры, программное обеспечение и другие компоненты.

Важным аспектом разработки и применения АСУ является обеспечение их надежности и безопасности. Для этого применяются различные методы и стандарты, включая стандарты по ИБ и надежности, а также методы анализа рисков.

В дальнейшем в рамках данной работы будет рассмотрен подход к оценке состояния автоматизированных систем управления на судне при помощи моделей машинного обучения, что позволит улучшить надежность и эффективность их работы.

### **1.1.6 Электронные системы**

Электронные системы широко используются на современных судах для управления и контроля различных систем и процессов. Они могут быть

разделены на несколько категорий в зависимости от функций, которые они выполняют.

Существует множество различных электронных систем, используемых на кораблях. К ним относятся:

- системы автоматического управления кораблем;
- системы дистанционного зондирования и контроля параметров окружающей среды;
- системы бортового электропитания;
- системы автоматического контроля и диагностики технического состояния оборудования;
- системы связи и навигации;
- системы контроля и управления радиационной обстановкой;
- системы управления балластом и т.д.

Каждая из этих систем выполняет свои специфические функции, но все они имеют общую цель - обеспечение безопасности и эффективности работы корабля. В связи с этим, разработка и совершенствование электронных систем является актуальной задачей для судостроительной отрасли.

Одной из самых важных категорий электронных систем являются системы автоматического управления судном. Они позволяют контролировать движение судна, управлять его курсом и скоростью, а также обеспечивать безопасность при маневрировании. Системы автоматического управления судном могут быть основаны на различных принципах, таких как гироскопический или акустический.

Еще одной важной категорией электронных систем на судне являются системы управления энергетическими ресурсами. Они могут включать в себя контроль генераторов, батарей и других источников энергии, а также системы автоматического переключения между источниками энергии в зависимости от текущей нагрузки.

Также электронные системы на судне могут включать в себя системы навигации и обнаружения объектов в окружающей среде, такие как радары,



эхолоты и системы определения местоположения. Эти системы необходимы для обеспечения безопасности судна и его экипажа, а также для выполнения различных задач, таких как поиск рыболовных мест или избежание столкновений с другими судами.

Кроме того, на судне могут быть установлены электронные системы контроля и управления механическим оборудованием, такие как системы автоматического управления двигателями, системы контроля топливного расхода и системы диагностики и ремонта.

## **1.2 Оценка состояния объектов морской техники**

### **1.2.1 Оценка состояния объектов морской техники**

Оценка состояния объектов морской техники является важным аспектом обеспечения безопасности плавания. Для этого используются различные методы и технологии, которые позволяют определить состояние систем и обнаружить возможные угрозы.

Одним из методов оценки состояния бортовых систем является мониторинг и диагностика, которые позволяют следить за работой систем в режиме реального времени и обнаруживать отклонения от нормы. Для этого могут применяться различные сенсоры, контроллеры и программное обеспечение, которые обеспечивают сбор данных и их анализ. Также оценить состояние бортовых систем на судне можно технического обслуживания и ремонта, которые проводятся с целью предотвращения возможных неисправностей и обеспечения надежной работы систем. Для этого используются различные методы диагностики и испытания, которые позволяют выявить неисправности и подобрать оптимальные методы и средства их устранения.

Еще одним важным аспектом оценки состояния систем является обучение и подготовка персонала, который должен обеспечивать их работу. Для этого проводятся различные тренинги и курсы, которые позволяют освоить необходимые знания и навыки, связанные с обслуживанием и ремонтом систем.

Таким образом, оценка состояния судоводных бортовых систем является важным аспектом обеспечения безопасности плавания и требует комплексного подхода, который включает в себя мониторинг и диагностику, техническое обслуживание и ремонт, а также обучение и подготовку персонала.

### **1.2.2 Понятие надежности и ее модели**

Понятие надежности – это свойство системы или устройства, которое определяет его способность сохранять работоспособность в течение определенного времени при заданных условиях эксплуатации. Оценка надежности морской техники является важным аспектом ее проектирования и эксплуатации, так как от нее зависит безопасность на море и экономическая эффективность.

Модели надежности – это математические модели, которые описывают вероятность отказа системы или устройства в зависимости от времени эксплуатации. Эти модели используются для оценки надежности и прогнозирования срока службы техники.

#### **1.2.2.1 Расчет надежности судоводных бортовых систем**

Рассчитать надежность морской техники можно с помощью анализа ее технических характеристик, исторических данных о ее эксплуатации и методов математического моделирования. Ниже представлено несколько способов.

1. Анализ технических параметров. Оцените надежность морской техники, изучив ее технические характеристики, такие как мощность, скорость, грузоподъемность, длина и ширина. Оцените, насколько эти параметры соответствуют условиям эксплуатации, и сравните их с параметрами аналогичной техники.

2. Исторический анализ. Изучите исторические данные о эксплуатации морской техники, такие как ее рабочий наработка, степень износа, частота ремонта и поломок. Оцените, как часто возникают проблемы с определенными компонентами, и какие меры были приняты для их устранения.

3. Математическое моделирование. Используйте математические модели для расчета надежности морской техники. Эти модели могут включать в себя статистические данные, такие как распределение вероятности времени наработки на отказ, или физические модели, которые учитывают влияние различных факторов, таких как вибрации, температура и влажность.

4. Экспертный анализ. Проведите экспертную оценку надежности морской техники, пригласив специалистов в отрасли. Это может включать в себя оценку качества производства, квалификацию операторов, уровень обслуживания и т.д.

### **1.2.3 Виды состояний бортовых систем**

Надежность бортовых систем на судне – это способность системы выполнять свои функции без сбоев и отказов на протяжении всего срока эксплуатации судна. В зависимости от функционального назначения и характера работы, бортовые системы могут быть подвержены различным видам состояний.

1. Нормальное состояние - когда система работает без сбоев и выполняет свои функции в соответствии с требованиями.

2. Критическое состояние - когда происходит сбой системы, который может привести к серьезным последствиям для безопасности судна и его экипажа.

3. Аварийное состояние - когда происходит полный отказ системы, что может привести к аварии на судне.

4. Предаварийное состояние - когда происходит сбой системы, который может привести к аварии, но еще не достиг критического уровня.

Для обеспечения надежности бортовых систем на судне необходимо проводить регулярное техническое обслуживание и проверку всех систем на соответствие требованиям безопасности и надежности. В случае обнаружения неисправностей, необходимо незамедлительно принимать меры по их устранению, чтобы избежать возможных аварийных ситуаций.

Оценка состояния процессов функционирования бортовых систем на судне является важной задачей для обеспечения безопасности плавания и сохранности жизни и здоровья людей на борту. Для этого используются различные методы и технические средства, такие как:

1. Системы мониторинга и диагностики (СМид). Это комплексное оборудование, позволяющее непрерывно контролировать работу всех систем и устройств на судне, а также выявлять и диагностировать возможные неисправности.

2. Визуальный осмотр и проверка состояния оборудования. Этот метод включает в себя регулярный осмотр и проверку состояния всех систем, узлов и устройств на судне, а также их техническое обслуживание и ремонт при необходимости.

3. Испытания и проверки систем и устройств на судне. Для этого могут применяться различные методы, такие как испытание на герметичность, проверка на прочность, испытание на нагрузку и т.д.

4. Анализ данных и статистических показателей. Путем анализа данных о состоянии оборудования и эксплуатации судна можно выявить возможные проблемы и улучшить процессы функционирования бортовых систем.

В целом, оценка состояния процессов функционирования бортовых систем на судне должна проводиться регулярно и комплексно, с использованием различных методов и средств контроля и диагностики. Это позволит обеспечить безопасность плавания и предотвратить возможные аварии и неисправности на борту судна.

Определение устойчивости работоспособности бортовых систем судна является комплексным процессом, который включает в себя следующие этапы:

1. Анализ технической документации на судно и его системы. Необходимо изучить характеристики систем, их назначение, принципы работы, технические характеристики и требования к их эксплуатации.

2. Инспекция бортовых систем судна. Необходимо проверить работоспособность каждой системы и ее компонентов, состояние оборудования, наличие повреждений, износа и других дефектов.

3. Оценка рисков. Необходимо оценить вероятность возникновения неполадок в работе систем и их последствий для судна и экипажа.

4. Разработка мероприятий по обеспечению устойчивости работоспособности систем. На основе результатов анализа и инспекции необходимо разработать мероприятия по устранению выявленных дефектов и профилактике возможных неполадок.

5. Проведение испытаний. После проведения мероприятий необходимо провести испытания систем, чтобы убедиться в их работоспособности и устойчивости.

6. Мониторинг и контроль. Для поддержания устойчивости работоспособности систем необходимо проводить регулярный мониторинг и контроль и проводить необходимые ремонтные работы в случае выявления дефектов.

#### **1.2.4 Контроль качества оценки судноходны бортовых систем**

Все эти этапы должны проводиться в соответствии с международными стандартами и требованиями безопасности судоходства.

Самым важным стандартом в Российской Федерации является ГОСТ. ГОСТ является основным документом, регулирующим качество продукции, услуг и процессов в Российской Федерации. Он устанавливает требования к продукции, услугам, процессам и системам управления, которые должны быть соблюдены производителями и поставщиками, чтобы обеспечить высокое качество и безопасность продукции и услуг.

Для оценки состояния процессов функционирования бортовых систем судна ГОСТы могут быть важными в следующих аспектах:

1. ГОСТы, регулирующие качество и безопасность судовых систем и оборудования, могут использоваться для оценки соответствия устройств бортовых систем судна требованиям стандартов. Это может быть полезно при

проведении идентификации состояния устройств и определении возможных проблем [2].

2. ГОСТ Р 56243-2014 «Внутренний водный транспорт. Система управления безопасностью судов. Требования по обеспечению надежности механизмов».

3. ГОСТ 32455-2013 «Глобальная навигационная спутниковая система. Морская навигационная аппаратура потребителей. Приемные устройства. Общие требования, методы и требуемые результаты испытаний» [3].

4. - ГОСТы, которые регулируют процессы управления и контроля качества, могут быть применены для разработки стратегии идентификации состояния устройств бортовых систем судна на основе беггинга. Эти стандарты могут помочь определить процедуры и требования для контроля качества и обеспечения соответствия требованиям.

5. ГОСТ Р ИСО 10006–2019 «Менеджмент качества. Руководящие указания по менеджменту качества в проектах».

6. ГОСТы, которые регулируют оборудование и технические требования, могут использоваться для оценки технической готовности устройств бортовых систем судна к выполнению своих функций. Это может быть особенно полезно при определении возможных проблем и уязвимых мест системы.

В целом, ГОСТы играют важную роль в идентификации состояния устройств бортовых систем судна, обеспечивая соответствие требованиям стандартов и помогая определить возможные проблемы и уязвимости системы.

### **1.3 Постановка задачи**

В рамках первой главы была изучена теоретическая составляющая базовых терминов и определений понятия бортовая система и ее устройств. Установлено, что бортовая система – это сложная составляющая любого судна независимо от его назначения. Следовательно, сбои и неточная работа системы мониторинга информационной безопасности внутренних бортовых

сетей могут привести к нарушению целостности данных хранимых и обрабатываемых данными системами, к их краже или к сбою работы всего судна. За этим следует, что идентифицировать состояния мониторинга ИБ объектов морской техники необходимо для любой из существующих бортовых систем.

Одним из проблемных вопросов классификации состояния информационной безопасности является использование моделей машинного обучения, которые повышают качество своей работы благодаря использованию уточненной локальной информации [4].

Обычно обучающая выборка рассматривается как единое множество. Однако на составляющие ее кортежи данных могут влиять различные факторы. Например, появление отдельных управляющих команд вызывает увеличение количества служебных сообщений в сетевом трафике, что может существенно повлиять на качество идентификации событий ИБ. Многие факторы, влияющие на значения переменных обучающей выборки, известны заранее. В связи с этим появляется возможность идентифицировать кортежи обучающей выборки, полученные в момент воздействия.

Пусть  $X^P$  выборка размером  $p$ ,  $a(x)$  – классифицирующий алгоритм.

Имеется множество факторов  $\{x_1, v_2, \dots, v_m\} \in \mathcal{V}$ , влияющих на значения целевых переменных в кортежах. Функция (1) представляет собой функции потерь в момент действия фактора  $v$ :

$$L(a, x, v) \tag{1}$$

Средняя величина ошибки с учетом действия фактора  $v$  определяется выражением:

$$Q(a, X^p, V) = \frac{1}{p} \sum_{x \in X} L(a, x, v) \tag{2}$$

Необходимо минимизировать функционал выражением:

$$Q(a, X^p, V) \rightarrow \min \tag{3}$$

Данная постановка позволяет учитывать воздействие известных факторов, которые могут вызывать эффекты, влияющие на разброс, смещение ответов классификаторов.

В предлагаемом методе изначально предполагается, что действуют факторы, которые оказывают влияние на свойства данных. В их качестве могут выступать команды, информационные воздействия, события, связанные с изменением среды.

Формируемая обучающая выборка состоит из кортежей, значения которых получены в условиях действия факторов или в результате искусственно смоделированных событий, вызывающих инциденты ИБ. Информация о внешних и внутренних воздействиях используется для разделения подмножеств таким образом, чтобы уменьшить количество зашумленных объектов и улучшить свойства различимости классов.

Применение метода можно рассмотреть для классификационных задач [5]. В качестве функции для измерения потерь алгоритма классификации  $a(x_i)$ , действующего на выборке  $X^p$ , рассмотрим индикатор ошибок формулой:

$$I(x, a) = [x_i \neq a(x_i)] \quad (4)$$

Частота ошибок алгоритма  $a(x)$  определяется выражением (5):

$$v(a, X^p) = \frac{1}{p} \sum_{i=1}^p I(x_i, a(x_i)) \quad (5)$$

На регистрируемые данные осуществляют воздействие факторы  $v$ . Разобьем множество  $X^p$  на подмножества с учетом влияния на данные  $m$  факторов  $v_i \in v, i = 1 \dots m$ . Зависимость представлена выражениями:

$$X^p = X_1^{p_1} \cup X_2^{p_2} \cup \dots \cup X_m^{p_m} \quad (6)$$

$$\sum_{i=1}^m p_i = p \quad (7)$$

Каждое образованное в результате анализа действия факторов подмножество  $X_i^{p_i} \in X^p$  делится на обучающие и контрольные выборки:

$$X_i^{p_i} = X_{n_i}^{l_i} \cup X_{n_i}^{k_i} \quad (8)$$

$$p_i = l_i + k_i \quad (9)$$

где  $n_i=1, \dots, N_i$  – варианты разбиения выборки;



$X_i^{p_i}$ ,  $l_i$  и  $k_i$  – длины обучающей и контрольной подвыборок.

Функционал эмпирического риска  $Q_i$  для выборки  $X_i^{p_i}$ , определяемой влиянием фактора  $v_i$ :

$$Q_i(a, X_i^{p_i}) = \frac{1}{N_i} \sum_{n_i=1}^{N_i} v(a(X_{n_i}^{l_i}), X_{n_i}^{k_i}) \quad (10)$$

Полученному с учетом влияния факторов подмножеству может быть назначен классификатор.

В условия предположения, что выборка является простой и повторяет свойства генеральной совокупности, возникает возможность рассмотреть версии алгоритмов  $a = \{a_1, a_2, \dots, a_r\}$  и выбрать для множества  $X_i^{p_i}$  классификатор  $a_j$  с учетом условия:

$$Q_i(a_j, X_i^{p_i}) \quad (11)$$

В этом случае алгоритм классификации можно обучать на каждом сегменте данных отдельно. За счет манипуляции выборками, в ряде случаев можно улучшить качественные показатели алгоритма.

В первой главе работы было проведено обзорное исследование по теме идентификации и актуализации информации о состоянии систем мониторинга информационный безопасности внутренних сетей судна. Были рассмотрены основные подходы к решению задачи идентификации состояния систем, а также приведены примеры их применения в судостроении.

Было отмечено, что идентификации состояния систем мониторинга ИБ бортовых систем является сложной и актуальной задачей, так как позволяет своевременно выявлять угрозы, предотвращать кражу или подмену данных и повысить безопасность судоходства.

## **ГЛАВА 2 ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИИ О ВОЗДЕЙСТВУЮЩИХ ФАКТОРАХ ДЛЯ ОЦЕНКИ СОСТОЯНИЯ ОБЪЕКТОВ МОРСКОЙ ТЕХНИКИ**

В рамках второй главы выпускной квалификационной работы рассматривается возможность применения методов машинного обучения для оценки состояния объектов морской техники. Машинное обучение является мощным инструментом, позволяющим извлекать информацию из данных и применять ее для прогнозирования и оценки различных параметров и состояний.

В разделе 2.1 исследуется применение методов машинного обучения в контексте мониторинга информационной безопасности внутренних сетей бортовых систем. Также приведена информация о различных методах машинного обучения и проведена их классификация, для того чтобы понять их особенности и применимость. Особое внимание будет уделено областям применения этих методов, включая морские информационные системы, где оценка состояния бортовых систем имеет важное значение для надежного и безопасного функционирования.

В разделе 2.2 приведен анализ преимуществ и ограничений существующих подходов к применению моделей машинного обучения для оценки состояния бортовых систем, учитывая такие факторы, как вычислительные затраты, доступность, объем данных, особенности и распределение данных, а также сложности при обработке и анализе больших объемов информации. Это позволит выявить наиболее эффективные методы машинного обучения в контексте оценки состояния бортовых систем.

В разделе 2.3 исследуется возможность повышения качества оценки состояния бортовых систем. Исследуется возможность повышения качества мониторинга информационной безопасности бортовых систем путем сегментации выборки данных. Сегментация данных позволяет адаптировать модели машинного обучения к различным сегментам системы и улучшить точность их прогнозов. Мы исследуем различные подходы к сегментации данных и их

применимость в контексте оценки состояния бортовых систем, особенно в условиях переменных условий и неопределенностей.

## **2.1 Применение методов машинного обучения для оценки состояния бортовых систем**

Машинное обучение — это наука о разработке алгоритмов и статистических моделей, которые используются компьютерными системами для выполнения задач без четких инструкций, полагаясь вместо этого на закономерности и результаты. Таким образом, системы могут прогнозировать результаты на основе ранее известного набора данных, что открывает широкий спектр возможностей, охватывающих многие области деятельности человека.

### **2.1.1 Области применения методов машинного обучения**

Методы машинного обучения с каждым годом находят все большее применение в различных областях, где требуется анализ больших объемов данных и прогнозирование будущих событий. В данной главе мы рассмотрим основные области применения методов машинного обучения. Каждая из этих областей имеет свои особенности и требует специфических подходов к решению задач, что делает использование методов машинного обучения еще более интересным и актуальным.

Основные области применения машинного обучения:

#### **1) Бортовые системы**

Методы машинного обучения нашли широкое применение в области бортовых систем, которые используются в авиации, судоходстве, автомобильной промышленности и других смежных отраслях. Эти системы отвечают за управление и мониторинг различных аспектов работы транспортных средств, обеспечивая безопасность, эффективность и комфорт пассажиров. Вот несколько областей, где методы машинного обучения применяются в бортовых системах:

Прогнозирование технического состояния различных систем и компонентов транспортных средств. На основе данных о прошлых поломках,

измерениях и других параметрах модели машинного обучения могут предсказывать вероятность возникновения отказов или неисправностей в будущем. Это позволяет проводить про активное техническое обслуживание, устраняя проблемы до их возникновения, и повышать надежность и безопасность транспортных средств.

1. Распознавание образов и детекция объектов на основе данных, полученных с помощью датчиков и камер. Например, в авиации они могут использоваться для автоматического распознавания и классификации навигационных маркеров или для обнаружения и отслеживания других воздушных судов. В автомобильной промышленности методы машинного обучения применяются для систем помощи водителю, таких как распознавание знаков дорожного движения и предупреждение о возможных столкновениях.

2. Оптимизация управления различными аспектами бортовых систем, такими как управление двигателем, автопилот или системы поддержания стабильности. Модели машинного обучения могут обучаться на основе исторических данных и адаптироваться к изменяющимся условиям для достижения оптимальной производительности и эффективности.

3. Распознавание и обработка речи: Методы машинного обучения применяются для распознавания и обработки речи в бортовых системах, позволяя взаимодействовать с транспортным средством через голосовые команды. Это может быть полезно, например, для управления навигацией, мультимедийными системами или системами управления климатом.

4. Адаптивные системы безопасности: Методы машинного обучения применяются для анализа и прогнозирования потенциальных опасностей и рисков на дороге, в воздухе или на воде. На основе данных о движении, окружающей среде и поведении других объектов, адаптивные системы безопасности могут предупреждать водителя или автоматически принимать меры для предотвращения аварийных ситуаций.

Методы машинного обучения в области бортовых систем значительно улучшают функциональность, безопасность и эффективность транспортных

средств. Они способствуют автоматизации процессов, улучшению прогнозирования и принятию решений на основе данных, повышают уровень комфорта и безопасности для пассажиров и участников движения.

#### 1) Медицина и здравоохранение

Методы машинного обучения имеют широкий спектр применений в медицине и здравоохранении, где они играют важную роль в диагностике, лечении, прогнозировании и улучшении качества здравоохранения. Методы машинного обучения применяются для диагностики и обработки медицинских изображений (рентгенограммы, маммограммы, КТ-снимки и МРТ), прогнозирование заболеваний и результатов лечения (анализ данных пациентов, генетические данные, результаты лабораторных исследований и медицинские истории, для прогнозирования риска развития заболеваний и результатов лечения), анализ медицинских текстов (электронные медицинские записи, научные публикации и медицинские журналы), разработка протезов и систем реабилитации, управление медицинскими данными, оптимизации процессов здравоохранения и принятии решений на основе данных. Это помогает улучшить точность диагностики, прогнозирование результатов лечения, оптимизацию процессов и принятие данных обоснованных решений. Они могут повысить эффективность и качество медицинской помощи, способствовать разработке новых методов лечения и улучшению здоровья и благополучия пациентов.

#### 2) Финансовый сектор

Методы машинного обучения играют важную роль в финансовом секторе, где они применяются для анализа данных, прогнозирования рынков, определения рисков и улучшения финансовых процессов. Благодаря использованию МО производится анализ финансовых данных, истории цен, объемов торговли и других показателей для прогнозирования движений рынка и принятия решений о торговле, Оценка кредитоспособности заемщиков и определения риска в финансовых операциях, управление портфелем и активами.

Машинное обучение значительно улучшает эффективность и точность анализа данных, принятия решений и прогнозирования в финансовом секторе, способствуя более эффективному управлению рисками, повышению доходности и обеспечению финансовой стабильности.

### 3) Промышленность и производство

Промышленность и производственные отрасли также не обходятся без применения машинного обучения, где они помогают оптимизировать производственные процессы, улучшать качество продукции, снижать издержки и повышать безопасность. Применение методов машинного обучения позволяет прогнозировать спрос на производимую продукцию и управлять запасами, управлять качеством и контролировать процессы при производстве продукции, планирования производственных операций, оптимизации распределения ресурсов и снижения издержек.

Методы машинного обучения играют значительную роль в современной промышленности и производстве, помогая компаниям повышать эффективность, снижать издержки и обеспечивать высокое качество продукции.

### 4) Морские информационные системы (МИС)

Методы машинного обучения также находят широкое применение в морских информационных системах (МИС), которые предоставляют поддержку и управление в морских отраслях. Вот некоторые области применения методов машинного обучения в МИС:

1. Повышение безопасности судоходства и оптимизации навигационных процессов. Алгоритмы машинного обучения могут анализировать данные о состоянии морской среды, погодных условиях, движении судов, радарных и судовых датчиках, чтобы предсказывать потенциальные риски, предотвращать столкновения и предлагать оптимальные маршруты.

2. Обнаружение и идентификация объектов различных объектов в морской среде, таких как суда, плавучие препятствия, рыболовные сети и др. Алгоритмы машинного обучения могут анализировать данные с радаров,

видеокамер, гидролокаторов и других датчиков для автоматического обнаружения и классификации объектов, что помогает улучшить морскую безопасность.

3. Прогнозирование морских условий, таких как волны, приливы, течения и погодные условия. Алгоритмы машинного обучения могут анализировать исторические данные, сенсорные измерения, данные о погоде и другие факторы для предсказания будущих изменений в морской среде. Это позволяет морским организациям принимать информированные решения и планировать действия, связанные с безопасностью и производственными операциями.

4. Мониторинг и управление морской экосистемой, включая океанские ресурсы, рыбные запасы и морскую биологическую разнообразность. Алгоритмы машинного обучения могут анализировать данные, полученные от датчиков, и обнаруживать паттерны, связанные с изменениями экосистемы, распространением вирусов или загрязнением морской среды. Это помогает принимать меры для сохранения морской жизни, рационального использования ресурсов и устойчивого развития морских отраслей.

Морские информационные системы, поддерживаемые методами машинного обучения, играют важную роль в обеспечении безопасности, эффективности и устойчивости морских операций. Они помогают принимать информированные решения, оптимизировать процессы и снижать риски, связанные с морскими деятельностью.

### **2.1.2 Основные методы машинного обучения**

Методы машинного обучения в свою очередь представляют собой алгоритмы и модели, которые позволяют компьютерам обучаться на основе данных и опыта, а не на основе жестко заданных правил. Они используются для решения различных задач, таких как классификация, регрессия, кластеризация, обработка естественного языка, распознавание образов и т.д.

Методы машинного обучения могут быть разделены на три категории: обучение с учителем, обучение без учителя и обучение с подкреплением.

#### 1) Методы обучения с учителем

Методы обучения с учителем предоставляют широкий набор алгоритмов, которые позволяют оценивать состояние бортовых систем на основе размеченных данных. В этом подразделе мы рассмотрим три популярных метода обучения с учителем: линейную регрессию, деревья решений и случайный лес.

Линейная регрессия является простым и интерпретируемым методом, который моделирует линейную зависимость между входными признаками и целевой переменной [6]. Она стремится найти оптимальные значения коэффициентов, которые минимизируют сумму квадратов разности между прогнозируемыми значениями и фактическими значениями целевой переменной. Линейная регрессия широко используется для прогнозирования количественных переменных и может быть применена в оценке состояния бортовых систем для предсказания параметров, таких как эффективность, потребление энергии и т.д.

Деревья решений являются графическим представлением правил принятия решений. Каждый узел дерева представляет собой тест на определенный признак, а каждое ребро соответствует возможному значению этого признака. Дерево решений разделяет пространство признаков на различные регионы и позволяет классифицировать или предсказывать целевую переменную на основе значений признаков. Деревья решений могут быть использованы для оценки состояния бортовых систем путем классификации состояний или предсказания качественных переменных. Например, дерево решений может быть применено для определения состояния системы (нормальное, аварийное, предаварийное) на основе различных признаков, таких как температура, давление, скорость и другие.

Случайный лес представляет собой ансамбль деревьев решений. Он объединяет несколько деревьев решений и использует метод баггинга для получения итоговых прогнозов. Каждое дерево в случайном лесу строится на



основе подвыборки данных с возвращением, а также случайного подмножества признаков. Это позволяет создавать разнообразные деревья, которые в совокупности образуют более точную модель. Случайный лес широко применяется в оценке состояния бортовых систем для предсказания параметров или классификации состояний. Он обладает хорошей устойчивостью к переобучению и способен обрабатывать большие объемы данных.

## 2) Методы обучения без учителя

Методы обучения без учителя предоставляют инструменты для анализа неразмеченных данных и выявления скрытых закономерностей или структуры в данных. В этом подразделе мы рассмотрим два популярных метода обучения без учителя: кластеризацию и метод главных компонент (РСА).

Кластеризация является методом группировки объектов внутри набора данных на основе их сходства. Целью кластеризации является обнаружение скрытых групп или кластеров, где объекты внутри каждого кластера более похожи друг на друга, чем на объекты из других кластеров. Алгоритмы кластеризации могут определять число кластеров автоматически или требовать задания количества кластеров. Кластеризация может быть полезна при оценке состояния бортовых систем для выявления аномалий или группировки подобных состояний. Например, можно использовать кластеризацию для выделения групп схожих рабочих режимов или состояний системы, что поможет в дальнейшем анализе и принятии решений.

Метод главных компонент (РСА) является статистическим методом, используемым для снижения размерности данных и извлечения наиболее значимых признаков. Он находит новые оси, называемые главными компонентами, которые являются линейными комбинациями исходных признаков и обладают свойством максимальной дисперсии. Главные компоненты упорядочены по убыванию вклада в объяснение общей дисперсии данных. Метод главных компонент позволяет сократить размерность данных и сжать информацию, сохраняя при этом наиболее важные характеристики данных. В контексте оценки состояния бортовых систем РСА может быть использован для сжатия данных

с большим числом признаков или для визуализации пространства признаков с помощью двух или трех главных компонент.

Данные методы обучения без учителя позволяют извлекать полезную информацию из данных без предварительного разметки и могут быть применены в задачах оценки состояния бортовых систем для обнаружения аномалий, группировки или сокращения размерности данных.

### 3) Методы обучения с подкреплением

Методы обучения с подкреплением (reinforcement learning) отличаются от методов обучения с учителем и без учителя тем, что агент взаимодействует с окружающей средой, принимая последовательность действий, и получает положительные или отрицательные награды в зависимости от результатов своих действий. Целью обучения с подкреплением является нахождение оптимальной стратегии действий для агента, которая максимизирует накопленную награду на протяжении времени.

Одним из популярных методов обучения с подкреплением является Q-обучение (Q-learning). В Q-обучении агент стремится научиться оценивать ценность состояний и действий, используя функцию Q. Функция Q представляет собой оценку ожидаемой награды агента при выполнении определенного действия в конкретном состоянии. Агент обновляет значения функции Q на основе полученных наград и использует их для принятия решений о следующем действии.

Глубокое Q-обучение (deep Q-learning) представляет собой комбинацию метода Q-обучения с глубокими нейронными сетями. Вместо использования таблицы значений Q, глубокое Q-обучение использует нейронную сеть для аппроксимации функции Q. Это позволяет агенту обрабатывать большие объемы данных и работать с непрерывным пространством состояний и действий. Глубокие нейронные сети в глубоком Q-обучении обучаются при помощи метода обратного распространения ошибки, который минимизирует разницу между оценками Q-значений, полученными от сети, и фактическими наградами агента.

Методы обучения с подкреплением, такие как Q-обучение и глубокое Q-обучение, могут быть полезными при оценке состояния бортовых систем, особенно при наличии динамического окружения и вариабельности входных данных. Они позволяют агенту изучать и приспосабливаться к изменяющимся условиям и выбирать оптимальные действия для достижения поставленных целей. Применение этих методов может помочь в анализе и определении состояния бортовых систем на основе их поведения и взаимодействия с окружающей средой.

### **2.1.3 Классификация методов машинного обучения для оценки состояния объектов морской техники:**

К методам машинного обучения для оценки состояния объектов морской техники относятся:

- статистические методы,
- нейронные сети и глубокое обучение,
- методы ансамблей,
- гибридные методы.

Статистические методы широко применяются в области оценки состояния бортовых систем для анализа данных, выявления зависимостей и построения прогностических моделей. Два распространенных статистических метода, используемых для оценки состояния, включают регрессию и анализ временных рядов.

Регрессия – это статистический метод, используемый для изучения взаимосвязи между зависимой переменной и одной или несколькими независимыми переменными. В контексте оценки состояния бортовых систем регрессия может использоваться для определения влияния различных факторов на состояние системы. Например, можно построить регрессионную модель, чтобы оценить, какие параметры или показатели оказывают наибольшее влияние на работу системы, и использовать эту информацию для предсказания состояния в будущем.

Анализ временных рядов – это метод статистического моделирования и прогнозирования данных, упорядоченных во временной последовательности. В контексте оценки состояния бортовых систем, временные ряды могут представлять данные о работе системы, полученные с течением времени. Анализ временных рядов позволяет исследовать временные зависимости и тренды, а также выявлять сезонные колебания или аномалии в данных. Это может помочь в определении текущего состояния системы и прогнозировании ее поведения в будущем.

Статистические методы, такие как регрессия и анализ временных рядов, предоставляют структурированный и формализованный подход к оценке состояния бортовых систем. Они основаны на математических и статистических принципах, которые позволяют анализировать данные, выявлять закономерности и делать прогнозы. Эти методы могут быть полезны для определения важных факторов, влияющих на состояние системы, и принятия решений на основе статистических выводов.

Нейронные сети и глубокое обучение стали важными инструментами в области оценки состояния бортовых систем благодаря их способности автоматически извлекать сложные закономерности из больших объемов данных. Нейронные сети моделируют работу мозга, состоящего из множества взаимодействующих нейронов, и способны обрабатывать информацию на разных уровнях абстракции.

Глубокое обучение является разделом машинного обучения, в котором используются нейронные сети с несколькими слоями для изучения представлений данных. Вместо ручного определения признаков или характеристик, глубокое обучение позволяет модели самостоятельно извлекать признаки из данных на разных уровнях. Это позволяет обнаруживать более сложные зависимости и позволяет более точно оценивать состояние бортовых систем.

Применение нейронных сетей и глубокого обучения в оценке состояния бортовых систем имеет несколько преимуществ. Во-первых, они могут обрабатывать различные типы данных, включая структурированные,

неструктурированные и временные ряды. Во-вторых, нейронные сети могут автоматически извлекать признаки из сырых данных, что освобождает от необходимости ручного определения признаков. В-третьих, глубокое обучение позволяет создавать модели с высокой степенью сложности, способные улавливать даже тонкие особенности и зависимости в данных.

Применение нейронных сетей и глубокого обучения в оценке состояния бортовых систем может включать различные архитектуры, такие как сверточные нейронные сети (CNN), рекуррентные нейронные сети (RNN) и глубокие нейронные сети (DNN). Эти архитектуры могут быть применены для задач классификации, регрессии, детектирования аномалий и прогнозирования состояния системы.

Нейронные сети и глубокое обучение имеют большой потенциал для улучшения точности и надежности оценки состояния бортовых систем. Однако их эффективное применение требует достаточного объема данных для обучения модели, а также вычислительных ресурсов для обучения и инференса. Кроме того, требуется тщательное проектирование и настройка модели для достижения оптимальных результатов.

Методы ансамблей являются мощным инструментом в машинном обучении, который объединяет несколько моделей для получения более точных и устойчивых прогнозов или оценок. Они основаны на принципе комбинирования прогнозов от нескольких базовых моделей с целью снижения ошибок и повышения обобщающей способности.

Один из наиболее популярных методов ансамблей – это случайный лес. Он состоит из набора решающих деревьев, где каждое дерево строится на случайной подвыборке данных и случайном подмножестве признаков. Конечный прогноз получается путем усреднения прогнозов всех деревьев. Случайный лес обладает способностью эффективно обрабатывать большие объемы данных, обнаруживать важные признаки и обеспечивать хорошую устойчивость к переобучению.

Еще один популярный метод ансамблей – это градиентный бустинг. Он основан на итеративном построении слабых моделей, которые последовательно добавляются в ансамбль, с каждой новой моделью исправляющей ошибки предыдущих моделей. Градиентный бустинг строит сильную модель, объединяя прогнозы слабых моделей с использованием градиентного спуска. Этот метод обладает высокой предсказательной способностью и может обрабатывать различные типы данных.

Преимущества методов ансамблей включают высокую точность прогнозирования, устойчивость к шуму и способность обрабатывать сложные зависимости в данных. Они также могут обнаруживать важные признаки и автоматически выбирать наиболее информативные признаки для моделирования.

Однако методы ансамблей могут быть требовательны к вычислительным ресурсам и времени обучения, особенно для больших наборов данных или сложных моделей. Также важно аккуратно настраивать гиперпараметры моделей ансамблей для достижения оптимальной производительности и избежание переобучения.

В итоге, методы ансамблей, такие как случайный лес и градиентный бустинг, предоставляют эффективные и мощные инструменты для оценки состояния бортовых систем, объединяя прогнозы нескольких моделей и улучшая качество прогнозов.

Гибридные методы машинного обучения представляют собой комбинацию двух или более различных методов с целью улучшения обобщающей способности и эффективности моделей. Эти методы объединяют различные подходы и техники, чтобы достичь более точных и надежных результатов, чем каждый отдельный метод в отдельности.

Гибридные методы машинного обучения могут использовать комбинацию методов обучения с учителем, без учителя и с подкреплением, а также других методов, в зависимости от конкретной задачи и требований. Вот несколько примеров гибридных методов:

1. Стекинг (stacking): метод, в котором прогнозы нескольких базовых моделей сначала объединяются, а затем используются в качестве входных данных для метамодели. Метамодель обучается использовать прогнозы базовых моделей для получения итогового прогноза. Стекинг позволяет комбинировать различные модели и использовать их сильные стороны для улучшения качества предсказаний.

2. Бэггинг и бустинг: гибридные подходы также могут комбинировать методы бэггинга и бустинга для создания более мощных моделей. Например, можно применить градиентный бустинг над ансамблем случайных лесов или использовать случайный лес в качестве базовых моделей для градиентного бустинга. Это позволяет сочетать преимущества обоих методов и достичь лучших результатов.

3. Гибридные нейронные сети: в глубоком обучении также применяются гибридные методы, которые комбинируют различные типы нейронных сетей или архитектуры для решения сложных задач. Например, можно использовать комбинацию сверточных нейронных сетей (Convolutional Neural Networks, CNN) и рекуррентных нейронных сетей (Recurrent Neural Networks, RNN) для обработки изображений и последовательных данных одновременно.

Гибридные методы машинного обучения позволяют использовать сильные стороны различных методов и подходов для достижения лучших результатов при оценке состояния бортовых систем. Они требуют тщательной настройки и адаптации для конкретных задач и данных, но при правильном применении могут значительно повысить качество прогнозов и анализа.

## **2.2 Достоинства, недостатки и ограничения различных методов машинного обучения**

### **2.2.1 Вычислительные затраты и ресурсоемкость**

При выборе методов машинного обучения для применения в морских информационных системах (МИС), необходимо учитывать вычислительные ограничения, связанные с ресурсоемкостью и доступностью вычислительной

мощности на борту судна или в ограниченном морском окружении. Вот некоторые аспекты, которые следует учесть при выборе методов МО с учетом вычислительных ограничений МИС:

1. Необходимо учитывать вычислительную эффективность алгоритмов и возможности доступной вычислительной инфраструктуры на борту судна или в морской среде.

2. Оптимизация моделей – сокращение размерности данных, удаление ненужных признаков или применение аппроксимации моделей.

3. Выбор методов МО должен учитывать специфику морской среды и требования МИС. Например, можно использовать методы с учетом временных рядов, пространственной корреляции или учетом неопределенности данных.

4. Возможность использования распределенных вычислительных ресурсов для улучшения производительности и эффективности МО – распределение вычислений между несколькими узлами, использование параллельных алгоритмов или применение облачных вычислений.

При применении методов машинного обучения в морских информационных системах (МИС), важным аспектом является оптимизация и ускорение вычислительных процессов. Это позволяет обеспечить более эффективное использование ресурсов и достичь более быстрой обработки данных, что особенно важно в реальном времени.

Вот некоторые подходы и методы, которые могут быть применены для оптимизации и ускорения вычислительных процессов в МИС:

1. Использование параллельных вычислений
2. Применение аппаратного ускорения, такого как графические процессоры (GPU) или специализированные процессоры (например, тензорные процессоры)

3. Предварительное вычисление и кэширование результатов с целью сокращения времени выполнения вычислений при повторных запросах или обработке данных.

4. Использование распределенных систем.



### 2.2.2 Недостаток данных и избыточность данных

Одной из распространенных проблем при применении методов машинного обучения в морских информационных системах (МИС) является нехватка данных для обучения моделей. Недостаток данных может возникать из-за различных причин, таких как сложность сбора данных, ограничения доступа к данным или недостаточное количество событий, которые требуется обучить модель.

Для решения проблемы нехватки данных в МИС можно применить различные аналитические и технические подходы, такие как анализ полноты и качества имеющейся выборки, увеличение объема путем генерации новых образцов данных, синтез данных и т.д.

В целом, анализ и решение проблемы нехватки данных требуют тщательного подхода и комбинации различных методов. Цель состоит в том, чтобы использовать имеющиеся данные максимально эффективно и обеспечить надежную оценку состояния бортовых систем в МИС.

Избыточность данных также может быть проблемой при использовании методов машинного обучения в морских информационных системах (МИС). Избыточные данные могут включать повторяющуюся информацию, выбросы, шумы или несбалансированность классов. Это может негативно сказаться на качестве моделей и точности оценки состояния бортовых систем.

Для работы с избыточными данными и улучшения качества моделей можно применить различные подходы, в том числе препроцессинг данных, выбор репрезентативных образцов, применение L1 или L2 регуляризации и т.д.

Работа с избыточными данными требует тщательного анализа и применения соответствующих методов обработки данных. Цель состоит в том, чтобы избежать искажений и неправильных выводов, а также обеспечить качественную оценку состояния бортовых систем в МИС.

### 2.2.3 Свойства данных и их распределение

При выборе и применении методов машинного обучения (МО) в морских информационных системах (МИС) необходимо учитывать свойства обрабатываемых данных, такие как мульти модальность, гетерогенность, особенности распределения, уделять особое внимание разметке данных.

При работе с морскими информационными системами (МИС) возникает необходимость адаптировать модели машинного обучения (МО) к различным типам данных и их распределениям. Различные типы данных в МИС могут иметь разные характеристики и требования к моделированию.

Некоторые методы и подходы к адаптации моделей МО к различным типам данных и их распределениям в МИС включают:

1. Преобразование данных. Например, для данных с неравномерным распределением можно применить логарифмическое преобразование или стандартизацию данных. Для временных рядов могут использоваться методы сглаживания или декомпозиции.

2. Инженерия признаков – создание новых признаков на основе исходных данных.

3. Разработка и применение специализированных моделей, учитывающих особенности и свойства обрабатываемых данных.

4. Оптимизация параметров модели МО.

Адаптация моделей МО к различным типам данных и их распределениям в МИС является важным шагом для достижения точных и надежных результатов оценки состояния бортовых систем. Это позволяет учесть особенности данных, повысить качество моделей и обеспечить более эффективное использование МО в МИС.

## **2.3 Сегментирования выборки данных для повышения показателей полноты и точности оценки состояния объектов морской техники**

### **2.3.1 Существующие подходы**

Развитие и совершенствование методов машинного обучения позволило внедрить и использовать их в таких областях как компьютерное зрение, медицинский анализ, ИБ [7]. Во многих случаях методы машинного обучения превосходят традиционные алгоритмы, основанные на правилах [8]. Эта тенденция также распространяется на область кибербезопасности, в частности безопасности объектов морской техники. Различные системы мониторинга ИБ, системы обнаружения вторжений, средства защиты информации модернизируются с использованием компонентов машинного обучения [9].

Большинство решений по анализу состояний системы и ее элементов базируется на традиционных алгоритмах классификации. Для адаптации методов «поверхностного обучения» требуется эксперт по предметной области, который определяет обучающие данные. Глубокое обучение основывается на многоуровневом представлении входных данных и может выполнять настройку автономно [10, 11].

Подходы поверхностного и глубокого обучения можно дополнительно охарактеризовать, разделяя на контролируемые и неконтролируемые алгоритмы.

Несмотря на довольно большой объем статических данных накопленных в ходе функционирования объектов морской техники сетей бортовых систем, существует ряд проблемных вопросов. Они связаны с тем, что большинство событий в области ИБ являются либо редкими, либо имеют изменяющийся сценарий последовательности действий [12-14]. А в ходе деструктивных воздействий может возникать трансформация свойств анализируемого явления.

В работающей информационной системе данные всегда имеют несовершенную форму. Могут возникать выбросы, появляться шумовые значения. В то же время появляются внешние воздействия, такие как управляющие команды, служебные сообщения, подключение и отключение устройств. Они

могут быть проанализированы заранее и изменять регистрируемые результаты в пределах некоторого заранее оцениваемого диапазона [15].

Использование информации о влияющих на диапазоны значений факторах дает возможность сегментировать выборку данных, что приводит к уменьшению влияния выбросов и сужения диапазонов целевых переменных. В результате получается несколько подмножеств, каждое из которых определено исходя из влияния выбранного фактора.

Одним из основополагающих факторов, определяющих результат алгоритмов машинного обучения, является формирование обучающего подмножества. Наличие качественных выборок данных во многих случаях гораздо важнее качества алгоритмов. В рассматриваемом случае для улучшения качественных показателей осуществляется разделение выборки на отдельные сегменты, где действовали разные факторы.

Количество подмножеств дата сета определяется влияющими факторами. Для каждого фактора определяется подмножество кортежей  $x$  из множества  $X$ , на которые им осуществлялось влияние. Необходимо получить все возможные комбинации разбиения множества на подмножества. Для этого используются индикаторные функции, показывающие присутствие влияния фактора на кортежи значений дата сета.

Количество  $n$  подмножеств определяется формулой:

$$n = \prod_{i=1}^m V_i, \quad (11)$$

где  $m$  – число факторов,

$V_i$  – число диапазонов, на которые делит каждый фактор множество  $X$ .

Используя влияющий на множество фактор возможно провести разбиение множества  $X = X_1 \cup X_2 \cup \dots \cup X_n$  при условиях  $X_j \cap X_i = \emptyset \quad \forall i \neq j$ . Объединение подмножеств совпадает с множеством  $X$ , подмножества не пересекаются.

Индикаторная функция  $\mu$  подмножества  $X_j \in X$ , где  $x$  – кортеж обучающей выборки, может быть представлена в виде:

$$\mu_{X_j}(x) = \begin{cases} 1, & x \in X_j \\ 0, & x \notin X_j \end{cases} \quad (12)$$

Однако на кортежи датасета влияют несколько факторов. В результате получается  $m$  разбиений одного и того же множества  $X$  при действии факторов  $v_1, v_2, \dots, v_m$ :

$$\begin{aligned} v_1: x &= x_{11} \cup x_{12} \cup \dots \cup x_{1n} \\ v_2: x &= x_{21} \cup x_{22} \cup \dots \cup x_{2q} \\ v_m: x &= x_{m1} \cup x_{m2} \cup \dots \cup x_{mp} \end{aligned} \tag{13}$$

Каждому подмножеству соответствует своя индикаторная функция  $\mu_{X_j}(x)$ . В результате при увеличении числа анализируемых факторов происходит рост количества подмножеств [16].

Границы разбиения определяются влиянием фактора. Один и тот же кортеж попадает в разные подмножества при влиянии различных факторов. Факторы могут вызывать эффекты, влияющие на разброс, смещение ответов классификаторов. Для повышения качественных показателей каждой выборке может назначаться свой классифицирующий алгоритм.

Каждому подмножеству соответствует своя индикаторная функция  $\mu_{X_j}(x)$ . при увеличении числа анализируемых факторов происходит рост количества подмножеств. Мощность подмножества определяется влиянием фактора. Один и тот же кортеж может попадать в разные подмножества при влиянии различных факторов [17].

На рисунке 1 показано формирование подмножеств с учетом факторов и их комбинаций.

| Подмножества | Факторы |       |     |       | Правила      |     |                     |     |
|--------------|---------|-------|-----|-------|--------------|-----|---------------------|-----|
|              | $V_1$   | $V_2$ | ... | $V_n$ | $V_1 \& V_2$ | ... | $V_1 \& V_2 \& V_n$ | ... |
| $X_1$        | x       | x     |     |       | x            |     |                     |     |
| $X_2$        | x       | x     |     | x     | x            |     | x                   |     |
| $X_3$        |         | x     |     |       |              |     |                     |     |
| $X_4$        | x       | x     |     | x     | x            |     | x                   |     |
| $X_5$        |         |       |     | x     |              |     |                     |     |
| $X_n$        | x       |       |     | x     |              |     |                     |     |

Рисунок 1 – Формирование подмножеств при влиянии различных факторов и их комбинаций

При формировании обучающей выборки для алгоритмов классификации используется информация о действии известного фактора. Если действует

заданный фактор, то выбирается соответствующее подмножество. Если действуют несколько факторов, то их пересечение, например:  $x \in X_i \cap X_j \leftrightarrow (x \in X_i) \wedge (x \in X_j)$ .

Различные факторы могут вызывать эффекты, влияющие на разброс, смещение ответов классификаторов, иметь явление дрейфа концепта [18]. Для повышения качественных показателей в каждой выборке, в случае вычисления ее свойств, может назначаться свой классифицирующий алгоритм.

Обучающее множество приобретает вид размеченной выборки

$$\{v_j \{(x_i C_i)\}_{i=1}^N\}_{j=1}^M \quad (14)$$

$i$  – количество кортежей в момент, когда оказывал влияние рассматриваемый  $j$ -ый фактор  $v_j$ .

В дальнейшем для повышения качественных показателей на каждую выборку может назначаться свой классифицирующий алгоритм.

Имеется  $m$  классификаторов  $\phi$  на  $n$  сегментов множества  $X_j \in X, j=1, \dots, n$ . Каждый классификатор  $\phi$  обладает своими свойствами  $s$  (полнота, точность, F-мера) на подмножествах  $X_j$

Классификатор  $\phi_k$  может быть назначен на любое подмножество  $X_j$ .

Требуется распределить классификатор таким образом, чтобы произвести классификацию максимально эффективно:

$$L = \sum_{i=1}^n \sum_{j=1}^m c_{ij} t_{ij} \rightarrow \max, \quad (15)$$

где  $c_{ij}$  – выбранный показатель качества при назначении  $i$ -го классификатора на  $j$ -й сегмент множества,

$t_{ij}$  – параметр, принимающий значения 1 или 0 в случае назначения классификатора.

Во второй главе работы было исследовано применение методов машинного обучения (МО) при оценке состояния объектов морской техники в контексте морских информационных систем (МИС). В рамках этой главы были рассмотрены различные аспекты, связанные с повышением качества оценки

состояния бортовых систем за счет сегментирования выборки данных, а также преимущества и ограничения существующих подходов и методов МО.

В пункте 2.2 были рассмотрены преимущества и ограничения существующих подходов. Были выделены вычислительные затраты и ресурсоемкость, проблема нехватки данных и избыточность данных, а также свойства данных и их распределение. Более детально были рассмотрены методы МО, такие как регрессия, кластеризация, нейронные сети и ансамбли, их достоинства и недостатки в контексте МИС. Была также обсуждена адаптация методов МО к МИС, включая применение предварительного обучения, переноса обучения и учет специфических требований и задач МИС при разработке новых методов.

В пунктах 2.3 было представлено разделение данных на сегменты и применение методов МО для каждого сегмента. Были рассмотрены автоматическое и ручное разбиение данных, обучение моделей для каждого сегмента, выбор и оптимизация методов МО, агрегация результатов моделей и учет взаимосвязей между сегментами. Этот подход позволяет улучшить точность и надежность оценки состояния бортовых систем, а также адаптировать модели к различным сценариям и условиям работы.

Основным результатом исследования второй главы является разработка и представление подхода, основанного на сегментировании выборки данных и адаптации методов МО к МИС, для повышения качества оценки состояния бортовых систем. Этот подход позволяет улучшить точность и надежность оценки состояния, а также адаптировать модели к различным сценариям и условиям работы.

Исследование преимуществ и ограничений существующих подходов в контексте МИС позволяет определить наиболее эффективные методы МО, учитывая вычислительные ограничения, доступность данных и их свойства. Это помогает выбрать подходящие методы МО и оптимизировать вычислительные процессы для применения в МИС.

Вторая глава работы содержит информацию о методах МО при оценке состояния объектов морской техники в МИС. Исследованные подходы и

методы предоставляют ценные инструменты для повышения качества оценки состояния, обеспечивая точность, надежность и адаптивность моделей в различных сценариях работы. Однако следует учитывать преимущества и ограничения каждого метода и подхода, а также особенности данных и требования МИС, чтобы выбрать наиболее эффективное решение.



## ГЛАВА 3 СЕГМЕНТИРОВАНИЕ ВЫБОРКИ ДАННЫХ ПРИ ОЦЕНКЕ СОСТОЯНИЯ ИБ БОРТОВЫХ СИСТЕМ

### 3.1 Предлагаемый подход

Для оценки предлагаемого подхода были выбраны различные алгоритмы машинного обучения [19]:

1. Линейный дискриминантный анализ (ЛДА) — это метод многомерного статистического анализа, который используется для определения различий между двумя или более группами объектов на основе их характеристик. ЛДА позволяет определить, какие переменные наиболее значимы для разделения групп, и создать линейную комбинацию этих переменных, которая максимально разделяет группы. ЛДА может быть использован для классификации объектов в соответствии с их характеристиками, а также для прогнозирования вероятности принадлежности объекта к определенной группе.

При ЛДА предполагается, что функции совместной плотности распределения вероятностей  $\rho(\vec{x}|y = 1)$  и  $\rho(\vec{x}|y = 0)$  - нормальны. В этих предположениях оптимальное байесовское решение - относить точки ко второму классу если отношение правдоподобия ниже некоторого порогового значения  $T$ .

Если не делается никаких дальнейших предположений, полученную задачу классификации называют квадратичным дискриминантным анализом.

В ЛДА делается дополнительное предположение о гомоскедастичности (т.е. предполагается, что ковариационные матрицы равны, и считается, что ковариационные матрицы имеют полный ранг. При этих предположениях задача упрощается и сводится к сравнению скалярного произведения с пороговым значением для некоторой константы  $c$ , где  $\vec{\omega} = \Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$ .

Это означает, что вероятность принадлежности нового наблюдения  $x$  к классу  $Y$  зависит исключительно от линейной комбинации известных наблюдений.

2. Квадратичный дискриминантный анализ (КДА) – это статистический метод, используемый для классификации объектов на основе набора их

характеристик или признаков. Он основан на построении квадратичной функции, которая разделяет объекты на различные классы. КДА используется для решения задач классификации, идентификации и прогнозирования. В рамках выпускной квалификационной работы КДА использован для анализа данных и определения влияния факторов на состояния объектов морской техники.

3. Линейная регрессия – это статистический метод, используемый для анализа связи между двумя переменными, где одна переменная является зависимой, а другая - независимой. Цель линейной регрессии заключается в том, чтобы построить математическую модель, которая наилучшим образом описывает зависимость между этими переменными. Модель линейной регрессии представляет собой уравнение прямой линии, которая проходит через набор точек данных. Эта линия используется для прогнозирования значений зависимой переменной на основе известных значений независимой переменной.

4. Наивный байесовский классификатор – специальный частный случай байесовского классификатора, основанный на дополнительном предположении, что объекты  $x \in X$  описываются  $n$  статистически независимыми признаками. Предположение о независимости означает, что функции правдоподобия классов представимы в виде:

$$p_y(x) = p_{y_1}(\xi_1) \dots p_{y_n}(\xi_n), \quad (17)$$

где  $p_{y_j}(\xi_j)$  – плотность распределения значений  $j$ -го признака для класса  $y$ .

Предположение о независимости существенно упрощает задачу, так как оценить  $n$  одномерных плотностей гораздо легче, чем одну  $n$ -мерную плотность. К сожалению, оно крайне редко выполняется на практике, отсюда и название метода.

Наивный байесовский классификатор может быть как параметрическим, так и непараметрическим, в зависимости от того, каким методом восстанавливаются одномерные плотности.

Основные преимущества наивного байесовского классификатора – простота реализации и низкие вычислительные затраты при обучении и

классификации. В тех редких случаях, когда признаки действительно независимы (или почти независимы), наивный байесовский классификатор (почти) оптимален.

Основной его недостаток – относительно низкое качество классификации в большинстве реальных задач.

5. Метод ближайших соседей – это алгоритм машинного обучения, который используется для классификации и регрессии. Он основывается на принципе близости объектов в пространстве признаков: если объекты имеют похожие признаки, то они скорее всего принадлежат к одному классу или имеют похожее значение целевой переменной.

Алгоритм k-NN работает следующим образом: для каждого нового объекта, который нужно классифицировать или для которого нужно предсказать значение целевой переменной, находятся k ближайших к нему объектов из обучающей выборки. Затем, используя метки классов или значения целевой переменной этих ближайших объектов, определяется класс или значение целевой переменной для нового объекта.

Выбор параметра k (количество ближайших соседей) является важным шагом при использовании метода k-NN. Если k выбрано слишком маленьким, то алгоритм может быть чувствителен к шуму в данных и приводить к переобучению. Если k выбрано слишком большим, то алгоритм может упускать важные детали и приводить к недополучению.

Для оценки качества полученных моделей были использованы следующие метрики:

1. Accuracy (точность) в машинном обучении — это метрика, которая измеряет, насколько точно модель классификации предсказывает правильный класс для тестовых данных. Она определяется как отношение числа правильно классифицированных объектов к общему числу объектов в тестовом наборе данных. Например, если модель правильно классифицирует 90 из 100 объектов, то ее точность составляет 90%. Accuracy является одной из наиболее распространенных метрик для оценки качества моделей классификации.

2. Площадь под ROC-кривой AUC (Area Under Curve) является агрегированной характеристикой качества классификации, не зависящей от соотношения цен ошибок. Чем больше значение AUC, тем «лучше» модель классификации. ROC-кривая в свою очередь показывает зависимость верных положительных классификаций TPR (True Positive Rate) от ложных положительных классификаций FPR (False Positive Rate) при варьировании порога  $\omega_0$ . Она проходит из точки (0, 0), соответствующей максимальному значению  $\omega_0$ , в точку (1, 1), соответствующую минимальному значению  $\omega_0$ . Чем выше лежит кривая, тем лучше качество классификации. На рисунке 2 приведен пример ROC-кривой.

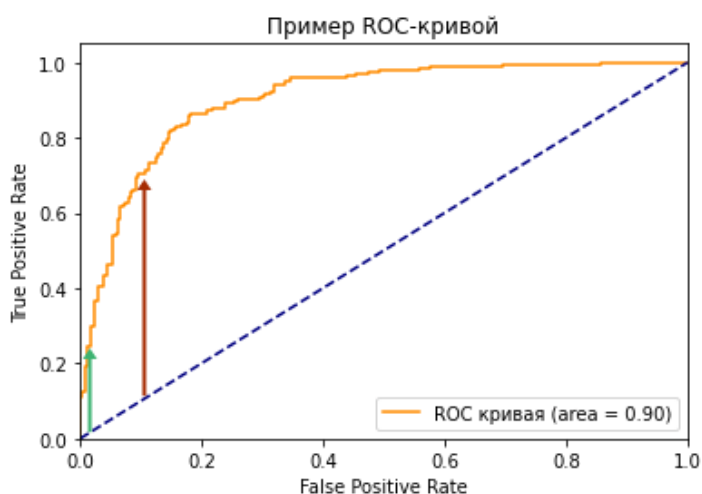


Рисунок 2 – Пример ROC-кривой

Однако ROC-кривая, вычисленная по обучающей выборке, является оптимистично смещенной влево-вверх вследствие переобучения. Величину этого смещения предсказать довольно трудно, поэтому на практике ROC-кривую всегда оценивают по независимой тестовой выборке.

### 3.2 Анализ полученных экспериментальных результатов бортовые телекоммуникационные системы

В рамках эксперимента рассматривалась информационно-телекоммуникационная система. Рассматривалась ситуация наличия или отсутствия паразитного трафика, генерируемого сканерами сетей, трассировщиками,

анализаторами, вызывающего появление различных служебных сообщений. В эксперименте рассматривалась задача идентификации паразитного трафика. В качестве воздействующего фактора определено время суток.

| <b>Этапы</b> | <b>Содержание</b>   | <b>Мат. обеспечение</b>  | <b>Инф. обеспечение</b>   |
|--------------|---|--|---|
| 1            | Конфигурация сети передачи данных, состоящей из нескольких приемо-передающих устройств на базе процессорной платы формата Eric.   | Процессорная плата CPC 805, накопитель данных, совместимые кабели для требуемых разъемов, устройство визуализации. | ОС Linux, сниффер, сопутствующее тестовое программное обеспечение.                |
| 2            | Анализ базы сигнатур сетевых атак NSL-KDD   |  | ОС Linux, датасет NSL-KDD, сопутствующее тестовое программное обеспечение.        |
| 3            | Применение методов машинного обучения в задаче классификации сетевого трафика   |  | ОС Linux, датасет NSL-KDD, сопутствующее тестовое программное обеспечение, MatLab |
| 4            | Система мониторинга в пассивном режиме слушает сеть, осуществляет распознавание сетевого трафика, выданного в эфир на определенном интервале времени. В активном режиме, кроме перечисленных действий, устройством осуществляется опрос | Приемно-передающих устройств на базе процессорной платы формата Eric   | ОС Linux, сниффер, сопутствующее тестовое программное обеспечение.                |

|   |  |  |                                 |
|---|--|--|---------------------------------|
|   | состояния приемо-передающих устройств. |  |                                 |
| 5 | Анализ полученных результатов          |  | ОС Linux, массив данных, MatLab |

Используя информацию об отдельных факторах, влияющих на значения, сегментируя выборку становится возможным уменьшить диапазон изменения данных.

На рисунках 3 и 4 на основе частот значений построена функция оценки плотности вероятности для паразитного (класс 0) и рабочего (класс 1) трафика по всему датасету и по множеству рабочих часов.

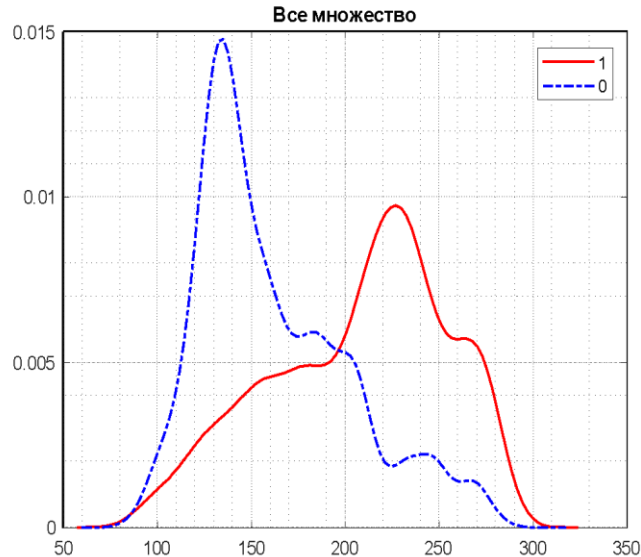


Рисунок 3 – Функция оценки плотности вероятности для классов по всем данным набора

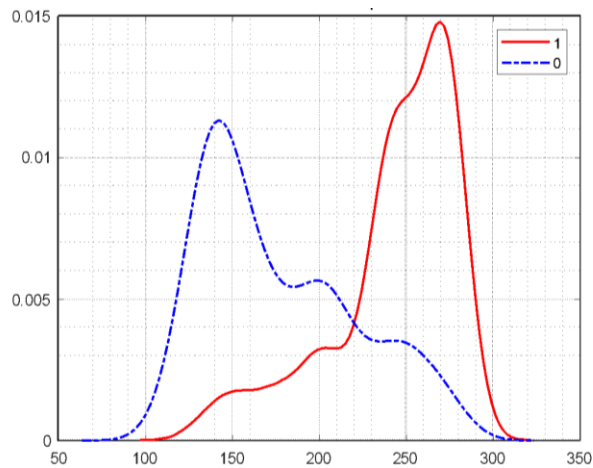


Рисунок 4 – Функция оценки плотности вероятности для классов по сегментированному множеству рабочих часов

Перекрывающиеся области дают возможность оценить вероятности возникновения ошибок первого и второго рода. Площадь пересечения данных всего множества больше, чем для подмножества рабочих дней. В ряде случаев сегментирование позволяет уменьшить вероятность возникновения ошибок.

Гипотеза компактности определяет, что последовательности, принадлежащие к одному целевому классу, будут близки друг к другу и далеки от объекта другого класса. Предполагается, что значения данных одного класса группируются рядом. Поэтому для оценки таких кластеров использовалась функция силуэта. С ее помощью возможна проверка согласованности данных в областях.

$$S(i) = \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (18)$$

где  $a_i$  – среднее расстояние от  $i$ -й точки до других точек в том же кластере, что и  $i$ ,

$b_i$  – минимальное среднее расстояние от  $i$ -й точки до точек в другом кластере.

Оценка всей кластерной структуры производилась, как усреднение показателей по элементам:

$$SWC = \frac{1}{N} \sum_{i=1}^N S(i) \quad (19)$$

На рисунках 5 и 6 показан график коэффициента силуэта для значений рабочего (класс 1) и нерабочего (класс 0) времени всего датасета и подмножества рабочих дней. Он позволяет определить, насколько близко каждая точка в одном классе находится к точкам в соседней области.

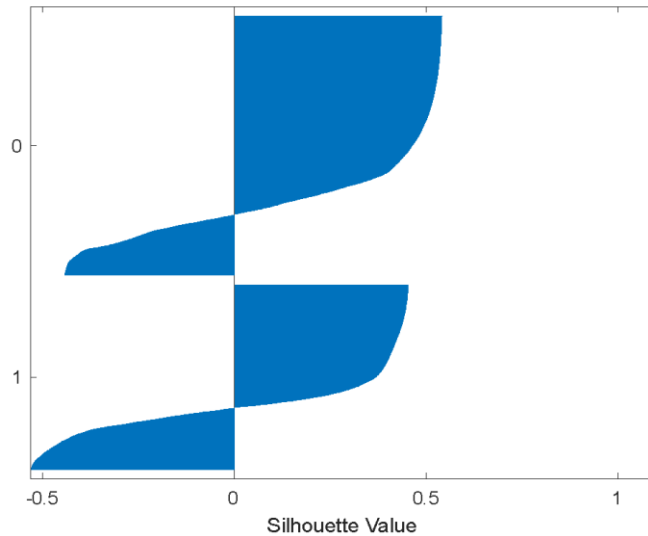


Рисунок 5 – Силуэт для значений всего множества

Графики показывают, что в случае сегментированного множества значения лучше сбалансированы и образуют более компактную область по сравнению данным всего множества.

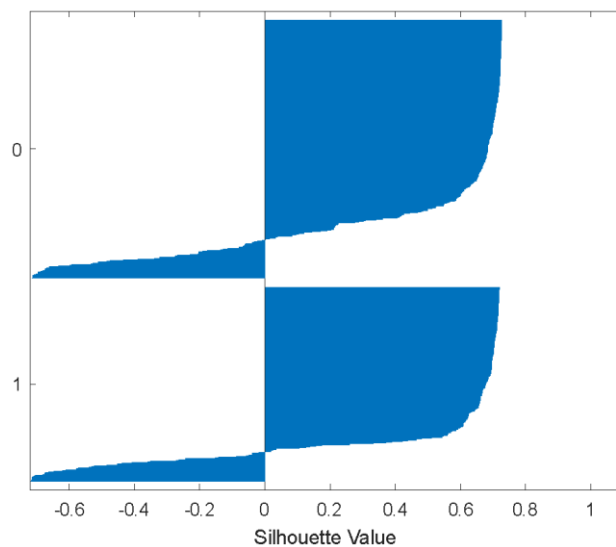


Рисунок 6 – Силуэт классов для значений рабочего времени



В приводимом эксперименте было произведено несколько разбиений датасета, содержащего 125973 записи. Набор данных содержал записи о наиболее типовых атаках, которым статистически чаще подвергались различные морские информационные системы [20-22], а также о штатном функционировании телекоммуникационной системы. В рамках эксперимента проводилась бинарная классификация состояний. При разбиении датасета в качестве управляющего воздействия учитывалась временная метка. Вся выборка X1, выборка рабочих дней X2, значения с 9 утра до 13 часов X3, с 13 часов до 18 часов X4, выходные дни X5, с 9 до 18 часов в выходные дни X6.

Значения класса группируются по областям. Оценивая качество каждой области с помощью коэффициента силуэта, можно дать априорную оценку классифицирующей модели [25,26].

Результаты коэффициентов «Силуэта» для различных разделений выборок приведены на рисунке 7.

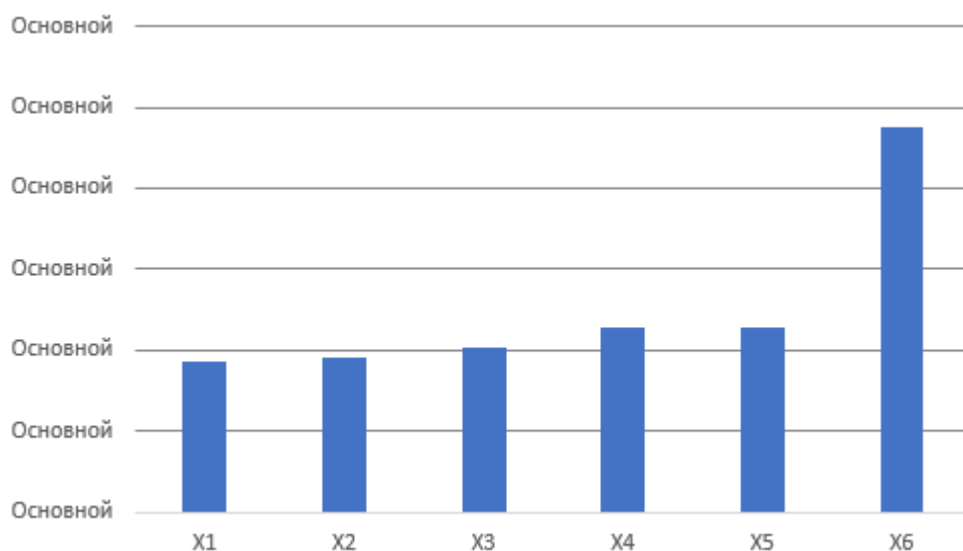


Рисунок 7 – Значения коэффициентов «Силуэта» для различных разделений выборок

Для оценки влияния подмножеств на качество результатов моделей машинного обучения были выбраны различные алгоритмы: линейного дискриминантного анализа (LD), квадратичного дискриминантного анализа (QD), наивного байесовского классификатора (NB), ближайших соседей (KNN).

Ниже в таблице 1 приводятся результаты тестирования классификаторов (AUC- площадь под ROC кривой, точность Accuracy, F-мера). В анализируемых подвыборках с ростом коэффициента силуэта наблюдался рост показателей точности.

Результаты тестирования показывают, что разбиение общей выборки на отдельные подмножества, в основном, позволяет повысить ряд показателей качества классификации для выбранных алгоритмов. Однако бывают ситуации, когда предложенный метод может проигрывать. Это вызывает большой разброс данных, сложность процессов разделения областей, необходимость формирования сложных разделяющих поверхностей для отдельных видов классификаторов.

Таблица 1 – Результаты классифицирующих алгоритмов

|           |                 | <b>X1</b> | <b>X2</b> | <b>X3</b> | <b>X4</b> | <b>X5</b> | <b>X6</b> |
|-----------|-----------------|-----------|-----------|-----------|-----------|-----------|-----------|
| <b>LD</b> | <b>Accuracy</b> | 66,4      | 71,6      | 73,8      | 74,6      | 75,1      | 85,7      |
|           | <b>AUC</b>      | 0,74      | 0,75      | 0,79      | 0,78      | 0,79      | 0,92      |
|           | <b>F-мера</b>   | 0,727273  | 0,768769  | 0,772864  | 0,780065  | 0,785542  | 0,873112  |
| <b>QD</b> | <b>Accuracy</b> | 65,4      | 72        | 73,6      | 74,8      | 74,9      | 85,7      |
|           | <b>AUC</b>      | 0,74      | 0,75      | 0,79      | 0,78      | 0,79      | 0,91      |
|           | <b>F-мера</b>   | 0,738786  | 0,778592  | 0,781377  | 0,794178  | 0,785657  | 0,879518  |
| <b>NB</b> | <b>Accuracy</b> | 70,2      | 71,2      | 73,8      | 74,8      | 75,9      | 86,9      |
|           | <b>AUC</b>      | 0,73      | 0,74      | 0,79      | 0,79      | 0,79      | 0,91      |
|           | <b>F-мера</b>   | 0,738739  | 0,767062  | 0,772589  | 0,784855  | 0,791569  | 0,886931  |
| <b>LR</b> | <b>Accuracy</b> | 70,76     | 70,8      | 70,8      | 74,2      | 75,5      | 86,9      |
|           | <b>AUC</b>      | 0,73      | 0,74      | 0,77      | 0,78      | 0,78      | 0,9       |
|           | <b>F-мера</b>   | 0,744373  | 0,754061  | 0,757308  | 0,794057  | 0,803519  | 0,883041  |

## ЗАКЛЮЧЕНИЕ

На сегодняшний день стандартные алгоритмы машинного обучения, представленные в специализированных библиотеках программных сред типа Weka, Matlab, могут быть легко применены для анализа различных данных. Основные проблемные вопросы методов машинного обучения лежат в области формирования обучающих выборок, позволяющих достичь заданного качества. Для этих целей применяются подходы, направленные на устранение шумов и удаление дисбаланса классов.

В данной выпускной квалификационной работе были исследованы различные аспекты мониторинга информационной безопасности бортовых систем в контексте морских информационных систем (МИС) и применение методов машинного обучения (МО) для достижения этой цели.

В первой главе были подробно описаны бортовые системы, их составляющие и способы оценки их состояния. Были рассмотрены различные типы систем, включая навигационные, электронные, системы безопасности и энергетические системы. Также было представлено понятие «мониторинг информационной безопасности и ее способы оценки, а также важные ГОСТы, связанные с оценкой надежности и информационной безопасности.

Заключительные пункты первой главы были посвящены постановке задачи, целям и методам исследования. Основной целью работы является разработка метода оценки состояния объектов морской техники с использованием моделей машинного обучения.

Во второй главе было исследовано применение методов МО при мониторинге информационной безопасности бортовых систем в контексте МИС. Были рассмотрены различные аспекты, связанные с повышением качества оценки состояния через сегментацию выборки данных. Были представлены различные методы разделения данных на сегменты, обучение моделей для каждого сегмента, выбор и оптимизация методов МО, агрегация результатов моделей и учет взаимосвязей между сегментами. В результате исследования

был предложен подход, основанный на сегментировании выборки данных и адаптации методов МО к МИС, для повышения качества оценки состояния бортовых систем.

Третья глава была посвящена применению сегментации данных в задачах классификации и регрессии. В работе были использованы различные алгоритмы классификации, такие как линейная регрессия, гауссова регрессия, деревья решений и метод опорных векторов. Были проведены оценки функции потерь для каждого из этих алгоритмов и сравнительный анализ их результатов. Также были рассмотрены алгоритмы регрессии, такие как логистическая регрессия и случайный лес, и была проведена оценка их точности. В результате исследования было показано, что сегментация данных может значительно улучшить качество мониторинга информационной безопасности состояния бортовых систем.

В заключительной части работы было представлено обобщенное исследование всех трех глав. Были подведены итоги проведенного исследования и выделены основные результаты. Отмечается, что применение методов МО при оценке состояния бортовых систем в контексте МИС позволяет повысить точность, надежность и адаптивность моделей в различных сценариях работы. Выводится заключение о значимости исследования для области морских информационных систем и подчеркивается его вклад в развитие технологий оценки состояния бортовых систем.

Таким образом, данная выпускная квалификационная работа представляет важный вклад в область оценки мониторинга информационной безопасности бортовых систем в контексте МИС. Исследования, проведенные в работе, позволяют лучше понять принципы и методы оценки состояния, а также предлагают эффективные подходы и методы на основе машинного обучения. Результаты работы могут быть использованы в инженерной практике для повышения надежности, безопасности и эффективности бортовых систем в морских информационных системах.

Практическая значимость и применимость исследования: В работе были предложены новые подходы и методы оценки состояния бортовых систем, основанные на применении методов машинного обучения и сегментации данных. Эти методы могут быть полезны и применимы в различных областях, включая судостроение, авиация, автомобильную промышленность и другие сферы, где требуется надежная оценка состояния систем для обеспечения безопасности и эффективности работы. Ограничения и дальнейшие направления исследования: в ходе работы могли выявиться определенные ограничения и ограничивающие факторы, которые могут повлиять на применимость предложенных методов. Например, ограниченная доступность данных, сложность взаимосвязей между различными сегментами данных, вычислительные ограничения и другие факторы. Дальнейшие исследования могут быть направлены на преодоление этих ограничений и разработку более точных и эффективных методов оценки состояния.

Значимость исследования в контексте актуальных тенденций: Отметьте, как исследование соответствует современным тенденциям и вызовам в области оценки состояния бортовых систем. Укажите на важность разработки надежных и безопасных систем в контексте развития автономных и интеллектуальных технологий, а также на потенциальный вклад вашей работы в решение данных проблем.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Дмитриев, Форафонов. Современные навигационные системы и безопасность судовождения <https://www.labyrinth.ru/books/329740/>
2. ГОСТ Р ИСО/МЭК 15408–1–2012, 15408–2–2012, 15408-3–2012. Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Ч. 1–3. Госстандарт России [Электронный ресурс]. – Электронный фонд правовых и нормативно-технических документов. – 2012. – Режим доступа: <https://docs.cntd.ru/document/1200101777>.
3. ГОСТ 27.310–95. Надежность в технике. Анализ видов, последствий и критичности отказов. Основные положения [Электронный ресурс]. – Электронный фонд правовых и нормативно-технических документов. – 1995. – Режим доступа: <https://docs.cntd.ru/document/1200001363>.
4. Takacs A., Toledano-Ayala M., Dominguez-Gonzalez A., Pastrana-Palma A., Velazquez D. T., Ramos J. M., Rivas-Araiza A. E. Descriptor generation and optimization for a specific outdoor environment. *IEEE Access*, 2020, vol. 8, pp. 2169–3536. doi:10.1109/ACCESS.2020.2975474
5. Park J, Kim S. Machine Learning-Based Activity Pattern Classification Using Personal PM2.5 Exposure Information. *International Journal of Environmental Research and Public Health*. 2020; 17(18):6573. <https://doi.org/10.3390/ijerph17186573>
6. J. Jin, J. Gubbi, S. Marusic, M. Palaniswami An information framework for creating a smart city through internet of things // *IEEE Internet of Things Journal*. Vol. 1 (2). 2014. P. 112-121. doi: 10.1109/IIOT.2013.2296516.
7. Дьяконов А. Методы решения задач классификации с категориальными признаками // *Прикладная математика и информатика. Труды факультета Вычислительной математики и кибернетики МГУ имени М.В. Ломоносова*. 2014. № 46. С. 103–127.

8. Гайфулина Д. А., Котенко И. В. Анализ моделей глубокого обучения для задач обнаружения сетевых аномалий интернета вещей. Информационно-управляющие системы, 2021, № 1, с. 28–37. doi:10.31799/1684-8853-2021-1-28-37
9. Z.-H. Zhou. Ensemble Methods: Foundations and Algorithms // New York: Chapman & Hall/CRC. 2012. 222 pp. ISBN: 978-1-4398-3003-1.
10. Y. Yu, Z. Zhou, K. M. Ting Cocktail Ensemble for Regression // Seventh IEEE International Conference on Data Mining (ICDM 2007). Omaha, NE. 2007. pp. 721-726. doi: 10.1109/ICDM.2007.60.
11. Шелухин О.И., Симонян А.Г., Ванюшина А.В. Влияние структуры обучающей выборки на эффективность классификации приложений трафика методами машинного обучения. Т-Comm: Телекоммуникации и транспорт, 2017, Т. 11, № 2, с. 25–31.
12. Lughofer E., Weigl E., Heidl W., Eitzinger C., Radauer T. Recognizing input space and target concept drifts in data streams with scarcely labeled and unlabelled instances. Information Sciences, 2016, vol. 35, pp.127–151
13. Sethi T., Kantardzic M. Handling adversarial concept drift in streaming data. Expert Systems with Applications, 2018, vol. 97, pp. 18–40.
14. Лебедев, И. С. Сегментирование множества данных с учетом информации воздействующих факторов. Информационно-управляющие системы, 2021, №3, 29-38. <https://doi.org/10.31799/1684-8853-2021-3-29-38>
15. Рзаев Б.Т., Лебедев И. С. Применение бэггинга при поиске аномалий сетевого трафика. Научно-технический вестник информационных технологий, механики и оптики, 2021. Т. 21, № 2, с.50-56. doi: 10.17586/2226-1494-2021-21-2-50-56
16. Сухопаров, М.Е. Сегментирование выборок данных при анализе состояния безопасности устройств интернета вещей / М.Е. Сухопаров, И.С. Лебедев // Проблемы информационной безопасности. Компьютерные системы. – 2021. – № 3. – С. 130–136.

17. Сухопаров, М.Е. Использование информации о влияющих факторах для разбиения выборок данных в методах машинного обучения для оценки состояния ИБ/ М.Е. Сухопаров, И.С. Лебедев // Проблемы информационной безопасности. Компьютерные системы. – 2022. – № 2. – С. 125–134.

18. Сухопаров, М.Е. Сегментирование выборок данных при анализе состояния безопасности устройств интернета вещей / М.Е. Сухопаров, И.С. Лебедев // Проблемы информационной безопасности. Компьютерные системы. – 2021. – № 3. – С. 130–136.

19. K. Wolsing. Network Attacks Against Marine Radar Systems: A Taxonomy, Simulation Environment, and Dataset / K. Wolsing, A. Sailard, J. Bauer etc. // 2022 IEEE 47th Conference on Local Computer Networks (LCN). – 2022. – pp. 114-122.

20. A. Androjna. Assessing Cyber Challenges of Maritime Navigation / A. Androjna, T. Brcko, I. Pavic, H. Greidanus // Journal of Marine Science and Engineering. – 2020. – Vol. 8. – pp. 776 – 797.

21. Meiseles and L. Rokach, Source Model Selection for Deep Learning in the Time Series Domain in IEEE Access, vol. 8, pp. 6190-6200, 2020, doi: 10.1109/ACCESS.2019.2963742

22. P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," J. Comput. Appl. Math., vol. 20, pp. 53\_65, Nov. 1987.