УДК 556:004.8

ОПТИМИЗАЦИЯ И МОДЕЛИРОВАНИЯ ЗАДАЧ ВОДНЫХ РЕСУРСОВ НА ОСНОВЕ МАШИННОГО И ГЛУБОКОГО ОБУЧЕНИЯ

Сабитов[1] Б.Р., Шейшенов[1] Ж., Мусалиева[2] С.К., Матраимов[2] С.К.
[1]КНУ им. Баласагына, [2]КГТУ им. М. Раззакова

В данной работе с инструментами машинного обучения построено регрессионная модель классификации и прогнозирования качества воды. В исследованиях водной среды модели и выводы, полученные с помощью машинного обучения, применяется для построения, мониторинга, моделирования, оценки и оптимизации различных систем очистки и управления водными ресурсами. Кроме того, машинное обучение может предоставлять решения для контроля загрязнения воды, улучшения качества воды и управления безопасностью экосистем водосборов.

Ключевые слова: машинное обучение, прогнозирование, загрязнение воды, качество воды.

МАШИНА ЖАНА ТЕРЕҢ ОКУТУУГА НЕГИЗДЕЛГЕН СУУ РЕСУРСТАРЫНЫН МАСЕЛЕЛЕРИН ОПТИМАЛДАШТЫРУУ ЖАНА МОДЕЛДӨӨ

Сабитов[1] Б.Р., Шейшенов[1] Ж., Мусалиева[2] С.К., Матраимов[2] С.К.
Баласагын атындагы [1]КМУ, И.Раззаков атындагы [2]КМТУ

Бул макалада машиналык окутуу куралдарын колдонуу менен суунун сапатын классификациялоо жана алдын ала айтуу үчүн регрессиялык модель курулган. Суу изилдөөлөрүндө машиналык окутууну колдонуу менен алынган моделдер жана тыянактар ар кандай сууну тазалоо жана башкаруу системаларын долбоорлоо, көзөмөлдөө, моделдөө, баалоо жана оптималдаштыруу үчүн колдонулат. Андан тышкары, машиналык окутуу суу бассейндеринде суунун булганышын көзөмөлдөө, суунун сапатын жакшыртуу жана экосистеманын коопсуздугун башкаруу үчүн чечимдерди бере алат.

Баштапкы сөздөр: машиналык окутуу, божомолдоо, суунун булганышы, суунун сапаты.

OPTIMIZATION AND MODELING OF WATER RESOURCE PROBLEMS BASED ON MACHINE LEARNING AND DEEP LEARNING

Sabitov[1] B.R., Sheishenov[1] Zh., Musalieva[2] S.K., Matraimov[2] S.K.
[1]J. Balasagyn Kyrgyz National University,
[2]I. Razzakov Kyrgyz State Technical University

This article presents a regression model for water quality classification and forecasting developed using machine learning tools. In aquatic environment research, models and insights derived from machine learning are applied to the design, monitoring, modeling, assessment, and optimization of various water treatment and resource management systems. Furthermore, machine learning can provide solutions for controlling water pollution, improving water quality, and managing the safety of watershed ecosystems.
Keywords: machine learning, forecasting, water pollution, water quality.

Введение. Создано модель для определения качества воды -WQI с применением нейронного моделирования DenseNet. На базе обученной модели построено веб система на фреймворке FLASK для определения пригодности питья воды. В данной работе с инструментами машинного обучения построено регрессионная модель классификации и прогнозирования качества воды. В отличие от традиционных моделей, используемых в исследованиях, связанных с водными ресурсами, модели на основе данных, основанные на моделях, и могут эффективно решать более сложные нелинейные задачи о водных проблемах. В исследованиях водной среды модели и выводы, полученные с помощью машинного обучения, применяется для построения, мониторинга, моделирования, оценки и оптимизации различных систем очистки и управления водными ресурсами. Кроме того, машинное обучение может предоставлять решения для контроля загрязнения воды, улучшения качества воды и управления безопасностью экосистем водосборов. В этом обзоре мы описываем случаи, в которых алгоритмы машинного обучения применялись для оценки качества воды в различных водных средах, таких как поверхностные воды, грунтовые воды, питьевая вода, сточные воды и морская вода.

Разработаны различные модели по контролю загрязнения воды. Исследовано сточные воды содержащие различные загрязняющие вещества, что представляет серьезную угрозу для естественной водной среды. Важно в данном случае анализ и оценка моделей качества воды, которые могут существенно повысить эффективность определения и контроля уровень загрязнения воды [1]. На сегодняшний день

разработано много методов для мониторинга и оценки качества воды во всем мире, таких как многомерный статистический метод [2], нечеткий вывод [3] и индекс качества воды (WQI) [4]. Что касается оценки качества воды, хотя большинство параметров качества воды можно контролировать в соответствии с процедурами, определенными в соответствующих стандартах, окончательные результаты оценки качества воды могут сильно различаться из-за выбора параметров [5]. Учет всех параметров качества воды нереалистичен, поскольку это не только дорого и технически сложно, но и не учитывает изменчивость качества воды [6]. Однако в последние годы, благодаря достижениям в методах машинного обучения, все больше исследователей полагают, что можно успешно собирать и анализировать огромные объемы данных для удовлетворения сложных и масштабных требований к оценке качества воды.

В машинном обучении, отрасли искусственного интеллекта, алгоритмы используются для анализа данных и попытки извлечь потенциальные закономерности из данных для прогнозирования новой информации [7, 8]. Как новый метод анализа и обработки данных, машинное обучение широко используется во многих областях благодаря своей высокой точности, гибкой настройке и удобной расширяемости [9]. Сложные нелинейные реляционные данные могут быть легко обработаны с помощью машинного обучения, что облегчает обнаружение базовых механизмов [10]. Отличная адаптивность машинного обучения продемонстрировала его потенциал как инструмента в областях экологической науки и техники в последние годы. Поэтому можно ожидать более точных результатов оценки, несмотря на сложность использования машинного обучения для анализа и оценки качества воды [11].

Типы воды, включая питьевую воду, сточные воды, грунтовые воды, поверхностные воды, морскую воду и пресную воду, являются сложными [12]. Эти различные типы воды имеют разные характеристики, что приводит к значительным проблемам при исследовании их качества. Основываясь на результатах предыдущих исследований, машинное обучение может быть эффективным подходом к решению этих проблем. Поэтому в этом обзоре мы суммируем преимущества и недостатки широко используемых алгоритмов машинного обучения и обсуждаем применение и производительность машинного обучения в поверхностных водах, грунтовых водах, питьевой воде, сточных водах и морской воде Кроме того, мы предлагаем возможные будущие применения подходов машинного обучения к водным средам.

Методы исследования и результаты. В настоящее время мощным методом анализа данных о водной среде является инструменты машинного обучения, широко используемые для выявления закономерностей или составления прогнозов на основе больших данных. Для разработки моделей машинного обучения на практике необходимо провести сбор данных, выбор подходящего алгоритма, обучение модели и ее валидацию. Среди этих процессов выбор алгоритма имеет решающее значение. Обучение с учителем и без учителя — это два основных класса технологий машинного обучения [13].

Основное различие между этими двумя классами заключается в наличии меток в наборах данных. Обучение с учителем выводит предиктивные функции из помеченных обучающих наборов данных. Каждый обучающий экземпляр включает входные значения и ожидаемые выходные значения. Алгоритмы обучения с учителем пытаются определить взаимосвязи между входными и выходными значениями и создать предиктивную модель для прогнозирования результата на основе соответствующих входных данных. Контролируемое обучение может использоваться для классификации и регрессии данных, при этом были разработаны различные алгоритмы, включая линейную регрессию, искусственную нейронную сеть (ИНС), дерево решений (ДР), машину опорных векторов (SVM), наивный байесовский алгоритм, метод k-ближайших соседей (KNN), случайный лес (RF) и т. д.

В отличие от этого, обучение без учителя обычно используется для обработки данных без меток, решая различные задачи распознавания образов на основе немаркированных обучающих наборов данных. Обучение без учителя классифицирует обучающие данные по различным категориям в соответствии с их различными характеристиками, в основном на основе снижения размерности и кластеризации [14]. Однако количество категорий неопределено, и значение каждой категории неясно. Поэтому обучение без учителя обычно используется для классификации и поиска ассоциаций [13]. Анализ главных компонент (PCA), метод k-средних и т. д. являются наиболее часто используемыми алгоритмами машинного обучения без учителя.

Загрузка и предварительная обработка данных: анализ данных, обработка пропущенных значений, нормализация. Создание модели DenseNet: адаптация архитектуры DenseNet для табличных данных.

Обучение двух моделей: с оптимизаторами Adam и RMSprop. Визуализация: Кривые обучения (accuracy, loss, precision, recall) Матрицы ошибок ROC и AUC кривыеPrecision-Recall кривыеСравнительный

анализ: Сравнение метрик accuracy и AUC Анализ переобучения Рекомендации по выбору модели. Интерпретация результатов: Подробный анализ производительности обеих моделей.

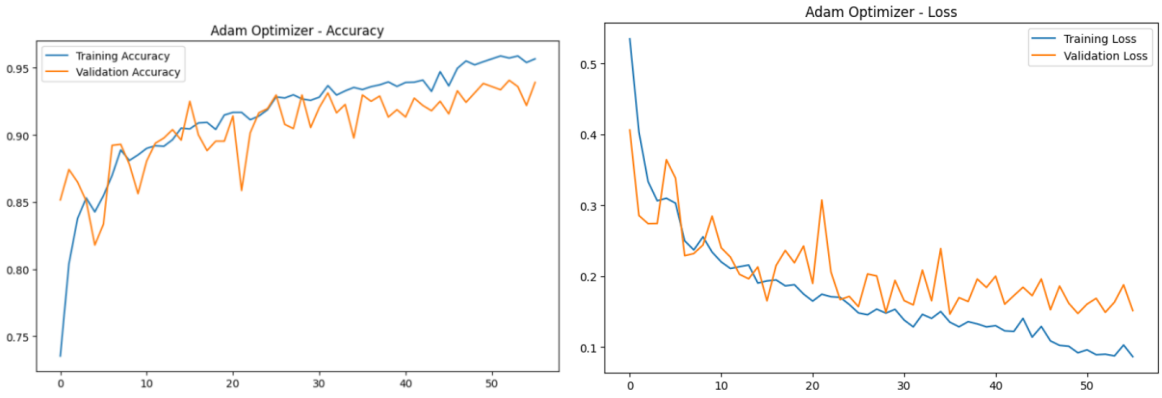
Ключевые особенности реализации: Использование transfer learning с DenseNet архитектурой. Балансировка классов для работы с несбалансированными данными Early stopping и reduce learning rate для улучшения сходимости. Полный набор метрик для комплексной оценки моделей.

Приведено также детальная визуализация результатов. Для получения наилучших результатов рекомендуется настроить гиперпараметры (learning rate, количество слоев, dropout rate) в зависимости от конкретных характеристик вашего набора данных. База данных для реализации использует следующую .csv файл. Признаки данного файла состоит из металлов и записано в виде файла

Построим нейронную сеть на основе открытой базы данных Dataset_WQI.csv с признаками aluminium, ammonia, arsenic, barium, cadmium, chloramine, chromium, copper, flouride, bacteria, viruses, lead, nitrates, nitrites, mercury, perchlorate, radium, selenium, silver, uranium и с целевой функцией качества воды is_safe построим модель нейронной сети трансферного обучения DenseNet с оптимизаторами Adam и RMSProp.

Результатом реализации будет также построение кривых обучения и ошибки моделей, матрицы ошибок, ROC, AUC кривые и сравнительный анализ моделей в различных метриках.

Результаты и обсуждения. Создано модель на основе архитектуры нейронной сети DenseNet. Создан базовый входной слой. Добавлен собственные слои. Выходной слой и процесс компиляции модели. Одной из особенностью данной модели является использования разных оптимизаторов, оптимизатор Adam и RMSProp.В процесс реализации использовано 100 эпох с разными скоростями обучения. Вот результат обучения с разными скоростями обучения.



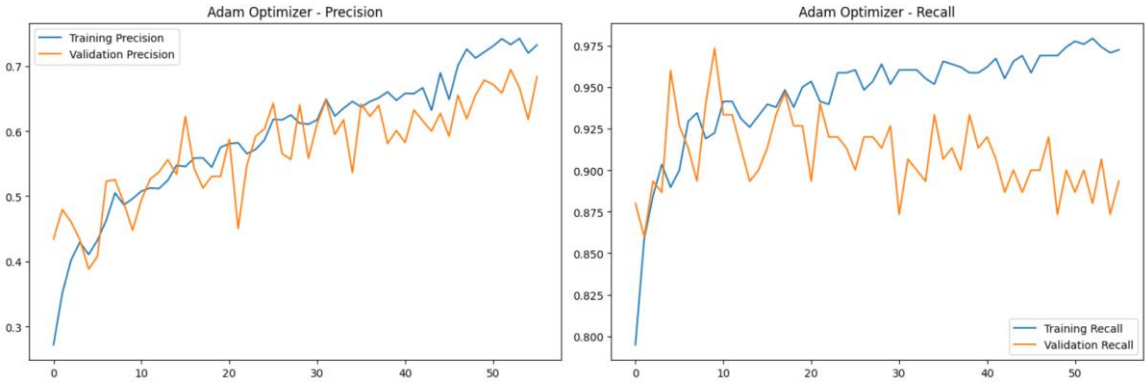


Рис.1. Результаты реализации модели прогнозирования качества воды с использованием оптимизатора Adam.

Результаты с оптимизатором RMSProp

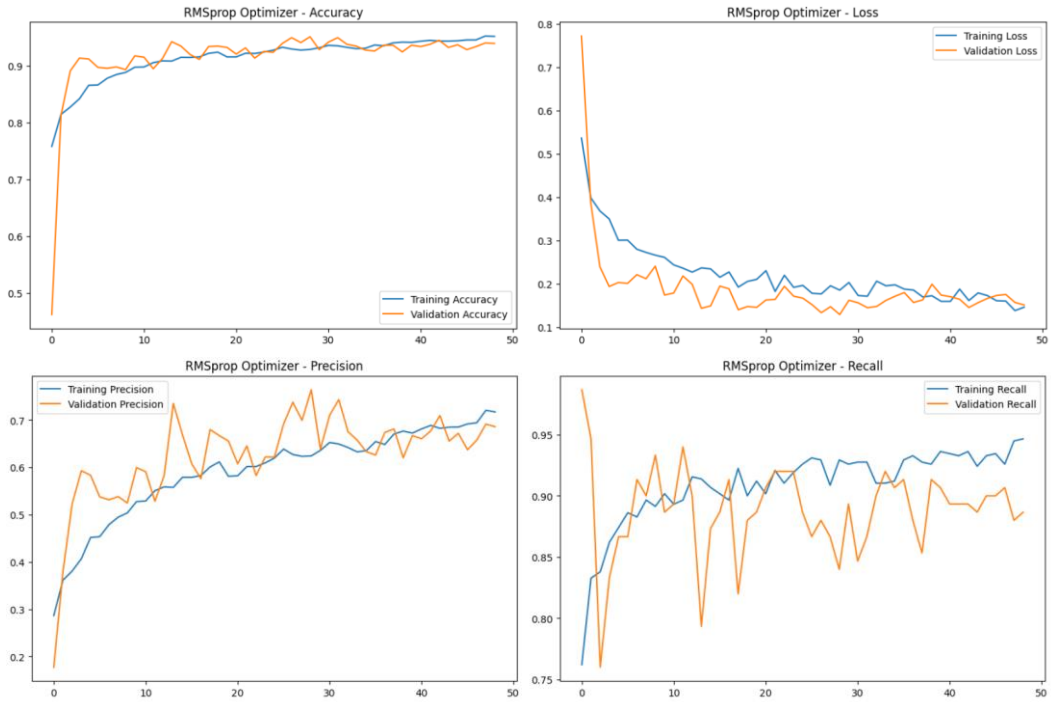


Рис.2. Результаты реализации модели прогнозирования качества воды с использованием оптимизатора RMSProp.

Вот результаты прогноза с оптимизатором Адам

```
Adam Optimizer - Accuracy: 0.9363
Adam Optimizer - Classification Report:
              precision    recall  f1-score   support

         0.0       0.98      0.94      0.96      1418
         1.0       0.67      0.88      0.76       182

    accuracy                           0.94      1600
   macro avg       0.82      0.91      0.86      1600
weighted avg       0.95      0.94      0.94      1600
```

Рис.3. Результаты прогнозирования качества воды с использованием оптимизатора Adam и RMSProp.

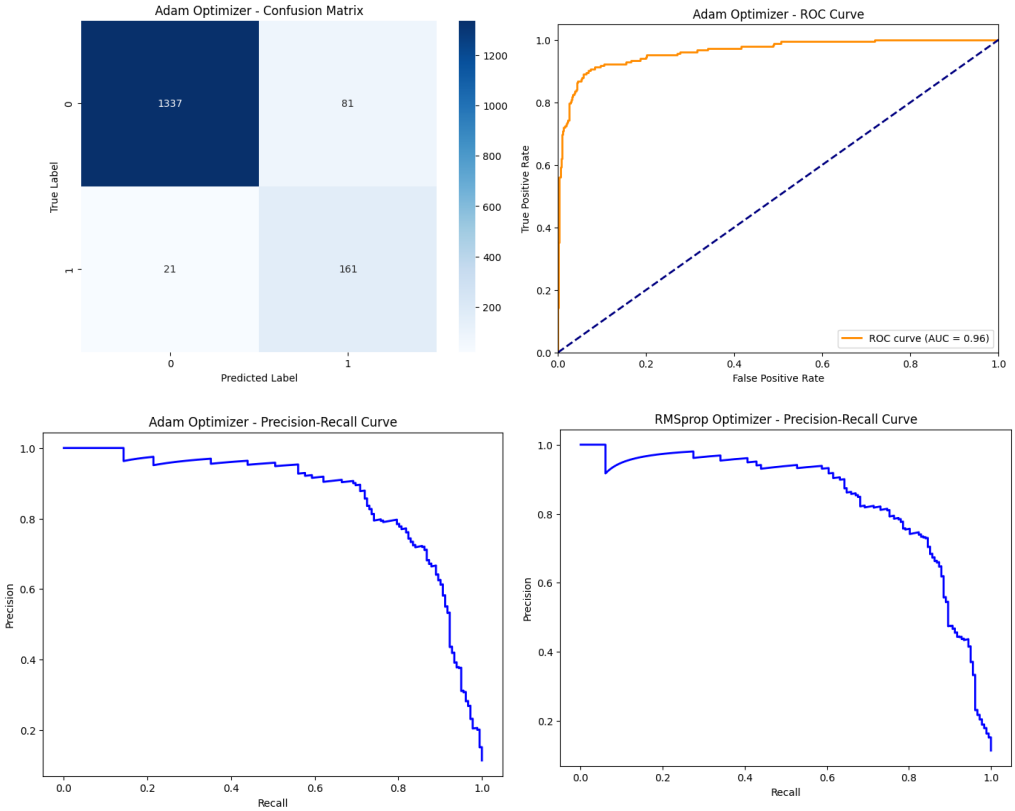


Рис.4. Матрица погрешностей Roc и Auc кривые прогнозирования качества воды с использованием оптимизатора Adam и RMSProp. Улучшенные результаты прогнозирования качества воды с использованием оптимизатора Adam и RMSProp.

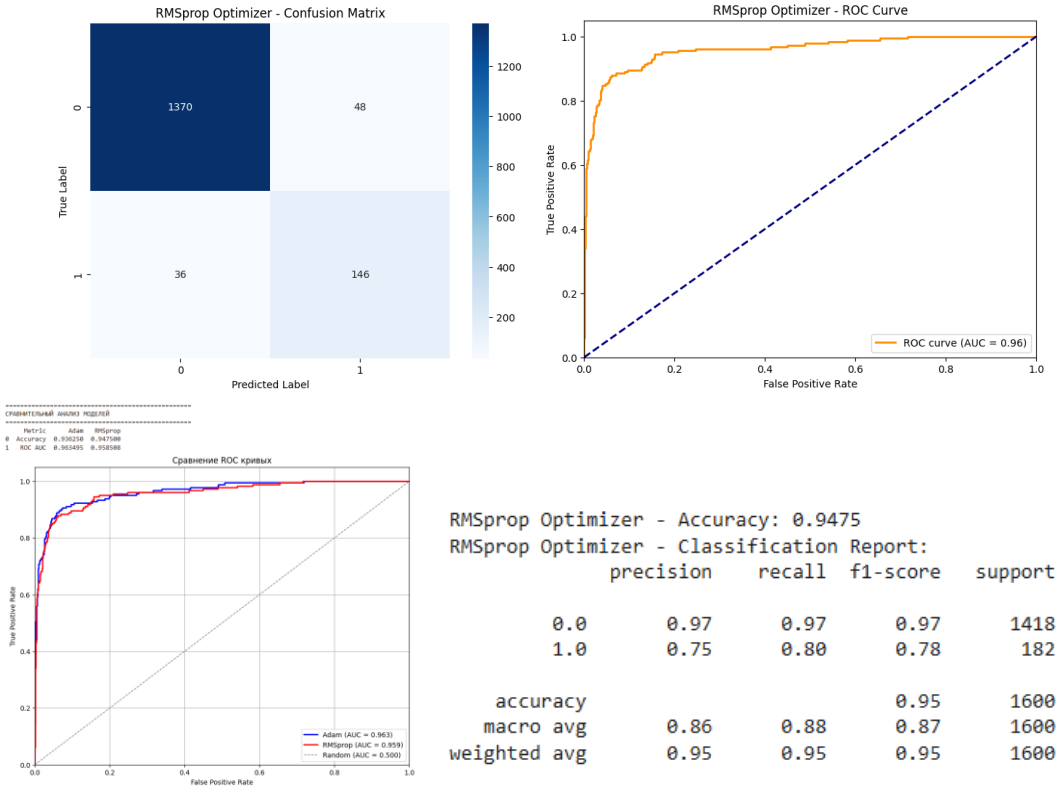


Рис. 5. Сравнительный анализ матрица погрешностей , Roc и Auc кривые прогнозирования качества воды с использованием оптимизатора Adam и RMSProp с различными скоростями обучения нейронной сети.

Листинг 1. Анализ процесса обучения нейронной сети по прогнозированию качества воды

```
АНАЛИЗ ПРОЦЕССА ОБУЧЕНИЯ:
Adam - Final Training Accuracy: 0.9566
Adam - Final Validation Accuracy: 0.9391
RMSprop - Final Training Accuracy: 0.9515
RMSprop - Final Validation Accuracy: 0.9391

==================================================
РЕКОМЕНДАЦИИ:
==================================================
✓ Модель с RMSprop оптимизатором показала лучшую точность
✓ Модель с Adam оптимизатором имеет лучшую AUC

Степень переобучения:
Adam: 0.0175
RMSprop: 0.0125
✓ Модель с RMSprop оптимизатором менее склонна к переобучению
```

В данной главе построено веб система для определения качества воды с применением моделей трансферного обучения. Построены модел DenseNET для прогнозирования пригодности питья воды. Архитектура модели для прогнозирования качества воды выглядит следующим образом. Пробный модель выглядит так.

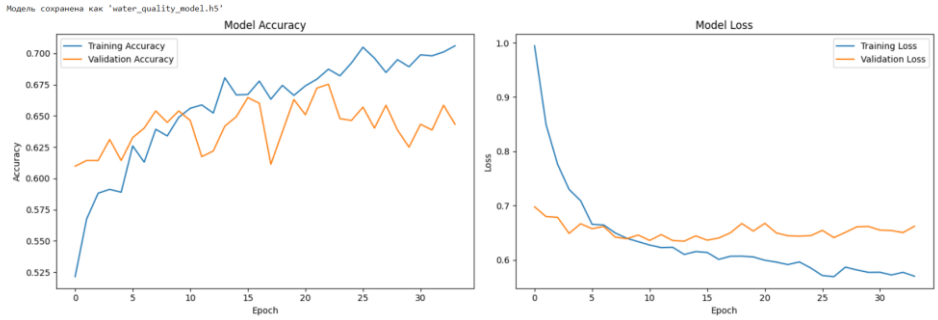


Рис. 6. Обучение модели о качестве воды на основе нейронной сети.

При реализации модели как веб система главная страница выглядит следущим образом для упрощенной модели.

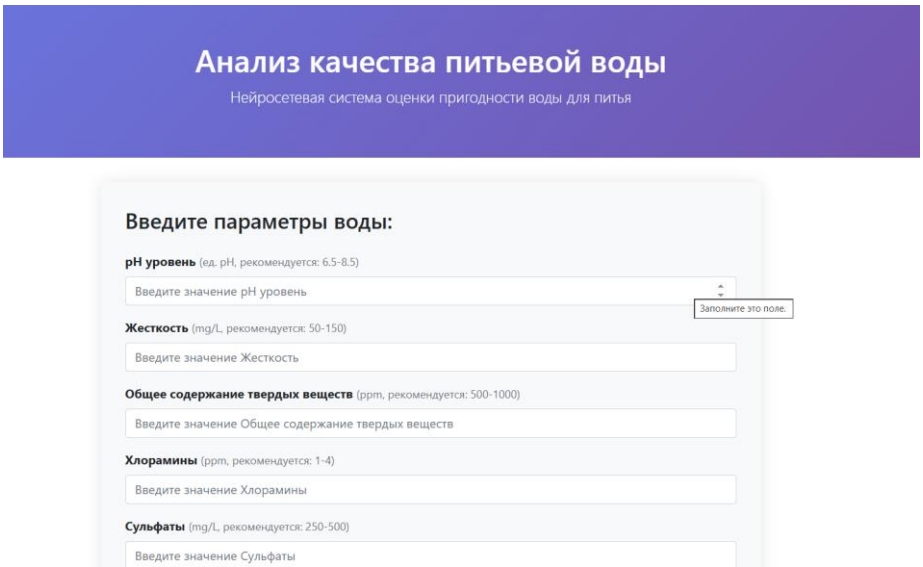
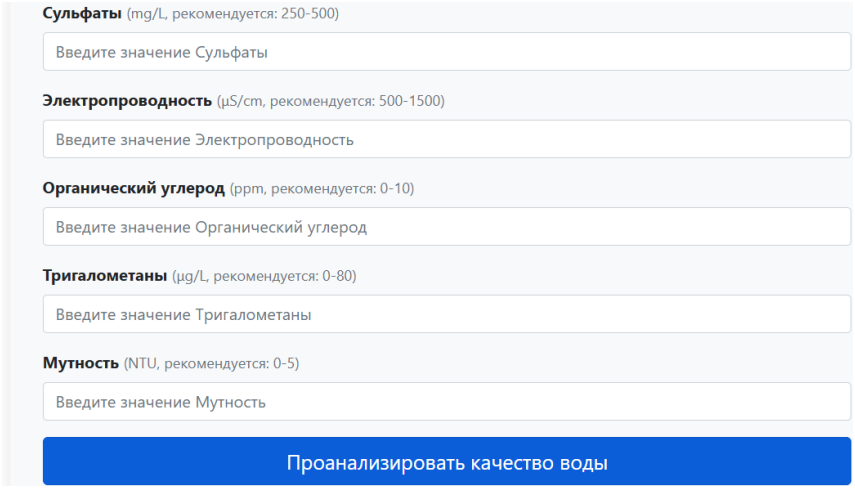


Рис. 7. Главная страница приложения

Технологии построения веб системы с базой знаний основывается на модели DenseNet и на фрейворке Flask. База данных представляет собой .csv файл параметры которой определяют качествао воды. Ниже представлены признаки и допустимый диапазон парметров входящие в модель.



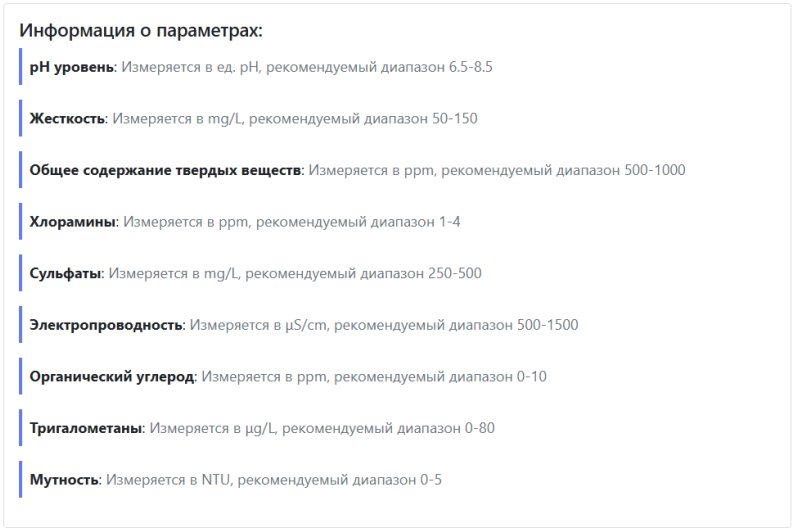


Рис. 8. Информация о параметрах.

Данная веб система имеет следующие особенности в реализациирелизации:

1. Модель DenseNet: используется архитектура, вдохновленная DenseNet, с плотными блоками и соединениями
2. Предобработка данных: Автоматическое заполнение пропущенных значений и масштабирование
3. Веб-интерфейс: Пользовательский интерфейс с Bootstrap
4. API endpoint: REST API для интеграции с другими системами
5. Анализ параметров: Детальный анализ каждого параметра воды
6. 6.Визуализация: Графики процесса обучения модели

Данную систему можно использовать для платформы фрейворков python и может быть развернута на любом хостинге, поддерживающем системы библиотек Python.Меню и главная страница модели с конечными результатами тестирования пригодности питья воды представлена в главной странице.

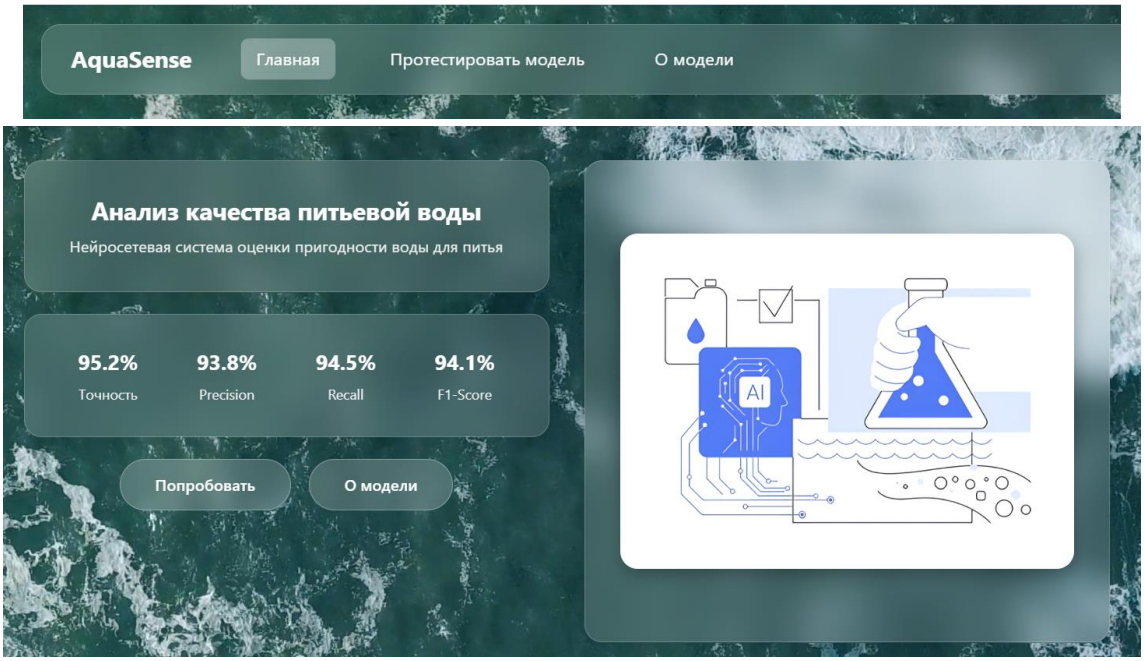


Рис.9. Главная страница

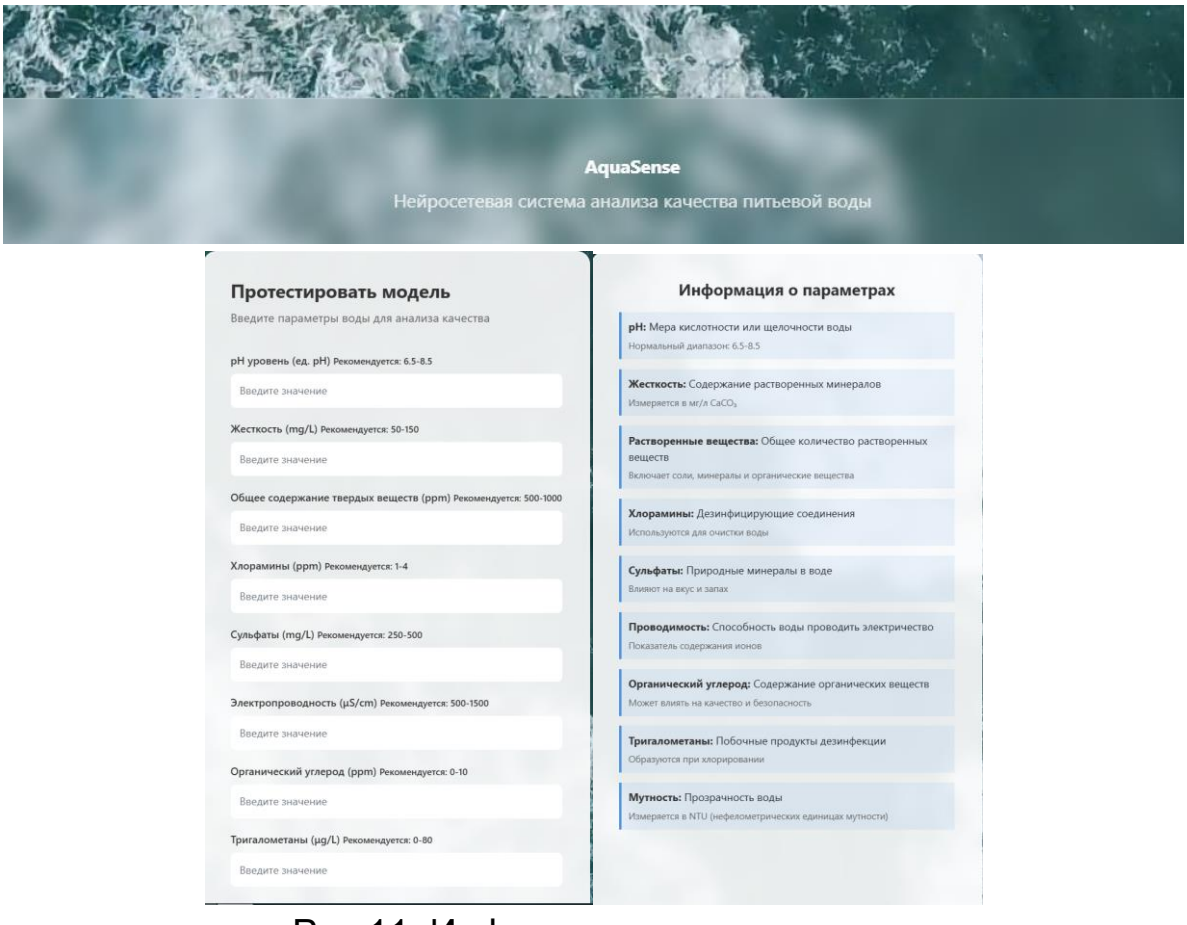


Рис 11. Информацияо параметрах

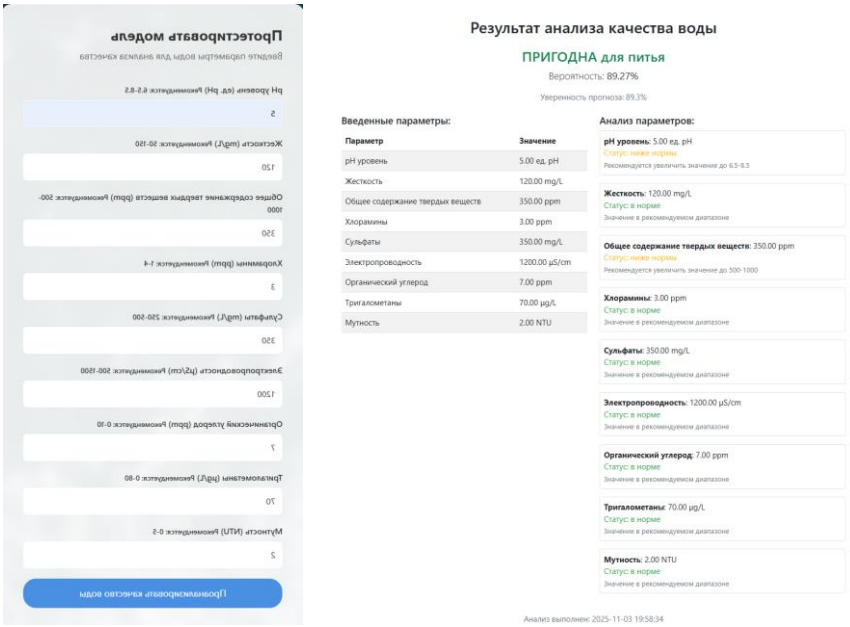


Рис 12. Результаты анализа



Рис 13. Информация о моделях



Рис. 14. О статистике обучения

Ниже представлен фрагмент кода данной веб-системы написанной на Python для Фреймворка Flask.

ЛИТЕРАТУРА

1. J. Pan, T. Leng, Y. Liu Shifosi reservoir water environmental assessment based on grey clustering Prog. Environ. Sci. Eng., 857 (2013), pp. 610-613 http://dio.org/10.4028/www.scientific.net/AMR.610-613.857 View in ScopusGoogle Scholar
2. R. Alam, Z. Ahmed, S.M. Seefat, K.T.K. Nahin Assessment of surface water quality around a landfill using multivariate statistical method, Sylhet, Bangladesh,Environ. Nanotechnol. Monit. Manag., 15 (2021), Article 100422, 10.1016/j.enmm.2020.100422 View PDFView articleView in ScopusGoogle Scholar
3. J.O. Oladipo, A.S. Akinwumiju, O.S. Aboyeji, A.A. Adelodun Comparison between fuzzy logic and water quality index methods: a case of water quality assessment in Ikare community, Southwestern Nigeria Environ. Chall., 3 (2021), Article 100038, 10.1016/j.envc.2021.100038 View PDFView articleView in ScopusGoogle Scholar
4. J. Wang, Z. Fu, H. Qiao, F. Liu Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China Sci. Total Environ., 650 (2019), pp. 1392-1402, 10.1016/j.scitotenv.2018.09.137 View PDFView articleView in ScopusGoogle Scholar
5. B. Wang, Y. Wang, S. Wang Improved water pollution index for determining spatiotemporal water quality dynamics: case study in the Erdao Songhua River Basin, China Ecol. Indicat., 129 (2021), Article 107931, 10.1016/j.ecolind.2021.107931 View PDFView articleView in ScopusGoogle Scholar
6. F.D. Simoes, A.B. Moreira, M.C. Bisinoti, S.M.N. Gimenez, M.J.S. Yabe
7. Water quality index as a simple indicator of aquaculture effects on aquatic bodies Ecol. Indicat., 8 (2008), pp. 476-484, 10.1016/j.ecolind.2007.05.002 View PDFView articleView in ScopusGoogle Scholar