

DOI: 10.25558/VOSTNII.2022.92.48.014

УДК 004: 504.4.054

© В.П. Потапов, Н.И. Юкина, Е.Л. Счастливцев, 2022

### **В.П. ПОТАПОВ**

д-р техн. наук, проф.,  
заведующий лабораторией  
ФИЦ ИВТ КФ  
e-mail: vadimptpv@gmail.com

### **Н.И. ЮКИНА**

канд. техн. наук,  
старший научный сотрудник  
ФИЦ ИВТ КФ  
e-mail: leonakler@mail.ru

### **Е.Л. СЧАСТЛИВЦЕВ**

д-р техн. наук,  
заведующий лабораторией  
ФИЦ ИВТ КФ  
e-mail: schastlivtsev@ict.sbras.ru

## **ОЦЕНКИ КАЧЕСТВА ВОД И АНСАМБЛЕВЫЕ МЕТОДЫ**

*Несмотря на активное развитие систем математического моделирования экологических процессов, их комплексные модели пока не нашли достойного отражения в научной литературе. В то же время ансамблевые методы находят все более широкое применение при обработке различных типов данных, включая мультимодальные. Основным преимуществом ансамблевого подхода является возможность интеграции различных методов, которые могут адаптироваться на определенную специфику проблемной области, за счет чего увеличивается достоверность анализа. В настоящей работе нами рассматриваются несколько подходов к оценке качества вод: классический, ансамблевый и эвристический. Предлагаемый подход к реализации вычислительного ансамбля для оценки качества вод включает стек новых технологий науки о данных, такие как: Neural networks, Xgboost, RandomForest, LogisticRegression, kNearestNeighbors (kNN). Все методы апробированы на гидрохимических данных водных объектов рек Ускал и Аба, протекающих по Кемеровской области и загрязняемых промышленными отходами различных предприятий, находящихся в их русле. В работе приводятся результаты применения ансамблирования к водным объектам на примере Кузбасских рек Аба и Ускал.*

**Ключевые слова:** РYTHON, АНСАМБЛЕВОЕ МОДЕЛИРОВАНИЕ, НЕЙРОННЫЕ СЕТИ, ГЛУБОКОЕ ОБУЧЕНИЕ, ВЫЧИСЛИТЕЛЬНЫЕ МОДЕЛИ, МАШИННОЕ ОБУЧЕНИЕ, ОЦЕНКА КАЧЕСТВА ВОД, ОЦЕНКА МОДЕЛЕЙ, ВОДНЫЕ ОБЪЕКТЫ.

### **ВВЕДЕНИЕ**

В настоящее время, все больше ученых и специалистов по всему миру (Китай [1], Индия [2], США [3], Бразилия [4], Сингапур [5] и др.) применяют ансамблевые методы в своих исследованиях [6–13]. Главная причина при-

менения ансамблирования — адаптивность к классу моделей, высокая достоверность результатов, достигаемая за счет обучения различных моделей с несколькими классами. При этом, мы получаем высококоррелированные выходные данные. С таким подходом

большая часть моделей оценивает значимость выбранного класса, который и обеспечивает высокую достоверность метода [14].

В работе рассматривается подход к созданию стека моделей на основе методов машинного обучения и тестирование его на конкретных водных объектах.

В качестве инструментария была использована специализированная программная среда на основе языка программирования Python и необходимые библиотеки машинного обучения: TensorFlow, Keras, NumPy, Pandas, Matplotlib, Seaborn, Scipy. На их основе для создания ансамбля моделей использовались следующие модели: Neural Network, Xgboost, RandomForest, LogisticRegression, kNearestNeighbors (kNN).

Применяемая нами модель Neural Network разработана авторами и подробно описана в [15]. Ее архитектура и параметры включают на первых двух слоях по 64 нейрона с функцией активации relu, на последнем слое 6 нейронов с функцией активации Softmax. В результате обучения нейронной сети была получена достаточно высокая точность на обучающей (98,96 %) и проверочной выборках (96,63 %).

kNN (k Nearest Neighbors или k Ближайших соседей) — один из самых простых методов классификации. В основе метода лежит определение новой категории данных на основе близлежащих категорий [16].

XGBoost — алгоритм машинного обучения, основанный на дереве поиска решений и использующий фреймворк градиентного бустинга. Хорошо работает на структурированных или табличных данных небольших размеров [17].

Random Forest является композицией (ансамблем) множества решающих деревьев, что позволяет снизить проблему переобучения и повысить точность в сравнении с одним деревом. Прогноз получается в результате агрегирования ответов множества деревьев. Тренировка деревьев происходит независимо друг от друга (на разных подмножествах), что не просто решает проблему построения одинаковых деревьев на одном и том же наборе данных, но и делает этот алгоритм весьма удобным для применения в системах распределённых вычислений [18].

LogisticRegression — это статистическая

модель, используемая для прогнозирования вероятности возникновения некоторого события путём его сравнения с логистической кривой. Эта регрессия выдаёт ответ в виде вероятности бинарного события (1 или 0) [19].

Для оценки качества вод в разработанном программном обеспечении предусмотрено 6 классов качества воды. Классификация классов качества вод, представленных в табл. 1, разработана на основе метода оценки качества вод по ассоциативным показателям, подробно описанных в [20–21].

Таблица 1

Классы качества вод

Класс качества	Название классов
1	Чистая
2	Умеренно загрязненная
3	Загрязненная
4	Грязная
5	Очень грязная
6	Чрезвычайно грязная

### ОПИСАНИЕ ВОДНЫХ ОБЪЕКТОВ

В качестве водных объектов выбраны две техногенные реки Аба и Ускат. Эти реки являются левыми притоками р. Томь (схема речной сети (рис. 1)) и находятся на территории Кемеровской области – Кузбасса. На качество рек Аба и Ускат сильное влияние оказывают как прямые сбросы загрязняющих веществ с промпредприятий, так и косвенные (смыв осадками с промплощадок). Топология бассейнов рек Ускат и Аба представлена на рис. 1 и 2.

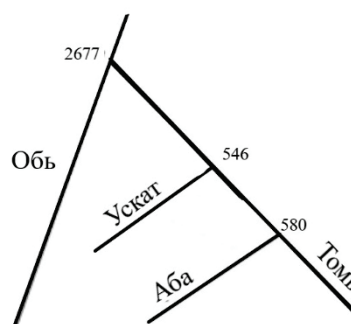


Рис. 1. Положение р. Аба и р. Ускат на линейной схеме речной сети (цифры — расстояние от устья, км)



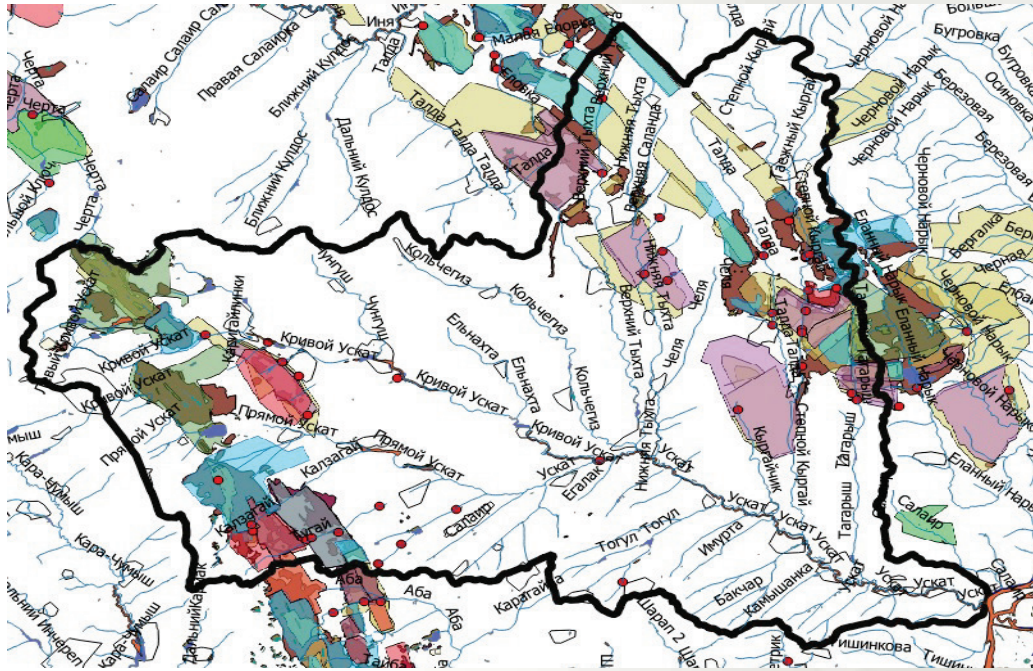


Рис. 2. Бассейн реки Ускат

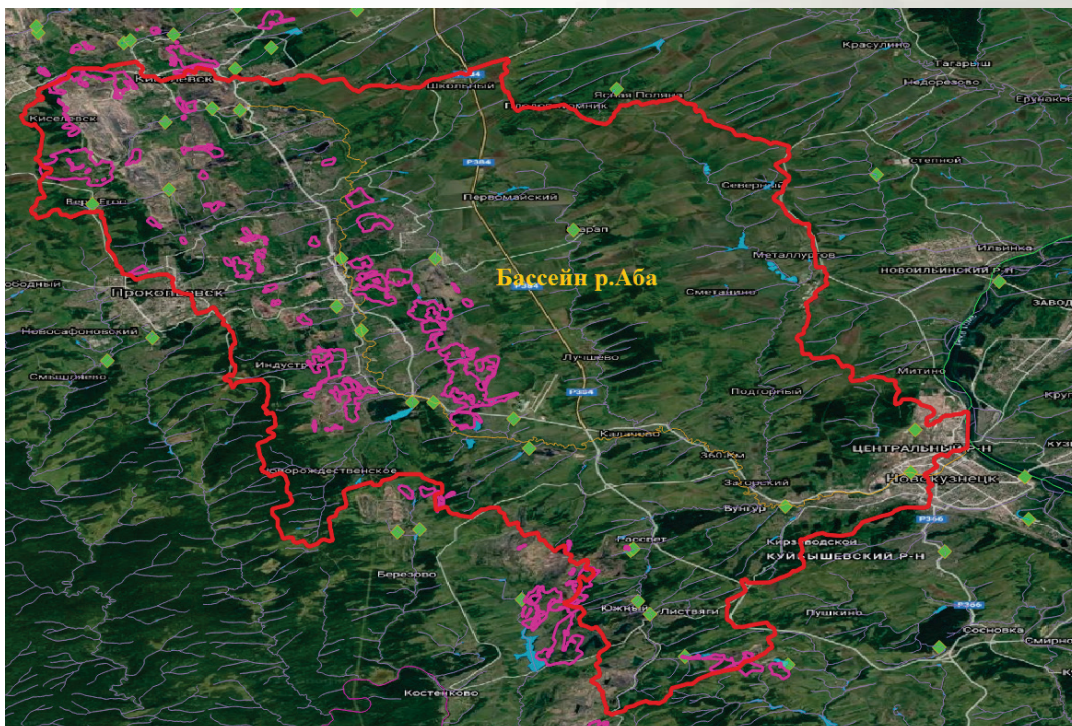


Рис. 3. Бассейн реки Аба

Примечание: на рис. 2 и 3 полигонами выделены угледобывающие предприятия, точками обозначены места сброса в реку сточных вод.

На территориях бассейнов рек Аба и Ускат расположены более 30 угледобывающих предприятий (УДП), свинокомплекс, канализационное хозяйство и другие промышленные

предприятия, осуществляющие сбросы сточных вод. Площадь нарушенных земель в бассейнах рек Аба и Ускат составляет 120 км<sup>2</sup> и 170 км<sup>2</sup> соответственно.

Оценку качества вод в реках Ускат (71 проба) и Аба (416 проб) определяли за период 1985–2003 гг. по 13 показателям: аммоний, БПКпол., взвешенные вещества, железо, марганец, медь, нефтепродукты, нитраты, нитриты, фенолы, хлориды, ХПК.

### МЕТОДЫ ОЦЕНКИ

Для оценки качества вод в реках Аба и Ускат предлагаются несколько подходов: классический, ансамблевый и эвристический (с применением энтропийного метода).

### Классический подход (рис. 4).

Выбираем несколько методов искусственного интеллекта (Neural Network [15], Xgboost, RandomForest, LogisticRegression, kNearestNeighbors (kNN)) и сравниваем их метрики, полученные в результате проведенных вычислений. Схема проведения вычислений представлена на рис. 4. Сначала проводим анализ каждым методом, затем сравниваем метрики качества работы методов (табл. 2) и выбираем лучший метод для определения качества водных объектов.

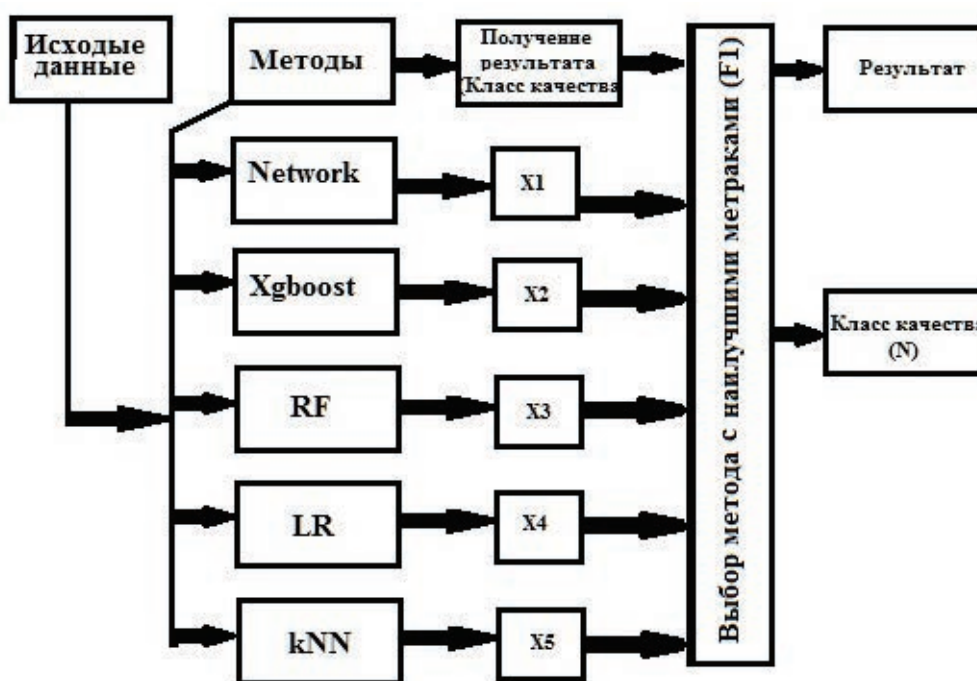


Рис.4. Классическая схема выбора метода и получение результата

Таблица 2

Метрики качества результативности работы методов

№ п/п	Река	Метод	F1	Accuracy	Precision	Recall
1	Ускат	Neural Network	0.97	0.96	0.97	0.97
2	Ускат	Xgboost	0.97	0.96	0.97	0.97
3	Ускат	RandomForest	0.67	0.84	0.68	0.68
4	Ускат	LogisticRegression	0.23	0.37	0.24	0.28
5	Ускат	kNearestNeighbors (kNN)	0.49	0.61	0.5	0.5
6	Аба	Neural network	0.96	0.99	0.97	0.95
7	Аба	Xgboost	0.99	0.99	0.99	0.99



№ п/п	Река	Метод	F1	Accuracy	Precision	Recall
8	Аба	RandomForest	0.88	0.97	0.88	0.88
9	Аба	LogisticRegression	0.39	0,69	0.45	0.43
10	Аба	kNearestNeighbors (kNN)	0.60	0,78	0.6	0,61

Выбираем метод с лучшими метриками и определяем качество воды в реках. Для рек Ускат и Аба лучшие значения метрик у методов Neural Network и Xgboost. В результате оценки был определен, класс качества воды: для р. Ускат — 2 (умеренно загрязненная), для р. Аба — 6 (чрезвычайно грязная). Название класса качества вод определяли по табл. 1.

#### Ансамблевый подход

Ансамбль — это несколько алгоритмов машинного обучения, собранных в единую систему. Такой подход часто используется для того, чтобы усилить «положительные качества» отдельно взятых алгоритмов, которые сами по себе могут работать слабо, а вот их комбинации в группе – ансамбле получать приемлемый результат. При использовании ансамблевых методов алгоритмы учатся одновременно и могут исправлять ошибки друг друга. Типичными примерами методов, направленных на объединение «слабых» учеников в группу сильных, являются:

- **Стекинг.** Могут рассматриваться разнородные отдельно взятые модели. Существует мета-модель, которой на вход подаются базовые модели, а выходом является итоговый прогноз.

- **Бэггинг.** Рассматриваются однородные модели, которые обучаются независимо и параллельно, а затем их результаты просто усредняются. Представителем данного метода является случайный лес.

- **Бустинг.** Рассматриваются однородные модели, которые обучаются последовательно, причем последующая модель должна исправлять ошибки предыдущей. Конечно, в качестве примера можно назвать градиентный бустинг [22].

В общем случае ансамблевая модель попадает в одну из двух категорий: последователь-

ные подходы и параллельные подходы.

Модель последовательного ансамбля оперирует тем, что базовые модели генерируются последовательно. Последовательные методы ансамбля обычно используются для повышения общей производительности, поскольку модель ансамбля может компенсировать неточные предсказания путем повторного взвешивания примеров, которые ранее были неправильно классифицированы. Примером этого является AdaBoost.

Параллельная модель — это методы, основанные на параллельном создании и обучении базовых учащихся. Параллельные методы направлены на снижение частоты ошибок путем параллельного обучения многих моделей и усреднения результатов. Примечательным примером параллельного метода является классификатор случайных лесов.

Таким образом, последовательные модели пытаются повысить производительность путем повторного взвешивания примеров, и модели генерируются последовательно, а параллельные модели работают путем усреднения результатов после обучения многих моделей одновременно [23].

Один из самых популярных способов ансамблирования алгоритмов — стекинг (Stacked Generalization или Stacking), т. е. использования нескольких алгоритмов для решения одной задачи машинного обучения [24]. Суть данного метода заключается в следующем: при обучении нескольких алгоритмов в задачах регрессии используется их среднее, а в задачах классификации — голосование большинству [9, 24].

Основная идея стекинга заключается в использовании базовых классификаторов для получения предсказаний и использовании их как признаков для некоторого «обобща-

ющего» алгоритма. Иными словами, основной идеей стекинга является преобразование исходного пространства признаков задачи в новое пространство, точками которого являются предсказания базовых алгоритмов. Предлагается сначала выбрать набор пар произвольных подмножеств из обучающей выборки, затем для каждой пары обучить базовые алгоритмы на первом подмножестве и предсказать ими целевую переменную для второго подмножества. Предсказанные значения и становятся объектами нового пространства [25].

Для создания ансамбля, нами выбрана совокупность моделей, представленных на рис. 4 (Neural network, Xgboost, RandomForest, LogisticRegression, kNN). Далее сравнили их метрики качества работы, представленные в

табл. 2. Затем из этой совокупности выбрали 3 лучшие модели, которые представлены в табл. 3. Итоговая модель формально изображена на рис. 5. Результат (рис. 5) определяем голосованием. Голосование проводится большинством голосов [9, 24] или мажоритарным голосованием [26], т. е. выбирается ответ большинства. Результат голосования представлен в табл. 4 в графе 7 «Большинство голосов».

Таким образом, в представленном ансамбле каждая модель определяет класс качества воды (табл. 4 графа 4 «Класс качества»), а далее большинством голосов выбираем сколько моделей проголосовала за тот или иной класс. И класс качества воды, за который проголосовали большинство моделей, отображается в табл. 4 в графе 7 «Большинство голосов».

Таблица 3

Метрики качества лучших моделей

№ п/п	Метод	Река	Класс качества	Количество проб	F1	Количество голосов
1	Neural network	Аба	6	206	0.96	6
2	Xgboost	Аба	6	206	0.99	
3	RandomForest	Аба	6	205	0.88	
6	Neural network	Ускат	2	25	0.97	2
7	Xgboost	Ускат	2	24	0.97	
8	RandomForest	Ускат	2	22	0.67	

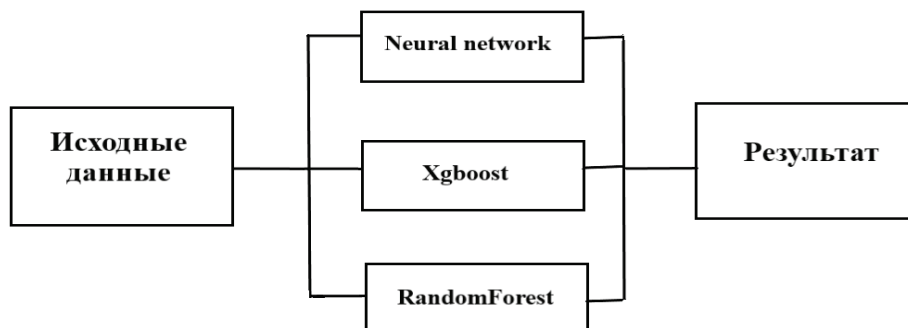


Рис. 5. Итоговая формальная модель ансамблевого метода

Таблица 4

Оценка качества воды в реках большинством голосов

№ п/п	Метод	Река	Класс качества	Количество проб	F1	Большинство голосов
1	2	3	4	5	6	7
1	Neural network	Аба	6	206	0.96	6
2	Xgboost	Аба	6	206	0.99	
3	RandomForest	Аба	6	205	0.88	
4	Neural network	Ускат	2	25	0.97	2
5	Xgboost	Ускат	2	24	0.97	
6	RandomForest	Ускат	2	22	0.67	

**Эвристический подход** (с применением энтропийного метода).

В энтропийном методе распределения примесей в пробах воды подменяются распределениями количества информации о концентрациях ингредиентов и отображаются в

пространстве состояний (фазовом пространстве). Алгоритм оценки загрязненности вод энтропийным методом представлен на рис. 6. Фазовый портрет общей загрязненности вод в реках Аба и Ускат представлен на рис. 7.

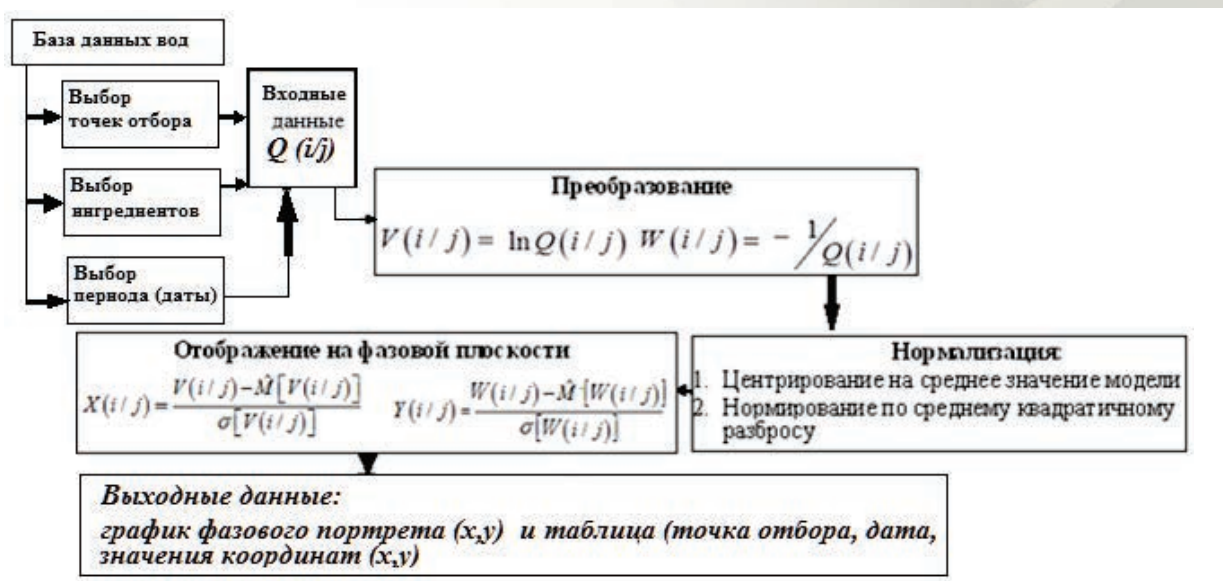


Рис. 6. Алгоритм оценки загрязненности вод энтропийным методом

Загрязненность водного объекта определяется в зависимости от того, в каком квадранте на фазовой плоскости отобразилась

проба воды (как правило, в I квадранте отображаются очень грязные, а в III — чистые пробы воды).

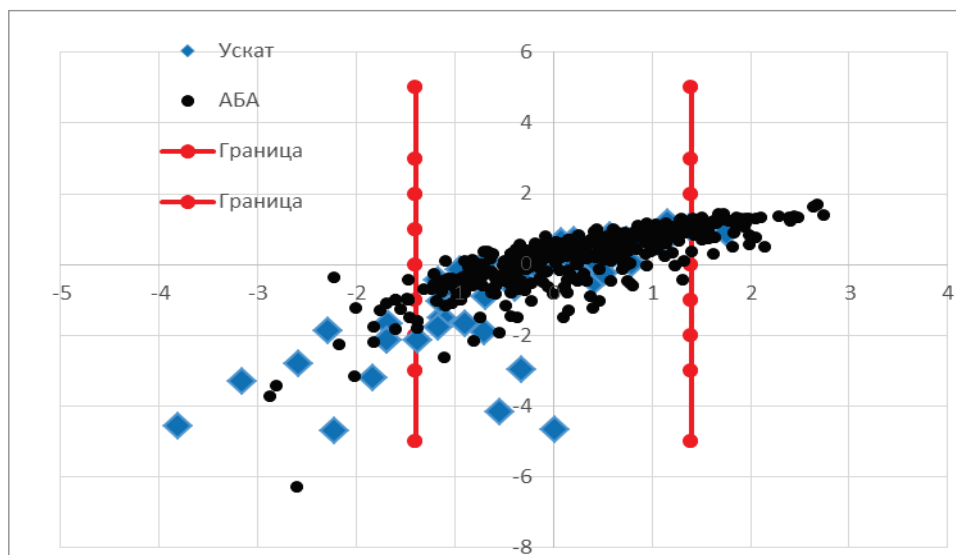


Рис. 7. Фазовый портрет общей загрязненности вод в реках Аба и Ускат

На фазовом портрете (рис. 7) в I квадранте значительно больше проб воды р. Аба, чем проб воды р. Ускат. И, соответственно, в III квадранте преобладают водные пробы р. Ускат. Поэтому факту можно сделать вывод о том, что качество воды в реке Аба хуже, чем в р. Ускат.

### ВЫВОДЫ

Применение алгоритма ансамблирования позволяет получать приемлемый результат. Выбранные для ансамбля методы, с лучшими

метриками качества: Neural Network, Xgboost и RandomForest. Разработанная и описанная в статье модель ансамблирования характеризуется высокими показателями качества прогнозирования. Результат, полученный ансамблевым методом, хорошо коррелирует с результатами других методов: классическим и эвристическим. Таким образом, ансамблевый метод позволяет использовать его для определения качества вод, что и было сделано на примере рек Ускат и Аба.

### СПИСОК ЛИТЕРАТУРЫ

1. Shen H., Shi J., Zheng J., Xu J., Zhu S. Wave forecasting algorithm with stacking ensemble machine learning method // Journal of Hohai University. China. Nanjing. 2020. Vol. 48. No. 4. P. 354–358.
2. Singla S.K., Garg R.D., Dubey O.P. Ensemble machine learning methods to estimate the sugarcane yield based on remote sensing information // Revue D'intelligence Artificielle. Roorkee, Uttarakhand. 2020. Vol. 34. No. 6. P. 731–743.
3. Tang L., Wang J., Li M., Martin J.F., Hill M.C. Predicting unrecognized enhancer-mediated genome topology by an ensemble machine learning model // Genome Research. USA. Houston. 2020. Vol. 31. No. 12. P. 1835–1845.
4. Blanes de Oliveira L.A., de Carvalho Carneiro C. Synthetic geochemical well logs generation using ensemble machine learning techniques for the brazilian pre-salt reservoirs // Journal of Petroleum Science and Engineering. Brazil. Rio de Janeiro. 2021. Vol. 196. P. 108080–108086.
5. Sarwar A., Manhas J., Sharma V., Ali M. Diagnosis of diabetes type-ii using hybrid machine learning based ensemble model // International Journal of Information Technology (Singapore). 2020. Vol. 12. No. 2. P. 419–428.
6. Алабугин С.К., Соколов А.Н. Обнаружение вторжений в автоматизированных системах управления технологическими процессами с использованием ансамбля моделей рекуррентной и двунаправленной генеративно-сопоставительной нейронных сетей // Вестник УрФО. Без-



опасность в информационной сфере. 2021. № 3 (41). С. 38–48.

7. Байков И.И., Семерова Е.А., Курмуков А.И. Метод ансамблирования алгоритмов кластеризации для решения задачи совместной кластеризации // Сенсорные системы. 2021. Т. 35. № 1. С. 43–49.

8. Желтова К.А. Ансамблирование моделей градиентного бустинга в задаче прогнозирования выживаемости пациентов // Актуальные проблемы авиации и космонавтики. 2020. Т. 2. С. 132–134.

9. Мисюра В.В., Пирко Д.В. Ансамблирование регрессионных алгоритмов машинного обучения на примере задачи house prices: advanced regression techniques с ресурса kaggle.com // Молодой исследователь Дона. 2019. № 6(21). С. 54–62.

10. Рысин Н.А. Разработка высоконагруженного решения прогнозирования временных рядов // Хроноэкономика. 2020. № 4 (25). С. 82–87.

11. Сидоренко Д.А., Богомолов Ю.В. Методы ансамблирования алгоритмов классификации на базе сбалансированных данных // Заметки по информатике и математике. Сборник научных статей. 2018. С. 153–160.

12. Янц А.Ю., Токарев А.А. Реализация параллельного алгоритма ансамблирования на GPGPU и CPU // Математическое моделирование в естественных науках. 2018. Т. 1. С. 356–359.

13. Konstantinov Andre, Utkin Lev, Muliukha Vladimir. Gradient boosting machine with partially randomized decision trees // Conference of open innovations association, Fruct association. 2021. No. 28. P. 167–173.

14. Ashish Patel .Ensemble Learning Relation with Bias and variance [Electronic resource] // Published in ML Research Lab. 2019. URL: <https://medium.com/ml-research-lab/ensemble-learning-relation-with-bias-and-variance-431cdc0a3fc9> (date of application: 28.06.2022).

15. Потапов В.П., Счастливец Е.Л., Юкина Н.И., Харлампенков И.Е. Глубокие нейронные сети для оценки качества вод // Горный информационно-аналитический бюллетень. 2019. № 11 (специальный выпуск 37). С. 569–577.

16. Python: реализация алгоритма KNN [Электронный ресурс]. URL: <https://russianblogs.com/article/15541296642/> (дата обращения: 15.09.2022).

17. Алгоритм XGBoost: пусть он царствует долго! [Электронный ресурс]. URL: <https://medium.com/nuances-of-programming/%D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC-xgboost-%D0%BF%D1%83%D1%81%D1%82%D1%8C-%D0%BE%D0%BD-%D1%86%D0%B0%D1%80%D1%81%D1%82%D0%B2%D1%83%D0%B5%D1%82-%D0%B4%D0%BE%D0%BB%D0%B3%D0%BE-dc8c4eca3fbc> (дата обращения: 15.09.2022).

18. Random Forest: прогулки по зимнему лесу [Электронный ресурс]. URL: <https://habr.com/ru/post/320726/> (дата обращения: 15.09.2022).

19. Логистическая регрессия [Электронный ресурс]. URL: [https://ru.wikipedia.org/wiki/%D0%9B%D0%BE%D0%B3%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F\\_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F](https://ru.wikipedia.org/wiki/%D0%9B%D0%BE%D0%B3%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F) (дата обращения: 15.09. 2022).

20. Счастливец Е.Л. Техногенное воздействие угледобывающих предприятий на окружающую среду (на примере Кузбасса): автореферат дис. ... д-р техн. наук: 25.00.36 / Счастливец Евгений Леонидович. Барнаул, 2006. 43 с.

21. Потапов В.П., Мазикин В.П., Счастливец Е.Л., Вашлаева Н.Ю. Геоэкология угледобывающих районов Кузбасса. Новосибирск: Наука, 2005. 660 с.

22. Методы сбора ансамблей алгоритмов машинного обучения: стекинг, бэггинг, бустинг [Электронный ресурс]. URL: <https://habr.com/ru/post/561732/> (дата обращения: 15.09.2022).

23. Ensemble/Voting Classification in Python with Scikit-Learn [Electronic resource]. URL: <https://stackabuse.com/ensemble-voting-classification-in-python-with-scikit-learn/> (date of

application: 15.09.2022).

24. Дьяконов А. Анализ малых данных. Стекинг (Stacking) и блендинг (Blending) [Электронный ресурс]. URL: <https://alexanderdyakonov.wordpress.com/2017/03/10/c%D1%82%D0%B5%D0%BA%D0%B8%D0%BD%D0%B3-stacking-%D0%B8-%D0%B1%D0%BB%D0%B5%D0%BD%D0%B4%D0%B8%D0%BD%D0%B3-blending/> (дата обращения: 15.09.2022).

25. Нагорный О. Построение ансамблей моделей [Электронный ресурс]. URL: <https://nagornyy.me/it/postroenie-ansamblei-modelei/> (дата обращения: 15.09.2022).

26. Ансамбль с мажоритарным голосованием на Python [Электронный ресурс]. URL: <https://shwanoff.ru/majority-voting-ensemble/> (дата обращения: 15.09.2022).

---

**DOI: 10.25558/VOSTNII.2022.92.48.014**

**UDC 004: 504.4.054**

**© V.P. Potapov, N.I. Yukina, E.L. Schastlivtsev, 2022**

**V.P. ПОТАПОВ**

Doctor of Engineering Sciences, Professor,

Head of Laboratory

Federal Research Center for Information and Computing Technologies, Kemerovo branch

e-mail: vadimptpv@gmail.com

**N.I. YUKINA**

Candidate of Engineering Sciences,

Senior Research

Federal Research Center for Information and Computing Technologies, Kemerovo branch

e-mail: leonakler@mail.ru

**E.L. SCHASTLIVTSEV**

Doctor of Engineering Sciences,

Head of Laboratory

Federal Research Center for Information and Computing Technologies, Kemerovo branch

e-mail: schastlivtsev@ict.sbras.ru

**WATER QUALITY ASSESSMENTS AND ENSEMBLE METHODS**

*Despite the active development of systems of mathematical modeling of environmental processes, their complex models have not yet found a worthy reflection in the scientific literature. At the same time, ensemble methods are increasingly used in the processing of various types of data, including multimodal. The main advantage of the ensemble approach is the ability to integrate various methods that can be built on a certain specificity of the problem area, due to which the reliability of the analysis is increased. In this work we consider several approaches to water quality assessment: classical, ensemble and heuristic. The proposed approach to implementing a computational ensemble to assess water quality includes a stack of new data science technologies such as: Neural Networks, Xgboost, RandomForest, LogisticRegression, kNearestNeighbors (kNN). All methods are tested on hydrochemical data of water bodies of the Uskat and Aba rivers flowing through the Kemerovo region and polluted by industrial waste of various enterprises located in their channel. The work presents the results of applying the ensemble to water bodies on the example of the Kuzbass rivers Aba and Uskat.*

Keywords: PYTHON, ENSEMBLE MODELING, NEURAL NETWORKS, DEEP LEARNING,

COMPUTATIONAL MODELS, MACHINE LEARNING, WATER QUALITY ASSESSMENT, MODEL ASSESSMENT, WATER BODIES.

#### REFERENCES

1. Shen H., Shi J., Zheng J., Xu J., Zhu S. Wave forecasting algorithm with stacking ensemble machine learning method // Journal of Hohai University. China. Nanjing. 2020. Vol. 48. No. 4. P. 354–358.
2. Singla S.K., Garg R.D., Dubey O.P. Ensemble machine learning methods to estimate the sugarcane yield based on remote sensing information // Revue D'intelligence Artificielle. Roorkee, Uttarakhand. 2020. Vol. 34. No. 6. P. 731–743.
3. Tang L., Wang J., Li M., Martin J.F., Hill M.C. Predicting unrecognized enhancer-mediated genome topology by an ensemble machine learning model // Genome Research. USA. Houston. 2020. Vol. 31. No. 12. P. 1835–1845.
4. Blanes de Oliveira L.A., de Carvalho Carneiro C. Synthetic geochemical well logs generation using ensemble machine learning techniques for the brazilian pre-salt reservoirs // Journal of Petroleum Science and Engineering. Brazil. Rio de Janeiro. 2021. Vol. 196. P. 108080–108086.
5. Sarwar A., Manhas J., Sharma V., Ali M. Diagnosis of diabetes type-ii using hybrid machine learning based ensemble model // International Journal of Information Technology (Singapore). 2020. Vol. 12. No. 2. P. 419–428.
6. Alabugin S.K., Sokolov A.N. Intrusion detection in automated process control systems using an ensemble of models of recurrent and bidirectional generative-adversarial neural networks // Bulletin of the Ural Federal District. Security in the information sphere [Vestnik URFU. Bezopasnost v informacionnoj sfere]. 2021. No. 3 (41). P. 38–48. [In Russ.].
7. Baykov I.I., Semenova E.A., Kurmukov A.I. Method of clustering algorithms for solving the problem of joint clustering // Sensor systems [Sensornye sistemy]. 2021. Vol. 35. No. 1. P. 43–49. [In Russ.].
8. Zheltova K.A. Ensembling of gradient boosting models in the task of predicting patient survival // Actual problems of aviation and cosmonautics [Aktualnye problemy aviacii i kosmonavtiki]. 2020. Vol. 2. P. 132–134. [In Russ.].
9. Misyura V.V., Pirko D.V. Ensembling regression algorithms of machine learning on the example of the problem house prices: advanced regression techniques from the resource kaggle.com // Young researcher of the Don [Molodoj issledovatel Dona]. 2019. No. 6 (21). P. 54–62. [In Russ.].
10. Rysin N.A. Development of a highly loaded solution for forecasting time series // Chronoeconomics [Hronoekonomika]. 2020. No. 4 (25). P. 82–87. [In Russ.].
11. Sidorenko D.A., Bogomolov Yu.V. Methods of ensembling classification algorithms based on balanced data // Notes on computer science and mathematics. Collection of scientific articles [Zametki po informatike i matematike. Sbornik nauchnyh statej]. 2018. P. 153–160. [In Russ.].
12. Yants A.Yu., Tokarev A.A. Implementation of a parallel algorithm of ensembling on GPGPU and CPU // Mathematical modeling in natural sciences [Matematicheskoe modelirovanie v estestvennyh naukah]. 2018. Vol. 1. P. 356–359. [In Russ.].
13. Konstantinov Andre, Utkin Lev, Muliukha Vladimir. Gradient boosting machine with partially randomized decision trees // Conference of open innovations association, Fruct association. 2021. No. 28. P. 167–173.
14. Ashish Patel .Ensemble Learning Relation with Bias and variance [Electronic resource] // Published in ML Research Lab. 2019. URL: <https://medium.com/ml-research-lab/ensemble-learning-relation-with-bias-and-variance-431cdc0a3fc9> (date of application 28.06.2022).
15. Potapov V.P., Lucky E.L., Yukina N.I., Kharlampenkov I.E. Deep neural networks for water quality assessment // Mining Information and Analytical Bulletin [Gornyj informacionno-analiticheskij byulleten]. 2019. No. 11 (Special issue 37). P. 569–577. [In Russ.].
16. Python: algorithm implementation KNN [Electronic resource]. URL: <https://russianblogs.com>.



com/article/15541296642/ (date of application: 15.09.2022). [In Russ.].

17. The XGBoost algorithm: May it reign long! [Electronic resource]. URL: <https://medium.com/nuances-of-programming/%D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC-xgboost-%D0%BF%D1%83%D1%81%D1%82%D1%8C-%D0%BE%D0%BD%D1%86%D0%B0%D1%80%D1%81%D1%82%D0%B2%D1%83%D0%B5%D1%82-%D0%B4%D0%BE%D0%BB%D0%B3%D0%BE-dc8c4eca3fbc> (date of application: 15.09.2022). [In Russ.].

18. Random Forest: winter forest walks [Electronic resource]. URL: <https://habr.com/ru/post/320726/> (date of application: 15.09.2022). [In Russ.].

19. Logistic regression [Electronic resource]. URL: [https://ru.wikipedia.org/wiki/%D0%9B%D0%BE%D0%B3%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F\\_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F](https://ru.wikipedia.org/wiki/%D0%9B%D0%BE%D0%B3%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F) (date of application: 15.09.2022). [In Russ.].

20. Shchastyantsev E.L. Technogenic impact of coal mining enterprises on the environment (on the example of Kuzbass): abstract of the dis. ... Doctor of Engineering Sciences: 25.00.36 / Lucky Evgeny Leonidovich. Barnaul, 2006. 43 c. [In Russ.].

21. Potapov V.P., Mazikin V.P., Shchastyantsev E.L., Vashlaeva N.Yu. Geoecology of coal-mining areas of Kuzbass. Novosibirsk: Nauka, 2005. 660 p. [In Russ.].

22. Methods of collecting ensembles of machine learning algorithms: stacking, bagging, boosting [Electronic resource]. URL: <https://habr.com/ru/post/561732/> (date of application: 15.09.2022). [In Russ.].

23. Ensemble/Voting Classification in Python with Scikit-Learn [Electronic resource]. URL: <https://stackabuse.com/ensemble-voting-classification-in-python-with-scikit-learn/> (date of application 15.09.2022).

24. Diakonov A. Analysis of small data. Stacking и Blending [Electronic resource]. URL: <https://alexanderdyakonov.wordpress.com/2017/03/10/c%D1%82%D0%B5%D0%BA%D0%B8%D0%BD%D0%B3-stacking-%D0%B8-%D0%B1%D0%BB%D0%B5%D0%BD%D0%B4%D0%B8%D0%BD%D0%B3-blending/> (date of application: 15.09.2022). [In Russ.].

25. Nagorny O. Building ensembles of models [Electronic resource]. URL: <https://nagornyy.me/it/postroenie-ansamblei-modelei/> (date of application: 15.09.2022). [In Russ.].

26. Ensemble with majority voting in Python [Electronic resource]. URL: <https://shwanoff.ru/majority-voting-ensemble/> (date of application: 15.09.2022). [In Russ.].