

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ И АВТОМАТИЗАЦИЯ ПРОЕКТИРОВАНИЯ / COMPUTER MODELING AND DESIGN AUTOMATION

DOI: <https://doi.org/10.60797/IRJ.2025.151.12>

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ У СОТРУДНИКОВ УНИВЕРСИТЕТА СТЕПЕНИ ГОТОВНОСТИ К УЧАСТИЮ В ПРОГРАММАХ ПО ЗДОРОВОМУ ОБРАЗУ ЖИЗНИ

Научная статья

Баженов Р.И.^{1,*}, Баженова Н.Г.², Воротилкина И.М.³

¹ORCID : 0000-0003-2668-1142;

²ORCID : 0000-0003-2659-4970;

^{1,2,3} Приамурский государственный университет им. Шолом-Алейхема, Биробиджан, Российская Федерация

* Корреспондирующий автор (r-i-bazhenov[at]yandex.ru)

Аннотация

Целью исследования является разработка модели машинного обучения, способной прогнозировать готовность сотрудников университетов лично участвовать в программах и мероприятиях, направленных на укрепление здоровья. Опрос проведён с использованием платформы Яндекс.Формы, в нём приняли участие 394 научно-педагогических и административных сотрудника вузов Дальнего Востока. Для обработки и анализа данных использовалась программа Orange. Протестировано несколько моделей машинного обучения, включая метод ближайших соседей, логистическую регрессию, дерево принятия решений, случайный лес, наивный Байес, градиентный бустинг и адаптивный бустинг. Модели оценивались по различным метрикам с акцентом на F1-меру для достижения баланса между точностью и полнотой при работе с несбалансированными классами. Метод «Адаптивный бустинг» показал наивысшее значение F1-меры (0,807) и продемонстрировал высокую точность в прогнозировании класса «Нет» (71,1%) и класса «Да» (85,7%), что делает его наиболее подходящей для поставленной задачи. Модель адаптивного бустинга рекомендована для прогнозирования готовности сотрудников участвовать в мероприятиях, направленных на укрепление здоровья. Полученные результаты могут быть использованы для разработки программ и мероприятий, направленных на продвижение здорового образа жизни среди сотрудников университетов, что повысит эффективность таких инициатив в образовательных организациях.

Ключевые слова: здоровый образ жизни, опрос, Orange, машинное обучение, Адаптивный бустинг.

APPLICATION OF MACHINE LEARNING TO PREDICT UNIVERSITY EMPLOYEES' WILLINGNESS TO PARTICIPATE IN HEALTHY LIFESTYLE PROGRAMMES

Research article

Bazhenov R.I.^{1,*}, Bazhenova N.G.², Vorotilkina I.M.³

¹ORCID : 0000-0003-2668-1142;

²ORCID : 0000-0003-2659-4970;

^{1,2,3} Sholom-Aleichem Priamursky State University, Birobidzhan, Russian Federation

* Corresponding author (r-i-bazhenov[at]yandex.ru)

Abstract

The aim of the study is to develop a machine learning model capable of predicting the willingness of university employees to personally participate in programmes and activities aimed at health improvement. The survey was conducted using the Yandex.Forma platform, and 394 scientific, pedagogical and administrative staff of universities in the Far East took part in it. Orange software was used for data processing and analysis. Several machine learning models were tested, including nearest neighbour method, logistic regression, decision tree, random forest, naive Bayes, gradient boosting and adaptive boosting. Models were evaluated on various metrics, with emphasis on F1-measure, to achieve a balance between accuracy and completeness when dealing with unbalanced classes. The Adaptive Boosting method showed the highest F1-measure value (0.807) and demonstrated high accuracy in predicting the 'No' class (71.1%) and 'Yes' class (85.7%), making it the most appropriate for the task. The adaptive boosting model is recommended for predicting employees' willingness to participate in health improvement activities. The results obtained can be used to develop programmes and activities aimed at promoting healthy lifestyles among university employees, which will increase the effectiveness of such initiatives in educational organizations.

Keywords: healthy lifestyle, survey, Orange, machine learning, Adaptive Boosting.

Введение

1.1. Актуальность

Применение методов машинного обучения (МО) позволяет автоматизировать анализ больших объемов данных, полученных из опросов, и выявлять скрытые закономерности, например, в предпочтениях и мотивациях сотрудников вузов. Использование МО способствует более точной сегментации участников и адаптации программ по здоровому образу жизни (ЗОЖ) под их реальные потребности.

Актуальность темы обусловлена растущей необходимостью внедрения программ по здоровому образу жизни в образовательных учреждениях, что связано с улучшением качества жизни сотрудников и повышением их

профессиональной эффективности. Однако успешная реализация таких программ во многом зависит от степени готовности сотрудников к участию в них, что требует точного анализа их мотивации, предпочтений и барьеров. Традиционные методы обработки данных опросов не всегда позволяют выявить скрытые закономерности, особенно при работе с большими объемами информации. Применение МО в данной области предоставляет возможность автоматизировать анализ данных, улучшить прогнозирование и создать персонализированные подходы к вовлечению сотрудников. Таким образом, использование современных технологий для решения данной задачи не только повышает эффективность программ ЗОЖ, но и способствует развитию цифровых инструментов в управлении образовательной средой.

1.2. Обзор исследований

Дж. Гриммер, М. Э. Робертс и Б. М. Стюарт рассматривают использование машинного обучения в социальных науках. Исследователи подчеркивают, что интеграция методов машинного обучения требует переосмысления как их применения, так и традиционных исследовательских практик, поскольку эти методы позволяют выявлять новые концепты, измерять их распространённость, оценивать причинно-следственные связи и делать прогнозы. Они предлагают агностический подход, ориентированный на задачи социальных наук, который способствует переходу от дедуктивного к более последовательному, интерактивному и индуктивному подходу, расширяя возможности для решения сложных исследовательских вопросов [1].

Исследование Н. Н. Леонова посвящено разработке методики применения машинного обучения для анализа социологических данных. В работе рассматриваются возможности использования машинного обучения для обработки и анализа информации, собранной в рамках социологических исследований. Автор акцентирует внимание на важности дальнейшего развития данной методологии и подробно описывает применение ряда классических методов машинного обучения в этом контексте [2].

М. Д. Молина и Ф. Гарип представили применение методов машинного обучения в социологии. Ученые подчеркивают, что машинное обучение, как область на стыке статистики и компьютерных наук, позволяет извлекать информацию и знания, что может помочь в решении традиционных вопросов в социологии [3].

В статье Х. Лейтгёб, Д. Пранднер и Т. Вольбринг показано влияние цифровой революции и технологий больших данных на социальные науки, с акцентом на использование машинного обучения. Рассматриваются ключевые аспекты цифровизации, такие как датафикация общества, развитие вычислительных мощностей и алгоритмов, а также их влияние на методы исследования и анализ социальных явлений. Авторы подчеркивают необходимость адаптации социальных наук к новым условиям, включая развитие теоретических основ, методологических подходов и образовательных программ для успешной интеграции технологий больших данных и машинного обучения [4].

Анализ данных медико-социологического мониторинга на основе методов машинного обучения показали Г. Г. Рапаков и др. Исследователи представили, как машинное обучение может помочь в интерпретации и использовании социологических и медицинских данных для принятия решений [5].

М. Б. Богданова и И. Б. Смирнова рассматривают возможности и ограничения использования цифровых следов и методов машинного обучения в социологии. Авторы выделяют новые источники данных и подчеркивают, как потенциал для более глубокого анализа социальных явлений, так и ограничения, связанные с этими методами, включая вопросы этики, качества данных и интерпретации результатов [6].

Исследование Р. Х. Хайбергера применяет машинное обучение в социологии для предсказания пола исследователей и выявления их предпочтений в исследованиях. В частности, он использует предсказанный пол в тематическом моделировании, чтобы подчеркнуть значительные тематические различия между работами мужских и женских ученых, которые зачастую были упущены из виду [7].

К. Б. Мухамадиева показал применение машинного обучения для обеспечения обратной связи с преподавателем после обработки данных в режиме реального времени для повышения эффективности преподавания и обучения. Основное внимание в работе уделяется тому, как современные технологии способствуют улучшению качества образовательного процесса [8].

Разработали технику в социологии для систем анализа настроений, основанную на методе машинного обучения (глубокие нейронные сети с использованием методов выбора признаков на основе хи-квадрат), М. Хуссейн и Ф. Озюрт. Эта комбинация показала улучшение точности классификации на стандартных наборах данных [9].

Р. Р. Халде рассмотрел различные способы применения алгоритмов машинного обучения в образовательных учреждениях. Автор показал, как машинное обучение может способствовать улучшению образовательной системы, в том числе через предсказание успеха студентов и оптимизацию учебных процессов [10].

1.3. Постановка задачи

Для проведения эксперимента по определению ценностно-мотивационных приоритетов научно-педагогических работников и сотрудников в разрезе здорового образа жизни (на примере вузов Дальнего Востока) составлен опрос [11].

Вопросы концентрируются возле нескольких тем:

- инфраструктура для занятий физической культурой и спортом в университете (спортивные залы, бассейны, тренажерные залы и т.д.);
- роль администрации университетов в продвижении здорового образа жизни среди научно-педагогических работников и сотрудников;
- вовлеченность научно-педагогических работников и сотрудников в программы по здоровому образу жизни, их мотивация и барьеры к участию;
- мотивация научно-педагогических работников и сотрудников к ведению здорового образа жизни;
- интеграция вопросов здорового образа жизни в образовательные программы университетов;

- системный подход к организации оздоровительных мероприятий, индивидуальных программ для научно-педагогических работников и сотрудников;
- предложения по формированию здорового образа жизни в условиях образовательной организации;
- информация о респонденте.

Важной составляющей в дальнейшем проектировании и реализации программ университетов, направленных на ЗОЖ, является готовность сотрудников к личному участию в таких мероприятиях. Поэтому для построения модели машинного обучения выбран вопрос «Готовы ли Вы лично участвовать в программах и мероприятиях, направленных на укрепление здоровья?».

1.4. Цель исследования

Цель исследования – разработать модель машинного обучения, способную прогнозировать степень готовности сотрудников к личному участию в программах и мероприятиях университета по ЗОЖ.

Материалы и методы

В теории социологических опросов используются термины: доверительная вероятность, доверительная погрешность. Доверительная вероятность – это показатель точности измерений. Доверительная погрешность – это возможная ошибка результатов исследования.

Например, при генеральной совокупности более 250000 тыс. человек выборка будет равняться 384 человека при доверительной вероятности 95% и погрешности 5% или при доверительном интервале $95 \pm 5\%$ [11]. Соответственно, получим репрезентативную выборку с минимальной вероятностью статистической ошибки.

Оценим генеральную совокупность научно-педагогических работников и сотрудников. По данным в Информационно-аналитических материалах по результатам проведения мониторинга деятельности образовательных организаций высшего образования за 2023 год в Дальневосточном федеральном округе 171 477 студентов бакалавриата, специалитета, магистратуры [12]. Нормативное соотношение между студентами и преподавателями 1:12. Соответственно примерно 14 000 (округление до тысяч).

Таким образом, генеральная совокупность научно-педагогических работников и сотрудников в Дальневосточном федеральном округе составляет порядка 14 тыс. человек. Получаем минимальную оценку репрезентативной выборки в 374 человек при доверительном интервале $95 \pm 5\%$. В анкетировании приняло участие 394 научно-педагогических работников и сотрудников вузов Дальнего Востока, что достаточно для дальнейшей работы.

Опрос проведен на платформе Яндекс.Формы [13].

Для построения модели машинного обучения выбрана компьютерная система Orange [14]. В связи с тем, что Orange не поддерживает вывод р-уровня значимости (p-value), то корреляционный анализ проводился отдельно с помощью написанной программы на python, в которой использовались библиотеки pandas и SciPy для расчета коэффициента Спирмена и р-уровня значимости.

Результаты и обсуждение

Для построения модели осуществлена предварительная обработка данных и конструирование признаков (Feature Engineering).

Начальное распределение ответов на вопрос «18. Готовы ли Вы лично участвовать в программах и мероприятиях, направленных на укрепление здоровья?» представлено в табл.1.

Таблица 1 - Распределение ответов в вопросе «18. Готовы ли Вы лично участвовать в программах и мероприятиях, направленных на укрепление здоровья?»

DOI: <https://doi.org/10.60797/IRJ.2025.151.12.1>

Варианты ответов	Доля ответов (%)
Да	65,65
Затрудняюсь ответить	21,09
Нет	13,26

Для более полного представления сконструирована новая целевая переменная «18.1 Готовы ли Вы лично участвовать в программах и мероприятиях, направленных на укрепление здоровья?», только в ней объединяются ответы «Нет» и «Затрудняюсь ответить» в «Нет». Такое объединение возможно, так как можно предположить, что при неопределённости будет в большинстве случаев «Нет».

Распределение по переменной «18.1 Готовы ли Вы лично участвовать в программах и мероприятиях, направленных на укрепление здоровья?» представлено в табл. 2.

Таблица 2 - Распределение ответов в новой переменной «18.1 Готовы ли Вы лично участвовать в программах и мероприятиях, направленных на укрепление здоровья?»

DOI: <https://doi.org/10.60797/IRJ.2025.151.12.2>

Варианты ответов	Доля ответов (%)
Да	65,65
Нет	34,35

Таким образом, в новой целевой переменной улучшилась балансировка классов.

Для корреляционного анализа выбран расчёт коэффициента Спирмена, так как переменные категориальные. Из-за того что среда Orange не показывает р-уровня значимости (p-value), анализ для предварительно закодированных переменных проведен с помощью отдельно написанной программы на Python в среде Google Colab.

При корреляционном анализе обнаружены достаточные сильные связи (значимые по p-value) между переменными (табл. 3).

Таблица 3 - Результаты корреляционного анализа (значимый по p-value коэффициент корреляции Спирмена больше по модулю 0,4)

DOI: <https://doi.org/10.60797/IRJ.2025.151.12.3>

Переменная 1	Переменная 2	Коэффициент корреляции Спирмена
1. Создана ли в Вашем университете инфраструктура для занятий физической культурой и спортом (спортивные залы, бассейны, тренажерные залы и т.д.).	2. Доступна ли для научно-педагогических работников и сотрудников университета созданная в Вашем университете инфраструктура для занятий физической культурой и спортом (спортивные залы, бассейны, тренажерные залы и т.д.).	0,64
2. Доступна ли для научно-педагогических работников и сотрудников университета созданная в Вашем университете инфраструктура для занятий физической культурой и спортом (спортивные залы, бассейны, тренажерные залы и т.д.).	3. Считаете ли вы, что в Вашем университете созданы условия для ведения здорового образа жизни (спортивные секции, медицинские кабинеты, пропаганда ЗОЖ и т.д.)?	0,61
3. Считаете ли вы, что в Вашем университете созданы условия для ведения здорового образа жизни (спортивные секции, медицинские кабинеты, пропаганда ЗОЖ и т.д.)?	4. Какие меры по поддержанию здорового образа жизни преподавателей и сотрудников предпринимает Ваш университет? / К) не предпринимает	0,53
1. Создана ли в Вашем университете инфраструктура для занятий физической культурой и спортом (спортивные залы, бассейны, тренажерные залы и т.д.).	3. Считаете ли вы, что в Вашем университете созданы условия для ведения здорового образа жизни (спортивные секции, медицинские кабинеты, пропаганда ЗОЖ и т.д.)?	0,48
1. Создана ли в Вашем университете инфраструктура для занятий физической культурой и спортом (спортивные залы, бассейны, тренажерные залы и т.д.).	4. Какие меры по поддержанию здорового образа жизни преподавателей и сотрудников предпринимает Ваш университет? / К) не предпринимает	0,43
4. Какие меры по поддержанию здорового образа жизни преподавателей и сотрудников предпринимает Ваш университет? / З) Ежедневные занятия фитнесом	4. Какие меры по поддержанию здорового образа жизни преподавателей и сотрудников предпринимает Ваш университет? / Д) Групповые занятия физической активности	0,40

Для предотвращения влияния мультиколлинеарности из входных переменных исключены вопросы: «2. Доступна ли для научно-педагогических работников и сотрудников университета созданная в Вашем университете

инфраструктура для занятий физической культурой и спортом (спортивные залы, бассейны, тренажерные залы и т.д.)» и «4. Какие меры по поддержанию здорового образа жизни преподавателей и сотрудников предпринимает Ваш университет?». Также были исключены личностные переменные: «Курите ли Вы?», «Употребляете ли Вы алкогольные напитки?», «Ваш пол», «Ваша должность».

Из-за ограниченного объема данных (394 строки) для построения моделей машинного обучения в Orange была применена кросс-валидация на 10 фолдов со стратификацией. Этот метод перекрестной проверки делит набор данных на 10 случайных частей, при этом сохраняя пропорции классов. Модель обучается 10 раз, каждый раз используя 9 частей данных для обучения, а оставшуюся одну часть – для проверки. Соответственно, каждый кусочек данных побывает в роли тестовой выборки ровно один раз.

Проведены эксперименты с несколькими моделями: Метод ближайших соседей (рис. 1); Логистическая регрессия (рис. 2); Дерево принятия решений (рис. 3); Случайный лес (рис. 4); Наивный Байес (рис.5); Градиентный бустинг (рис.6); Адаптивный бустинг (Адабуст) (рис.7).

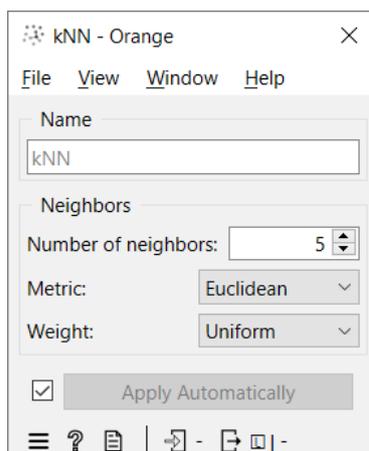


Рисунок 1 - Параметры модели: Метод ближайших соседей
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.4>

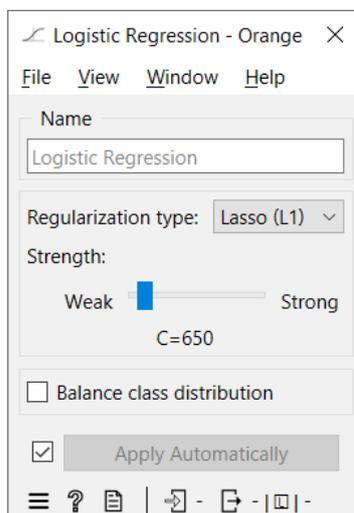


Рисунок 2 - Параметры модели: Логистическая регрессия
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.5>

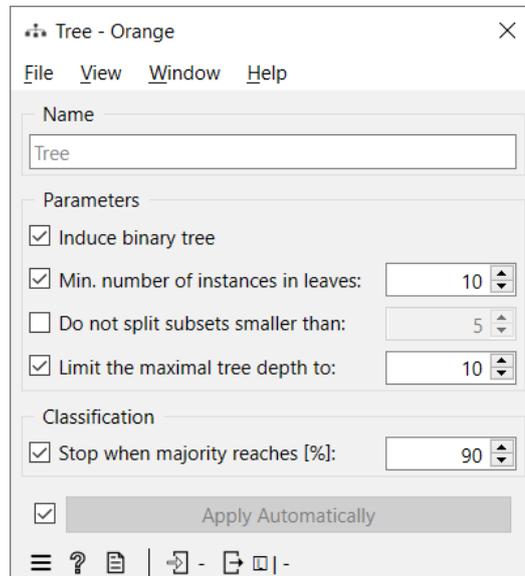


Рисунок 3 - Параметры модели: Дерево принятия решений
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.6>

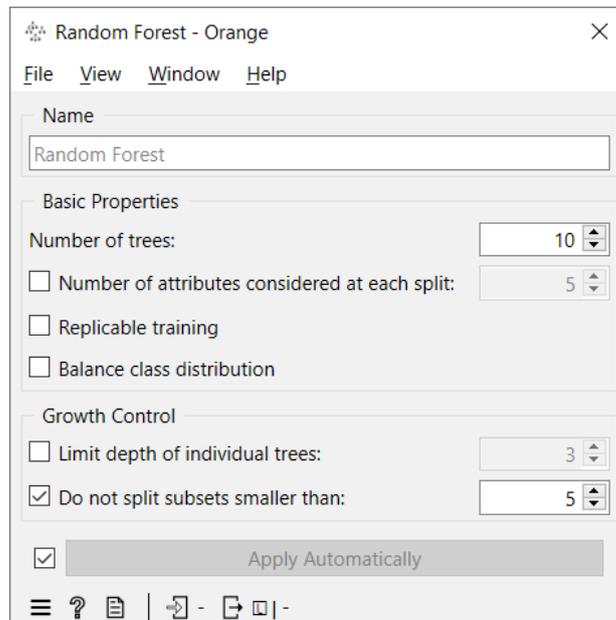


Рисунок 4 - Параметры модели: Случайный лес
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.7>

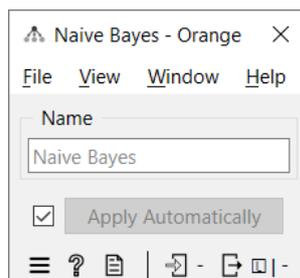


Рисунок 6 - Параметры модели: Наивный Байес
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.8>

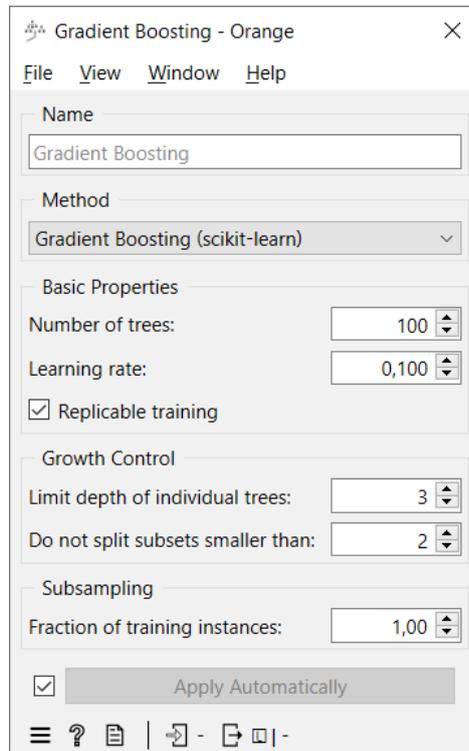


Рисунок 6 - Параметры модели: Градиентный бустинг
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.9>

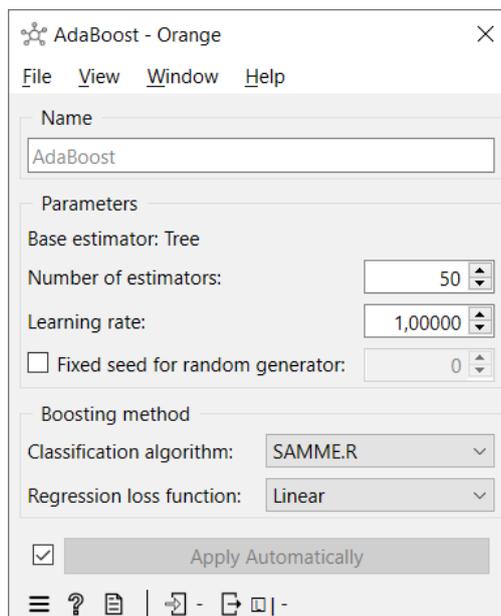


Рисунок 7 - Параметры модели: Адаптивный бустинг
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.10>

Итоговая схема моделирования в Orange показана на рис.8.

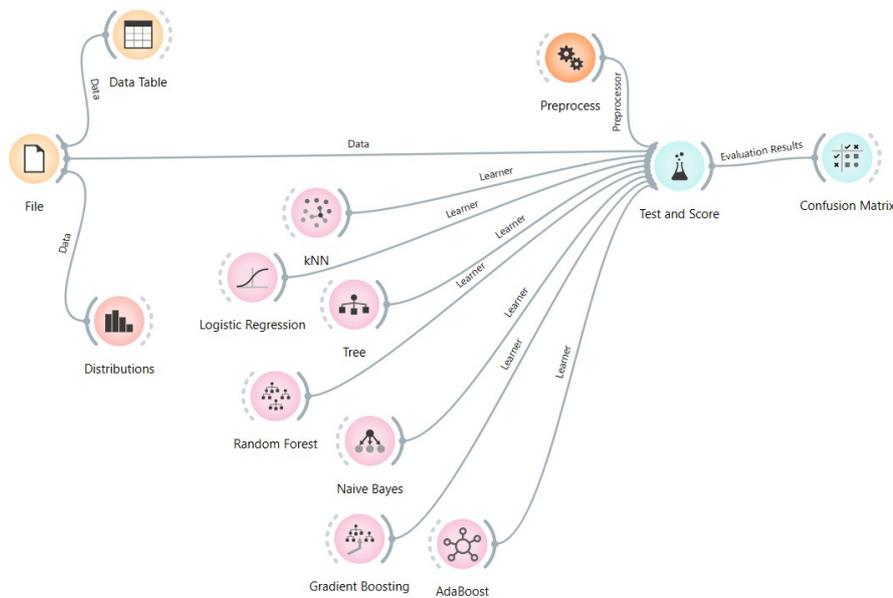


Рисунок 8 - Схема моделирования в Orange
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.11>

После обучения моделей полученные результаты представлены на рис. 9.

Model	AUC	CA	F1	Prec	Recall
Logistic Regression	0.729	0.703	0.695	0.692	0.703
kNN	0.727	0.711	0.694	0.697	0.711
Tree	0.776	0.744	0.740	0.739	0.744
Random Forest	0.871	0.807	0.801	0.804	0.807
Gradient Boosting	0.827	0.777	0.772	0.772	0.777
Naive Bayes	0.747	0.693	0.697	0.705	0.693
AdaBoost	0.784	0.807	0.807	0.806	0.807

Рисунок 9 - Результаты моделирования
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.12>

На рис. 9 показаны значения метрик:

- AUC (Area under ROC) – Площадь под ROC-кривой;
- CA (Classification accuracy) – Доля правильно классифицированных примеров;
- F1 (F1-score) – Взвешенное гармоническое precision и recall;
- Prec (Precision) – Доля истинных положительных результатов среди случаев, классифицированных как положительные (точность);
- Recall (Recall) – Доля истинных положительных результатов среди всех положительных случаев в данных (полнота).

Для оценки качества моделей выбрана метрика F1. F1-мера является метрикой для оценки качества бинарной классификации, объединяющей точность (precision) и полноту (recall) в одно значение. Метрика полезна, когда требуется сбалансировать точность и полноту, особенно в случаях с несбалансированными классами.

По метрике F1 лучшей моделью можно признать модель «Адаптивный бустинг» со значением 0,807.

В процессе моделирования необходимо добиваться уменьшения ошибок второго рода, такие ошибки для целевой переменной получаются в классе (ответе) «Нет».

На рис. 10 представлена матрица ошибок для модели Адабуст.

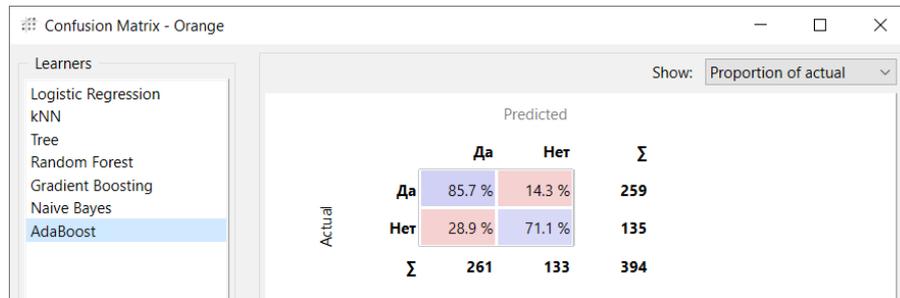


Рисунок 10 - Матрица ошибок модели Адабуст
DOI: <https://doi.org/10.60797/IRJ.2025.151.12.13>

Для метода «Адаптивный бустинг» (рис. 10) получится достаточно хороший результат: модель дает правильный прогноз в 71,1% значения «Нет» (вторая строка в матрице) и ошибается в 28,9%, то есть правильно классифицирует «Нет» в 71,1% случаев. В то же время модель по ошибкам первого рода дает правильный прогноз в 85,7% значения «Да» (первая строка в матрице) и ошибается в 14,3%, то есть правильно классифицирует «Да» в 85,7% случаев. У других моделей результаты хуже.

Таким образом, лучшей моделью для прогнозирования качества у научно-педагогического работника, с кем требуется провести различные мероприятия по убеждению его в укреплении здоровья, является адаптивный бустинг.

Заключение

В ходе исследования была проведена предварительная обработка данных для построения модели машинного обучения, в том числе улучшение балансировки классов целевой переменной.

Выявлены значимые и достаточно высокие корреляции между переменными, что привело к исключению некоторых факторов (мультиколлинеарных) для минимизации их влияния на качество модели.

Для построения и оценки моделей использовались несколько алгоритмов машинного обучения: Метод ближайших соседей, Логистическая регрессия, Дерево решений, Случайный лес, Наивный Байес, Градиентный бустинг, Адаптивный бустинг. Результаты моделирования сравнивались по метрике F1, позволяющей сбалансировать точность и полноту.

Лучшая F1-мера (0,807) была достигнута моделью Адаптивного бустинга, показавшую достаточно высокую точность в прогнозировании класса «Нет» (71,1%) и класса «Да» (85,7%), что делает её наиболее подходящей для задачи.

Модель Адаптивного бустинга рекомендована для использования при прогнозировании готовности научно-педагогических работников участвовать в мероприятиях, направленных на укрепление здоровья, с целью дальнейшего формирования программ убеждения и поддержки здорового образа жизни.

Финансирование

Научное исследование проведено при поддержке средств государственной программы стратегического академического лидерства «Приоритет 2030. Дальний Восток».

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Funding

The scientific research was carried out with the support of the funds of the state program of strategic academic leadership "Priority 2030. The Far East".

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

- Grimmer J. Machine learning for social science: An agnostic approach / J. Grimmer, M.E. Roberts, B.M. Stewart // Annual Review of Political Science. — 2021. — № 1. — P. 395–419. — DOI: 10.1146/annurev-polisci-053119-015921.
- Леонов Н.Н. О методике применения машинного обучения в анализе социологических данных / Н.Н. Леонов // Социологический альманах. — 2019. — № 10. — С. 56–64.
- Molina M. Machine learning for sociology / M. Molina // Annual Review of Sociology. — 2019. — № 1. — P. 27–45. — DOI: 10.1146/annurev-soc-073117-041106.
- Leitgöb H. Big data and machine learning in sociology / H. Leitgöb, D. Prandner // Frontiers in Sociology. — 2023. — № 8. — P. 1173155. — DOI: 10.3389/fsoc.2023.1173155.
- Рапаков Г.Г. Анализ данных медико-социологического мониторинга на основе методов машинного обучения / Г.Г. Рапаков, Г.Т. Банщиков, В.А. Горбунов [и др.] // Вестник Череповецкого государственного университета. — 2019. — № 1. — С. 27–38.

6. Богданов М.Б. Возможности и ограничения цифровых следов и методов машинного обучения в социологии / М.Б. Богданов, И.Б. Смирнов // Мониторинг общественного мнения: экономические и социальные перемены. — 2021. — № 1. — С. 304–328.
7. Heiberger R.H. Applying machine learning in sociology: How to predict gender and reveal research preferences / R.H. Heiberger // KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie. — 2022. — № 1. — P. 383–406. — DOI: 10.1007/s11577-022-00839-2.
8. Мухамадиева К.Б. Машинное обучение в совершенствовании образовательной среды / К.Б. Мухамадиева // Образование и проблемы развития общества. — 2020. — № 4. — С. 70–77.
9. Hussein M. A new technique for sentiment analysis system based on deep learning using Chi-Square feature selection methods / M. Hussein, F. Özyurt // Balkan Journal of Electrical and Computer Engineering. — 2021. — № 4. — P. 320–326.
10. Halde R.R. Application of Machine Learning algorithms for betterment in education system / R.R. Halde. — Pune: IEEE, 2016. — P. 1110–1114. — DOI: 10.1109/ICACDOT.2016.7877759.
11. Как правильно рассчитать объем выборки? // Анкетолог. — 2010 — URL: <https://blog.anketolog.ru/2015/12/vyborka/> (дата обращения: 01.12.2024)
12. Дальневосточный федеральный округ. Характеристика системы высшего образования // Информационно-аналитические материалы по результатам проведения мониторинга деятельности образовательных организаций высшего образования. — 2024 — URL: https://monitoring.miccedu.ru/iam/2023/_vpo/material.php?type=1&id=7 (дата обращения: 01.12.2024)
13. Ценностно-мотивационные приоритеты и запросы научно-педагогических работников и сотрудников в разрезе здорового образа жизни // Яндекс Формы. — 2024 — URL: <https://forms.yandex.ru/u/66cfdbe0d04688535c86cc57/> (дата обращения: 01.12.2024)
14. Suggested Download. — 2024 — URL: <https://orangedatamining.com/download/> (accessed: 01.12.2024)

Список литературы на английском языке / References in English

1. Grimmer J. Machine learning for social science: An agnostic approach / J. Grimmer, M.E. Roberts, B.M. Stewart // Annual Review of Political Science. — 2021. — № 1. — P. 395–419. — DOI: 10.1146/annurev-polisci-053119-015921.
2. Leonov N.N. O metodike primenenija mashinnogo obuchenija v analize sotsiologicheskikh dannyh [On the methodology of using machine learning in the analysis of sociological data] / N.N. Leonov // The Sociological Almanac. — 2019. — № 10. — P. 56–64. [in Russian]
3. Molina M. Machine learning for sociology / M. Molina // Annual Review of Sociology. — 2019. — № 1. — P. 27–45. — DOI: 10.1146/annurev-soc-073117-041106.
4. Leitgöb H. Big data and machine learning in sociology / H. Leitgöb, D. Prandner // Frontiers in Sociology. — 2023. — № 8. — P. 1173155. — DOI: 10.3389/fsoc.2023.1173155.
5. Rapakov G.G. Analiz dannyh mediko-sotsiologicheskogo monitoringa na osnove metodov mashinnogo obuchenija [Analysis of medical and sociological monitoring data based on machine learning methods] / G.G. Rapakov, G.T. Banschikov, V.A. Gorbunov [et al.] // Bulletin of Cherepovets State University. — 2019. — № 1. — P. 27–38. [in Russian]
6. Bogdanov M.B. Vozможности i ogranichenija tsifrovyyh sledov i metodov mashinnogo obuchenija v sotsiologii [Possibilities and limitations of digital traces and machine learning methods in sociology] / M.B. Bogdanov, I.B. Smirnov // Monitoring Public Opinion: Economic and Social Changes. — 2021. — № 1. — P. 304–328. [in Russian]
7. Heiberger R.H. Applying machine learning in sociology: How to predict gender and reveal research preferences / R.H. Heiberger // KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie. — 2022. — № 1. — P. 383–406. — DOI: 10.1007/s11577-022-00839-2.
8. Muhamadiev K.B. Mashinnoe obuchenie v sovershenstvovanii obrazovatel'noj sredy [Machine learning in improving the educational environment] / K.B. Muhamadiev // Education and Problems of Social Development. — 2020. — № 4. — P. 70–77. [in Russian]
9. Hussein M. A new technique for sentiment analysis system based on deep learning using Chi-Square feature selection methods / M. Hussein, F. Özyurt // Balkan Journal of Electrical and Computer Engineering. — 2021. — № 4. — P. 320–326.
10. Halde R.R. Application of Machine Learning algorithms for betterment in education system / R.R. Halde. — Pune: IEEE, 2016. — P. 1110–1114. — DOI: 10.1109/ICACDOT.2016.7877759.
11. Kak pravil'no rasschitat' ob'em vyborki? [How to calculate the sample size correctly?] // The questionnaire specialist. — 2010 — URL: <https://blog.anketolog.ru/2015/12/vyborka/> (accessed: 01.12.2024) [in Russian]
12. Dal'nevostochnyj federal'nyj okrug. Harakteristika sistemy vysshego obrazovaniya [The Far Eastern Federal District. Characteristics of the higher education system] // Information and analytical materials on the results of monitoring the activities of educational institutions of higher education. — 2024 — URL: https://monitoring.miccedu.ru/iam/2023/_vpo/material.php?type=1&id=7 (accessed: 01.12.2024) [in Russian]
13. Tsennostno-motivatsionnye prioritety i zaprosy nauchno-pedagogicheskikh rabotnikov i sotrudnikov v razreze zdorovogo obraza zhizni [Value-motivational priorities and requests of scientific and pedagogical workers and employees in the context of a healthy lifestyle] // Яндекс Формы. — 2024 — URL: <https://forms.yandex.ru/u/66cfdbe0d04688535c86cc57/> (accessed: 01.12.2024) [in Russian]
14. Suggested Download. — 2024 — URL: <https://orangedatamining.com/download/> (accessed: 01.12.2024)