

КЛАССИФИКАЦИЯ ЭПИДЕМИОЛОГИЧЕСКИХ ПРОГНОЗОВ И ПРИМЕНЕНИЕ ГРАДИЕНТНОГО БУСТИНГА В ПРОФИЛАКТИЧЕСКОЙ МЕДИЦИНЕ

Голубков А.В.¹, Гаврилова М.П.²

*1 - Федеральное государственное казенное учреждение «Главный центр
государственного санитарно-эпидемиологического надзора (специального назначения)»*

Министерства обороны Российской Федерации, 105066, г. Москва,

ул. Спартаковская, дом 2Б

Тел. 8(495) 693-18-87, e-mail: golubkov_av@mail.ru

*2 - Федеральное государственное бюджетное учреждение «Санкт-Петербургский
научно-исследовательский институт физической культуры», 191040, г. Санкт-*

Петербург, Лиговский пр. 56, литера Е

Тел. 8(812) 600-41-16, e-mail: mgavrilova@spbniifk.ru

Резюме: в статье предложена новая обоснованная классификация эпидемиологических прогнозов острых респираторных инфекций. Приведены сведения об использовании алгоритмов машинного обучения в различных сферах деятельности человека. Показаны примеры применения метода градиентного бустинга в области клинической медицины с целью прогнозирования соматической заболеваемости. Освещена актуальность применения градиентного бустинга в прогнозе уровней заболеваемости инфекционными болезнями и болезнями органов дыхания. Предложена разработка и осуществление прогноза острых респираторных инфекций в организованных коллективах Вооруженных Сил Российской Федерации с предварительным определением предикторов.

Ключевые слова: машинное обучение, градиентный бустинг, ретроспективный эпидемиологический анализ, прогнозирование, прогноз заболеваемости, острые респираторные инфекции, организованные коллективы, Вооруженные Силы Российской Федерации.

CLASSIFICATION OF EPIDEMIOLOGICAL FORECASTS AND APPLICATION OF GRADIENT BOOSTING IN PREVENTIVE MEDICINE

Golubkov A.V.¹, Gavrilova M.P.²

1 - Federal Government State Establishment «Main Center for State Sanitary and Epidemiological Surveillance (Special Purpose)» of The Russian Federation Defense Ministry, 105066, Moscow, Spartakovskaya st., 2B

Phone number: 8(495) 693-18-87, e-mail: golubkov_av@mail.ru

2 - Federal State Budget Institution «Saint_Peterburg Scientific-Research Institute for physical culture», 191040 St.-Petersburg, Ligovsky pr. 56, LET. «E»

Phone number: 8(812) 600-41-16, e-mail: mgavrilova@spbniifk.ru

Abstract: The article proposes a new justified classification of epidemiological forecasts of acute respiratory infections. An overview of machine learning methods is presented. Examples of the use of the gradient boosting method in the field of clinical medicine for the purpose of predicting somatic morbidity are shown. The relevance of the use of gradient boosting in predicting the incidence of infectious diseases and respiratory diseases is highlighted. The development and implementation of the forecast of acute respiratory infections in organized teams of the Armed Forces of the Russian Federation with a preliminary definition of predictors is proposed.

Key words: machine learning, gradient boosting, retrospective epidemiological analysis, forecasting, prognosis of morbidity, acute respiratory infection, organized collectives, and the Armed Forces of the Russian Federation.

Одной из наиболее важных задач профилактической медицины и основным мероприятием эпидемиологического надзора является прогнозирование инфекционной заболеваемости, осуществляемое на основании данных ретроспективного анализа многолетней и внутригодовой динамики определенной нозологии, с учетом оценки комплексного воздействия различных факторов [33].

В отечественной науке понятие прогноза было рекомендовано Комитетом научно-технической терминологии Академии наук СССР в 1975-1990 гг. [27]. При этом авторы оговорили факт постоянного развития прогностики как науки, таким образом предложенная терминология не должна быть статичной, однако по прошествии сорока

пяти лет изменений в классификации прогнозов не произошло. Кроме того, предложенная классификация унифицирована для прогнозирования развития объектов любой природы. Необходимо так же отметить, что методы прогнозирования за время, прошедшее с момента выхода в свет рекомендаций, заметно прогрессировали.

Отечественная эпидемиология до настоящего времени придерживается указанных рекомендаций, которые нашли свое отражение в печатных изданиях.

В понятийном аппарате эпидемиологического прогнозирования в настоящее время выделяют *поисковый прогноз*, предполагающий определение будущих изменений в эпидемиологической ситуации с учетом воздействия факторов риска, *нормативный прогноз*, определяющий пути и сроки изменения (улучшения) неблагоприятной эпидемиологической ситуации после элиминации действия факторов риска, и *комплексный прогноз*, являющийся синтезом двух предыдущих.

Кроме того, описывается *одномерный прогноз*, определяющий какую-либо одну качественную или количественную характеристику эпидемиологического процесса, и *многомерный прогноз*, учитывающий несколько характеристик.

По временным характеристикам, эпидемиологические прогнозы разделяют на *точечный* и *интервальный*, который, учитывая временной период, подразделяют на *оперативный* - до 1 месяца, *краткосрочный* - от 1 месяца до 1 года, *среднесрочный* - от 1 года до 5 лет, *долгосрочный* - от 5 до 15 лет и *дальнесрочный* - более 15 лет.

Выделяют также *качественный* и *количественный* прогнозы в эпидемиологии.

Качественный прогноз основывается на ретроспективных данных, показывающих динамику и характер уровней заболеваемости, и определяет вероятность событий, которые произойдут в будущем на основе некоторых количественных характеристик.

Количественный прогноз определяется с помощью математических (статистических) методов [17,26,27].

В 2013 году М.А. Кондратьевым была предпринята попытка совершенствовать прогнозы в эпидемиологии, разделив их на три вида по временным промежуткам (срокам), на которые они выполняются и целям, для которых они предназначены:

краткосрочные – на несколько недель, применяются с целью оперативного реагирования на предполагаемые повышения уровня заболеваемости, при выявлении эпидемических вспышек заболеваемости с целью управления ими;

среднесрочные - от 2 до 6 месяцев, менее точные, чем краткосрочные и используемые в тактическом управлении. Показано, что среднесрочные прогнозы считаются наиболее полезными в связи с достаточным временем для подготовки к

возможным чрезвычайным ситуациям, ухудшению санитарно-эпидемиологической ситуации и возможности проведения санитарно-противоэпидемических (профилактических) мероприятий, направленных на предупреждение негативных факторов.

долгосрочные - на год и более, низкой точности и используются при принятии стратегических решений (оценка необходимых объемов производства лекарственных препаратов и вакцин, оснащение медицинских учреждений, подготовка персонала) [16].

Однако, следует отметить, что в данной работе, так же, как и в рекомендациях почти полувековой давности, не указаны причины, на которых автор строит свою классификацию, неясен выбор временных интервалов эпидемиологических прогнозов.

В последнее десятилетие наблюдается прогрессивный рост методов моделирования прогнозов в различных отраслях науки и хозяйственной деятельности, повышается их точность. Мы считаем, что назрела необходимость введения новой классификации эпидемиологических прогнозов, основанных на привязке уровня заболеваемости определенной нозологией к одному или нескольким наиболее значимым предикторам.

Учитывая тот факт, что основными предикторами заболеваемости болезнями органов дыхания, в частности острыми респираторными инфекциями (далее – ОРИ), являются такие показатели, как климатические (температура, влажность, скорость движения воздуха) состояние возбудителя (возбудителей) заболевания, состояние и возрастная группа населения, а также текущая заболеваемость в районе проживания, мы считаем, что в прогнозный период (до 7-10 суток) состояние микро- и микроорганизмов, не будет иметь выраженных колебаний (будет относительно стабилен), и целесообразно классифицировать интервальные прогнозы уровней заболеваемости с наиболее точно прогнозируемыми предикторами – прогнозом погоды, основанном на сроках метеорологических прогнозов [18,28], адаптированном с учетом точности, и уровнем текущей заболеваемости, определяемой еженедельно для каждой возрастной категории населения.

Основываясь на вышеизложенном, предлагается следующая классификация прогнозов заболеваемости ОРИ:

краткосрочный (оперативный) – до трех суток;

среднесрочный – от трех до семи суток;

долгосрочный – от восьми суток и более.

В *долгосрочном* прогнозе заболеваемости предлагается выделять: *месячный*, *квартальный* (прогноз на три месяца), *полугодовой*, *сезонный* (соответствует

эпидемическому сезону определенной нозологии), *годовой, трехлетний, пятилетний, десятилетний* и т. д.

Безусловным является тот факт, что наиболее точный прогноз заболеваемости можно сделать на наименее короткий временной интервал. И, соответственно, с увеличением прогнозного временного интервала уменьшается и его точность. На данном этапе развития метеорологических прогнозов, наиболее точными являются прогнозы до трех суток.

Авторы не исключают некоторой коррекции *среднесрочного* и *долгосрочного* прогнозируемых интервалов (от семи до десяти, и от одиннадцати и более суток соответственно) после проведения планируемого исследования.

Прогнозирование инфекционных заболеваний с наибольшей вероятностью в настоящее время становится возможным благодаря современному подходу и междисциплинарному взаимодействию на стыке интересов разных научных изысканий. В частности, исследования показывают эффективность применения методов машинного обучения в области здравоохранения, как в России так за рубежом.

Data Mining [29] – обширная область, которая возникла и развивается на стыке таких наук, как статистика, искусственный интеллект, машинное обучение и др. и сочетает в себе такие методы и алгоритмы как: искусственные нейронные сети, деревья решений, корреляционный, кластерный анализ, линейная регрессия, байесовские сети и множество других. Решаются такие задачи как классификация, кластеризация, прогнозирование. Статистические методы чаще всего сводятся к решению линейных регрессионных уравнений. Однако найти зависимость при таком подходе не всегда возможно. В таких случаях применяются методы машинного обучения [27].

Прогнозирование инфекционной заболеваемости с помощью математических моделей ведет свое начало с 1760 года, когда швейцарским физиком, математиком и физиологом Бернули (Daniel Bernoulli) с целью изучения инфекционного заболевания и эффективности различных способов вакцинации против оспы был впервые применен математический анализ. Спустя восемьдесят лет, в 1840 году Уильям Фарр (W. Farr) применил кривую нормального распределения для описания данных в 1837-39 гг. в Англии и Уэльсе по смертности от оспы. Продолжение данной работы нашло свое отражение в трудах Браунли (John Brownlee) в 1906 году, который сопоставил ряды эпидемиологических данных на основе распределения Пирсона и описал результаты в своей статье «Статистический подход к иммунной защите: теория эпидемий» [36].

Совершенствование методик в прогнозе инфекционной заболеваемости с начала XX века привело к их стремительному росту. В области прогнозов, основанных на математическом моделировании и использовании статистических данных, можно выделить несколько основных подходов: статистические методы, методы на основе машинного обучения, методы на базе фильтрации и математическое моделирование распространения инфекции. Фактически, это условное деление, так как большинство методик представляют собой синтез подходов, имея самостоятельное значение и особенности применения в практической плоскости.

В последние годы алгоритмы машинного обучения стали одними из основных методов в области интеллектуального анализа больших данных, хранение и обработка которых благодаря развитию технологий и снижению затрат становится все более доступным и позволяет использовать новые методы интеллектуального анализа знаний.

Существуют разные математические подходы в построении моделей прогнозирования, в основе которых лежат определенные методы классификации, более подробно описанные нами ранее [4]. В настоящей работе будет рассмотрен частный случай подхода математического моделирования, а именно – машинное обучение, который используется для обнаружения нелинейности и структуры взаимодействия в данных.

Методы машинного обучения основаны на разработке высокоуровневых алгоритмов обработки данных, учитывающих типы неконтролируемого, полуправляемого и контролируемого обучения, предлагая алгоритмы, которые доводят данные до исчерпания для поиска знаний.

Однако, поскольку эти методы разрабатываются с высоким уровнем данных и сложными алгоритмами, известными как «черный ящик», валидация точности, предсказания и интерпретируемости все еще широко обсуждается и ставится под сомнение.

Каждый из алгоритмов машинного обучения предназначен для решения конкретных задач. Среди этого множества выделяют *алгоритмы контролируемого машинного обучения («обучения с учителем»)*, такие как *логистическая регрессия, методы опорных векторов, и градиентный бустинг*.

Логистическая регрессия оценивает вероятность результата взвешенной суммы входных переменных (учитывая смещение). Положительным качеством модели является быстрое обучение и быстрый прогноз, особенно при большом количестве признаков по сравнению с количеством наблюдений в выборке, и достаточно простом обучении по

сравнению с более сложными моделями. Однако, в случаях, когда набор данных имеет сильно коррелированные признаки, коэффициенты могут быть трудно интерпретируемы. Кроме того, при меньшей размерности набора данных, другие модели могут демонстрировать лучшую обобщающую способность с более точными прогнозами [44,42].

Метод опорных векторов (SVM – Support Vector Machine) способен выполнять линейную или нелинейную классификацию. Наиболее эффективен для обработки сложных малых или средних по размеру наборов данных.

Представляя собой подмножество обучающей выборки, опорные векторы позволяют строить различные обучаемые машины со своими собственными нелинейными поверхностями решений. Из недостатка метода следует отметить, что для классификации используется небольшая часть образцов, находящихся на границах.

С учетом того, что линии и гиперплоскости имеют ограниченную гибкость, в низкоразмерных пространствах линейные модели накладывают весьма жесткие ограничения, поскольку линии и гиперплоскости имеют ограниченную гибкость. Одним из способов сделать *линейную модель* более гибкой – добавление новых признаков, [15,50].

Градиентный бустинг (далее – ГБ) является частным случаем алгоритмов машинного обучения для регрессионных и классификационных задач, который создает модель прогнозирования в виде ансамбля слабых моделей (слабых учеников, weak learners) прогнозирования, обычно деревьев решений небольшой глубины. Однако этот ансамбль слабых предикторов при добавлении все большего количества деревьев дает сильную модель с очень высокой точностью [47]. ГБ, основанный на построении линейной комбинации простых алгоритмов путем изменения веса входных данных [41,46] строит модель поэтапно, как и другие методы повышения, и обобщает их, позволяя оптимизировать произвольную дифференцируемую функцию потерь [2,3].

Впервые градиентный бустинг был представлен в 1999 году профессором Стэнфордского университета Джеромом Х. Фридманом (Jerome H. Friedman) и был дополнительно разработан и усовершенствован в нескольких пакетах с открытым исходным кодом, таких как Scikit-Learn и R. Крупные разработчики программного обеспечения, такие как SAS и IBM, разработали свои собственные реализации, а оригинальная машина градиентного бустинга Фридмана (с уточнениями) была частью продукта Salford Systems SPM с 2001 года [54].

Сильной стороной ГБ, относящегося, как было сказано ранее, к методам последовательного построения композиции алгоритмов машинного обучения, является

стремление каждого следующего алгоритма компенсировать недостатки композиции всех предыдущих алгоритмов, рекурсивно подгоняя слабого ученика к остаточному так, чтобы улучшить производительность модели с постепенно увеличивающимся числом итераций [49]. Считается, что ГБ над решающими деревьями — это один из самых универсальных и эффективных методов машинного обучения, известных на сегодняшний день. При этом сами решающие деревья хорошо зарекомендовали себя в практике решения задач классификации. ГБ способен автоматически обнаруживать сложную структуру данных, включая нелинейность и взаимодействия высокого порядка, даже в контексте сотен, тысяч или десятков тысяч потенциальных предикторов [14,34,47].

К слабой стороне ГБ следует отнести трудоемкость метода, который, зачастую требует построения сотен или даже тысяч базовых алгоритмов, а также необходимость дополнительных модификаций, без которых он имеет свойство полностью подстраиваться под данные, в том числе под ошибки и выбросы в них. Немаловажным является и время, затраченное на построение большого количества алгоритмов ГБ.

Частным случаем ГБ является экстремальный градиентный бустинг (XGBoost) при построении модели на основе деревьев решений. В данном алгоритме размер дерева и величина весов контролируются стандартными параметрами регуляризации, а также используется дополнительный параметр рандомизации для уменьшения корреляции между деревьями, что приводит к более высокой устойчивости и скорости модели. Недостатками XGBoost считаются необходимость тщательной настройки параметров, значительное время обучения и плохая работа с высокоразмерными разреженными данными [40].

Еще одним способом ГБ является категориальный бустинг (CatBoost), разработанный компанией Яндекс. Основным отличием данной модели является совмещение стандартной модели ГБ с использованием данных после преобразования категорий в целевую статистику, в результате которого получился новый алгоритм обработки категориальных признаков, названный CatBoost, превосходит существующие современные реализации деревьев решений с ГБ по разнообразному набору популярных задач машинного обучения [48].

Следует отметить, что основой для успешного применения ансамблевых методов прогнозирования является разнообразие базовых моделей, как базис для итогового прогноза. Следовательно, чем разнообразнее будут базовые модели, тем точнее прогноз за счет потенциально большего числа возможностей исправления ошибок друг друга.

Использование генерации случайных поворотов позволяет увеличить разнообразие базовых алгоритмов и, следовательно, повысить качество прогноза [38].

Методы ГБ получили широкое применение для решения задач классификации и регрессии в широком спектре прикладных областей [5]. В первую очередь в сфере финансов, где ГБ применяется в прогнозировании кризисов на рынках [39] и кредитном скорринге [52]. В политической сфере с помощью ГБ анализируют значимость факторов, влияющих на социально-политическую нестабильность в государстве [35]. Практическое применение ГБ показано в вопросе прогнозирования значений показателей эффективности рекламных акций в продвижении продуктов первой необходимости [43]. В компьютерном программировании ГБ используется для идентификации программного обеспечения операционных систем для снижения числа уязвимостей при установке несанкционированного программного обеспечения [29]. Кроме того, модель ГБ применяется в области энергетики (в частности, ветровой и солнечной радиации) [37,51,53]. Существуют работы по использованию ГБ а прогнозе погоды, в частности для восстановления пропущенных значений в данных об осадках [5].

В здравоохранении градиентный бустинг в настоящее время в первую очередь используется как метод построения прогнозных моделей соматических заболеваний [54]. Также, показано применение ГБ в фармакологии, где с помощью модели предсказывают биологическую активность лекарственных препаратов [45].

ГБ в клинической медицине моделирует ситуацию, когда каждый следующий врач тренируется фокусировать внимание на ошибочно поставленных диагнозах предыдущих врачей, ускоряя процесс и повышая качество диагностики.

Наиболее широко методы машинного обучения, и, в частности, ГБ, применяются в диагностике заболеваний системы кровообращения [4]. Так, использование ансамбля деревьев решений ГБ позволяет в существенной мере повысить точность прогноза эффективности лечения пациентов с хронической сердечной недостаточностью [1]. Лучший результат прогнозирования на основе случайного леса составил 90,5 %, а доли ошибочной классификации и выборок (обучающей и экзаменационной), практически равнозначны.

В прогнозировании диагнозов заболеваний (железодефицитная анемия, инсулиннезависимый сахарный диабет, другие нарушения обмена углеводов, нарушения обмена липопротеидов и другие липидемии) также можно увидеть положительные результаты [30]. С помощью ГБ исследовали совокупность результатов 200 лабораторных исследований (7 918 случаев), разделенных впоследствии на обучающую и тестовую,

выборки. Точность распознавания моделью выбранных диагнозов составила более 89 %, при средней уверенности модели в каждом прогнозируемом диагнозе в 92 %.

Прогнозирование развития заболевания у пациентов с сахарным диабетом показало 77,95 % качество, с учетом небольшого количества данных [14].

Однако, в профилактической медицине, как с целью прогнозирования инфекционной заболеваемости, так и с целью прогнозирования заболеваемости острых респираторных инфекций (далее – ОРИ), его использование еще не изучено.

За последние десять лет наибольший финансовый ущерб среди нозологий наносят ОРИ, неуклонно оставаясь на первом ранговом месте среди первичной заболеваемости в Российской Федерации [6,8,9,10,11,12,13,19,20,21,22,23,24,25,32,]. В организованных воинских коллективах тенденция течения заболеваний ОРИ, как наиболее значимой военно-эпидемиологической нозологии, в целом соответствует общегосударственному тренду [7].

Учитывая высокую эпидемиологическую и экономическую значимость ОРИ, авторы считают, что проведение исследований по применению в прогнозе уровней заболеваемости методов машинного обучения, и, в первую очередь, метода градиентного бустинга, с целью своевременного упреждающего профилактического воздействия на подъемы заболеваемости, приведет к снижению финансовой нагрузки на бюджет страны и повышению здоровья граждан.

В настоящее время такая работа проводится совместно с ФГБУ «НИИ гриппа им. А.А. Смородничева» Минздрава России в рамках договора о научном сотрудничестве.

Литература

1. Абрамович, М.С. Применение ансамблей деревьев решений для прогнозирования эффективности лечения хронической сердечной недостаточности / М.С. Абрамович, Е.С. Атрощенко, Д.В. Романовский, О.В. Кошлатая // Материалы международного конгресса по информатике: информационные системы и технологии Минск, 24–27 октября 2016 года. Издательство: Белорусский государственный университет, Минск. — 2016. — С. 406–410.
2. Бочкарев, Б.В. Исследование применимости методов машинного обучения на основе решающих деревьев для анализа электрокардиосигналов / Б.В. Бочкарев, А.А. Ракитский // Труды XVII Международной конференции DICR-2019, Новосибирск, 3-6 декабря 2019 г. — Новосибирск — 2019. — С.42–47.
3. Ветров, Д.П. Байесовские методы машинного обучения: учебное пособие по спецкурсу / Д.П. Ветров, Д.А. Кропотков. — 2007.
4. Голубков, А.В. Применение искусственных нейронных сетей в профилактической и клинической медицине (научный обзор) / А.В. Голубков, М.П. Гаврилова // Профилактическая и клиническая медицина. — 2020. — № 4 (77). — С. 30–39. DOI: 10.47843/2074-9120_2020_4_30.
5. Горшенин, А.К. Гибридные модели экстремального градиентного бустинга для восстановления пропущенных значений в данных об осадках / А.К. Горшенин, О.П. Мартынов // Информ. и ее примен. — 2019, — Т. 13. — Вып. 3. — С. 34-40. DOI: 10.14357/19922264190306
6. Демография в России: Сборник Федеральной службы государственной статистики. <https://rosstat.gov.ru/folder/12781/> (дата обращения: 09.03.2021)
7. Евдокимов, В.И. анализ первичной заболеваемости военнослужащих по призыву военно-морского флота и сухопутных войск России за последние 10 лет (2010-2019 гг.) / В.И. Евдокимов, П.П. Сивашенко // Морская медицина. — 2020. — Т. 6. — № 4. — С. 63–72. DOI: 10.22328/2413-5747-2020-6-4-63-72.
8. Здравоохранение в России — 2011 г. — URL: https://rosstat.gov.ru/bgd/regl/B11_34/Main.htm/ (дата обращения: 15.04.2021)
9. Здравоохранение в России — 2013 г. — URL: https://rosstat.gov.ru/bgd/regl/B13_34/Main.htm/ (дата обращения: 15.04.2021)
10. Здравоохранение в России — 2015 г. — URL: https://rosstat.gov.ru/bgd/regl/B15_34/Main.htm/ (дата обращения: 15.04.2021)

11. Здравоохранение в России – 2017 г. – URL: https://rosstat.gov.ru/bgd/regl/B17_34/Main.htm/ (дата обращения: 15.04.2021)
12. Здравоохранение в России – 2019 г. – URL: https://rosstat.gov.ru/bgd/regl/b19_34/Main.htm/ (дата обращения: 15.04.2021)
13. Здравоохранение в России: Сборник Федеральной службы государственной статистики. <https://rosstat.gov.ru/folder/210/document/13218/> (дата обращения: 09.03.2021)
14. Ильясов, Э.С. Применение модели градиентного бустинга для прогнозирования развития диабета/ Э.С. Ильясов // Молодой ученый. — 2016. — № 27. — С. 1-5. – URL: <https://moluch.ru/archive/131/36581/> (дата обращения: 20.03.2021)
15. Китова, О.В. Метод машин опорных векторов для прогнозирования показателей инвестиций / О.В. Китова, И.Б. Колмаков, И.А. Пеньков // Экономика, Статистика и Информатика. - 2016. - № 4. – С. 27-30.
16. Кондратьев, М.А. Методы прогнозирования и модели распространения заболеваний. / М.А. Кондратьев. // Компьютерные исследования и моделирование. — 2013. — Т. 5. — № 5. — С. 863–882.
17. Мельниченко, П.И. Ретроспективный эпидемиологический анализ и прогнозирование заболеваемости личного состава Вооруженных сил Российской Федерации: метод. указания / П.И. Мельниченко [и др.]. — М.: Воениздат, 2006. — 143 с.
18. Наставление по Глобальной системе обработки данных и прогнозирования. Дополнение IV к Техническому регламенту ВМО. ВМО-№ 485. — Женева, Швейцария. — 2017. — 123 с.
19. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2012 году: Государственный доклад. — М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2013. — 176 с. – URL: https://www.rospotrebnadzor.ru/upload/iblock/7cd/gosudarstvennyy-doklad-o-sostoyanii-sanitarno_epidemiologicheskogo-blagopoluchiya-naseleniya-v-rossiyskoy-federatsii-v-2012-godu.pdf/ (дата обращения: 15.04.2021)
20. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2013 году: Государственный доклад. — М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2014. — 191 с. – URL: https://www.rospotrebnadzor.ru/upload/iblock/3b8/gd_2013_dlya-sayta.pdf/ (дата обращения: 15.04.2021)
21. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2014 году: Государственный доклад. — М.: Федеральная служба

по надзору в сфере защиты прав потребителей и благополучия человека, 2015. — 206 с. — URL: https://www.rospotrebnadzor.ru/upload/iblock/22c/gd_2014_seb_dlya-sayta.pdf/ (дата обращения: 15.04.2021)

22. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2015 году: Государственный доклад. — М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2016. — 200 с. — URL: https://www.rospotrebnadzor.ru/upload/iblock/486/gd_2015_ds.pdf/ (дата обращения: 15.04.2021)

23. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2016 году: Государственный доклад. — М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2017. — 220 с. — URL: <https://www.rospotrebnadzor.ru/upload/iblock/0b3/gosudarstvennyy-doklad-2016.pdf/> (дата обращения: 15.04.2021)

24. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2017 году: Государственный доклад. — М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2018. — 268 с. — URL: https://www.rospotrebnadzor.ru/upload/iblock/d9d/gd_2017_seb.pdf/ (дата обращения: 15.04.2021)

25. О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2018 году: Государственный доклад. — М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2019. — 254 с. — URL: https://www.rospotrebnadzor.ru/upload/iblock/798/gosudarstvennyy-doklad-o-sostoyanii-sanitarno_epidemiologicheskogo-blagopoluchiya-naseleniya-v-rossiyskoy-federatsii-v-2018-godu.pdf/ (дата обращения: 15.04.2021)

26. Прогнозирование заболеваемости геморрагической лихорадкой с почечным синдромом (Пуумала) с использованием статистических методов: учеб.-метод. пособие для практ. занятий / Ю.Л. Минаев, В.Г. Морозов, С.И. Коломинов, Г.Д. Коробов; НОУ ВПО СМи «РЕАВИЗ» — Самара: НОУ ВПО СМи «РЕАВИЗ», 2011. — 44 с.

27. Прогностика: терминология / Акад. наук СССР, Ком. науч.-техн. терминологии; отв. ред. В. И. Сифоров. - Москва: Наука, 1990. 2019 — 54 с. - ISBN 5-02-006645-1.

28. РД 52.27.724–2019. Наставление по краткосрочным прогнозам погоды общего назначения: дата введения 2019-06-25 / Федеральная служба по

гидрометеорологии и мониторингу окружающей среды. Изд. Официальное. – Москва: ФГБУ «Гидрометцентр России», 2019. — 66 с.

29. Салахутдинова, К.И. Алгоритм градиентного бустинга деревьев решений в задаче идентификации программного обеспечения / К.И. Салахутдинова, И.С. Лебедев, И.Е. Кривцова // Научно-технический вестник информационных технологий, механики и оптики. — 2018. — Т. 18. — № 6. — С. 1016–1022. DOI: 10.17586/2226-1494-2018-18-6-1016-1022.

30. Сахибгареева, М.В. разработка системы прогнозирования диагнозов заболеваний на основе искусственного интеллекта / М.В. Сахибгареева, А.Ю. Заозерский // Вестник РГМУ. — 2017. — № 6. — С.

31. Слободенюк, А.В. Эпидемиологический анализ: учеб. пособие / А.В. Слободенюк, А.А. Косова, Р.Н. Ан; ГБОУ ВПО УГМУ. – Екатеринбург: изд-во ГБОУ ВПО УГМУ, 2015. – 36 с.

32. Социально-экономическое положение России. Январь 2020 (1): Сборник Федеральной службы государственной статистики. — URL: <https://rosstat.gov.ru/storage/mediabank/osn-01-2020.pdf/> (дата обращения: 24.04.2021)

33. Тусаева, А.Р. Применение корреляционно-регрессионного метода в оценке уровня заболеваемости Курской области. / А.Р. Тусаева, А.В. Гаврилова. // Молодой ученый. — 2017. — № 4 (138). — С. 541-544. — URL: <https://moluch.ru/archive/138/38666/> (дата обращения: 02.05.2021).

34. Фролов, А.И. Анализ алгоритма градиентного бустинга для целей прогнозирования / А.И. Фролов // Альманах научных работ молодых ученых Университета ИТМО. XLVIII научная и учебно-методическая конференция Университета ИТМО / Издательство: ФГАОУ ВО «Национальный исследовательский университет ИТМО» — Санкт-Петербург, 2019. — С. 285—288.

35. Шульгин, С.Г. Отбор переменных для анализа и прогнозирования нестабильности с помощью моделей градиентного бустинга / С.Г. Шульгин // Системный мониторинг глобальных и региональных рисков. Социально-политическая и экономическая дестабилизация: анализ страновых и региональных ситуаций в мир-системном аспекте / Издательство «Учитель» — Волгоград, 2018. — Т. 9. — С. 115—153.

36. Янчевская, Е.Ю. Математическое моделирование и прогнозирование в эпидемиологии инфекционных заболеваний / Е.Ю. Янчевская, О.А. Меснянкина // Вестник Российского университета дружбы народов. Серия: Медицина. 2019. — Т. 23. — № 3 — С. 328-334. DOI: 10.22363/2313-0245-2019-23-3-328-334.

37. Aler R., Galvan I. M., Ruiz-Arias J. A., Gueymard C. A. Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting // *Sol. Energy*, 2017. Vol. 150. P. 558–569.
38. Blaser R., Fryzlewicz P. Random Rotation Ensembles // *Journal of Machine Learning Research* №17 / 2016. с. 1–26.
39. Chatzis, S.P., Siakoulis V., Petropoulos A., Stavroulakis E., Vlachogiannakis N. Forecasting stock market crisis events using deep and statistical machine learning techniques // *Expert Syst. Appl.*, 2018. Vol. 112. P. 353–371.
40. Chen, T., and Guestrin, C. (2016). «XgBoost: A Scalable Tree Boosting System» in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. doi: 10.1145/2939672.2939785
41. Friedman, J.H. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* 2001; 29(5): 1189–1232, — URL: <https://doi.org/10.1214/aos/1013203451/> (дата обращения: 15.03.2021)
42. Harrell FE. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer, New York; 2001
43. Henzel, J. Gradient Boosting Application in Forecasting of Performance Indicators Values for Measuring the Efficiency of Promotions in FMCG Retail / J. Henzel, M. Sikora // *arXiv:2006.04945v1 [cs.CY]* 30 May 2020
44. Hosmer, D.W., Lemeshow, S. *Applied Logistic Regression*. New York:Wiley 1989.
45. Mustapha, I. B., and F. Saeed. Bioactive molecule prediction using extreme gradient boosting. *Molecules* 2016. 21(8):983.
46. Natekin A., Knoll A. Gradient boosting machines, a tutorial. *Front Neurobot* 2013; 7: 21, <https://doi.org/10.3389/fnbot.2013.00021>.
47. Paulo Celio Di Cellio Dias, Melissa Forti, Marc Witarsa // A comparison of Gradient Boosting with Logistic Regression in Practical Cases. – URL: <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1857-2018.pdf/> (дата обращения: 19.04.2021)
48. Prokhorenkova, L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. CatBoost: Unbiased boosting with categorical features. 2018. *Adv. Neur. In.* 31:6638–6648
49. Shalev-Shwartz S, Ben-David S. *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press; 2014. 410 p.

50. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Prediction, Inference and Data Mining (Second Edition). Springer Verlag, New York, 2009.
51. Torres-Barran, A., Alonso A., Dorronsoro J. R. Regression tree ensembles for wind energy and solar radiation prediction // Neurocomputing, 2018. Vol. 326. P. 151–160.
52. Xia, Y., Liu C., Li Y., Liu N. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring // Expert Syst. Appl., 2017. Vol. 78. P. 225–241.
53. Zhang, D. A data-driven design for fault detection of wind turbines using random forests and XGboost / D. Zhang [et al.] // IEEE Access, 2018. Vol. 6. P. 21020-21031.
54. Zhang, Z. Predictive analytics with gradient boosting in clinical medicine / Z. Zhang [et al.] // Ann Transl Med 2019;7(7):152. doi: 10.21037/atm.2019.03.29.