

# МЕТОДИКА ОБУЧЕНИЯ КОМПОЗИЦИЙ МОДЕЛЕЙ КЛАССИФИКАЦИИ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ВЫХОДНЫХ ДАННЫХ О ДЕЯТЕЛЬНОСТИ ОБРАЗОВАТЕЛЬНЫХ ОРГАНИЗАЦИЙ

**METHODOLOGY FOR TEACHING  
COMPOSITIONS OF CLASSIFICATION  
MODELS TO INCREASE THE ACCURACY  
OF OUTPUT DATA ON THE ACTIVITIES  
OF EDUCATIONAL ORGANIZATIONS**

**R. Lomovtsev  
O. Romashkova**

*Summary.* The article is devoted to the method of applying the composition of classification models on the activities of secondary schools in the regional education authorities. In addition, the article describes the problems of initial data on the effectiveness of educational organizations and suggests a methodology for preparing initial data for training models.

*Keywords:* data processing models, classification models, educational organization, composition of models.

**Ломовцев Роман Сергеевич**

Аспирант, Московский городской педагогический  
Университет»

feedback.roman@gmail.com

**Ромашкова Оксана Николаевна**

Доктор технических наук, профессор

Профессор, Российская академия народного  
хозяйства и государственной службы при Президенте

РФ, г. Москва, Россия

ox-rom@yandex.ru

*Аннотация.* Статья посвящена методу применения композиции моделей классификации для данных о деятельности образовательных организаций на региональном уровне управления образованием. Кроме того, в статье описаны проблемы исходных данных об эффективности образовательных организаций и предложена методика подготовки исходных данных для обучения моделей.

*Ключевые слова:* модели обработки данных, модели классификации, образовательная организация, композиция моделей.

## Введение

Средства интеллектуального анализа данных все чаще применяются в различных сферах экономики. С использованием таких средств сегодня производится сегментация клиентской базы для формирования наилучших клиентских предложений, прогнозирования спроса, выявления ассоциативных правил.

Вместе с тем, в сфере образования, например, преобладают государственные образовательные организации, существует большое количество нормативно-правовых актов, регулирующих деятельность образовательных организаций и органов государственной власти регионального и муниципального уровней.

Ввиду этого сегодня очевидна сравнительно низкая степень инновационности деятельности по управлению сферой образования, эффективности внутренних процессов.

Процесс информатизации образования (применения новых информационных технологий) развивается

медленнее, чем во многих других сферах экономики. Это объясняется, кроме прочего, особой ролью образования в жизни общества. За годы своего развития сфера образования приобрела устойчивые традиции, принципы и концепции осуществления своей деятельности, которые передавались из поколения в поколение, сформировав устойчивую консервативную систему. В таких условиях процесс цифровой трансформации образования протекает особенно медленно.

За последние годы учебная литература стала дублироваться в цифровом виде, появились новые средства обучения (цифровые интерактивные доски, проекторы), традиционные метрики оценки качества образования поэтапно уступили место единым проверочным работам — для выпускников школ введены единый государственный экзамен (ЕГЭ) в 11-х классах, государственная итоговая аттестация (ГИА) в 9-х классах, все-российские проверочные работы (ВПР).

Введение подобных единых метрик оценки качества образования позволило упростить процесс сравнения образовательных организаций, а значит, и введения рейтинговых шкал и классификаций для последних.

## 1. Постановка задачи классификации данных о деятельности образовательных организаций

Сегодня в рамках применения интеллектуального анализа данных выделяют целый комплекс задач:

1. Классификация;
2. Кластеризация;
3. Поиск ассоциативных правил;
4. Последовательные шаблоны;
5. Регрессия;
6. Поиск отклонений.

Перечисленные задачи интеллектуального анализа данных можно поделить на две группы: обучения с учителем и обучения без учителя. Рассматриваемая в данной статье задача классификации относится к классу задач обучения с учителем.

Сущность задачи классификации заключается в группировке данных, содержащихся в выборке, в соответствии с некоторыми признаками, характерными для данной группы. В отличие от задачи кластеризации, правила отнесения объектов к той или иной группе сформулированы заранее неявным образом в специальной обучающей выборке.

Постановка задачи классификации для данных о деятельности образовательных организаций зависит от целей анализа данных и набора параметров в выборке.

Для целей оценки эффективности деятельности отдельных педагогических работников образовательных организаций выборка должна содержать данные о результатах обучающихся (в рамках диагностических мероприятий и экзаменов), чьим учителем является педагогический работник.

Для целей анализа эффективности работы сотрудников образовательных организаций по профилактике правонарушений выборка должна содержать массив данных, связанных с количеством и долей обучающихся (от общего числа обучающихся в образовательной организации), состоящих на внутреннем школьном учете, количеством и долей обучающихся (от общего числа обучающихся в образовательной организации), состоящих на учете в органах внутренних дел МВД России, численность и доля обучающихся, состоящих на учете и не совершивших за очередной отчетный год правонарушения.

Основная цель анализа данных образовательных организаций на региональном уровне связана с оценкой качества образования в образовательных органи-

зациях как самостоятельных единицах. В целях оценивания образовательных организаций выборка данных должна содержать информацию о результатах независимых диагностических мероприятий и государственных экзаменов. Примерный перечень таких проверок и экзаменов включает:

1. Государственную итоговую аттестацию (ГИА-9);
2. Единый государственный экзамен (ЕГЭ);
3. Всероссийские проверочные работы;
4. Региональные диагностики.

В рамках данной задачи классы модели классификации могут быть сопоставлены учебным профилям образовательных организаций либо определенным диапазоном результатов обучающихся [1].

## 2. Формирование выборки для обучения моделей классификации

Обучающая выборка должна отвечать двум основным требованиям:

1. Выборка должна состоять из агрегированных данных об образовательных организациях;
2. Выборка должна основываться на мотивированной классовой принадлежности.

Под агрегированными данными подразумевается, что исходное множество накопленных в хранилище данных фактов (транзакций) должно быть суммировано (агрегировано) до таких величин, как:

1. Средний балл/средняя оценка;
2. Минимальный балл/минимальная оценка;
3. Максимальный балл/максимальная оценка;
4. Доля учеников, получивших определенную оценку;
5. Успеваемость (доля учеников, получивших оценку «3», «4», «5»);
6. Качество (доля учеников, получивших оценку «4», «5»).

Под мотивированной классовой принадлежностью подразумевается, что та или иная образовательная организация должна быть отнесена к тому или иному классу на основании экспертных заключений/оценок/рейтинговой принадлежности/профильной направленности программ обучения в образовательных организациях.

Кроме того, для эффективности обучения композиции моделей классификации размер выборки должен быть больше, чем при обучении единственной модели классификации. Количество образцов для каждого из представленных классов не должно отличаться в значительной степени.

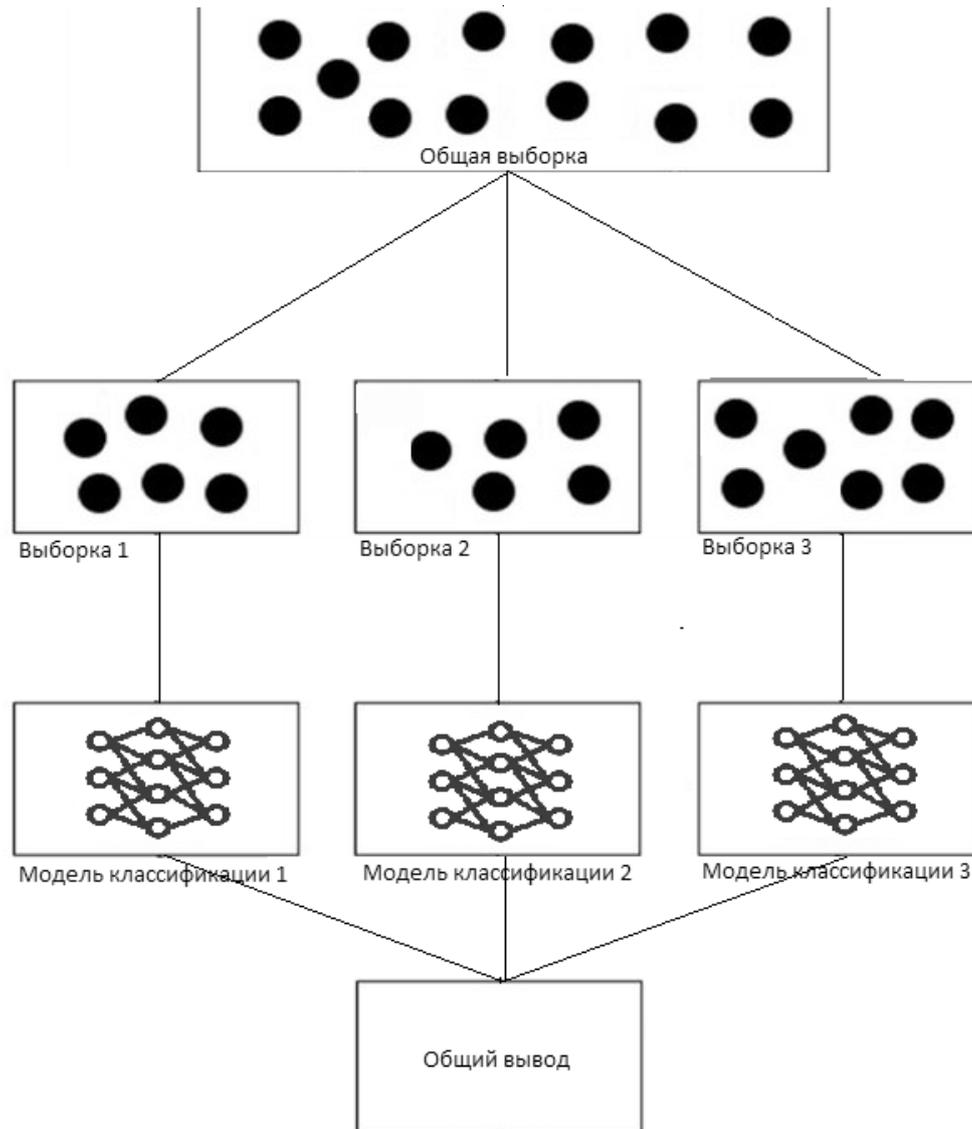


Рис. 1. Схема работы ансамбля моделей классификации

### 3. Сегментация выборки и обучение композиции моделей классификации

Бэггинг (бутстрэп-агрегирование) — это метаалгоритм обучения композиций моделей классификации. Идея метаалгоритма заключается в автономном обучении моделей классификации и их последующем обобщении, за счет чего повышается общая стабильность моделей, а, следовательно, и точность будущей классификации (рисунок 1).

Концепция метаалгоритма не подразумевает наличие обязательных требований к исходной выборке данных, сферам применения, а также не содержит правил и условий обучения моделей. Таким образом, концепция закладывает лишь основные принципы мно-

жественного обучения моделей, их комбинирования. На основе концепции возможно создание собственных алгоритмов, адаптированных к применению в конкретной предметной области.

В рамках концепции (метаалгоритма) бэггинга предусмотрено:

1. Разбиение исходной выборки на подмножества произвольным образом;
2. Обучение моделей автономно на каждой из выборок;
3. Объединение результатов классификации.

Целесообразность применения концепций данного метаалгоритма при анализе данных об эффективности образовательных организаций объясняется миними-

Таблица 1. Частота сдачи предметов по выбору и доля результатов ниже допустимого уровня для допуска к участию в конкурсе в вузах.

Предмет	Сдало ЕГЭ-2022, тыс. чел.	%	Доля результатов за пределами диапазона среднего отклонения от среднего значения,%
Русский язык (100%)	701	100	0
Математика (профиль)	337	48	4
Обществознание	333	48	0
Биология	143	20	15
Информатика и ИКТ	129	18	4
Физика	124	18	4
История	117	17	15
Иностранные языки	112,5	16	4
Химия	95	14	27
Литература	62	9	12
География	22	3	Н/Д, >50%

зацией влияния выбросов на итоговую оценку образовательной организации. Отдельные образовательные организации могут иметь аномальные значения по одному показателю. Это может возникнуть, в частности, в случае, когда внутренняя выборка образовательной организации невелика.

Внутренняя выборка образовательной организации — набор данных, который используется для формирования агрегированного показателя в отчетной документации.

Агрегированный показатель — это показатель, представленный в обобщенном виде, предназначенный для общего представления частных данных.

Примерами широко используемых в отчетной документации образовательных организаций агрегированных показателей являются данные о среднем, медианном, минимальном, максимальном баллах, полученных в ходе массовой диагностики уровня знаний обучающихся, например, в ходе всероссийских проверочных работ среди обучающихся одного класса.

Примером внутренней выборки образовательной организации малого объема могут быть результаты государственных экзаменов (итоговой государственной аттестации и единого государственного экзамена) по необязательным предметам. Ядро проблемы заключается в том, что в образовательных организациях, в которых численность обучающихся выпускных классов ниже среднего показателя, велика вероятность того, что некоторые предметы по выбору не будут выбраны ни одним обучающимся, либо предмет по выбору будет

сдавать лишь один обучающийся образовательной организации. Общий результат образовательной организации, полученный на основе данных лишь по одному обучающемуся, может являться выбросом (аномально малым или большим значением, не коррелирующим с общим уровнем подготовки в образовательной организации).

Рассмотрим наиболее и наименее часто сдаваемые предметы по выбору среди выпускников образовательных организаций среднего образования (таблица 1).

В ходе проведенного сбора данных о результатах сдачи единого государственного экзамена по предметам по выбору была установлена зависимость между долей выбросов (отсутствующими значениями и значениями, за пределами допустимых для участия в конкурсе при поступлении в высшие учебные заведения) и федеральными данными о популярности предметов по выбору среди выпускников средних образовательных организации в Российской Федерации [2, 3].

Как видно из таблицы 1, частота сдачи предмета по выбору в некоторой степени связана с вероятностью возникновения значения, которое ниже допустимого порога для поступления в высшие учебные заведения.

Следует заметить, что в выборке отсутствуют данные о сдаче экзамена по географии ввиду абсолютной редкости этого показателя (вероятность отсутствия показателя в отчетной документации образовательной организации значительно больше 50%). Что делает его

непригодным для сравнения показателей образовательных организаций.

Использование разных сегментов общей выборки данных позволяет минимизировать влияние выбросов на общий результат классификации, при этом в качестве основы (обучающей выборки) могут выступать:

1. Данные о классовой принадлежности образовательных организаций, составленные на основании оценок по аналогичным показателям с участием экспертов и/или специальных алгоритмов за прошедшие отчетные периоды;
2. Данные о классовой принадлежности, составленные на основании данных формулы;
3. Данные о классовой принадлежности, составленные на основании формулы и экспертных оценок для определения ценностных коэффициентов показателей.

В любой рейтинговой оценке могут быть выделены градации — специальные диапазоны, определяющие классовую принадлежность образовательных организаций, например, образовательные организации, получившие высокий или низкий общий балл, могут быть отнесены к отдельным классам.

#### 4. Объединенный вывод композиции моделей классификации

Завершающей процедурой классификации образовательных организаций с применением концепции бутстрэп-агрегирования является формирование объединенного вывода композиции моделей классификации.

Среди методов формирования объединенного вывода можно выделить следующие:

1. Метод большинства;
2. Метод абсолютного согласия;
3. Метод среднего класса.

Идея методов большинства и абсолютного согласия заключается в простом голосовании моделей классификации. Согласно указанным методам, если все или

большинство моделей классификации проголосовали за принадлежность образовательной организации к одному и тому же классу, то общий вывод моделей также указывает на этот же класс.

Метод среднего класса не предлагается к использованию в рамках метаалгоритма бутстрэп-агрегирования, однако, в том случае, когда классы образовательных организаций были образованы на основании рейтинга образовательных организаций путем их ранжирования по качеству обучения, т.е. в порядке убывания или возрастания рейтинговой оценки, тогда классы представляют собой последовательную структуру, при которой каждая образовательная организация, принадлежащая более высокому классу, с большой вероятностью будет иметь более высокую общую оценку деятельности либо более высокие образовательные результаты.

Благодаря такой последовательности общий вывод моделей в бутстрэп-агрегировании можно рассматривать как средний класс, предложенный несколькими моделями классификации. Таким образом, модификация процесса анализа данных, основанная на совмещении моделей согласно концепции бутстрэп-агрегирования, способна повысить качество принятия управленческих решений в органах управления образованием на региональном уровне [4, 5].

## Заключение

В статье описана методика применения ансамблей моделей классификации согласно метаалгоритму бутстрэп-агрегирования. Предложенная модификация алгоритма позволяет повысить степень точности итогового результата классификации образовательных организаций за счет проведения итогового голосования моделей классификации. Кроме того, выявлены проблемы исходных данных об эффективности образовательных организаций [6, 7] и предложена методика подготовки исходных данных для обучения моделей.

## ЛИТЕРАТУРА

1. Ромашкова О.Н., Ермакова Т.Н. Применение информационных технологий для анализа показателей качества обучения образовательного комплекса // В сборнике: Технологии информационного общества. X Международная отраслевая научно-техническая конференция: сборник трудов. 2016. С. 388–389
2. Gaidamaka, Y.V., Romashkova, O.N., Ponomareva, L.A., Vasilyuk, I.P. Application of information technology for the analysis of the rating of university // В сборнике: CEUR Workshop Proceedings 8. Sep. "ITMM 2018 — Proceedings of the Selected Papers of the 8th International Conference "Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems"" 2018. С. 46–53.
3. Ломовцев, Р.С. Особенности организации хранения данных о состоянии региональной образовательной среды / Р.С. Ломовцев // Открытая наука 2021: Сборник материалов научной конференции с международным участием, Москва, 22 апреля 2021 года. — Москва: Издательство «Aegitas», 2021. — С. 343–348.

4. Ломовцев, Р.С. Процесс интеллектуальной оценки деятельности региональных образовательных организаций среднего образования / Р.С. Ломовцев, О.Н. Ромашкова // Международная молодежная научная школа-конференция «Цифровая трансформация реального сектора экономики»: Сборник тезисов докладов, Москва, 21 мая 2021 года. — Москва: Национальный исследовательский ядерный университет «МИФИ», 2021. — С. 14–16.
5. Ломовцев, Р.С. Сводная модель поддержки принятия решений в управлении образовательной средой / Р.С. Ломовцев // #ScienceJuice2021: Сборник статей и тезисов, Москва, 22–26 ноября 2021 года / Составители: Е.В. Страмнова, С.А. Лепешкин. — Москва: Издательство ПАРАДИГМА, 2021. — С. 681–688.
6. Ромашкова О.Н., Ермакова Т.Н. Методика выбора информационной модели для оценки показателей качества обучения // Вестник Российского университета дружбы народов. Серия: Информатизация образования. 2015. № 2. С. 14–20.
7. Ромашкова О.Н., Пономарева Л.А., Василюк И.П. Применение инфокоммуникационных технологий для анализа показателей рейтинговой оценки вуза // В книге: Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем. Материалы Всероссийской конференции с международным участием. 2018. С. 65–68

---

© Ломовцев Роман Сергеевич ( [feedback.roman@gmail.com](mailto:feedback.roman@gmail.com) ), Ромашкова Оксана Николаевна ( [ox-rom@yandex.ru](mailto:ox-rom@yandex.ru) ).

Журнал «Современная наука: актуальные проблемы теории и практики»



Московский городской педагогический университет