Письма в ЭЧАЯ. 2025. Т. 22, № 1(258). С. 62–74

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ В ФИЗИКЕ

ДЕРЕВЬЯ И ЛЕСА РЕШЕНИЙ В ЗАДАЧЕ КЛАССИФИКАЦИИ КРЕДИТНЫХ ОРГАНИЗАЦИЙ

В. В. Иванов, Е. П. Акишина, А. С. Приказчикова [1]

Объединенный институт ядерных исследований, Дубна

Проведена классификация кредитных организаций с использованием методов деревьев и лесов решений для идентификации высокорисковых объектов. В ходе исследований разработаны модели дерева решений CART и лесов решений Random Forest, Adaboost и Xgboost, позволяющие идентифицировать высокорисковые кредитные организации. Для проведения расчетов использовались две вычислительные системы: облачный сервис Google Colaboratory (Google Colab) и гетерогенная платформа HybriLIT.

The purpose of this paper is to classify credit institutions using trees and decision forests to identify high-risk objects. During the research, the CART decision tree models and Random Forest, Adaboost and Xgboost decision forests were developed, allowing us to identify high-risk credit institutions. Two computing systems were used for the calculations: the Google Colaboratory cloud service (Google Colab) and the HybriLIT heterogeneous platform.

PACS: 02.50.Le; 02.50.Sk

ВВЕДЕНИЕ

Целью настоящей работы является проведение классификации кредитных организаций с целью идентификации высокорисковых объектов. Авторами поставлена задача — с использованием методов деревьев и лесов решений провести классификацию кредитных организаций на благонадежные и высокорисковые с точки зрения потенциального отзыва у них лицензий. Рассматриваются два типа объектов: благонадежные кредитные организации (класс «0») и неблагонадежные (класс «1»). Ранее авторами уже рассматривалась задача идентификации высокорисковых кредитных организаций, а именно с использованием нейронных сетей [1] и многомерного статистического анализа [2].

Метод деревьев решений позволяет оценить обстановку в финансовом секторе, а также визуализировать отдельные аспекты деятельности кредитных организаций. Деревья решений основаны на логических схемах, позволяющих получить решение о классификации объектов с помощью ответов на иерархически организованную систему вопросов. Основная идея данного метода заключается в построении бинарного дерева, в котором каждый узел представляет собой условие на признак, а листья — конечный результат работы алгоритма — принадлежность к определенному классу. Для того, чтобы провести классификацию, необходимо двигаться по дереву от корня до листа.

[1] E-mail: aska4.92@mail.ru

В начале исследования была сформирована рабочая выборка данных о 536 организациях (из которых 202 объекта с отозванной лицензией). Для неблагонадежных организаций выборка содержала 23 финансовых показателя на дату за три месяца до даты отзыва у них лицензии. Для всех благонадежных (действующих по настоящее время) организаций использовались данные на 31 января 2019 г. Используемые показатели:

1) население региона регистрации банка;
2) уставный капитал;
3) чистые активы;
4) счета в Банке России;
5) корсчета (НОСТРО);
6) ценные бумаги;
7) кредиты (общий показатель);
8) кредиты организациям;
9) средства организаций на расчетных счетах;
10) вклады физических лиц;
11) векселя;
12) капитал;
13) кредиты физическим лицам;
14) кредиты другим банкам (МБК);
15) основные средства;
16) учтенные векселя;
17) прибыль (убыток) до налогообложения;
18) корсчета (ЛОРО);
19) кредиты других банков;
20) средства клиентов (физические лица);
21) депозиты юридических лиц;
22) облигации;
23) резервы на возможные потери.

1. ДЕРЕВЬЯ РЕШЕНИЙ В ЗАДАЧЕ КЛАССИФИКАЦИИ КРЕДИТНЫХ ОРГАНИЗАЦИЙ

Рассмотрим три наиболее распространенных вида бинарных деревьев решений: CART, C5, CHAID. По результатам применения указанных моделей к исследуемым данным в Python [3] установлено, что максимальная точность (Accuracy) классификации кредитных организаций достигается при использовании алгоритма CART и составляет 70 %, C5 — 68 %, CHAD — 68 %. В связи с этим алгоритм CART будет детально проанализирован далее. Точность классификации Accuracy рассчитывалась по формуле

$$\text{Accuracy} = \frac{K}{G}, \tag{1}$$

где K — количество правильно классифицированных объектов, а G — общее количество объектов в выборке [4].

CART является алгоритмом построения бинарного дерева решений — дихотомической классификационной модели. Решающее дерево бинарное, если каждый узел

дерева при разбиении имеет только два потомка. Алгоритм на каждом шаге построения дерева последовательно сравнивает все возможные разбиения для всех атрибутов и выбирает наилучший [5].

В алгоритмах классификации, чаще всего основанных на деревьях решений, в том числе и в алгоритме CART, используется показатель Gini. Он трактуется как «примесь», «неоднородность» и характеризуется долей «примеси» представителей из других классов (признаков) в текущем варианте разбиения. Данный показатель называется неопределенностью Джини, так как связан с показателем информационной энтропии и находится как

$$I(A_k) = 1 - \sum_{k=1}^{m} p_k^2, \tag{2}$$

где m — количество классов целевой переменной при номерах классов $k = 1, 2, \ldots, m$; p_k — доля признаков в выборке A, принадлежащих классу k [5].

В общем случае задача классификации анализируемых объектов формулируется так: для некоторого вектора $\mathbf{x} = (\{x_1, \ldots, x_m\})$, где m — число признаков, необходимо провести процедуру принятия решения о принадлежности данного вектора к классу из множества $\mathbf{f} = (\{f_1, \ldots, f_n\})$, где n — число классов, основываясь на обучающей выборке данных.

По результатам применения алгоритма CART в Python были рассчитаны значения точностей для обучающей и тестовой выборок, составившие 100 и 70 % соответственно. Так как точность классификации на тестовой выборке оказалась значительно ниже точности на обучающей выборке, можно сделать предположение о переобучении нашего дерева решений. Деревья часто «страдают» подобным недостатком, они могут показывать хороший результат на обучающей выборке, но быть неэффективными на новых данных. Для борьбы с переобучением применяется принудительная остановка построения дерева такими методами, как ранняя остановка алгоритма после достижения заданного значения критерия, отсечение ветвей, ограничение глубины дерева, настройка гиперпараметров для модели, задание минимально допустимого количества примеров в узле, применение лесов решений. Основная задача в подобной ситуации — поиск наиболее выгодного баланса между сложностью и точностью дерева.

2. ОПТИМИЗАЦИЯ МОДЕЛИ ДЕРЕВА CART

Для оптимизации работы алгоритма, повышения точности модели и борьбы с переобучением используем процедуру настройки гиперпараметров дерева решений CART. В библиотеке scikit-learn пакета Python реализованы, в частности, два алгоритма для подбора оптимальных гиперпараметров: решетчатый подход (Grid Search) и случайный поиск (Random Search) [6]. При этом первый алгоритм реализует полный перебор по всей совокупности гиперпараметров, а второй отвечает за случайный перебор.

На первом шаге с помощью алгоритма Random Search локализуется область оптимальных значений гиперпараметров, отвечающая наилучшей точности модели CART. Затем по найденной области с помощью алгоритма Grid Search выполняется поиск комбинации гиперпараметров, отвечающей максимальной точности модели CART.

Таблица 1. **Оптимальные гиперпараметры для CART**

Гиперпараметр	Поисковый диапазон	Оптимальное значение
Максимальная глубина	(3, 11)	6
Максимальное количество признаков	(4, 24)	11
Максимальное количество листьев	(3, 30)	11
Минимальное количество объектов в листе	(1, 21)	10
Минимальное количество объектов в узле для его разделения	(2, 21)	2
Минимальное уменьшение неоднородности	[0,0001, 0,0005, 0,001, 0,002, 0,003, 0,004, 0,005, 0,006, 0,007, 0,008, 0,009, 0,01, 0,02, 0,03, 0,04, 0,05, 0,06, 0,07, 0,08, 0,09, 0,1, 0,2, 0,3, 0,5, 0,7]	0,001
Точность модели, %	78	

При этом алгоритмом Grid Search анализируется каждая комбинация гиперпараметров. В связи с этим второй шаг требует гораздо бóльших вычислительных и временных ресурсов, чем первый, при котором число итераций поиска задается пользователем вручную.

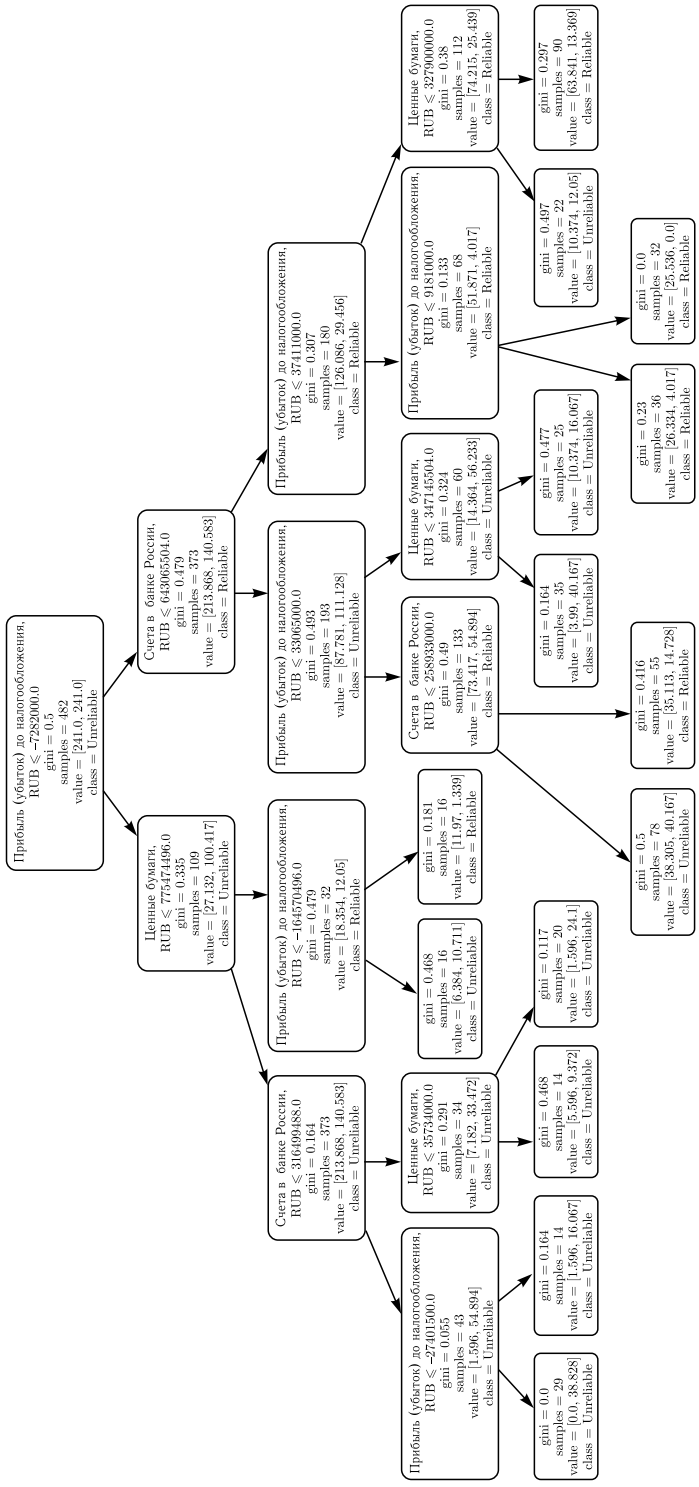
Результаты поиска оптимальных гиперпараметров для модели дерева CART на тестовых данных приведены в табл. 1.

Как видно из табл. 1, точность классификации модели CART на тестовой выборке с использованием оптимальных гиперпараметров составила 78 %, что выше точности без оптимизации.

Одним из достоинств дерева CART является возможность отбора наиболее важных показателей, которые в первую очередь влияют на отнесение объекта к одному из классов кредитных организаций: благонадежные и неблагонадежные. В связи с этим в Python была реализована процедура оценки значимости показателей. Наиболее важными для модели CART оказались следующие: «Прибыль», «Счета в Банке России» и «Ценные бумаги». В результате использования указанных показателей была построена новая модель дерева решений CART, точность которой составила 85 %, что на 7 % больше, чем для предыдущей модели. Дерево решений CART для новой модели представлено на рисунке.

Перейдем к анализу структуры построенного дерева CART. Рассмотрим листы (результирующие классы) дерева CART, индекс Gini которых не превышает 0,2. Листовой элемент № 1 содержит 29 неблагонадежных кредитных организаций. Индекс Gini для такого листа равен 0. Данный лист (класс) является абсолютно неблагонадежным, так как он содержит 100 % организаций, у которых были отозваны лицензии:

$$k_1 = \frac{29}{29} \cdot 100\,\% = 100\,\%. \qquad (3)$$



Дерево решений CART (точность классификации — 85 %)

Таблица 2. **Правила идентификации высокорисковых объектов с использованием дерева CART**

Номер листового элемента	Доля неблагонадежных организаций, %	Правила обхода дерева для попадания в листовой элемент
1	100	1. Прибыль $\leqslant -7282000$ & 2. Ценные бумаги $\leqslant 775474496$ & 3. Счета в ЦБ $\leqslant 316499488$ & 4. Прибыль $\leqslant -27401500$
2	94	1. Прибыль $\leqslant -7282000$ & 2. Ценные бумаги $\leqslant 775474496$ & 3. Счета в ЦБ $\leqslant 316499488$ & 4. Прибыль > -27401500
4	95	1. Прибыль $\leqslant -7282000$ & 2. Ценные бумаги $\leqslant 775474496$ & 3. Счета в ЦБ > 316499488 & 4. Ценные бумаги > 35734000
7	91	1. Прибыль > -7282000 & 2. Счета в ЦБ $\leqslant 643065504$ & 3. Прибыль > 33065000 & 4. Ценные бумаги $\leqslant 347145504$

Листовой элемент № 4 содержит 20 кредитных организаций, из которых 19 неблагонадежны. Индекс Gini такого листа равен 0,1. Данный лист (класс) является условно неблагонадежным, так как содержит 95 % организаций с отозванной лицензией:

$$k_2 = \frac{19}{20} \cdot 100\,\% = 95\,\%. \tag{4}$$

При классификации нового случая (анализируемой кредитной организации) любая организация, попавшая в лист № 4, является высокорисковой и подлежит дальнейшему анализу.

По аналогии с перечисленными выше листьями № 1 и 4 листовые элементы № 2 и 7 дерева CART содержат 91 и 94 % неблагонадежных объектов соответственно.

В табл. 2 представлены условия попадания в листовые элементы № 1, 2, 4, 7 дерева CART. В ходе анализа новой кредитной организации необходимо проводить сверку ее показателей с заданными условиями, в случае их соответствия одному из условий кредитную организацию следует отнести к числу высокорисковых. Если условия попадания в листовые элементы № 1, 2, 4, 7 не выполняются, кредитная организация считается благонадежной.

Для оценки качества работы модели CART в целом и на каждом из классов по отдельности были рассчитаны метрики классификации рассматриваемых объектов (5)–(7) [4] с использованием матрицы ошибок [7], представленной в табл. 3. Из этой таблицы видно, как часто классификатор выдает точные результаты и как часто его предсказания неверны.

В табл. 3 использованы следующие обозначения: TP — количество верноположительных результатов, TN — количество верноотрицательных результатов, FP — ко-

Таблица 3. **Матрица ошибок классификации кредитных организаций**

Данные прогнозов	Фактические данные	
	Благонадежный банк	Неблагонадежный банк
Благонадежный банк	TP	FP
Неблагонадежный банк	FN	TN

личество ложноположительных результатов, FN — количество ложноотрицательных результатов. Метрики классификации рассматриваемых объектов рассчитываются по следующим формулам:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{6}$$

$$\text{F1} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}} = 2\frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} = \frac{\text{TP}}{\text{TP} + \frac{\text{FP} + \text{FN}}{2}}, \tag{7}$$

где Precision — точность классификации благонадежных (неблагонадежных) объектов. Precision показывает долю объектов, названных моделью благонадежными (неблагонадежными), при этом действительно являющимися благонадежными (неблагонадежными); Recall — показывает, какую долю благонадежных (неблагонадежных) объектов из всех благонадежных (неблагонадежных) объектов нашел алгоритм/модель; F1 — среднее гармоническое пары Precision–Recall. F1 показывает одновременно, насколько хорошо модель находит объекты разных классов из всех объектов рассматриваемых классов и какую долю из тех, которые алгоритм назвал благонадежным (неблагонадежным) объектом, составляют действительно благонадежные (неблагонадежные).

В табл. 4 представлены итоговые метрики качества модели CART, построенной на 23 показателях, и модели CART, реализованной на трех наиболее важных показателях, где «0» — класс благонадежных объектов, «1» — класс неблагонадежных объектов.

Высокие значения метрик модели CART Accuracy и Precision, Recall, F1, рассчитанные отдельно для благонадежных и неблагонадежных классов объектов, свидетельствуют о целесообразности применения модели CART для выявления высокорисковых кредитных организаций (максимальная точность классификации неблагонадежных объектов составила 79 %).

Таблица 4. **Метрики качества моделей дерева CART**

Метрика	CART (23 показателя)		CART (3 показателя)	
	0	1	0	1
Precision, %	82	71	90	79
Recall, %	78	73	84	86
F1, %	80	70	87	83
Accuracy, %	**78**		**85**	

3. АНСАМБЛИ ДЕРЕВЬЕВ РЕШЕНИЙ

Одним из самых перспективных методов, позволяющих понизить уровень переобучения, является ансамблирование деревьев решений в леса решений. Леса решений строят множество деревьев решений и усредняют их результаты. Рассмотрим леса, показавшие наибольшую эффективность при решении задач финансового мониторинга: Random Forest, Adaboost и Xgboost.

Random Forest (случайный лес) — один из способов объединения деревьев решений в ансамбль. В задаче классификации выбирается решение по большинству результатов, выданных совокупностью деревьев [8]. Random Forest сочетает в себе идеи метода бэггинга (сокр. от bootstrap aggregation) и метода случайных подпространств [9, 10]. В алгоритме Random Forest, как и в методе бэггинга, обучение классификаторов происходит независимо друг от друга на разных подмножествах обучающей выборки, что решает проблему построения одинаковых деревьев на одном и том же множестве данных. Как и в методе случайных подпространств, в алгоритме Random Forest решающие деревья обучаются по случайным подмножествам показателей.

При реализации случайного леса важнейшим этапом является определение его гиперпараметров. При этом в качестве критерия разделения узла при решении задачи классификации может быть использован индекс Gini. В табл. 5 представлены результаты поиска оптимальных гиперпараметров для модели Random Forest.

Следующие алгоритмы — Adaboost и Xgboost — используют идею бустинга, заключающуюся в последовательном построении ансамбля базовых алгоритмов, при котором каждый следующий алгоритм добавляется в композицию с целью достижения наилучшей компенсации ошибок на текущем этапе.

Adaboost является первой моделью бустинга. Adaboost минимизирует экспоненциальную функцию потерь, что может сделать алгоритм чувствительным к выбросам. Adaboost выполняет взвешивание анализируемых объектов, которые были неправильно классифицированы ранее, и на следующих этапах производит «работу над ошибками» по выявленным инцидентам. Недостатки деревьев решений определяются по точкам с большим весом [11].

Таблица 5. **Оптимальные гиперпараметры для Random Forest**

Гиперпараметр	Поисковый диапазон	Оптимальное значение
Количество деревьев	(100, 1000)	300
Максимальная глубина	(3, 10)	6
Максимальное количество признаков	(4, 23)	17
Максимальное количество листьев	(3, 15)	5
Минимальное количество объектов в листе	(1, 10)	1
Минимальное количество объектов в узле для его разделения	(2, 10)	2
Точность модели, %	83	

Xgboost включает в себя идеи метода градиентного спуска и бустинга. На каждом этапе вычисляется градиент функции потерь по предсказаниям предыдущих деревьев решений. Каждое следующее дерево корректирует предсказания предыдущих [12]. С алгоритмом Xgboost можно использовать любую дифференцируемую функцию потерь, а следовательно, Xgboost более устойчив к выбросам, чем Adaboost.

Таблица 6. **Оптимальные гиперпараметры для Adaboost**

Гиперпараметр	Поисковый диапазон	Оптимальное значение
Количество деревьев	[100, 200, 300, 500, 1000, 1500, 2000, 3000, 5000]	2000
Алгоритм	['SAMME', 'SAMME.R']	SAMME.R
Максимальная глубина дерева	(3, 11)	4
Максимальное количество листьев	(4, 25)	18
Максимальное количество признаков	4, 24)	8
Минимальное количество объектов в листе	(1, 11)	1
Минимальное количество объектов в узле для его разделения	(2, 11)	6
Скорость обучения	[0,0001, 0,0005, 0,001, 0,005, 0,01, 0,05, 0,1, 0,5, 1,0]	0,1
Точность модели, %	80	

Таблица 7. **Оптимальные гиперпараметры для Xgboost**

Гиперпараметр	Поисковый диапазон	Оптимальное значение
Количество деревьев	[100, 200, 300, 500, 1000, 1500, 2000, 3000, 5000]	3000
Метод построения дерева	['exact', 'approx', 'hist']	hist
Максимальная глубина	(3, 11)	6
Максимальное количество листьев	(4, 25)	10
Скорость обучения	[0,0001, 0,001, 0,005, 0,01, 0,05, 0,1, 0,2, 0,3]	0,1
Параметр gamma	[0, 0,1, 0,5, 1, 1,5, 2]	1,5
Параметр reg_lambda	(0, 0,8)	0,6
Параметр reg_alpha	(0, 0,6)	0,5
Доля выборки для обучения	(0,5, 1)	0,8
Доля признаков для обучения	(0,5, 1)	0,5
Точность модели, %	80	

Таблица 8. **Метрики качества моделей лесов решений на 23 показателях**

Метрика	Random Forest		Adaboost		Xgboost	
	0	1	0	1	0	1
Precision, %	79	93	76	92	76	92
Recall, %	97	64	98	55	97	56
F1, %	87	76	86	69	85	69
Accuracy, %	**83**		**80**		**80**	

Продемонстрируем классификацию кредитных организаций по подготовленным данным с помощью Adaboost и Xgboost. Оптимальные гиперпараметры для указанных моделей представлены в табл. 6 и 7.

В табл. 8 представлены итоговые метрики качества моделей Random Forest, Adaboost и Xgboost, где «0» — класс благонадежных объектов, «1» — класс неблагонадежных объектов.

Точность модели Random Forest, реализованной на исходных 23 показателях, составила 83 %. При этом значение точности классификации неблагонадежных объектов составило 93 %, оно является максимальным среди рассматриваемых в данной работе моделей. Следует также отметить, что модель Random Forest показала лучшие результаты классификации по сравнению с другими моделями лесов решений: Adaboost и Xgboost.

Точность модели Adaboost, реализованной на исходных 23 показателях, составила 80 %. При этом значение точности классификации неблагонадежных объектов составило 92 %, что позволяет применять модель Adaboost для идентификации высокорисковых кредитных организаций. Аналогичная ситуация возникает и с моделью Xgboost. Схожесть результатов, которые показали модели Adaboost и Xgboost, обуславливается схожестью их алгоритмов, основанных на идее бустинга.

4. РЕАЛИЗАЦИЯ ДЕРЕВЬЕВ РЕШЕНИЙ НА GOOGLE COLABORATORY И ПЛАТФОРМЕ HYBRILIT

Для целей настоящего исследования были задействованы две вычислительные среды: облачный сервис Google Colaboratory [13] и гетерогенная платформа HybriLIT [14]. Облачный сервис Google Colaboratory (Google Colab) использует для расчетов центральный процессор сервера Google. Платформа HybriLIT является частью Многофункционального информационно-вычислительного комплекса (МИВК) Лаборатории информационных технологий ОИЯИ (Дубна).

В табл. 9 представлены аппаратные характеристики указанных систем, а в табл. 10 приведено время вычислений для дерева CART и лесов решений Random Forest, Adaboost, Xgboost. По данным табл. 10 видно, что для поиска оптимальных гиперпараметров с использованием платформы HybriLIT требуется существенно меньше времени, чем при использовании сервиса Google Colab. Из представленных в работе результатов следует, что использование гибридных вычислительных архитектур позволяет существенно ускорить решение научно-прикладных задач, а гетерогенный

Таблица 9. **Аппаратные характеристики сервера Google и HybriLIT**

Характеристика	Сервер Google	Платформа HybriLIT
Общий объем оперативной памяти, ГБ	13,3	24,7
Дисковое пространство	108 ГБ	665 ТБ
Информация о процессорах	Количество физических/виртуальных процессоров — 2 Intel(R) Xeon(R) CPU @ 2.20GHz Архитектура процессора — 32 или 64 бит Количество ядер на процессор — 1 Суммарное количество процессорных ядер — 2 Частота процессора — 2199,998 МГц Объем кеш-памяти — 56320 кБ	Количество физических/виртуальных процессоров — 4 Intel Core Processor (Broadwell, IBRS) Архитектура процессора — 32 или 64 бит Количество ядер на процессор — 1 Суммарное количество процессорных ядер — 4 Частота процессора — 2599,996 МГц Объем кеш-памяти — 16384 кБ

Таблица 10. **Результаты отработки систем для генерации деревьев и лесов решений**

Модель	Время поиска оптимальных гиперпараметров		Отношение t_1/t_2	Количество обработанных объектов
	Сервер Google (t_1)	Платформа HybriLIT (t_2)		
CART	28 мин 49 с	8 мин 47 с	3,3	576 480
Random Forest	58 мин 6 с	1 мин 14 с	47	3252
Adaboost	2 ч 54 мин 39 с	4 мин 5 с	43	2196
Xgboost	65 мин 23 с	1 мин 31 с	43	5388

вычислительный кластер HybriLIT является эффективным средством для достижения этой цели.

ЗАКЛЮЧЕНИЕ

В данной работе поставлена задача классификации кредитных организаций с помощью моделей дерева решений CART и лесов решений Random Forest, Adaboost и Xgboost. Расчеты для этих моделей проводились на двух вычислительных системах: на облачном сервисе Google Colab и на гетерогенной платформе HybriLIT. Показано, что скорость вычислений на платформе HybriLIT существенно превосходит скорость вычислений на Google Colab.

По результатам применения модели дерева CART к исходным 23 показателям финансовой отчетности № 101 была рассчитана точность классификации анализируемых объектов, которая составила 78 %. С использованием модели CART удалось опреде-

лить ключевые показатели деятельности кредитных организаций, а именно «Прибыль», «Счета в Банке России», «Ценные бумаги». При использовании ключевых показателей точность классификации для модели CART составила 85 %, а точность классификации неблагонадежных объектов — 79 %.

В расчетах моделей Random Forest, Adaboost и Xgboost использовались все 23 показателя финансовой отчетности № 101. При этом точность классификации кредитных организаций для указанных моделей составила 83, 80 и 80 % соответственно, а точность классификации неблагонадежных объектов составила 93, 92 и 92 % соответственно.

Таким образом, в результате проведенного исследования показана эффективность применения дерева решения CART и лесов решений Random Forest, Adaboost и Xgboost к задаче классификации кредитных организаций. С учетом полученных значений метрик качества классификации неблагонадежных объектов рассматриваемые модели целесообразно применять для идентификации высокорисковых кредитных организаций и прогноза отзыва у них лицензий.

Финансирование работы. Никаких грантов на проведение или руководство данным конкретным исследованием получено не было.

Конфликт интересов. Авторы данной работы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Акишина Е. П., Иванов В. В., Крянев А. В., Приказчикова А. С.* Сравнительный анализ методов деревьев решений и нейронных сетей в задаче классификации кредитных организаций // Вестн. НИЯУ МИФИ. 2022. Т. 11, № 6. С. 442–449; https://doi.org/10.26583/vestnik.2022.12.
2. *Акишина Е. П., Иванов В. В., Крянев А. В., Приказчикова А. С.* Многомерный анализ данных в задаче прогнозирования попадания кредитных организаций в зону риска // Вестн. НИЯУ МИФИ. 2024. Т. 13, № 1. С. 22–29; https://doi.org/10.26583/vestnik.2024.302. EDN: HUDHFW.
3. Python: [Электронный ресурс]. https://www.python.org (дата обращения: 19.05.2024).
4. *Демидова Л. А., Клюева И. А.* Алгоритм подбора значений параметров bSMOTE-алгоритма в задаче SVM-классификации на основе несбалансированных наборов данных // Вестн. Рязан. гос. радиотехн. ун-та. 2017. № 61. С. 67–77.
5. *Jiawei Han, Micheline Kamber, Jian Pei.* Data Mining. Concepts and Techniques. Elsevier, 2012.
6. *Демидова Л. А., Клюева И. А.* Разработка и исследование гибридных версий алгоритма роя частиц на основе алгоритмов поиска по сетке // Вестн. Рязан. гос. радиотехн. ун-та. 2016. № 57. С. 105–116.
7. *Паклин Н. Б., Орешков В. И.* Бизнес-аналитика: от данных к знаниям: Учеб. пособие. 2-е изд., испр. СПб.: Питер, 2013. 704 с.
8. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Ch. 15. Random Forests. Springer-Verlag, 2009.
9. *Breiman L.* Random Forests // Machine Learning. 2001. V. 45, No. 1. P. 5–32.
10. *Кашницкий Ю. С., Игнатов Д. И.* Ансамблевый метод машинного обучения, основанный на рекомендации классификаторов // Интеллектуал. системы. Теория и приложения. 2015. Т. 19, № 4. С. 1–32.

11. *Freund Y., Schapire R. E.* A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting // J. Comput. Syst. Sci. 1997. No. 55. P. 119–139; doi: 10.1006/jcss.1997.1504.
12. *Chen T., Guestrin C.* XGBoost: A Scalable Tree Boosting System // Comput. Sci. Machine Learning. 2016. No. 1; doi: 10.48550/arXiv.1603.02754; https://arxiv.org/abs/1603.02754 (дата обращения: 01.05.2024).
13. Google Colaboratory: [Электронный ресурс]. https://colab.google. (дата обращения: 21.05.2024).
14. Гетерогенная платформа «HybriLIT»: [Электронный ресурс]. http://hlit.jinr.ru.(дата обращения: 28.07.2024).

Получено 12 августа 2024 г.