ВЕСТНИК НАЦИОНАЛЬНОГО ИССЛЕДОВАТЕЛЬСКОГО ЯДЕРНОГО УНИВЕРСИТЕТА «МИФИ», 2024, т. 13, № 1, с. 22–29

УДК 004.62

МНОГОМЕРНЫЙ АНАЛИЗ ДАННЫХ В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ПОПАДАНИЯ КРЕДИТНЫХ ОРГАНИЗАЦИЙ В ЗОНУ РИСКА

***Е.П. Акишина*[1], *В.В. Иванов*[1,2], *А.В. Крянев*[1,2*], *А.С. Приказчикова*[2]**
[1]*Объединённый институт ядерных исследований, Дубна, 141980, Россия*
[2]*Национальный исследовательский ядерный университет «МИФИ», Москва, 115409, Россия*
**e-mail: AVKryanev@mephi.ru*

Поступила в редакцию: 10.12.2023
После доработки: 10.12.2023
Принята к публикации: 26.12.2023

Исследование экономических процессов базируется на изучении большого числа параметров. В связи с чем для проведения анализа исследуемых явлений и решения прогностических задач возникает необходимость применения методов многомерного анализа данных. В статье рассматривается проблема идентификации подозрительных, с точки зрения финансовой состоятельности, кредитных организаций, осуществляющих свою деятельность на российском рынке. Настоящее исследование нацелено на разработку методики многомерного анализа данных для выявления подозрительных кредитных организаций и прогнозирования отзыва у них лицензий. Для решения указанной задачи предлагается использовать иерархические и итерационные методы кластерного анализа, а также метод главных компонент. На основе этих методов разработана методика формирования зоны риска, позволяющая прогнозировать отзыв лицензий у кредитных организаций. Для определения количества кластеров применялся метод кластеризации Варда, а также метод локтя, метод силуэтов, метод Дэвиса–Боулдина. Совместное использование методов кластерного анализа и метода главных компонент позволили продемонстрировать робастность предложенной методики. В настоящем исследовании использовались данные формы банковской отчетности № 101.

Ключевые слова: методы кластерного анализа, метод *k*-средних, метод главных компонент, евклидово расстояние, кредитная организация, отзыв лицензии.
DOI: 10.26583/vestnik.2024.302
EDN HUDHFW

ВВЕДЕНИЕ

Исследование экономических процессов базируется на изучении большого числа параметров. В связи с чем для проведения анализа исследуемых явлений и решения прогностических задач возникает необходимость применения методов многомерного анализа данных [1].

В настоящей работе для исследования деятельности банков на российском рынке использовались методы кластерного анализа и метод главных компонент.

Можно выделить два вида методов кластерного анализа: иерархические и итерационные. Иерархические методы предназначены для разбиения исходных данных на группы (кластеры) таким образом, чтобы элементы, входящие в одну группу, были максимально «схожи», а элементы из разных групп были максимально «отличными» друг от друга. К итерационным методам относится, в частности, метод *k*-средних, с помощью которого вычисляются характеристики кластеров [2].

Авторами работы [3] показано, что совместное применение методов кластерного анализа и метода главных компонент [4] позволяет существенно сократить размерность анализируемых данных без заметной потери качества информации.

Настоящее исследование нацелено на разработку методики многомерного анализа данных для выявления подозрительных кредитных организаций и прогнозирования отзыва у них лицензий.

1. ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА

Ранее, в работах [5, 6], нами анализировалась банковская сфера в России, а именно незаконная активность отдельных ее звеньев. Признаковое пространство, характеризующее операционную деятельность банков, исследовалось с

помощью искусственных нейронных сетей и деревьев решений.

В настоящей работе анализировались данные о 469 кредитных организациях (из которых 111 банки с отозванной лицензией) с января 2019 по декабрь 2021 г., выборка содержит 22 финансовых показателя из формы банковской отчетности № 101. Кроме того, в выборке присутствует категориальный признак «Отзыв» («1» – у банка отозвана лицензия, «0» – действующий банк) и поле «Причина», в котором указана причина отзыва лицензии у банка в случае прекращения его деятельности.

1.1. Оценка количества кластеров

На первом шаге исследования для проведения кластеризации всех объектов определим подходящее количество кластеров. Для этого применим метод иерархической кластеризации Варда [7]. В этом методе используется мера сходства между объектами. Примем в качестве такой меры – евклидово расстояние, которое является геометрическим расстоянием в многомерном пространстве между объектами, и вычисляется следующим образом (1):

$$S(x,y) = \sum_i \sqrt{(x_i - y_i)^2}, \qquad (1)$$

где x_i, y_i – координаты точек (объектов), между которыми рассчитывается расстояние, i = 1, 2, ..., k, k – количество измерений [8].

В результате применения метода Варда в программе Statistica [9] ко всем объектам исследования построена дендрограмма (рис. 1), на которой видно, что при межкластерном расстоянии 60 (красная линия) объекты образуют четыре кластера.

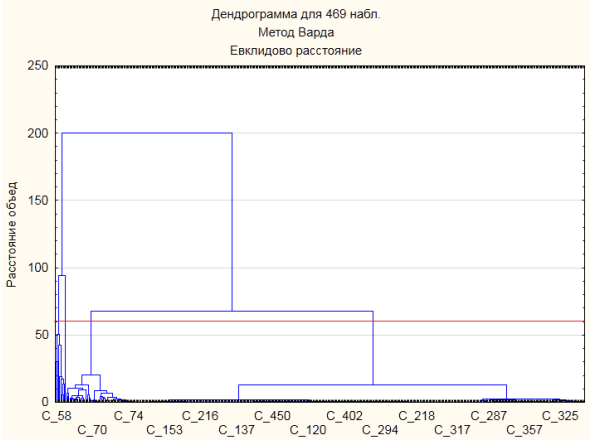

Рис. 1. Применение метода Варда

Кроме метода Варда, нами были рассмотрены метод локтя, метод силуэтов, метод Девиса-Боулдина [10]. Результаты применения этих методов и метода Варда представлены в табл. 1.

Таблица 1. Оценка числа кластеров разными методами

Метод оценки числа кластеров	Количество кластеров
Метод Варда	4
Метод локтя	6
Метод силуэтов	2
Метод Девиса-Боулдина	5

Так как методы локтя, силуэтов и Девиса–Боулдина не дали однозначного ответа для оценки числа кластеров, остановим свой выбор на методе Варда и примем это число за 4. Далее перейдем к методу k-средних.

1.2. Применение метода кластеризации k-средних и формирование зоны риска

В методе k-средних кластерного анализа необходимо заранее задать число кластеров, для чего используем результат метода Варда – четыре. Из исходного множества данных случайным образом выбираются k-наблюдений, которые будут служить начальными центрами кластеров. Для каждого наблюдения исходных данных определяется ближайший к нему центр. Далее вычисляется центр масс каждого кластера. Затем объекты разбиваются на группы вновь в соответствии с тем, к какому из новых центров каждый объект оказался ближе, оценив это по метрике евклидового расстояния. Таким образом, минимизируется суммарное квадратичное отклонение объектов кластеров от центров этих кластеров:

$$V = \sum_{i=1}^{k} \sum_{x \in S_i} (x - \mu_i)^2, \qquad (2)$$

где k – число кластеров; S_i – полученные кластеры, i = 1, 2, ..., k, а μ_i – центры масс всех векторов x из кластера S_i [11].

После применения метода k-средних оценим качество разбиения на кластеры, для чего используем критерий различия (t-критерий Стьюдента). Результаты оценки представлены на рис. 2. Значение p (p-уровень статистической значимости) менее чем 0.05 говорит о значимом различии между полученными кластерами.

перемен.	Дисперсионный анализ					
	Между SS	cc	Внутри SS	cc	F	значим. p
Уставный капитал, RUB	280,4523	3	187,5477	465	231,782	0,000000
Чистые активы, RUB	447,0295	3	20,9705	465	3304,147	0,000000
Счета в Банке России, RUB	452,8242	3	15,1758	465	4624,965	0,000000
Коррсчета (НОСТРО), RUB	266,3827	3	201,6173	465	204,791	0,000000
Ценные бумаги, RUB	435,4265	3	32,5735	465	2071,965	0,000000
Кредиты, RUB	444,3893	3	23,6107	465	2917,329	0,000000
Кредиты организациям, RUB	450,2507	3	17,7493	465	3931,915	0,000000
Кредиты физическим лицам, RUB	437,7573	3	30,2427	465	2243,592	0,000000
Кредиты другим банкам (МБК), RUB	229,5782	3	238,4218	465	149,251	0,000000
Основные средства, RUB	379,9812	3	88,0188	465	669,142	0,000000
Учтённые векселя, RUB	344,9544	3	123,0456	465	434,538	0,000000
Прибыль (убыток) до налогообложения, RUB	376,0106	3	91,9894	465	633,569	0,000000
Коррсчета (ЛОРО), RUB	311,1045	3	156,8955	465	307,346	0,000000
Кредиты других банков (МБК), RUB	146,5341	3	321,4659	465	70,654	0,000000
Средства клиентов, RUB	451,0007	3	16,9993	465	4112,242	0,000000
Средства организаций на расчетных счетах, RUB	444,8062	3	23,1938	465	2972,566	0,000000
Депозиты юридических лиц, RUB	444,2574	3	23,7427	465	2900,262	0,000000
Вклады физических лиц, RUB	450,8257	3	17,1744	465	4068,742	0,000000
Облигации, RUB	340,2688	3	127,7312	465	412,911	0,000000
Векселя, RUB	362,2844	3	105,7156	465	531,181	0,000000
Резервы на возможные потери, RUB	378,6928	3	89,3072	465	657,252	0,000000
Капитал, RUB	431,6939	3	36,3061	465	1843,012	0,000000

Рис. 2. Результаты дисперсионного анализа в пакете Statistica при реализации метода *k*-средних

В табл. 2 представлен состав кластеров после проведения первого этапа кластеризации, включающего метод Варда и метод *k*-средних. Из табл. 2 видно, что в кластеры 1, 2 и 3 попали крупные и благонадежные организации.

Таблица 2. Состав кластеров после первого этапа кластерного анализа

№ кластера	Состав
1	ПАО «Сбербанк»
2	АО «Газпром», АО «ВТБ»
3	АО «Альфа-банк», ПАО «Росбанк», АО «Россельхозбанк», ПАО Банк «ФК Открытие», ПАО «Промсвязьбанк», ПАО «Московский кредитный банк», НКО АО «НРД», НКО АО «НКЦ»
4	Остальные банки

Анализ средних значений финансовых показателей объектов для разных кластеров (рис. 3) показал, что средние значения показателей для объектов четвертого кластера значительно ниже значений показателей для других кластеров.

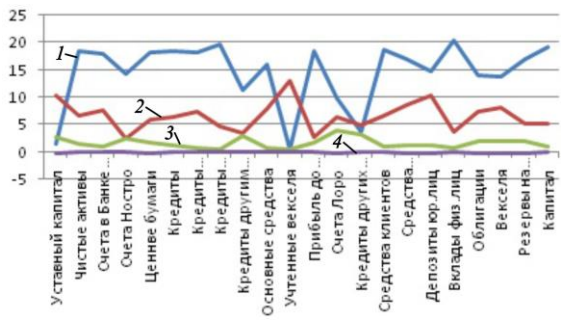

Рис. 3. График средних значений показателей объектов каждого кластера после первого этапа кластеризации: *1* – первый кластер; *2* – второй кластер; *3* – третий кластер; *4* – четвертый кластер

Из табл. 3 видно, что все неблагонадежные объекты сосредоточились в четвертом кластере (111 объектов).

Таблица 3. Количество благонадежных и неблагонадежных объектов после применения методов кластерного анализа на первом этапе

№ кластера	Количество благонадежных	Количество неблагонадежных	Общее количество
1	1	0	1
2	2	0	2
3	8	0	8
4	335	111	446

Таким образом, на первом этапе кластеризации мы можем выделить четвертый кластер, как целевой для дальнейшего анализа, и сосредоточиться на объектах, содержащихся в нем.

Реализуем аналогичную процедуру кластерного анализа теперь для объектов четвертого кластера: применим метод Варда, определим число кластеров и затем применим метод *k*-средних. Указанные ранее шаги кластеризации мы будем последовательно применять к самому многочисленному кластеру до тех пор, пока не получим кластер, который содержит соизмеримое количество благонадежных и неблагонадежных объектов.

В настоящем исследовании было проведено пять этапов кластеризации, по окончании которых осталось 194 объекта, разделенных на семь кластеров. Из табл. 4 видно, что кластер № 1 содержит сбалансированное количество благонадежных и неблагонадежных объектов.

Таблица 4. Количество благонадежных и неблагонадежных объектов после пятого этапа кластеризации

№ кластера	Количество благонадежных	Количество неблагонадежных	Общее количество
1	62	42	104
2	0	3	3
3	14	4	18
4	26	6	32
5	4	2	6
6	7	7	14
7	13	4	17

2. АНАЛИЗ ЗОНЫ РИСКА

На следующем шаге исследования нами проводился анализ первого кластера, полученного на пятой итерации. В данном кластере содержатся 62 благонадежных объекта и 42 неблагонадежных. В связи с этим, в данный кластер попало большинство неблагонадежных объектов, мы назвали объекты первого кластера высокорисковыми, и именно они формируют зону высокого риска.

Далее на основе показателей 42 неблагонадежных объектов строится центроид зоны риска. Центроидом назовем точку в многомерном пространстве, каждая координата которой представляет собой математическое ожидание по каждой переменной. На рис. 4 приведена диаграмма Евклидовых расстояний от центроида зоны риска до всех объектов.

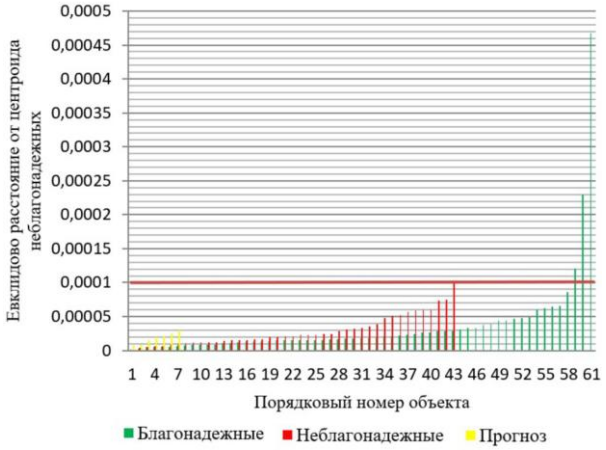

Рис. 4. Диаграмма евклидовых расстояний от центроида зоны риска до всех объектов

Из рис. 4 можно сделать предположение, что все благонадежные объекты, находящиеся на расстоянии меньшем, чем 10^{-4}, являются подозрительными. В связи с этим именно указанные объекты должны быть подвергнуты экспертной проверке в первую очередь.

На рис. 5 приведены распределения евклидовых расстояний от центроида зоны риска до всех объектов.

Из рисунка видно, что большинство неблагонадежных объектов сосредоточено в первых пяти интервалах (всего 26 организаций) распределения. Это дает основание полагать, что благонадежные объекты из этих интервалов являются наиболее подозрительными с точки зрения их возможной несостоятельности или вовлеченности в противоправную деятельность.

Рис. 5. Распределение евклидовых расстояний в области центроида (исходные данные)

Далее рассматривались два множества организаций: множество всех благонадежных организаций зоны риска и множество организаций, у которых по решению Банка России была отозвана лицензия в период с января 2022 по август 2023 г., всего 15 организаций. В результате пересечения указанных множеств были идентифицированы семь организаций:

$$k_1 = \frac{p_1}{n}, \qquad (3)$$

где p_1 – идентифицированные организации; n – организации с отозванной в заданный период лицензией.

Таким образом, с использованием предложенной методики спрогнозировано 46.6 % фактов отзыва банковских лицензий.

Из диаграммы на рис. 4 видно, что все семь идентифицированных организаций (отмечены желтым цветом) оказались в зоне неблагонадежных организаций.

3. ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА К ДАННЫМ, ПОЛУЧЕННЫМ НА ОСНОВЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ В ЭЛЕКТРОННОМ ВИДЕ И ФОРМИРОВАНИЕ ЗОНЫ РИСКА

В работе [3] показана практическая эффективность методов оценки количества кластеров после проведения метода главных компонент и формирования соответствующих главных компонент (ГК), на основе которых и осуществлялся кластерный анализ.

В основе метода ГК лежат всевозможные линейные преобразования исходных признаков. В результате применения метода к исходным данным сгенерированы 22 компоненты. На рис. 6 представлен график «каменистой осыпи»,

который демонстрирует вклад каждой главной компоненты в совокупную дисперсию. Из рисунка видно, что наибольший вклад дает первая главная компонента, а начиная с шестой компоненты вклад остальных незначителен. Поэтому можно использовать пять главных компонент (из сгенерированных 22), общая совокупность дисперсии которых составляет 96.4 %.

С целью оценки количества кластеров, на которое следует разбить совокупность анализируемых данных, применим к отобранным ГК метод Варда (рис. 7), метод Девиса–Боулдина (рис. 8), метод силуэтов (рис. 9) и метод локтя (рис. 10).

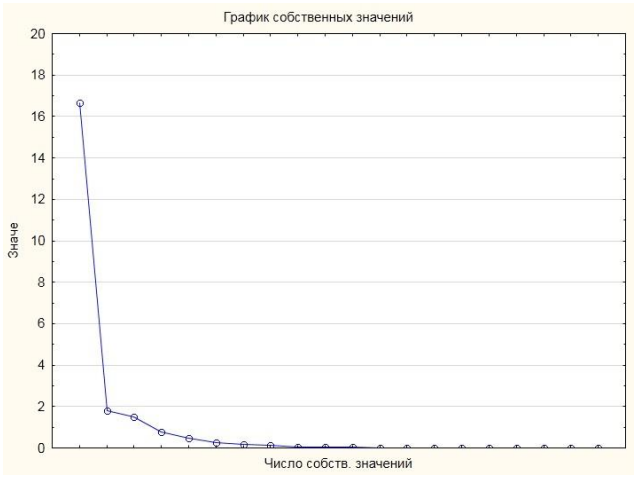

Рис. 6. График «каменистой» осыпи

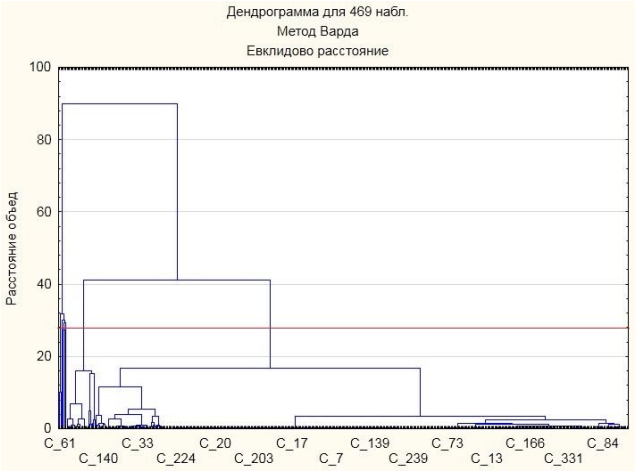

Рис. 7. Применение метода Варда

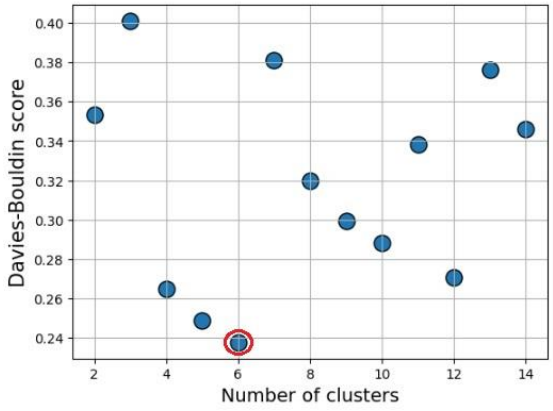

Рис. 8. Применение метода Девиса–Боулдина

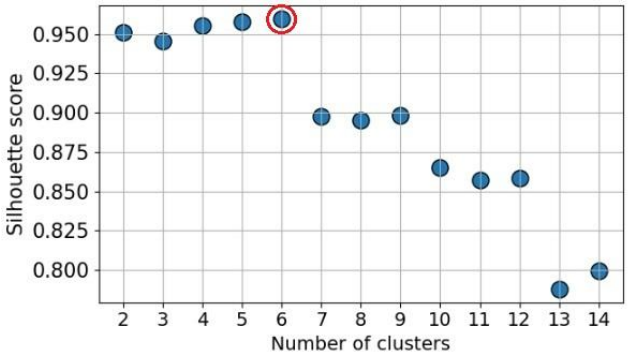

Рис. 9. Применение метода силуэтов

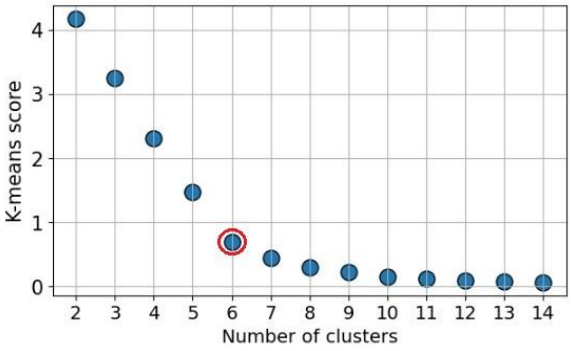

Рис. 10. Применение метода локтя

Результаты отработки методов оценки кластеров с использованием ГК представлены в табл. 5.

Исходя из результатов оценки, проведена кластеризация методом *k*-средних с числом кластеров 6. Результаты кластерного анализа первого этапа представлены в табл. 6.

Таблица 5. Оценка числа кластеров разными методами (ГК)

Метод оценки числа кластеров	Количество кластеров
Метод Варда	6
Метод локтя	6
Метод силуэтов	6
Метод Девиса–Боулдина	6

Таблица 6. Количество благонадежных и неблагонадежных объектов после применения методов кластерного анализа на первом этапе (ГК)

№ кластера	Количество благонадежных	Количество неблагонадежных	Общее количество
1	2	0	2
2	4	0	4
3	337	123	460
4	1	0	1
5	1	0	1
6	1	0	1

На рис. 11 представлен график средних значений пяти главных компонент для каждого из шести кластеров, из которого видно, что средние значения главных компонент для объектов третьего кластера расположены в зоне нуля.

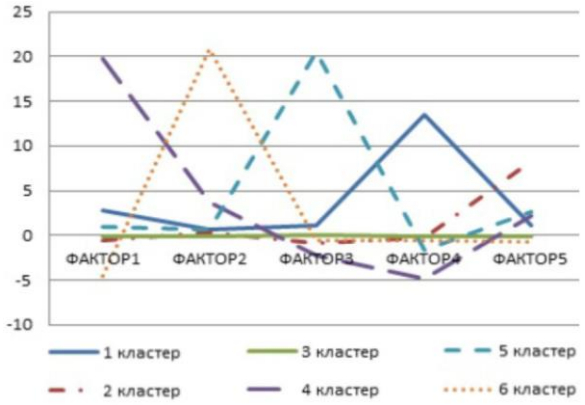

Рис. 11. График средних значений ГК объектов каждого кластера после первого этапа кластеризации

Из табл. 6 видно, что в третий кластер попало максимальное число неблагонадежных кредитных организаций. Повторим вышеописанную процедуру (оценка числа кластеров и кластеризация методом *k*-средних) для объектов третьего кластера. Указанные этапы кластеризации будем последовательно применять к самому многочисленному кластеру (аналогично тому, как делали с исходными данными) до тех пор, пока не получим кластер, который содержит соизмеримое количество благонадежных и неблагонадежных объектов.

Всего было проведено шесть этапов кластеризации, по окончании которых получено четыре кластера. Кластер № 1 (табл. 7) содержит сбалансированное количество благонадежных и неблагонадежных объектов.

Таблица 7. Количество благонадежных и неблагонадежных объектов после шестого этапа кластеризации

№ кластера	Количество благонадежных	Количество неблагонадежных	Общее количество
1	53	28	81
2	18	4	22
3	46	21	67
4	2	1	3

4. АНАЛИЗ ЗОНЫ РИСКА УСЕЧЕННЫХ ДАННЫХ

На заключительном этапе исследования проводился анализ выделенного первого кластера, полученного в третьем разделе настоящей работы. В данном кластере содержатся 53 благонадежных объекта и 28 неблагонадежных. В связи с тем, что 53 благонадежных объектов попали в кластер с большим количеством неблагонадежных объектов, будем считать эти 53 объекта подозрительными. Они и будут формировать так называемую зону риска.

С использованием данных о неблагонадежных объектах строился центроид зоны риска. На рис. 12 приведена диаграмма евклидовых расстояний от центроида до всех объектов. В данном случае в зону высокого риска попадают все объекты (из них 37 благонадежных организаций), евклидово расстояние которых до центроида неблагонадежных менее чем $3 \cdot 10^{-3}$. Отметим, что неблагонадежный объект с евклидовым расстоянием 0.0068 был признан аномальным и не учитывался при формировании зоны риска.

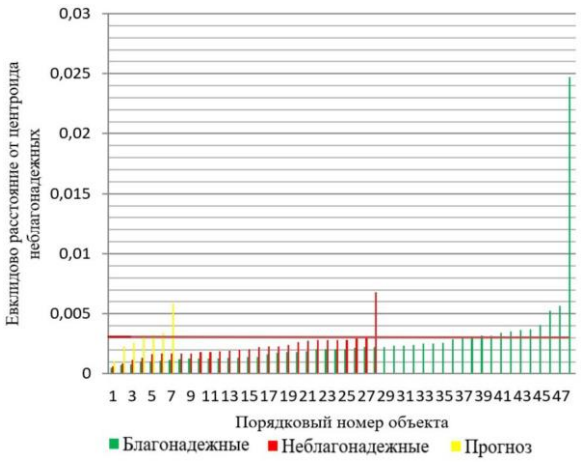

Рис. 12. Диаграмма евклидовых расстояний от центроида до всех объектов (ГК)

В ходе анализа состава зоны риска рассматривались два множества: множество всех благонадежных организаций зоны риска и множество организаций, у которых по решению Банка России была отозвана лицензия в период с января 2022 по август 2023 г., всего 15. В результате пересечения указанных множеств были идентифицированы шесть организаций.

Таким образом, применение разработанной методики с использованием данных, построенных с помощью метода главных компонент, позволило спрогнозировать 40 % фактов отзыва банковских лицензий:

$$k_2 = \frac{p_2}{n}, \quad (4)$$

где p_2 – идентифицированные организации; n – организации с отозванной в заданный период лицензией.

АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ И ВЫВОДЫ

В заключение подведем итоги выполненного исследования:

- на основе методов кластерного анализа и метода главных компонент разработана методика для прогнозирования попадания кредитных организаций в зону риска;
- совместное использование методов кластерного анализа и метода ГК позволило продемонстрировать робастность предложенной методики;
- сформирована зона риска, в которую попало заметное количество благонадежных организаций, на которые, в первую очередь, необходимо обратить особое внимание;
- разработанная методика позволила спрогнозировать 40–47 % фактов отзыва лицензий у благонадежных кредитных организаций.

ФИНАНСИРОВАНИЕ

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (проект государственного задания № FSWU-2023-0031).

СПИСОК ЛИТЕРАТУРЫ

1. *Андерсон Т.* Введение в многомерный статистический анализ. М.: Государственное издательство физико-математической литературы, 1963. 500 с.

2. *Дюран Б., Оделл П.* Кластерный анализ. М.: Статистика, 1977. 128 с.

3. *Dr. Tirthajyoti Sarkar*. Clustering and dimensionality reduction techniques combined. [Электронный ресурс]. URL: https://github.com/ tirthajyoti/Machine-Learning-with-Python/blob/master/Clustering-Dimensionality-Reduction/Clustering_with_dim_reduction.ipynb (дата обращения: 07.07.2023).

4. *Айвазян С.А., Бухштабер В.М., Енюков И.С.* Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с. ISBN 5-279-00054-X.

5. *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 // Вестник НИЯУ МИФИ, 2022. Т. 11. № 6. С. 442–449. https://doi.org/10.26583/vestnik.2022.12.

6. *Акишина Е.П., Иванов В.В., Приказчикова А.С.* Применение нейронных сетей и метода главных компонент для идентификации кредитных организаций, потенциально вовлеченных в процесс по легализации преступных доходов // Известия Иссык-Кульского форума бухгалтеров и аудиторов стран Центральной Азии, 2022. № 2(37). С. 294–296.

7. *Murtagh F.* Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion. // Journal of Classification, 2014. № 31. P. 274–295.

8. *Айвазян С.А.* Многомерный статистический анализ в социально-экономических исследованиях // Экономика и математические методы, 1977. № 13 (5). С. 968–985.

9. StatSoft Inc., 2011. STATISTICA (data analysis software system), version 10. [Электронный ресурс]. URL: www.statsoft.com (дата обращения: 07.07.2023).

10. *Баймуратов И.Р.* Методы автоматизации машинного обучения. СПб.: Университет ИТМО, 2020. 40 с.

11. *Айвазян С.Л., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика. Исследование зависимостей. М.: Издательство «Финансы и статистика», 1985. 488 с.

Vestnik Natsional'nogo Issledovatel'skogo Yadernogo Universiteta «MIFI», 2024, vol. 13, no. 1, pp. 21–29

MULTIDIMENSIONAL DATA ANALYSIS IN THE TASK OF PREDICTING THE ENTRY OF CREDIT INSTITUTIONS INTO THE RISK ZONE

***E.P. Akishina*[1], *V.V. Ivanov*[1,2], *A.V. Kryanev*[2,*], *A.S. Prikazchikova*[2]**
[1]*Joint Institute for Nuclear Research, Dubna, 141980 Russia*
[2]*National Research Nuclear University «MEPhI», Moscow, 115409 Russia*
**e-mail: AVKryanev@mephi.ru*

Received December 10, 2024; revised December 10, 2024; accepted December 26, 2024

The study of economic processes is based on the study of a large number of parameters. In this connection, in order to analyze the phenomena under study and solve prognostic problems, there is a need to use methods of multidimensional data analysis. The article examines the problem of identifying suspicious, from the point of view of financial solvency, credit institutions operating in the Russian market. This study is aimed at developing a methodology for multidimensional data analysis to identify suspicious credit institutions and predict the revocation of their licenses. To solve this problem, it is proposed to use hierarchical and iterative methods of cluster analysis, as well as the principal component method. Based on these methods, a methodology for forming a risk zone has been developed that makes it possible to predict the revocation of licenses from credit institutions. To determine the number of clusters, the Ward clustering method was used, as well as the elbow method, the silhouette method, and the Davis-Bouldin method. The combined use of cluster analysis methods and the principal component method made it possible to demonstrate the robustness of the proposed methodology. Data from Bank Reporting Form No. 101 were used in this study.

Keywords: cluster analysis methods, *k*-means method, principal component method, Euclidean distance, credit institution, license revocation.

REFERENCES

1. *Anderson T.* Vvedenie v mnogomernyj statisticheskij analiz. [Introduction to multivariate statistical analysis]. Moscow, State Publishing House of Physical and Mathematical Literature Publ., 1963. 500 p.

2. *Durand B., Odell P.* Klasternyj analiz. [Cluster analysis]. Moscow, Statistika Publ., 1977. 128 p.

3. *Dr. Tirthajyoti Sarkar.* Clustering and dimensionality reduction techniques combined.

Available at: https: //github.com/tirthajyoti/Machine-Learning-with-Python/blob/master/Clustering-Dimensionality-Reduction/Clustering_with_dim_reduction.ipynb (accessed 07.07.2023).

4. *Ajvazyan S.A., Buhshtaber V.M., Enyukov I.S.* Prikladnaya statistika. Klassifikaciya i snizhenie razmernosti [Applied statistics. Classification and dimension reduction]. Moscow: Finansy i statistika Publ., 1989. 607 p. ISBN 5-279-00054-X.

5. *Akishina E.P., Ivanov V.V., Kryanev A.V., Prikazchikova A.S.* Cravnitel'nyj analiz metodov derev'ev reshenij i nejronnyh setej v zadache klassifikacii kreditnyh organizacij [Comparative analysis of decision tree and neural network methods in the problem of classification of credit institutions]. Vestnik NIYaU MIFI, 2022. Vol. 11. No. 6. Pp. 442–449. https://doi.org/10.26583/vestnik.2022.12 (in Russian)

6. *Akishina E.P., Ivanov V.V., Prikazchikova A.S.* Primenenie nejronnyh setej i metoda glavnyh komponent dlya identifikacii kreditnyh organizacij, potencial'no vovlechennyh v process po legalizacii prestupnyh dohodov [Application of neural networks and the principal component method for identifying credit institutions potentially involved in the process of money laundering]. Izvestiya Issyk-Kul'skogo foruma buhgalterov i auditorov stran Central'noj Azii, 2022. No. 2(37). Pp. 294–296 (in Russian).

7. *Murtagh F.* Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion. Journal of Classification, 2014. No. 31. Pp. 274–295.

8. *Ayvazyan S.A.* Mnogomernyj statisticheskij analiz v social'no-ekonomicheskih issledovaniyah [Multivariate statistical analysis in socio-economic research]. Ekonomika i matematicheskie metody, 1977. No. 13 (5). Pp. 968–985 (in Russian).

9. StatSoft Inc., 2011. STATISTICA (data analysis software system), version 10. Available at: www.statsoft.com (accessed 07.07.2023).

10. *Baimuratov I.R.* Metody avtomatizacii mashinnogo obucheniya [Methods for automating machine learning[. St. Petersburg, Universitet ITMO Publ., 2020. 40 p.

11. *Ajvazyan S.L., Enyukov I.S., Meshalkin L.D.* Prikladnaya statistika. Issledovanie zavisimostej [Applied statistics. Dependency research]. Moscow, Izdatel'stvo «Finansy i statistika» Publ., 1985. 488 p.