

ОЦЕНКА ИНВЕСТИЦИОННЫХ РИСКОВ НА ОСНОВЕ КЛАСТЕРНОГО АНАЛИЗА

2019 О. Ю. Лавлинская, Т. С. Гурьева

Воронежский институт высоких технологий (Воронеж, Россия)

В статье рассмотрены подходы к оценке рисков инвестиционных проектов для компаний-застройщиков на основе кластерного анализа. Интегральный показатель риска определяется как «центр тяжести» медиан полученных в результате расчета Евклидова расстояния между объектами-конкурентами в пространстве признаков, определяющих инвестиционное положение объектов. Рассчитывается интегральный показатель оценки проекта. Интегральный показатель сопоставляется со шкалой кластеризации, полученной при анализе рисков на рынке недвижимости. На основе сопоставления принимается решение об осуществлении проекта по застройке высотных домов.

Ключевые слова: кластерный анализ, Евклидово расстояние, принятие решения.

Анализ рисков является практическим инструментом получения информации, необходимой для принятия решения в различных сферах деятельности. Существуют различные методы анализа рисков: качественные и количественные. Качественные методы основываются на экспертных оценках, количественные методы требуют применения подходов, основанных на статистических выборках, больших объемах данных и методах машинного обучения, регрессионного и кластерного анализа.

В статье рассмотрим задачи кластерного анализа рисков для объектов анализа на данных, взятый из оценки евклидовой меры близости различных признаков и последующего их ранжирования на основе методов кластеризации. В качестве предметной области будем использовать строительную сферу, сегмент деятельности компаний застройщиков. Рассмотрим риски, которые существенны для Воронежа.

Группа 1 Экономические внешние риски

- Риск снижения покупательской способности населения

- Прогнозы ухудшения уровня жизни и экономического благосостояния страны и региона

- Ипотечные риски

- Риски недоступности ипотечных кредитов для различных категорий граждан

- Земельные риски

- Удорожание стоимости земельных участков под строительство

- Уровень инфляции

Группа 2 Правовые риски

- Правовые обеспечение

- Отсутствие необходимой правовой базы для развития высотного строительства (вопросы собственности земли и объектов недвижимости)

Группа 3. Инвестиционные риски

- Обеспеченность капиталом

- Рыночный риск

- Сложность выхода на новые рынки строительной продукции, динамика стоимости объектов недвижимости

- Временной риск

- Длительность возведения объекта

Группа 4. Технологические риски

- Проектные риски

- Сложность возведения

- Сложность обслуживания

Группа 5. Эксплуатационные риски

- Аварии и отказы систем

- Возгорание и задымление

- Взрывоопасность

Группа 6. Экологические риски

Для каждого вида риска в соответствии со статистическими данными [1], получены исходные множества за пять последних лет и произведен расчет статистической вероятности их возникновения для сегмента Воронежа и Воронежской области.

Далее, на основе методов машинного обучения проведена кластеризация рисков.

Кластеры сформированы таким образом, что в кластер попадают риски независимо от природы возникновения, но прибли-

Лавлинская Оксана Юрьевна – Воронежский институт высоких технологий, к. т. н., доцент, lavlin2010@yandex.ru.

Гурьева Татьяна Сергеевна – Воронежский институт высоких технологий, студентка магистратуры.

зительно одинаковые по количественным характеристикам [2]. Таким образом, выделены три основных кластера, которым даны новые групповые определения.

Кластеризацию объектов по уровням: умеренный риск при ($0 \leq p < 0.3$), допустимый риск ($0.3 \leq p < 0.6$), повышенный риск ($0.6 \leq p < 1$), где p – вероятность возникновения риска.

Интегральную оценку риска проекта проведем по следующему алгоритму.

1 этап. Подготовка исходных данных

В качестве исходных данных будем использовать матрицу статистических вероятностей возникновения по всем группам рисков, где каждому риску присвоен код в соответствии с группой и номером в группе, например, группа 6, подмножество 2, код – 62.

Проведена оценка шести групп рисков и рассчитана статистическая вероятность их возникновения для 15 исследуемых объектов. На рисунке 1 представлен результат статистической оценки по группам совокупности рисков

Вероятность						
0,8-1,0	11	12	43			
0,6-0,8	13				60	42
0,4-0,6						41
0,2-0,4	22			51		
0-0,2	23	31		33		
	0-20	20-40	40-60	60-80	80-90	90-100
	Последствия для объектов, %					

Рисунок 1. Матрица рисков.

Последствия риска – это цена, которую необходимо учесть при принятии решения по инициации проекта компанией застройщиков, если в проекте участвуют инвесторы, то они также должны оценивать риски для своих инвестиций.

Последствия оцениваются в процентах удорожания от стоимости проекта и вероятности возникновения. Эти два критерия влияют на результаты кластеризации.

2 этап. Кластеризация на основе метода НСА.

В области интеллектуального анализа данных и статистики иерархическая кластеризация (также называемая иерархическим кластерным анализом или НСА) является методом кластерного анализа, который направлен на построение иерархии кластеров [3]. Стратегии иерархической кластеризации обычно делятся на два типа:

Agglomerative: Это подход «снизу вверх»: каждое наблюдение начинается в своем кластере, а пары кластеров объединяются по мере движения вверх по иерархии.

Divisive: Это подход "сверху вниз" подход: все наблюдения начинаются в одном кластере, и расколы выполняются рекурсивно, как один движется вниз по иерархии.

В общем, слияния и «расколы» определяются жадным образом.

Результаты иерархической кластеризации обычно представлены в дендрограмме.

Расстояние между элементами представляет собой расстояния Евклида.

После выбора метрики расстояния необходимо определить, как вычисляется расстояние. Например, его можно вычислить между двумя наиболее похожими частями кластера (односвязная связь), двумя наименее похожими разделами кластера (полная связь), центром кластеров (средняя связь) или каким-либо другим критерием [3].

Кластеризацию выполним с помощью библиотеки Scipy.

Подготовка исходных данных проводилась следующим образом.

Была подготовлена матрица, которая содержит евклидовы расстояния для всех оцениваемых объектов по рискам и их последствиям. Вопрос получения значений евклидова расстояния подробно рассмотрены в [4].

На рисунке 2 показана матрица расстояний для элементов, градация которых осуществлялась по матрице, фрагмент которой представлен на рисунке 1 с учетом градации последствий для всех объектов.

```
[132]: import pandas as pd
import numpy as np
from scipy.cluster.hierarchy import dendrogram, ward, linkage

[133]: file = pd.read_csv('/www/ana/Documents/file2.csv')

[134]: data= file.set_index(' ')

[135]: data

[136]:
```

	101	102	103	211	212	213	221	222	231	\
101	0.00	0.00	1.00	56.30	56.30	49.00	1.00	9.00	36.00	
102	0.00	0.00	1.00	56.30	56.30	49.00	1.00	9.00	36.00	
103	1.00	1.00	0.00	42.30	42.30	36.00	4.00	16.00	25.00	
211	56.30	56.30	42.30	0.00	0.00	0.25	72.30	110.00	2.25	
212	56.30	56.30	42.30	0.00	0.00	0.25	72.30	110.00	2.25	
213	49.00	49.00	36.00	0.25	0.25	0.00	64.00	100.00	1.00	
221	1.00	1.00	4.00	72.30	72.30	64.00	0.00	4.00	49.00	
222	9.00	9.00	16.00	110.00	110.00	100.00	4.00	0.00	81.00	
231	36.00	36.00	25.00	2.25	2.25	1.00	49.00	81.00	0.00	
232	49.00	49.00	36.00	0.25	0.25	0.00	64.00	100.00	1.00	
233	72.30	72.30	56.30	1.00	1.00	2.25	90.30	132.00	6.25	
301	1.00	1.00	4.00	72.30	72.30	64.00	0.00	4.00	49.00	
302	2.25	2.25	0.25	36.00	36.00	30.30	6.25	20.30	20.30	
303	0.00	0.00	1.00	56.30	56.30	49.00	1.00	9.00	36.00	
304	49.00	49.00	64.00	210.00	210.00	196.00	36.00	16.00	169.00	
401	72.30	72.30	90.30	256.00	256.00	240.00	56.30	30.30	210.00	
402	12.30	12.30	20.30	121.00	121.00	110.00	6.25	0.25	90.30	
403	9.00	9.00	4.00	20.30	20.30	16.00	16.00	36.00	9.00	
501	81.00	81.00	100.00	272.00	272.00	256.00	64.00	36.00	225.00	
502	81.00	81.00	100.00	272.00	272.00	256.00	64.00	36.00	225.00	
503	36.00	36.00	49.00	182.00	182.00	169.00	25.00	9.00	144.00	
601	182.00	182.00	210.00	441.00	441.00	420.00	156.00	110.00	380.00	
602	25.00	25.00	16.00	6.25	6.25	4.00	36.00	64.00	1.00	
	232	-	303	304	401	402	403	501	502	\

Рисунок 2. Матрица расстояний.

Размерность матрицы определяется как произведение сегментов последствий наступления рисков на количество сегментов градации вероятности наступления риска и на количество рисков [5].

Таким образом, в рассматриваемом примере размерность матрицы 520x520.

Расстояние по Евклиду характеризует положение объекта в пространстве всех признаков, которые его характеризуют. Для оценки инвестиционных проектов мы оцениваем компании, которые являются застройщиками в строительном сегменте г. Воронежа, а признаки, которые характеризуют объект – это ранжированные риски с учетом последствий их возникновения для каждого конкретного объекта.

Результаты кластеризации представлены на рисунке 3.

По оси X мы видим индексы исходных данных, по оси Y мы видим евклидовы расстояния, шкалированные [5] таким образом, чтобы визуализация отображала кластеризацию. Иначе мы бы не разглядели «маленькие» элементы дендрограммы. Визуально видно, что существуют пересечения между сегментами и объекты выделены в различные диапазоны.

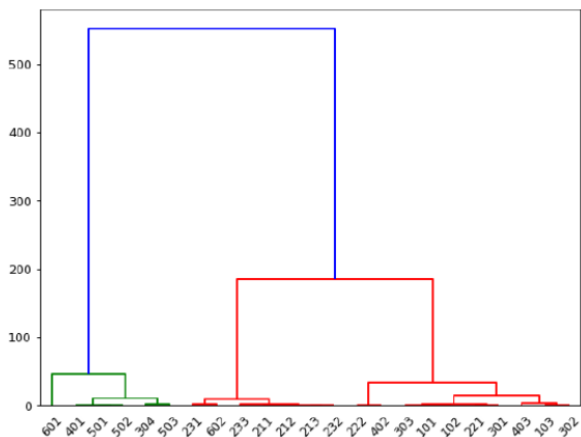


Рисунок 3. Результаты кластеризации для матрицы расстояний.

В отчете о результатах кластерного анализа представлены три группы рисков (синий – повышенный риск, красный – умеренный риск, зеленый – допустимый риск), объекты выделены в различные сегменты.

Определим расстояние между кластерами на основе полученных результатов (табл.)

Таблица
Матрица кластерных расстояний

Наименование кластеров	(1)	(2)	(3)
Умеренный (1)	0	22,5	38,5
Допустимый (2)	22,5	0	31,5
Повышенный (3)	38,5	31,6	0

Получим возможность оценки каждого отдельного объекта, компании застройщика по треугольнику рисков – геометрическая интерпретация [6].

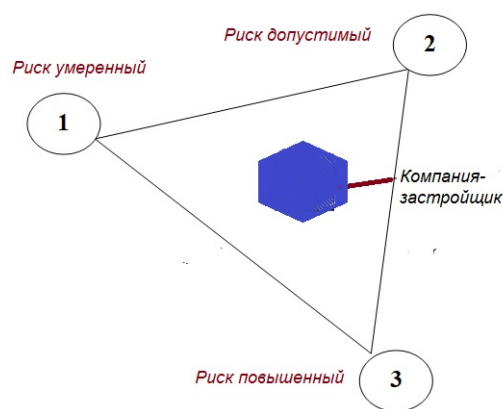


Рисунок 4. Геометрическая интерпретация положения компании в плоскости.

Геометрическая интерпретация ситуации рисков, представленная на рисунке 4, трактуется следующим образом. Объект находится в центре тяжести треугольника. Как известно, центр тяжести или центр масс – это точка пересечения медиан треугольника. Степень влияния каждого кластера рисков (вершины треугольника) определена геометрически как $2/3$ медианы.

Центр тяжести треугольника находится на пересечении медиан, причем медианы в этой точке имеют отношение 2:1. Следовательно, степень влияния кластера рисков на объект можно определить как $2/3$ длины медианы соответствующей кластерной вершины.

Таким образом положение объекта в плоскости треугольника и его интегральная оценка в проекции рисков, как формула вычисления «центра тяжести» дает возможность получить интегральную количественную оценку объекта и ранжировать все объ-

екты с точки зрения инвестиционной привлекательности.

ЛИТЕРАТУРА

1. Единый ресурс застройщиков (ЕРЗ) [электронный ресурс] \\<https://erzrf.ru> (дата обращения 02.12.2019)

2. Фам Хак Ксюань Модели управления процессами строительства высотных объектов в условиях рисков: дис. ... канд. техн. наук: 05.13.10. – ВГАСУ, 2013 - 180 с.

3. Терехина, А. Ю. Анализ данных методами многомерного шкалирования / А. Ю. Терехина. – М.: Наука, 1986. 168 с.

4. Библиотека ScyPy [электронный ресурс] \\<https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html>

5. Krause, Eugene F. (1987). Taxicab Geometry. Dover. ISBN 978-0-486-25202-5.

6. Эллерман, Д. Математические методы в оценке недвижимости / Д. Эллерман, // Институт Экономического Развития, Всемирный Банк. – М., 1994. – С. 35.

EXPERT INVESTMENT RISKS ON THE CLUSTER ANALYSIS

2019 O. U. Lavlinskaya, T. S. Guryeva,

Voronezh Institute of High Technologies (Voronezh, Russia)

The article considers approaches to assessing the risks of investment projects for developers based on cluster analysis. The integral risk indicator is defined as the "center of gravity" of the median resulting from the calculation of Euklidov's distance between the competing objects in the space of traits determining the investment position of the objects. An integral indicator of project evaluation is calculated. The integral indicator is compared to the clustering scale obtained in the real estate market risk analysis. On the basis of comparison, a decision is made on the implementation of the project for the development of high-rise buildings.

Keywords: cluster analysis, Euclied distance, decision-making.