

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

**ПОСТРОЕНИЕ ГРАФА НА ОСНОВЕ ПОКАЗАТЕЛЕЙ  
ОБАНКРОТИВШИХСЯ БАНКОВ РОССИИ И ЕГО АНАЛИЗ**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

Студентки 4 курса 451 группы

направления 38.03.05 — Бизнес-информатика

механико-математического факультета

Маловой Ольги Дмитриевны

Научный руководитель

доцент, к. ф.-м. н.

\_\_\_\_\_

О. А. Мыльцина

Заведующий кафедрой

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Саратов 2023

## ВВЕДЕНИЕ

**Актуальность темы.** Банки относятся к общественно значимым экономическим субъектам, поскольку они являются важными элементами инфраструктуры безналичных расчетов. В стабильности их работы заинтересован самый широкий круг: их клиенты, работники, государство, партнеры по бизнесу и социальным программам и др. Нарастание кризисных явлений в экономике, санкционные меры в отношении России влияют в том числе и на ответственный финансовый сектор. Возникает острая потребность в оценке банков как надежных бизнес-партнеров, гарантов сохранности денежных средств населения. Одним из основных источников информации для такой оценки являются результаты анализа их финансового состояния.

Период с 2008 по 2014 год стал переломным для российской банковской системы. Более трети всех банков обанкротилось. Претерпели дефолт даже такие крупные банки из десятки самых успешных, как «Открытие». Каждое банкротство приводит к потере вкладов для тысяч людей. Таким образом, крайне важно предсказать банкротство до его наступления и спасти банк, сделав его дефолт менее тяжелым. С помощью различных методов анализа финансовых показателей банка появляется возможность определения ухудшения финансовой ситуации.

Выполнить анализ взаимосвязанных характеристик можно с использованием графов, которые представляют собой мощный инструмент визуализации данных и связей в них. Кроме того, в ходе осуществления анализа графа можно вычислить характеристики функционирования рассматриваемой системы по присутствующим в них взаимосвязям. Значения параметров экономических показателей как узлов графа позволят создать основу для построения модели оценки состояния компании для своевременного устранения угроз.

**Целью бакалаврской работы** является выявление наиболее значимых финансовых показателей банков для определения их благонадежности.

**Объект исследования** - рынок коммерческих банков России.

**Предмет исследования** - набор данных о финансовом состоянии обанкротившихся российских банков в период 2010 - 2017 гг.

Для достижения поставленной цели в работе необходимо выполнить следующие задачи:

- определить основные понятия, связанные с финансовыми показателями деятельности банков;
- рассмотреть наиболее значимые показатели банков;
- определить основные понятия и характеристики, используемые в теории графов;
- изучить набор данных о российских банках;
- рассмотреть необходимые инструменты работы с данными с помощью графов;
- проанализировать различные вариации графа на основе имеющихся данных;
- выявить наиболее значимые характеристики с помощью построения моделей классификации.

**Практическая значимость** работы состоит в том, что на основании построения и анализа графа и регрессионных моделей по данным об экономических показателях российских обанкротившихся банков, можно проводить исследования современных банковских систем и выявлять их отклонения от экономического безопасного состояния. По результатам этих операций можно делать выводы об эффективности деятельности банка.

**Структура и содержание бакалаврской работы.** Работа состоит из введения, трех разделов, заключения, списка использованных источников, содержащего 21 наименование, на которые в тексте работы приведены ссылки и приложений, в которых приведены описание параметров используемого набора данных, результаты построения графов, вычисления их характеристик и код программы.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Введение** посвящено обоснованию актуальность выбранной темы работы, формулируется цель работы, приводится постановка необходимых задач, отмечается практическая значимость полученных результатов.

В **первом разделе** приводятся основные понятия, связанные с показателями деятельности банков. Банки относятся к категории общественно

значимых экономических субъектов, поскольку в стабильности их работы заинтересован самый широкий круг: их клиенты, работники, государство, партнеры по бизнесу и социальным программам и др. Регулярные обострения кризисных экономических ситуаций сильно влияют на банки, вплоть до полного прекращения их деятельности.

Для поддержания устойчивости банка его руководству необходима качественная информационная основа принятия управленческих решений. Одним из основных источников такой информации являются результаты анализа финансового состояния банков, которое характеризуется степенью обеспеченности достижения целей банков необходимыми финансовыми ресурсами.

Анализ состояния банка можно осуществлять основываясь на информации из открытых источников данных, таких как:

1. Публикуемая бухгалтерская отчетность банка, на основе которой можно оценить динамику его активов, пассивов, финансовых результатов, достаточность капитала, кредитный, рыночный риски, риск ликвидности.

2. Годовые отчеты, ежеквартальные отчеты, в которых характеризуются особенности корпоративного управления в банке, менеджмента, рыночное положение банка, приводятся его стратегические показатели, бухгалтерская (финансовая) отчетность, финансовые показатели дочерних банков и прочая важная для анализа информация.

Финансовые показатели деятельности банков можно разделить на несколько групп. Показатели *ликвидности* позволяют определить размер долговых обязательств фирмы, который может быть покрыт за счет денежных средств, высоко ликвидных ценных бумаг и депозитов а также, имеются ли у предприятия ресурсы, способные удовлетворить требования кредиторов в критической ситуации.

Показатели *финансовой устойчивости*, или благоприятной структуры капитала, отражают соотношение собственных и заемных средств в источниках финансирования компании, то есть характеризуют степень финансовой независимости компании от кредиторов. Показатели *рентабельности* определяют прибыльность компании. Показатели *деловой активности* определяют уровень использования компанией средств, полученной из различных источников.

**Второй раздел** содержит обзор основных понятий теории графов, а также характеристики графов. Кроме того, присутствует описание используемого набора данных, построение различных вариантов графов и вычисления указанных характеристик на языке  $R$  и проведен анализ полученных результатов.

Граф  $G$  можно определить как пару  $G = (V, E)$ , состоящую из множества узлов  $V(G)$  и множества ребер  $E(G)$ . Множество ребер  $E(G)$  индуцирует симметричное бинарное отношение на  $V(G)$ , которое называется отношением смежности  $G$ . Если  $e_{ij} \in E(G)$ , то ребро  $e$  соединяет вершины  $i$  и  $j$ , которые при этом называются смежными.

Если у ребер графа имеются направления от одной вершины к другой, то такой граф называется ориентированным. Ребра ориентированного графа называются дугами. Соответствующие узлы ориентированного графа называют началом и концом. Если направления ребер не указываются, то граф называется неориентированным (или просто графом). Граф, имеющий как ориентированные, так и неориентированные ребра, называется смешанным. Ориентированные графы широко используются для отображения связей в сетях.

Степенью  $d_i$  вершины  $i \in V$  в графе  $G = (V, E)$  называется число инцидентных ей ребер. Если  $d_i = 0$ , то вершина  $i$  называется *изолированной*, если  $d_i = 1$ , то вершина  $i$  называется *висячей*. Кроме того, в случае направленного графа различают степень  $d_i^-$  и степень  $d_i^+$  узла  $i$ .

Матрица смежности  $A(G)$  графа  $G$  представляет собой матрицу размерности  $n \times n$ , в которой запись  $a_{ij}$  равен 1, если ребро  $e_{ij} \in E(G)$ , в противном случае  $a_{ij}$  равен 0. Для неориентированного графа  $A$  симметрична, т. е.  $a_{ij} = a_{ji} \forall i, j \in V(G)$ .

Распределение степеней  $p_d$  показывает вероятность того, что случайно выбранный узел в графе имеет степень  $d$ . Поскольку  $p_d$  - это вероятность, она должна быть нормализована, т.е.  $\sum_{d=1}^{\infty} p_d = 1$

Для сети с  $N$  узлами распределение степеней в виде нормализованной гистограммы имеет вид  $p_d = \frac{N_d}{N}$ , где  $N_d$  - число узлов со степенью  $d$ . Следовательно, количество узлов степени  $d$  может быть получено из распределения степеней  $N_d = N p_d$ . Для экономических показателей распределение степеней

показывает насколько сильно они взаимосвязаны.

Распределение по степеням стало играть центральную роль в теории сетей после открытия безмасштабных сетей. Одна из причин заключается в том, что для расчета большинства свойств сети необходимо знать  $p_d$ . Например, средняя степень сети может быть записана как  $\langle d \rangle = \sum_{d=0}^{\infty} dp_d$ .

**Центральность вершин** в графе – это вектор, сопоставляющей каждой вершине графа некоторое число (индекс). Центральность измеряет важность узла на основе его положения в сети. Степень центральности это количество ссылок или связей, выпадающих на узел. Существуют разные подходы расчёта индексов центральности в графе. Каждый индекс имеет свою интерпретацию. Существует несколько мер центральности:

- центральность по *степени*;
- центральность по *близости*;
- центральность по *посредничеству*;
- центральность по *собственному вектору*.

Для изучения темы был выбран набор данных о 1000 российских банках, который содержит значения финансовых показателей и коэффициентов за разные отрезки времени. Всего данные содержат 70483 наблюдения, для каждого определено 44 признака. Данные собраны за период с 2010 по 2017 год.

Основными признаками объектов являются дата наблюдения, номер лицензии, обанкротился ли объект, размер чистых активов, рентабельность активов, объемы кредитов и вкладов организаций, физических лиц, размер кредита самих банков, наличие иностранных активов. Кроме того, в наборе представлены данные о работе с акциями, облигациями и другими ценными бумагами. Также набор данных содержит такие макроэкономические показатели, как курс доллара к рублю, размер ВВП, значение ключевой ставки Центробанка. Данные собраны на основе финансовой отчетности банков. Такой широкий ряд критериев позволяет составить полную картину деятельности банка. Среди них такие критерии, как: рентабельность активов, коэффициенты ликвидности, уровень ключевой ставки, кредитный резерв.

Для построения графа были выявлены взаимозависимости критериев и построена тепловая карта на основе корреляционной матрицы. Корреля-

ция - важнейший фактор, лежащий в основе анализа данных, показывающий степень связи между параметрами и ее вид. Для построения различных вариаций графов были выбраны пороговые значения корреляции, при достижении которых параметр попадал в рассмотрение. Такими значениями были 0, 0,3, 0,5 и 0,8.

Для полученных графов были вычислены их основные характеристики и рассмотрено их изменение. В результате проделанных операций были сделаны промежуточные выводы о значимости и взаимосвязи определенных параметров.

**Третий раздел** посвящен выявлению значимых показателей банков с помощью моделей классификации на языке *R*. Рассмотрены два метода построения моделей, их статистические параметры. С использованием имеющихся данных были построены различные варианты моделей, сделан вывод о значимых объясняющих признаках.

Задача предсказания относится к задаче классификации. В общем смысле алгоритм классификации – это моделирование, цель которого разделить классы данных на основе некоторого множества входных признаков, отвечая на вопрос «какого вида».

Формальная постановка задачи классификации звучит следующим образом. Предположим, что имеется  $n$  объектов, каждый из которых описывается  $m$  признаками. Объекты будут нумероваться индексом  $i = 1, \dots, n$ , а признаки (значения которых могут быть получены непосредственным измерением) - индексом  $j = 1, \dots, m$ . Для объекта с номером  $i$  обозначим через  $x_{ij}$  - значение признака  $j$ , а  $y_i$  - значение зависимого объекта  $i$ .

Для рассматриваемых данных объектами будут состояния банков в определенный момент времени. Наблюдаемый признак  $x$  - финансовые показатели банка, зависимым, или целевым, признаком  $y$  будет *default*. Необходимо спрогнозировать, потерпит ли дефолт банк при том или ином состоянии.

**Задача** состоит в том, чтобы определить наиболее важные факторы, влияющие на состояние банка и сформулировать правило, позволяющее по заданным значениям признаков определить наиболее вероятное значение переменной отклика (то есть наиболее вероятное состояние банка).

Для того, чтобы определить, какие переменные должны быть включены

в модель множественной регрессии, проводится содержательный анализ, в ходе которого факторы рассматриваются на качественном уровне.

Предварительная обработка данных включает в себя проверку типов данных, распределения значений, поиск выбросов в рассматриваемых параметрах. Проведено выявление пропущенных значений в каких-либо параметрах, а строки, имеющие их, были удалены. В ходе анализа плотности распределения данных выяснилось, что практически все параметры не имеют нормального распределения, что обусловлено смысловой составляющей факторов и их изменений, а также методами их расчетов.

Рассматриваемый набор данных имеет некоторые особенности. Во-первых, он содержит большое количество наблюдений (70483), что делает работу с ними более долгой и сложной. В связи с этим было отобрано 70% банков, наблюдения для которых использовались в построении модели. Полученные результаты проверяются также и на второй части данных. Во-вторых, данные не являются сбалансированными по признаку *default*, что представляет сложность для построения его прогнозных значений, однако не мешает выявить значимые в прогнозировании факторы. В-третьих, набор данных содержит большое количество признаков, простой подбор факторов не является эффективным. Поэтому подбор объясняющих переменных модели будет осуществляться на основе анализа графов.

**Множественная регрессия.** Для набора признаков  $X = (X_1, X_2, \dots, X_m)$ , состоящего из  $m$  объясняющих переменных уравнение множественной линейной регрессии будет выглядеть следующим образом:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon, \quad (1)$$

где  $y$  - зависимая переменная;

$x_i, i = 1, 2, \dots, m$  - объясняющие переменные (регрессоры);

$\beta_i, i = 1, 2, \dots, m$  - коэффициенты регрессии, показывают на сколько единиц в среднем изменится значение при увеличении значения фактора  $x_i$  на одну единицу измерения данного фактора;

$\beta_0$  - свободный член уравнения, коэффициент, которому будет равна объясняемая переменная при отсутствии воздействия на нее других факторов;

$\varepsilon$  - случайные ошибки.

Уравнение линейной регрессии описывает совокупное влияние изменения множества объясняющих переменных на результирующий признак [?].

В отличие от линейной регрессии, логистическая регрессия вместо того, чтобы предсказывать значения  $Y$  в зависимости от значений  $X$ , предсказываются вероятность  $Y$  при реализации тех или значений  $X$ . Используется для предсказания категориальных значений, а ее уравнение подбирается методом максимального правдоподобия. Уравнение логистической регрессии имеет вид:

$$P(Y) = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m)}}. \quad (2)$$

Множественная регрессия на языке  $R$  строится с помощью функций  $lm()$  или  $glm()$ . Первая из указанных функций используется для подбора линейной регрессии, вторая же создана для подбора обобщенных регрессионных моделей. Для функции  $glm()$  можно указать статистическое семейство модели, она используется для создания более сложной и точной модели.

Для рассматриваемого набора данных выбрана функция  $glm()$  с параметром семейства *binomial*, то есть для построения биномиальной логистической регрессии.

Модель улучшалась путем отбора наиболее значимых переменных. Были построены и проанализированы варианты модели, включающие как низкоррелирующие параметры (то есть не входящие в граф), так и высокоррелирующие (то есть входящие в граф). Итоговая модель включала в себя параметры обеих категорий.

В улучшенной версии модели все объясняющие переменные являются значимыми, для всех из них  $p$ -значение меньше 0,05. Значения  $\chi^2$  и информационного критерия улучшились по сравнению с промежуточными вариантами моделей, однако не являются идеальными, что может быть обусловлено преобладанием одного из классов.

Основным результатом построения моделей является определение наиболее значимых финансовых показателей банков. Ими оказались:

— *рентабельность активов* - характеризует эффективность использова-

- ния активов всех видов, показывает способность банка генерировать прибыль без учета структуры его капитала;
- *показатель абсолютной ликвидности* - отображает платежеспособность компании, оценивает возможность закрытия долгов собственными средствами организации;
  - *кредитный резерв* - позволяет гарантировать выплаты клиентам в случае возникновения финансовых проблем у банка;
  - *кредитный залог* - показывает объем средств, которые могут вернуть банку при невыплате кредитов;
  - *коэффициент достаточности собственного капитала* - оценивает способность банка обеспечивать свою деятельность средствами основного капитала
  - *расположение в Москве или Санкт-Петербурге* - функционирование в развитых районах является более надежным;
  - *ключевая ставка* - макроэкономический фактор, оказывает прямое влияние на уровень экономической активности

Таким образом, наиболее значимыми признаны факторы, раскрывающие надежность функционирования банка с разных сторон, учтены как экономические показатели самих организаций, так и внешние факторы. В модели не были отобраны показатели объемов ценных бумаг в портфеле, это может быть связано с тем, что они имеют не первостепенную роль в деятельности банка, а значит не оказывают сильного влияния на благосостояние. Были выбраны значимыми критерии, составляющие основу в результатах моздания финансовой отчетности. Указанные критерии в действительности имеют большое влияние на финансовую устойчивость банка, а значит требуют постоянного отслеживания и тщательного контроля.

Часто при построении моделей множественной регрессии недостаточно просто выявить значимые параметры, важно также провести анализ этих параметров для выявления связей на более глубоком уровне, а значит, и получения более полного и эффективного решения поставленных задач. Для контроля значений показателей также важно знать предпосылки их изменения.

Далее было изучено, с помощью каких параметров можно предсказать

уровень некоторых из указанных выше финансовых показателей банков. Эти объясняемые переменные, в отличие от уже рассмотренной *default*, принимают непрерывные значения, поэтому необходимо построить модель линейной регрессии. Для этого использована функция *glm()* с параметром семейства, равном *gaussian*, то есть модель регрессии гауссовского процесса, при котором наблюдаемые значения случайных величин изменяются со временем.

Были выявлены наиболее значимые факторы, влияющие на уровень рентабельности активов и абсолютной ликвидности, так как они являются ключевыми финансовыми показателями. Таким образом, в ходе работы был создан подход к прогнозированию неблагонадежных состояний банка, при котором строится модель как для выявления значимых факторов, так и для выявления факторов, влияющих на них.

В **заключении** приведены результаты бакалаврской работы.

### Основные результаты

1. Определены основные показатели финансово-хозяйственной деятельности банков. Рассмотрены основные современные тенденции их изменения.

2. Изучены основные понятия теории графов. Проведено построение графов, вычислены и проанализированы их основные характеристики.

3. Изучены основные понятия, связанные с методами бинарной классификации, а также алгоритмы их построения.

4. Построена модель классификации с помощью метода логистической регрессии, в результате чего выявлены и проанализированы наиболее значимые финансовые показатели банков в оценке их благонадежности.

5. Построены модели линейной регрессии для определения факторов, наиболее влияющих на уровень ключевых финансовых показателей, полученных при построении логистической регрессии.

6. Разработан подход к анализу данных, имеющих большое количество признаков и наблюдений, начальный этап которого реализуется с использованием теории графов.