

5. Khodan L. A. The legal basis of the risk-based approach in the implementation of budget control. Master's thesis. Moscow. The Institute of Legislation and Comparative Law under the Government of the Russian Federation, 2018. (In Russ.).

6. GOST R 51897-2011/ISO Guide 73:2009 Risk management. Terms and definitions. Approved and put into effect by the Federal Agency for Technical Regulation and Metrology dated November 16, 2011 No. 548-art. (In Russ.).

7. GOST R 50779.10-2000 (ISO 3534-1-93). Statistical technology. Probability and fundamentals of statistics. Terms and definitions. / Adopted and put into effect by the resolution of the state standard of Russia dated December 29, 2000 No. 429-art. (In Russ.).

8. Pyt'ev Yu. P. The possibility. Elements of theory and application. Moscow. Editorial URSS, 2000. 192 p. (In Russ.).

9. Krichevsky M. L. Financial risks. 2nd ed. Moscow, KNORUS Publ., 2017. 244 p. (In Russ.).

10. GOST R ISO 31000-2010. Risk management. Principles and guidance. Approved and put into effect by the Order of the Federal Agency for Technical Regulation and Metrology dated December 21, 2010 No. 883-art. (In Russ.).

11. The Budget Code of the Russian Federation dated July 31, 1998 No. 145-FZ. *Collection of legislation of the Russian Federation*, 1998, no. 31, article. 3823. (In Russ.).

12. Povetkina N. A. On formation of conceptual framework of the risk-oriented approach in the financial and budgetary control. *Financial law*, 2015, no. 12, pp. 15–17. (In Russ.).

Как цитировать статью: Ходан Л. А. Сущностные основы риска // Бизнес. Образование. Право. 2019. № 2 (47). С. 255–262. DOI: 10.25683/VOLBI.2019.47.231.

For citation: Khodan L. A. Essential basis of risk. *Business. Education. Law*, 2019, no. 2, pp. 255–262. DOI: 10.25683/VOLBI.2019.47.231.

УДК УДК 336
ББК 65

DOI: 10.25683/VOLBI.2019.47.271

Chelyshev Dmitriy Sergeevich,
post-graduate of program 08.00.13 “Mathematical
and instrumental methods in economics”,
Department of Mathematical Methods in Economics,
Plekhanov Russian
University of Economics,
Moscow,
e-mail: mr.chelyshev@gmail.com

Чельшев Дмитрий Сергеевич,
аспирант направления 08.00.13 «Математические
и инструментальные методы экономики»,
кафедра математических методов в экономике,
Российский экономический
университет им. Г. В. Плеханова,
г. Москва,
e-mail: mr.chelyshev@gmail.com

МОДЕЛИРОВАНИЕ ВЕРОЯТНОСТИ ДЕФОЛТА РОССИЙСКИХ БАНКОВ

MODELLING OF RUSSIAN BANKS' PROBABILITY OF DEFAULT

08.00.13 – Математические и инструментальные методы экономики
08.00.13 – Mathematical and instrumental methods in economics

В данной статье рассмотрен подход к моделированию вероятности дефолта банков на основе их внутренних показателей отчетности с помощью модели случайного леса. В качестве показателей были использованы показатели из годовых отчетов российских банков из разделов «Бухгалтерский баланс» и «Отчет о движении денежных средств». Для оценки вероятности дефолта российских банков были использованы данные долгосрочного международного рейтинга в иностранной валюте Moody's Investors Services. Кредитные рейтинги представляют собой мнения о кредитном риске. Рейтинг выражает мнение агентства о способности и готовности эмитента, например корпорации, государства или муниципального образования, выполнять финансовые обязательства своевременно и полностью. Каждое рейтинговое агентство имеет свою методологию оценки кредитного рейтинга, однако они, как правило, носят описательный характер, то есть невозможно сказать, какой фактор имеет большее влияние на фи-

нансовую устойчивость банка. Не все методы классификации могут быть применены на исходной выборке ввиду того, что количество банков, которые имеют кредитный рейтинг, ограничено. Это накладывает ограничение на методы, в основе которых лежит алгоритм обучения, так как многие из них подразумевают использование большого массива данных (более 10 000 объектов). Модель случайного леса позволяет выявить, какие факторы из внутренних показателей отчетности российских банков имеют наибольшее влияние на формирование кредитного рейтинга. Модель случайного леса в ходе вероятностного моделирования показала высокую предсказательную способность. Точность модели случайного леса на валидационной выборке выше, чем модели, построенной с помощью метода опорных векторов.

This article discusses an approach to modeling the probability of banks default based on their internal reporting indicators using the random forest model. Figures from

the annual reports of the Russian banks in the balance sheet section and cash flow statement section were used as indicators for the model. In order to assess the probability of default of Russian banks, Moody's Investors Services's long-term international foreign currency rating was used. Credit ratings are opinions about the credit risk. The rating expresses the opinion of agencies on the ability and readiness of the issuer, for example, a corporation, a state or a municipality, to fulfill financial obligations in a timely and complete manner. Each rating agency has its own methodology of assessment of the credit rating; however, it has, as a rule, a descriptive nature, and as a result, it is impossible to say, which factor has a greater impact on the financial stability of the bank. Not all classification methods can be applied to the testing sample due to the fact that the number of banks that have a credit rating is limited. This imposes a restriction on the methods based on the learning algorithm, since many of them require the use of a large data array (more than 10,000 objects). The random forest model makes it possible to identify, which factors from the reporting internal indicators of the Russian banks have the greatest influence on the formation of a credit rating. The random forest model showed a high predictive ability. The accuracy of the random forest model for the validation sample is higher than the model constructed using the support vector machine.

Ключевые слова: оценка дефолта банка, обучение с учителем, классификация, модель случайного леса, деревья решений, финансовая устойчивость, государственные российские банки, частные российские банки, методы многомерного статистического анализа, кредитный рейтинг.

Key words: bank's default estimation, supervised training, classification random forest model, decision tree, financial sustainability, public Russian banks, private Russian banks, multi-variable analysis, credit rating.

Введение

Прогнозирование кредитного риска является одной из актуальных тем в современной финансовой сфере, поскольку банковский регулятор увеличивает значимость внутренних и внешних кредитных рейтингов.

Процесс кредитного скоринга очень важен для инвесторов в процессе разграничения «хороших» и «плохих» банков в терминах вероятности их дефолта.

Статистические методы кредитного скоринга созданы для предсказания вероятности того, что банк не сможет выполнить финансовые обязательства перед своими заемщиками. Исторически дискриминантный анализ и логистическая регрессия были широко использованы с целью построения системы оценки банков, см. Хэнд и Хэнли [1], и Хэнд и др. [2]. Альтман [3] был первым, кто предложил использовать статистические методы для предсказания вероятности дефолта организации, рассчитывая его оценку с помощью стандартного дискриминантного анализа. Спустя десять лет Альтман [4] модифицировал его оценку путем расширения размера обучающей выборки. Кроме этого базового метода, более точные методы, такие как логистическая регрессия, нейронные сети и экспертные системы, были разработаны и широко применяются как для практических, так и для теоретических целей анализа вероятности дефолта организации [5; 6].

Цель данной работы — построение классификационного правила, позволяющего определить вероятность дефолта российских банков.

Задачи данной работы:

- отбор факторов, влияющих на финансовую устойчивость российских банков;
- классификация государственных и частных российских банков на основе экономических показателей ввнутренней отчетности;
- сравнение методов классификации на выборке российских банков.

Научная новизна исследования заключается в разработке оригинальных подходов к оценке кредитного рейтинга российских банков на основе модели случайного леса.

Практическая и теоретическая значимость заключается в том, что предложенные оценки кредитных организаций на основе строго математического аппарата, в отличие от оценок рейтинговых агентств, в которых учитываются в некотором смысле субъективные мнения экспертов, могут быть взяты за основу в принятии решений по инвестированию средств.

Анализируемая выборка

Для оценки финансовой устойчивости российских банков в данной работе были выбраны данные долгосрочного международного рейтинга в иностранной валюте Moody's Investors Services.

В качестве показателей были использованы показатели из годовых отчетов банков. Показатели бухгалтерского баланса:

- активы: денежные средства, средства кредитных организаций в Центральном банке Российской Федерации и др.;
- пассивы: кредиты, депозиты и прочие средства Центрального банка Российской Федерации, средства кредитных организаций и др.;
- источники собственных средств: средства акционеров (участников), собственные акции (доли), выкупленные у акционеров (участников), и др.;
- внебалансовые обязательства: безотзывные обязательства кредитной организации, выданные кредитной организацией гарантии и поручительства, условные обязательства некредитного характера.

Показатели для моделирования из отчета о движении денежных средства:

- процентные доходы всего, в том числе: от размещения средств в кредитных организациях, от ссуд, предоставленных клиентам (некредитным организациям), от оказания услуг по финансовой аренде (лизингу), от вложений в ценные бумаги;
- процентные расходы всего, в том числе: по привлеченным средствам кредитных организаций, по привлеченным средствам клиентов (некредитных организаций), по выпущенным долговым обязательствам;
- чистые процентные доходы (отрицательная процентная маржа);
- чистые доходы от операций с ценными бумагами;
- чистые доходы от операций с иностранной валютой;
- чистые доходы от переоценки иностранной валюты;
- комиссионные доходы;
- комиссионные расходы и др.

Были собраны годовые данные по 81 банку, которые получали долгосрочный международный рейтинг в иностранной валюте Moody's Investors Service за период с 2011 по 2017 г. [7].

Модель случайного леса

Основная идея модели случайного леса заключается в использовании большого комитета решающих деревьев, каждое из которых само по себе дает очень невысокое качество классификации, но за счет их большого количества точность результата значительно возрастает [8; 9].

Предположим, что в выборке A присутствуют объекты разных типов. Доля объектов типа i в выборке A равна f_i .

Энтропия определяется как (при соглашении $0 \cdot \ln = 0$)

$$I_E = -\sum f_i \ln f_i \quad (1)$$

Энтропия измеряет неоднородность выборки. Энтропия будет максимальной, если все типы равновероятны. Если же все объекты принадлежат одному типу, то энтропия равна нулю.

При каждом ветвлении дерево разбивает выборку на несколько частей (обычно на две): A_1, A_2, \dots энтропия для выборки, разбитой на несколько частей, определяется как взвешенные по частям:

$$I_E(Tree) = \sum_j P(A_j) I_E(A_j), \quad (2)$$

где $P(A_j)$ размер выборки A_j деленный на общее число наблюдений.

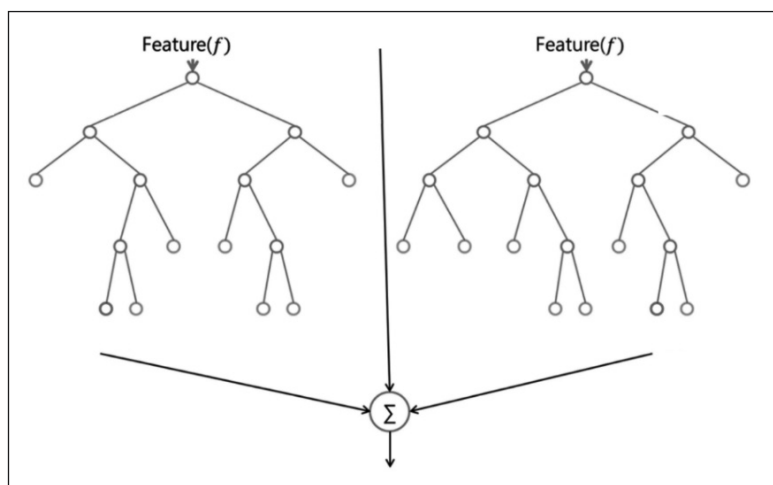


Рис. 1. Пример двух деревьев решений, формирующих случайный лес

Алгоритм формирует дерево так, чтобы падение энтропии было максимальным на каждом шаге. Проблема данного алгоритма состоит в том, что на обучающей выборке он показывает очень хороший результат, но при попытке прогнозирования за пределами обучающей выборки качество прогнозов становится плохим. Для решения данной проблемы применяется метод случайного леса.

Допустим, в общей выборке всего n наблюдений. Модель случайного леса строит из 500 деревьев [10]. Для каждого дерева:

- случайным образом с повторениями выбирается n наблюдений из m исходных. Поскольку наблюдения выбира-

ются с повторениями, то при построении каждого конкретного дерева часть наблюдений не будет использоваться, а часть наблюдений будет использоваться несколько раз;

- при делении каждой веточки на две случайным образом выбирается два регрессора из множества всех регрессоров, а уже затем из них выбирается переменная, использование которой дает наибольшее падение коэффициента энтропии.

Модель случайного леса относится к семейству робастных методов, ввиду этого из обучающей выборки требуется исключить коррелированные признаки. Исключим факторы с корреляцией более 75 % и построим корреляционную матрицу исходного пространства (табл. 1).

Таблица 1

Корреляционная матрица признаков

	X1	X2	X3	X4
X1	1	0,39	0,09	-0,49
X2	0,39	1	0,1	-0,57
X3	0,09	0,1	1	-0,06
X4	-0,49	-0,57	-0,06	1

Источник: [11–13].

Факторы, которые были использованы в качестве входных данных в модели случайного леса:

- X1 — неиспользованная прибыль (убыток) за отчетный период;
- X2 — процентные доходы от размещения средств в кредитных организациях;
- X3 — комиссионные расходы;
- X4 — изменение резерва на возможные потери.

Модель обучалась на выборке, равной 80 % объема на-

чальной выборки банков. Валидационная выборка выбиралась случайным образом из начальной выборки.

Результаты

Наибольшее влияние на формирование рейтинга (рис. 2 на стр. 253) оказал фактор X3 — комиссионные расходы, наименьшее влияние оказали факторы X1 — неиспользованная прибыль (убыток) за отчетный период и X4 — изменение резерва на возможные потери [14; 15].

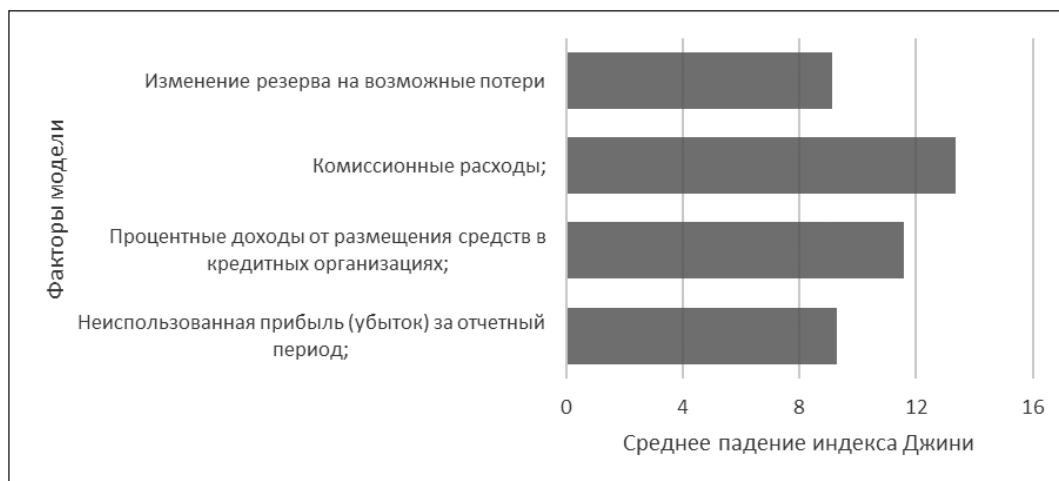


Рис. 2. Степень влияния факторов модели

Сравним предсказанные моделью рейтинги с фактическими данными (табл. 2). Точность модели на валидационной выборке составила 71 %. При этом прогноз и фактическое значение по банкам «Еврофинанс Моснарбанк» и «Кредит Европа Банк» находятся в одной подгруппе рейтинга — подгруппа В. С учетом этого точность модели в разрезе подгрупп составляет 86 %.

Данный метод позволил повысить точность прогноза, который был получен автором в статье «Применение метода опорных векторов для классификации российских банков» [16]. Рассмотрим алгоритм реализации метода опорных векторов. Основная идея метода опорных векторов — перевод исходных векторов в пространство

более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора. Метод опорных векторов позволил получить точность предсказания на уровне 29 % в разрезе рейтинга и 79 % в разрезе подгрупп соответственно.

Таблица 2

Сравнение фактических данных и прогноза

Наименование банка	Фактические данные	Прогноз	Прогноз совпал с фактическими данными
Акбарс Банк	B2	B2	TRUE
Акибанк	B3	B3	TRUE
Банк «Зенит»	B1	Ba2	FALSE
Банк «Уральский финансовый дом»	B3	B3	TRUE
ВТБ	Ba2	Ba2	TRUE
ВТБ 24	Ba2	Ba2	TRUE
Еврофинанс Моснарбанк	B1	B3	FALSE
Кредит Европа Банк	B1	B2	FALSE
МетКомБанк	B3	B3	TRUE
НК Банк	B3	B3	TRUE
Новикомбанк	B2	Ba2	FALSE
РГС Банк	B2	B2	TRUE
Россельхозбанк	Ba2	Ba2	TRUE
Транскапиталбанк	B1	B1	TRUE

Заключение

В статье был рассмотрен подход к оценке вероятности дефолта российских банков на основе модели случайного леса. В качестве показателя вероятности дефолта российских банков были использованы кредитные рейтинги, которые присваиваются международным рейтинговым агентством Moody’s Investors Services. Метод случайного леса характеризуется высокой точностью прогноза и позволил улучшить результат, полученный

с помощью метода опорных векторов. Модель была построена на основе четырех факторов бухгалтерской отчетности российских банков. Ввиду того, что рейтинг дефолтности может быть объяснен при помощи небольшого количества факторов, а также из-за потенциальной ненадежности рейтинговых агентств целесообразно включить большее количество факторов для повышения надежности оценки для принятия решений потенциальными инвесторами.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Hand D. J., Henley W. E. Statistical Classification Methods in Consumer Credit Scoring: A Review // *Journal of Royal Statistical Society*. 1997. Vol. 160. Pp. 523–541.
2. Hand D. J. Data Mining: New Challenges for Statisticians // *Social Science Computer Review*. Vol. 18. No. 4. Pp. 442–449.
3. Edward I. Altman Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy // *The Journal of Finance*. 1968. Vol. 23. No. 4. Pp. 589–609.
4. Edward I. Altman, Robert G. Haldeman, P. Narayanan. ZETATM analysis A new model to identify bankruptcy risk of corporations // *Journal of Banking & Finance*. 1977. Vol. 1. Iss. 1. Pp. 29–54.
5. Бирюков А. Н. Разработка показателей авторской модели в методике оценки угрозы банкротства предприятия // *Современные научные исследования и разработки*. 2018. № 1 (18). С. 55–65.
6. Lepetit L., Strobel F. Bank insolvency risk and Z-score measures: A refinement // *Finance Research Letters*. 2015. Vol. 13. Pp. 214–224. doi: 10.1016/j.frl.2015.01.001
7. Информационное агентство Spark. URL: <http://www.spark-interfax.ru/>
8. Use of Random forest in the identification of important variables / B. P. O. Lovatti, M. H. C. Nascimento, Á. C. Neto, E. V. R. Castro, P. R. Filgueiras // *Microchemical Journal*. 2019. Vol. 145. Pp. 1129–1134.
9. Random Forests for Big Data / R. Genuera, J.-M. Poggib, Ch. Tuleau-Malot, N. Villa-Vialaneix // *Big Data Research*. 2017. Vol. 9. Pp. 28–46.
10. Пальмов С. В., Денискова А. О. Случайный лес: основные особенности // *Наука сегодня: теоретические и практические аспекты : материалы международной науч.-практич. конф.* 2017. С. 51–53.
11. Статистический пакет caret. URL: <https://cran.r-project.org/web/packages/caret/index.html>
12. Статистический пакет corrplot. URL: <https://cran.r-project.org/web/packages/corrplot/index.html>
13. Тихомиров Н. П., Тихомирова Т. М., Ушмаев О. С. Методы эконометрики и многомерного статистического анализа. М. : Экономика, 2011. 647 с.
14. Статистический пакет randomForest. URL: <https://cran.r-project.org/web/packages/randomForest/index.html>
15. Статистический пакет e1071. URL: <https://cran.r-project.org/web/packages/e1071/index.html>
16. Чельшев Д. С. Применение метода опорных векторов для классификации российских банков // *News of Science and Education*. 2018. Т. 10. № 3. С. 21–26.

REFERENCES

1. Hand D. J., Henley W. E. Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of Royal Statistical Society*, 1997, 160, pp. 523–541.
2. Hand D. J. Data Mining: New Challenges for Statisticians. *Social Science Computer Review*, 18 (4), pp. 442–449.
3. Edward I. Altman Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 1968, 23 (4), pp. 589–609.
4. Edward I. Altman, Robert G. Haldeman, P. Narayanan. ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1977, 1 (1), pp. 29–54.
5. Biryukov A. N. Development of indicators of the author’s model in the methodology for assessing the threat of bankruptcy of an enterprise. *Modern research and development*, 2018, 1 (18), pp. 55–65. (In Russ.).
6. Lepetit L., Strobel F. Bank insolvency risk and Z-score measures: A refinement. *Finance Research Letters*, 2015, vol. 13, pp. 214–224. doi: 10.1016/j.frl.2015.01.001
7. Information Agency Spark. (In Russ.). URL: <http://www.spark-interfax.ru/>
8. Lovatti B. P. O., Nascimento M. H. C., Neto Á. C., Castro E. V. R., Filgueiras P. R. Use of Random forest in the identification of important variables. *Microchemical Journal*, 2019, vol. 145, pp. 1129–1134.
9. Genuera R., Poggib J.-M., Tuleau-Malot Ch., Villa-Vialaneix N. Random Forests for Big Data. *Big Data Research*, 2017, vol. 9, pp. 28–46.
10. Palmov S. V., Deniskova A. O. Random forest: key features. *Science today: theoretical and practical aspects*. Materials of the int. sci. and pract. conf. 2017, pp. 51–53. (In Russ.).
11. Statistical package caret. (In Russ.). URL: <https://cran.r-project.org/web/packages/caret/index.html>
12. Statistical package corrplot. (In Russ.). URL: <https://cran.r-project.org/web/packages/corrplot/index.html>
13. Tihomirov N. P., Tihomirova T. M., Ushmaev O. S. Methods of econometrics and multidimensional statistical analysis. Moscow, Ekonomika Publ., 2011. 647 p. (In Russ.).
14. Statistical package randomForest. (In Russ.). URL: <https://cran.r-project.org/web/packages/randomForest/index.html>
15. Statistical package e1071. (In Russ.). URL: <https://cran.r-project.org/web/packages/e1071/index.html>
16. Chelyshev D. S. Application of support vectors machine for classification of the Russian banks. *News of Science and Education*, 2018, 10 (3), pp. 21–26. (In Russ.).

Как цитировать статью: Чельшев Д. С. Моделирование вероятности дефолта российских банков // *Бизнес. Образование. Право*. 2019. № 2 (47). С. 262–266. DOI: 10.25683/VOLBI.2019.47.271.

For citation: Chelyshev D. S. Modelling of Russian banks’ probability of default. *Business. Education. Law*, 2019, no. 2, pp. 262–266. DOI: 10.25683/VOLBI.2019.47.271.