

Интернет-журнал «Наукоедение» ISSN 2223-5167 <https://naukovedenie.ru/>

Том 9, №6 (2017) <https://naukovedenie.ru/vol9-6.php>

URL статьи: <https://naukovedenie.ru/PDF/82EVN617.pdf>

Статья опубликована 01.02.2018

Ссылка для цитирования этой статьи:

Перерва О.Л., Степанов С.Е., Незимова С.С. Сравнение эконометрических моделей и методов бизнес-аналитики предсказания банкротства предприятий // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 9, №6 (2017) <https://naukovedenie.ru/PDF/82EVN617.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

УДК 330.4+519.86

Перерва Ольга Леонидовна

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
Калужский филиал, Россия, Калуга¹
Заместитель директора
Доктор экономических наук, профессор
E-mail: pol@bmstu-kaluga.ru
РИНЦ: http://elibrary.ru/author_profile.asp?id=406482

Степанов Сергей Евгеньевич

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
Калужский филиал, Россия, Калуга
Декан факультета
Кандидат физико-математических наук, доцент
E-mail: stepanov@bmstu-kaluga.ru
РИНЦ: http://elibrary.ru/author_profile.asp?id=9853
Researcher ID: <http://www.researcherid.com/rid/U-9190-2017>
SCOPUS: <http://www.scopus.com/authid/detail.url?authorId=8328785000>

Незимова София Сергеевна

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
Калужский филиал, Россия, Калуга
Магистрант
E-mail: stepanova_s@adm.kaluga.ru

Сравнение эконометрических моделей и методов бизнес-аналитики предсказания банкротства предприятий

Аннотация. В статье проведено сравнение классических эконометрических моделей банкротства предприятия и моделей, получаемых с использованием современных методов бизнес-аналитики, таких как логистическая регрессия, случайные леса, машины опорных векторов.

Авторами рассматриваются пять эконометрических моделей, как зарубежных, так и отечественных, и три метода бизнес-аналитики – логистическая регрессия, а также случайные леса и машины опорных векторов, являющиеся одними из наиболее точных методов

¹ 248000, г. Калуга, ул. Баженова, д. 2

машинного обучения. Во втором случае задача прогнозирования банкротства представляет собой одну из задач машинного обучения с учителем – задачу бинарной классификации.

Сравнение моделей проводилось авторами на данных о более чем десяти тысячах польских предприятий. Выбор этого набора объясняется тем, что и Польша, и Россия относятся к странам с переходными экономиками и имеют немало общих черт, а также наличием достаточно современных данных о польских предприятиях в свободном доступе.

Результаты анализа показали, что эконометрические модели, особенно классические зарубежные, не вполне применимы для расчета вероятности банкротства. В то же время, методы анализа, основанные на данных, дают более корректные результаты.

Ключевые слова: модели банкротства; эконометрические модели; бизнес-аналитика; случайные леса

Введение

Одной из актуальных проблем, возникающих при анализе деятельности предприятий, является как можно более ранняя диагностика риска банкротства. Важность максимально ранней диагностики состоит в том, чтобы иметь достаточный промежуток времени для принятия управленческих решений, позволяющих уменьшить риск, либо наиболее безболезненно пройти процедуру банкротства [10].

Перед более детальным анализом хозяйственного риска предприятия, который требует значительных трудовых и финансовых затрат, представляется разумным провести предварительный анализ, что позволит оценить вероятность банкротства и даст возможность сосредоточиться на самых проблемных показателях.

Для этих целей в финансово и политически устойчивых странах с начала 1970-х годов используются различные эконометрические модели, которые в настоящий момент можно считать классическими.

В то же время применение этих моделей к странам с так называемой переходной экономикой (к которым в определённой степени можно отнести и Россию) не даёт объективной картины и не позволяет получить заслуживающие доверие результаты [7, 8]. В работе [8] предложена модель, учитывающая особенности экономики России, но построенная на данных 1993-1996 года, уже утративших актуальность, из-за чего его применение также сильно ограничено.

Информационные возможности современной техники позволяют применить подход, в корне отличающийся от применявшихся ранее. В этом случае основой принятия решения служат «большие данные», для которых можно использовать методы, не требующие предварительного построения и анализа моделей. Сами данные непосредственно влияют на ход моделирования и позволяют даже избегать такого построения. Для работы с большими данными разработаны различные методы и алгоритмы машинного обучения и бизнес-аналитики: логистическая регрессия, деревья решений, случайные леса, бустинг, машины опорных векторов и т. д. Использование этих методов позволяет сделать прогноз, используя актуальные данные, адаптируя модели.

Сравнение качества классических эконометрических моделей банкротства и моделей, получаемых с помощью методов бизнес-аналитики, рассматривается в данной статье.

Методы и модели

а) Эконометрические модели банкротства

Двухфакторная модель Альтмана

$$Z = -0,3877 - 1,073x_1 + 0,0579x_2,$$

где: x_1 – коэффициент текущей ликвидности (отношение оборотных активов к краткосрочным обязательствам), x_2 – коэффициент капитализации (отношение суммы долгосрочных обязательства и краткосрочных обязательств к собственному капиталу).

Если величина Z меньше нуля, то вероятность банкротства менее 50 % и уменьшается по мере уменьшения значения Z ; если же Z больше нуля, то вероятность банкротства более 50 %.

Пятифакторная модель Альтмана

$$Z = 0,717x_1 + 0,847x_2 + 3,107x_3 + 0,42x_4 + 0,998x_5,$$

где: x_1 – отношение оборотного капитала к активам (working capital/total assets), x_2 – отношение нераспределённой прибыли к активам (retained earnings/total assets), x_3 – отношение операционной прибыли к активам (ЕБИТ/total assets), x_4 – отношение рыночной стоимости акций к обязательствам (book value of equity/total liabilities), x_5 – отношение выручки к активам (sales/total assets).

Эта модель была предложена в 1983 году для оценки уровня банкротства акционерных обществ США, чьи акции не котируются на фондовом рынке [1].

Оценка финансовой устойчивости компаний основана на следующих сравнениях:

- если $Z < 1,23$, то предприятие находится в зоне финансового риска, высока вероятность банкротства;
- если $1,23 < Z < 2,89$, то предприятие находится в «серой зоне», вероятность банкротства не очень высока;
- если $Z > 2,89$, то предприятие финансово устойчиво, вероятность банкротства низкая.

Модель Лиса

$$Z = 0,063x_1 + 0,057x_2 + 0,092x_3 + 0,001x_4,$$

где: x_1, x_2, x_3, x_4 имеют тот же смысл, что и в предыдущей модели.

Если $Z < 0,037$, то вероятность банкротства высокая, если же $Z > 0,037$, то вероятность банкротства низкая.

Эта модель является одной из первых европейских моделей и предложена в 1972 году [4].

Модель Альтмана для развивающихся рынков:

$$Z = 6,56x_1 + 3,26x_2 + 6,72x_3 + 1,05x_4 + 3,25,$$

где: x_1, x_2, x_3, x_4 имеют тот же смысл, что и в предыдущей модели.

Интерпретировать полученное значение Z можно следующим образом:

- если $Z < 1,1$, то с высокой долей вероятности предприятие обанкротится в ближайшей перспективе;
- если $1,1 < Z < 2,6$, то ситуация неопределённая;
- если $Z > 2,6$, то предприятие финансово устойчиво, вероятность банкротства низкая.

Данная модель предложена в 2003 году и является интерпретацией модели Альтмана для непроизводственных компаний [2].

Модель Беликова-Давыдовой:

$$R = 8,38x_1 + x_2 + 0,054x_3 + 0,63x_4,$$

где: x_1 – отношение оборотного капитала к активам (working capital/total assets), x_2 – отношение чистой прибыли к собственному капиталу (retained earnings/book value of equity), x_3 – отношение выручки к активам (sales/total assets), x_4 – отношение чистой прибыли к себестоимости (retained earnings/total sales).

Вероятность банкротства осуществляется с использованием таблицы 1.

Таблица 1

Вероятность банкротства в соответствии со значением R

Значение R	Вероятность банкротства, %
Меньше 0	Максимальная (90-100)
0-0,18	Высокая (60-80)
0,18-0,32	Средняя (35-50)
0,32-0,42	Низкая (15-20)
Больше 0,42	Минимальная (до 10)

Кроме перечисленных выше, применяются и другие модели, как зарубежные – Фулмера [12], Таффлера, Спрингейта, так и отечественные – Зайцевой, Савицкой и т. д. [7]

б) Методы статистического обучения и бизнес-аналитики

С точки зрения статистического обучения задача прогнозирования банкротства предприятия представляет собой одну из задач «обучения с учителем» – задачу бинарной классификации. Для решения такой задачи в последнее время разработаны разнообразные методы обучения. В данной работе применены некоторые из них. Выбор методов основан на рейтингах решения подобных задач.

Логистическая регрессия

Логистическая регрессия является одним из видов множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми переменными и зависимой переменной. Бинарная логистическая регрессия используется в том случае, когда зависимая переменная может принимать только два значения (как правило, 0 и 1).

Логистическую регрессию часто интерпретируют как вероятность того, что событие наступит в конкретном эксперименте или для конкретного испытуемого.

Уравнение логистической регрессии имеет вид

$$p(x_1, x_2, \dots, x_n) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}},$$

где: оценки коэффициентов $\beta_0, \beta_1, \dots, \beta_n$ находятся исходя из принципа максимального правдоподобия.

Для определения значения бинарной переменной применяют пороговое отсечение, т. е. в случае, когда значение p меньше порога, считают, что прогнозируемое значение выходной переменной равно нулю, в противном случае – единице. В случае отсутствия априорных предположений о данных, обычно пороговое значение полагают равным 0,5.

Деревья решений

В отличие от методов, использующих статистический подход, деревья решений являются одним из самых популярных методов машинного обучения и не требуют каких-либо первоначальных предположений. К преимуществам этого метода относятся быстрота и эффективность в обучении, универсальность, возможность легко объяснить и интерпретировать результаты.

Наиболее применяемым из деревьев решений является дерево классификации и регрессии (CART). При работе этого алгоритма всё множество наблюдений рекурсивно разбивается на подмножества вдоль различных ветвей так, что похожие наблюдения группируются вместе в конечных листьях. На каждом шаге алгоритма используется сравнение по одной переменной, приводящее к увеличению «чистоты» дерева [11]. Для любого нового наблюдения, используя соответствующий путь по дереву, можно получить предсказание результата классификации.

Случайные леса

Недостатком деревьев решений является их неустойчивость. Малые изменения в обучающем множестве могут привести к значительно различающимся деревьям. Обучение групп деревьев и объединение их в «случайный лес» позволяют значительно повысить устойчивость [3].

При использовании алгоритма случайного леса для построения каждого дерева используется случайная выборка с повторениями (такой способ называется «баггинг»). Также случайным образом выбираются переменные для проведения разбиения, что позволяет снизить систематическую ошибку из-за использования только сильно влияющих факторов. Кроме того, часто при использовании деревьев решений возникает эффект переобучения. Усреднение предсказанных вероятностей для большого числа деревьев позволяет получить более устойчивую оценку по сравнению с одиночным деревом [9].

Машины опорных векторов

Машины опорных векторов (support vector machines, SVM) представляют собой модель бинарного классификатора, который строит гиперплоскость наилучшим образом разделяющую множество наблюдений на два класса. Как правило, такое разделение в чистом виде

невозможно, поэтому приходится вводить функцию штрафов за неверно классифицированные наблюдения из обучающего множества. Ещё одним способом более точного разделения на классы является использование ядерных функций, отличных от линейных (функции Гаусса, вейвлеты и т. д.) [9].

Результаты

Сравнение эффективности использования классических эконометрических моделей банкротства и методов бизнес-аналитики для классификации данных было проведено на наборе данных о банкротствах малых и средних предприятий Польши [5, 6]. Выбор этого набора объясняется несколькими причинами:

- наличием большого набора данных (10503 предприятий) в свободном доступе;
- данные являются достаточно новыми (2007-2012 годы);
- в наборе присутствуют 64 экономических показателя, что позволяет применять модели бизнес-аналитики более качественно;
- Польша является, как и Россия, страной с переходной экономикой, что позволяет строить аналогии в применении методов.

К сожалению, аналогичного объема данные по российским предприятиям отсутствуют.

Из 10503 исследованных предприятий 495 являлись банкротами, а остальные продолжали деятельность. Для более корректного анализа из действующих предприятий случайным образом было отобрано 495 для пропорционального представительства. Множество из 990 предприятий было разделено на два: 75 % отнесены к обучающему (742), 25 % – к тестовому (248). Такое деление необходимо в методах машинного обучения для оптимального подбора коэффициентов моделей. Проверка всех моделей как эконометрическими методами, так и методами бизнес-аналитики проводилась на тестовом множестве. Для каждой модели была построена так называемая матрица классификации (таблица 2), где 0 соответствует действующим предприятиям, а 1 – предприятиям банкротам.

Таблица 2

Вид матрицы классификации

		Прогнозируемые значения	
		0	1
Реальные значения	0	Действующее предприятие прогнозируется как не банкрот	Действующее предприятие прогнозируется как банкрот
	1	Предприятие – банкрот прогнозируется как не банкрот	Предприятие – банкрот прогнозируется как банкрот

Точность модели определяется как доля верно определенных типов предприятий к общему числу всех предприятий в тестовом множестве.

Результаты анализа эконометрических моделей приведены в таблицах 3, 4.

Таблица 3

Матрицы классификации эконометрических моделей

106	18	66	58	39	85
92	32	46	78	14	110
Двухфакторная модель Альтмана		Пятифакторная модель Альтмана		Модель Лиса	
109	15	100	24		
96	32	68	56		
Модель Альтмана для развивающихся рынков		Модель Беликова-Давыдовой			

Таблица 4

Точность эконометрических моделей

Название модели	Точность, %
Двухфакторная модель Альтмана	55,6
Пятифакторная модель Альтмана	58,1
Модель Лиса	60,1
Модель Альтмана для развивающихся рынков	56,9
Модель Беликова-Давыдовой	62,9

Построение моделей бизнес-аналитики проведено в программной среде вычислений R, широко используемой как статистическое программное обеспечение для анализа данных. Для построения модели логистической регрессии использовалась функция `glm`, для нахождения случайного леса – пакет `randomForest`, для использования машин опорных векторов – пакет `e1071`.

Результаты анализа моделей бизнес-аналитики моделей приведены в таблицах 5, 6.

Таблица 5

Матрицы классификации моделей бизнес-аналитики

102	22	96	28	81	43
60	64	37	87	37	87
Логистическая регрессия		Случайные леса		Машины опорных векторов	

Таблица 6

Точность моделей бизнес-аналитики

Название модели	Точность, %
Логистическая регрессия	67,7
Случайные леса	73,8
Машины опорных векторов	67,7

Выводы

Основываясь на анализе полученных результатов, можно сделать следующие выводы.

Во-первых, из всех эконометрических моделей наибольшая точность получается в результате применения модели Беликова-Давыдовой, хотя и разработанной для российских предприятий, но дающей приемлемые результаты и для других стран с переходной экономикой.

Во-вторых, методы бизнес-аналитики дают значительно более точные и более сбалансированные результаты. К тому же точность может быть увеличена путём подбора

параметров в методах, например, используя предварительно процедуру на обучающем множестве перекрёстной проверки (в работе использовались параметры по умолчанию).

В-третьих, в методах бизнес-аналитики наибольшую роль играют не те показатели, что применяются в эконометрических моделях, а такие как отношение валовой прибыли к активам, отношение суммы чистой прибыли и амортизации к заёмному капиталу, прибыльность продаж.

В заключение, можно сделать вывод, что методы, основанные на статистическом и машинном обучении, дают хорошие результаты и при наличии большого количества постоянно получаемых данных применение таких методов является весьма перспективной областью бизнес-анализа.

ЛИТЕРАТУРА

1. Altman E. I. A further empirical investigation of the bankruptcy cost question // *The Journal of Finance*. – 1984. – Т. 39. – № 4. – С. 1067-1089.
2. Altman E. I. Managing credit risk: A challenge for the new millennium // *Economic Notes*. – 2002. – Т. 31. – № 2. – С. 201-214.
3. Breiman L. Random forests // *Machine learning*. – 2001. Т. 45. – № 1. – С. 5-32.
4. Charitou A., Neophytou E., Charalambous C. Predicting corporate failure: empirical evidence for the UK // *European Accounting Review*. – 2004. – Т. 13. – № 3. – С. 465-497.
5. Polish companies bankruptcy data Data Set // UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data>.
6. Zięba M., Tomczak S. K., Tomczak J. M. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction // *Expert Systems with Applications*. – 2016. – Т. 58. – С. 93-101.
7. Афанасьева А. Н., Ефимова Н. Ф. Применение моделей оценки степени банкротства предприятия // *Синергия Наук*. – 2017. – № 8. – С. 129-139.
8. Давыдова Г. В., Беликов А. Ю. Методика количественной оценки риска банкротства предприятий // *Управление риском*. – 1999. – Т. 3. – С. 13-20.
9. Джеймс Г. и др. Введение в статистическое обучение с примерами на языке R // М.: ДМК Пресс. – 2017. – 456 с.
10. Лаврухина Н. В., Перерва О. Л. Стоимостная концепция и оценочные технологии управления инновационными предприятиями: учебное пособие // М.: МГТУ им. Н.Э. Баумана. – 2013. – 243 с.
11. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям // СПб.: Питер, 2013. – 704 с.
12. Степанов С. Е., Хамер Г. В. Модифицированная модель Фулмера как инструмент диагностики вероятности наступления банкротства компании // *В мире научных открытий*. – 2013. – № 4.2. – С. 236-247.

Pererva Olga Leonidovna

Bauman Moscow state technical university
Kaluga branch, Russia, Kaluga
E-mail: pol@bmstu-kaluga.ru

Stepanov Sergei Evgen'evich

Bauman Moscow state technical university
Kaluga branch, Russia, Kaluga
E-mail: stepanov@bmstu-kaluga.ru

Nezimova Sofiia Sergeevna

Bauman Moscow state technical university
Kaluga branch, Russia, Kaluga
E-mail: stepanova_s@adm.kaluga.ru

Comparison of econometric models and methods of business analytics for prediction of bankruptcy of enterprises

Abstract. The article compares classical econometric models of enterprise bankruptcy and models obtained using modern methods of business analytics, such as logistic regression, random forests, support vector machines.

The authors consider five econometric models, both foreign and domestic, and three methods of business analytics – logistic regression, as well as random forests and support vector machines, which are one of the most accurate methods of machine learning. In the second case, the task of predicting bankruptcy is one of the tasks of machine learning with the teacher – the task of binary classification.

Comparison of models was carried out by the authors on data on more than ten thousand Polish enterprises. The choice of this set is explained by the fact that both Poland and Russia belong to countries with transitional economies and have many common features, as well as the availability of fairly modern data on Polish enterprises in free access.

The results of the analysis showed that econometric models, especially classical foreign ones, are not entirely applicable to calculating the probability of bankruptcy. At the same time, data-based analysis methods give more correct results.

Keywords: models of bankruptcy; econometric models; business analytics; random forests