

# ОЦЕНКА ПЛАТЕЖЕСПОСОБНОСТИ ПРЕДПРИЯТИЯ НА ОСНОВЕ МЕТОДА ДЕРЕВЬЯ КЛАССИФИКАЦИИ

Е.Ю. Пелипенко, А.А. Халафян

*ГОУ ВПО Кубанский государственный университет, г.Краснодар*

## Аннотация

Проблема определения платежеспособности предприятия в настоящее время является актуальной для кредитных организаций страны. В работе предложено решение этой проблемы статистическим методом дерева классификации. Для удобства использования данного метода был создан программный модуль, позволяющий на основе значений некоторых экономических показателей классифицировать предприятие по уровню неплатежеспособности.

Вопрос платежеспособности заемщиков с каждым годом приобретает все большую значимость. Современные условия выдачи денежных средств кредитными организациями характеризуются высокой степенью информационной неопределенности заемщиков. Для кредитных организаций весьма важно точно определить уровень риска неплатежеспособности клиента в условиях нехватки информации, что является достаточно сложной задачей.

Существующие методы, как отмечалось в статье [1], являются неустойчивыми к вариациям исходных данных, так как эти данные обладают сильной динамикой изменчивости как во времени, так и в пространстве.

Наиболее удачным способом решения проблемы информационной неопределенности является использование метода комплексного финансового анализа предприятий [2], в основу которого легли 16 устойчивых во времени и пространстве экономических показателей. Использование традиционных статистических методов классификации (кластерный анализ, дискриминантный анализ, и т.д.) эффективно при выполнении более строгих теоретических предположений и выборочное распределение обладает некоторыми специальными свойствами. Когда же «отказывают» все традиционные методы, деревья классификации, по мнению многих исследователей, дают вполне приемлемый результат.

В данной работе использован метод дерева классификации для определения принадлежности клиента к тому или иному уровню риска неплатежеспособности в зависимости от соответствующих значений наиболее весомых предикторных переменных (экономических показателей). Для удобства использования метода в среде программы Statistica был создан модуль, позволяющий автоматизировать процесс классификации предприятий.

В качестве наиболее весомых экономических показателей были рассмотрены 16 параметров предприятий, образующих четыре группы [2]:

1. Рентабельность: **R1** – общая рентабельность; **R2** – рентабельность активов; **R3** – рентабельность собственного капитала; **R4** – рентабельность продукции; **R5** – рентабельность оборотных активов.

2. Ликвидность и платежеспособность: **L1** – быстрый коэффициент ликвидности; **L2** – коэффициент покрытия запасов; **P1** – текущий коэффициент ликвидности.

3. Деловая активность: **A2** – оборачиваемость активов; **A4** – оборачиваемость кредиторской задолженности; **A5** – оборачиваемость дебиторской задолженности; **A6** – оборачиваемость запасов.

4. Финансовая устойчивость: **F1** – коэффициент финансовой зависимости; **F2** – коэффициент автономии собственных средств; **F3** – обеспеченность запасов собственными оборотными средствами; **F4** – индекс постоянного актива.

Для построения дерева классификации была использована виртуальная база [1] для 1000 предприятий пяти уровней риска (по 200 предприятий каждого класса) по всем шестнадцати экономическим показателям: *1 группа* – предприятия очень высокого уровня риска; *2 группа* – предприятия, находящиеся в зоне высокого риска; *3 группа* – предприятия, находящиеся в зоне среднего риска; *4 группа* – предприятия, находящиеся в зоне низкого риска; *5 группа* – предприятия, находящиеся в зоне очень низкого риска (кредитоспособные предприятия).

Процесс построения дерева классификации включил в себя три основных этапа:

1. Выбор критерия точности прогноза;
2. Выбор типа ветвления;
3. Определение «подходящего размера» дерева.

Критерий точности прогноза определяется наименьшей ценой (т.е. долей неверно классифицированных наблюдений). Наиболее «важным» классом является 5 группа риска неплатежеспособности, так как ошибка отнесения клиента к группе кредитоспособных предприятий говорит о высокой вероятности невыплаты денежных средств по долговому обязательству заемщика. Поэтому цену ошибок для предприятий, относящихся к группе очень низкого уровня риска неплатежеспособности, принимаем равную 0. А для остальных групп – равную по 0,25. Все внедиагональные элементы матрицы цен ошибок полагаются равными 1 (таблица 1).

Таблица 1 – Матрица цен ошибок

Уровень риска	Очень высокий	Высокий риск	Средний риск	Низкий риск	Очень низкий
Очень высокий	0,25	1	1	1	1
Высокий риск	1	0,25	1	1	1
Средний риск	1	1	0,25	1	1
Низкий риск	1	1	1	0,25	1
Очень низкий	1	1	1	1	0

В модуле дерева классификации программы Statistica реализованы алгоритмы QUEST и CART.

При глобальной кросс-проверке при построении дерева с типом ветвления по методу CART три предприятия класса *Низкий риск* неверно классифицированы как одно предприятие класса *Средний риск* и два предприятия *Очень низкий риск* (таблица 2). Это составляет менее 1% ошибки классификации от общего числа рассматриваемых предприятий.

Таблица 2 – Матрица ошибок классификации глобальной кросс-проверки

Предсказанные классы	Исходные классы				
	Очень высокий	Высокий риск	Средний риск	Низкий риск	Очень низкий
	Цена глобальной кросс-проверки = 0,003 Стандартное отклонение цены = 0,00172				
Очень высокий		0	0	0	0
Высокий риск	0		0	0	0
Средний риск	0	1		0	2
Низкий риск	0	0	0		0
Очень низкий	0	0	0	0	

Неверная классификация двух предприятий к группе *Очень низкий риск* говорит о возможности выдачи денежных средств этим предприятиям, когда вероятность невозврата кредита превышает допустимую норму. При этом цена глобальной кросс-проверки составила 0,003, стандартное отклонение цены – 0,00172. Эти величины незначительно отличаются от цены кросс-проверки (0,0) и ее стандартной ошибки (0,0), что означает: методом CART построено вполне приемлемое дерево классификации.

Анализ метода QUEST (метод дискриминантного многомерного ветвления по линейной комбинации) показал более качественную классификацию, чем метод CART. Для него удалось построить дерево классификации (рисунок 1) с нулевой матрицей глобальной кросс-проверки (цена глобальной кросс-проверки и стандартное отклонение цены получили равными 0).

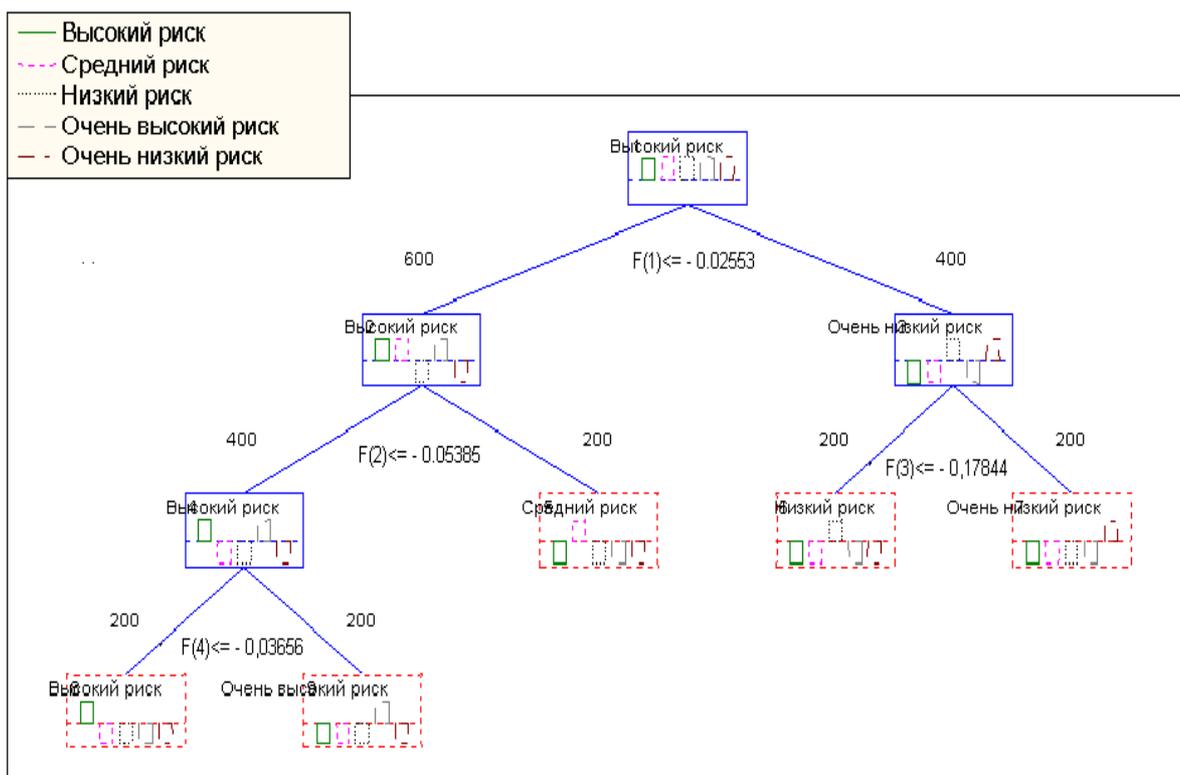


Рисунок 1 – Граф дерева классификации при выборе типа ветвления по методу QUEST

Так как дерево бинарное, то из каждой вершины выходят две ветви. Под вершинами указаны условия перехода по ветвям (левая ветвь отвечает за выполнение условия, правая – за нарушение). В верхней части вершин указано название класса, над вершиной – количество предприятий этого класса. Красным пунктиром обозначены терминальные вершины (листья) дерева, дальнейшее ветвление из которых невозможно. Высота гистограмм, изображенных в вершинах дерева, соответствует количеству предприятий определенного класса в данной вершине.

Для выбора дерева классификации «подходящего размера» все значения параметров были выбраны согласно методу автоматического построения дерева Бримана [4]: стандартная ошибка равна 1, остановка определяется ошибкой классификации. Минимальное число неправильно классифицированных наблюдений в терминальной вершине было принято равным 5. Сделанные установки для выбора размера дерева классификации соответствуют методу автоматического построения дерева Бримана [4].

Таблица результатов классификации (таблица 3) отражает количество объектов, отнесенных по результатам классификации к тому или иному классу. Из таблицы видно, что исходные и предсказанные классы полностью совпали.

Таблица 3 – Таблица принадлежности объектов классам

Предсказанные классы	Исходные классы				
	Очень высокий	Высокий риск	Средний риск	Низкий риск	Очень низкий
Очень высокий	200	0	0	0	0
Высокий риск	0	200	0	0	0
Средний риск	0	0	200	0	0
Низкий риск	0	0	0	200	0
Очень низкий	0	0	0	0	200

Числовые значения из таблицы решающих правил (условий ветвления) были использованы в качестве коэффициентов для составления линейных функций  $F(1)$ ,  $F(2)$ ,  $F(3)$ ,  $F(4)$ , представляющих собой условия перехода алгоритма программного модуля анализа данных (см. рисунок 1), написанного в среде STATISTICA VISUAL BASIC (SVB). По введенным значениям финансовых показателей предприятий модуль классифицирует их по пяти группам уровня риска неплатежеспособности. Для удобства использования модуля были созданы пользовательские диалоговые окна ввода и вывода данных, содержащие результаты классификации.

Созданный программный модуль автоматизирует процесс классификации предприятий по уровню кредитоспособности, тем самым значительно сокращая затраты времени пользователя, предупреждая ошибки кредитования. Используя данный модуль можно оперативно прогнозировать уровень риска неплатежеспособности реально существующего предприятия. От пользователя не требуется специальных знаний в области современных технологий анализа данных, так как программа имеет простой интерфейс, состоящий из окон ввода и вывода данных.

### Список литературы

1. Шевченко И.В., Халафян А.А., Васильева Е.Ю. «Создание виртуальной клиентской базы для анализа кредитоспособности российских предприятий»// Финансы и кредит, №1, 2010, с.13-18.
2. Недосекин А.О., Бессонов Д.Н., Лукашев А.В. «Сводный финансовый анализ российских предприятий за 2000-2003 г.г.»
3. Bartlett M.S. Properties of sufficiency of statistical tests // Proc. Roy. Soc., 1937, Series A, Vol. 31. – P. 268-282.
4. Breiman L.J. Random Forests–Random Features // UC Berkeley, Statistics Department, Technical Report N. 567, 1999.