

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ И СЕМАНТИЧЕСКОГО АНАЛИЗА ДЛЯ ПРОГНОЗИРОВАНИЯ БАНКРОТСТВА

Макеева Е.Ю.¹, Аршавский И.В.²

Предсказание финансовой неустойчивости компании является дискуссионной темой в экономической литературе и остается таковой уже на протяжении многих лет. Авторами было предложено большое количество разнообразных финансовых показателей, оказывающих существенное влияние на вероятность банкротства компании. Однако часть информации о деятельности компании является скорее качественной, нежели количественной, и не отражается в финансовых показателях. Примером может служить качество корпоративного управления или степень раскрытия информации в открытых источниках. Поскольку годовая отчетность компании содержит существенные факты о текущей и будущей эффективности компании, необходимо учитывать ее в качестве источника нефинансовой информации.

В данной работе представлена методология предсказания финансовой неустойчивости на основе применения семантического анализа корпоративной годовой отчетности и ансамбля искусственных нейронных сетей. Полученные результаты подтверждают значимость использования текстовой информации, содержащейся в годовой отчетности компании, для предсказания финансовой неустойчивости, а также демонстрируют повышение предсказательной способности модели.

JEL: G32, G33

Ключевые слова: Предсказание банкротств; финансовая неустойчивость; нейросети; раскрытие информации; семантический анализ

Индикаторы несостоятельности компании

Предсказание несостоятельности компаний достаточно продолжительное время представляет собой актуальную исследовательскую проблему. Наиболее простым и популярным способом является дискриминантный анализ. В 90-х годах прошлого века особую популярность приобрели методы машинного обучения, такие как искусственные нейронные сети, рекурсивные деревья и метод опорных векторов (Anandarajan et al., 2004; Huang, Chen, Hsu, Chen, & Wu, 2004; Kwon, Han, & Lee, 1996; Tam & Kiang, 1992). Пример того, как смещался фокус авторов с использования одних моделей на другие, представлен в таблице 1.

Таблица 1

Модели предсказания банкротств, наиболее часто используемые в период 1966-2004 гг.

	1960	1970	1980	1990	2000
Дискриминантный анализ	2	22	28	9	2
Логит / Пробит	0	2	19	19	3
Машинное обучение	0	0	1	35	4
Прочие	1	4	7	11	3
Всего	3	28	55	74	12

Источник: (Bellovary et al., 2007)

На протяжении 90-х годов метод искусственных нейронных сетей присутствовал более чем в 50% работ, посвященных анализу финансовой неустойчивости компании. Такие преимущества нейронных сетей, как отсутствие чувствительности к характеру входных данных и возможность аппроксимирования функции практически любой сложности, позволили ряду исследователей получить результаты, доказывающие превосходство данного рода моделей над статистическими (Fletcher, Goss, 1993; Carlos, 1996; Liang, Wu, 2005; Rafiei et al., 2011).

¹ Выпускник магистерской программы «Стратегическое управление финансами», Национальный исследовательский университет «Высшая школа экономики».

² Выпускник магистерской программы «Стратегическое управление финансами», Национальный исследовательский университет «Высшая школа экономики».

Результаты исследований зависят не только от способа формирования выборки, но и от тех показателей, что используются в виде входных данных модели предсказания. Большинство моделей прогнозирования банкротств компании основываются на финансовых показателях, полученных из бухгалтерской отчетности компании. Традиционно эти показатели характеризуют компанию с таких сторон, как рентабельность, ликвидность, генерация денежных потоков и структура капитала.

В то время как предсказательная способность таких факторов доказана эмпирически, все больше современных авторов оспаривают модели прогнозирования банкротств, основанные исключительно на финансовых показателях. Основным аргументом служит тот факт, что бухгалтерская отчетность компании отражает уже произошедшие изменения и, таким образом, является малоприменимой для целей прогнозирования (Amir, Lev, 1996; Kloptchenko et al., 2004; Riley et al., 2003; Shirata et al., 2011). Поэтому в современных исследованиях часто вместе с финансовыми детерминантами используются различные небалансовые показатели, отражающие рыночную оценку, уровень корпоративного управления, жизненный цикл продукта и другие аспекты деятельности компании.

Основным источником полезной информации о деятельности компании является ее годовая отчетность, содержащая на порядок больше информации, чем финансовая отчетность. В том числе годовая отчетность может содержать сигналы, говорящие о потенциальной неустойчивости компании. Такие сигналы можно обнаружить путем глубинного анализа текста. Основной идеей такого анализа является выявление специфичных слов, способных в составе фраз или индивидуально указывать на принадлежность исследуемого объекта (в данном исследовании – отчетности) к тому или иному кластеру.

Например, в статье Клатворфи и Джонс (Clatworthy, Jones, 2006) авторы исследуют ряд текстовых характеристик отчетов совета директоров 50 прибыльных и 50 неприбыльных публичных компаний Британии. В своем исследовании авторы обнаруживают, что склонность менеджеров ассоциировать себя с финансовым результатом деятельности компании в действительности связана с ее фактической эффективностью. Также, авторы заметили, что убыточные компании в своих отчетах употребляют больше фраз, описывающих будущую деятельность компании, нежели ее историческую эффективность. Смит и Тафлер (Smith, Taffler, 2000), также анализируя отчеты совета директоров, показали, что сам факт предоставления подобного рода отчета значимо связан с долгосрочной устойчивостью компаний.

Основным недостатком таких исследований является недостаточный уровень анализа контекста, в котором применяются ключевые слова. Простой подсчет количества повторений ключевых слов часто не может предоставить информацию, необходимую для проведения различий между компаниями. Так, например, одно лишь упоминание слова «выручка» в тексте годового отчета ничего не говорит о фактическом размере выручки или ее росте (падении) относительно уровня прошлых лет.

Последующий анализ предполагает дальнейшее изучение фраз, в которых упоминается ключевое слово, на предмет упоминания в отчете сигнальных фраз, придающих дополнительных «окрас» ключевому слову. Одним из ключевых исследований данной тематики является работа Ширата и соавторов (Shirata et al., 2011). В ней авторы применяют метод обработки текста с использованием естественного языка (natural language processing). На выборке из 90 компаний банкротов и 90 финансово здоровых компаний Японии авторам удалось показать, что существует ряд сигнальных слов, которые в сочетании с ключевым словом существенно повышают вероятность корректного предсказания банкротства.

Методология исследования

Методология исследования представлена двумя частями анализа компаний на предмет выявления признаков несостоятельности:

1. Семантический анализ годовой отчетности.
2. Построение адекватной устройству данных системы прогнозирования банкротства компании, агрегирующей совокупность финансовых и семантических показателей.

Обе части состоят из нескольких этапов и схематично представлены на рис. 1

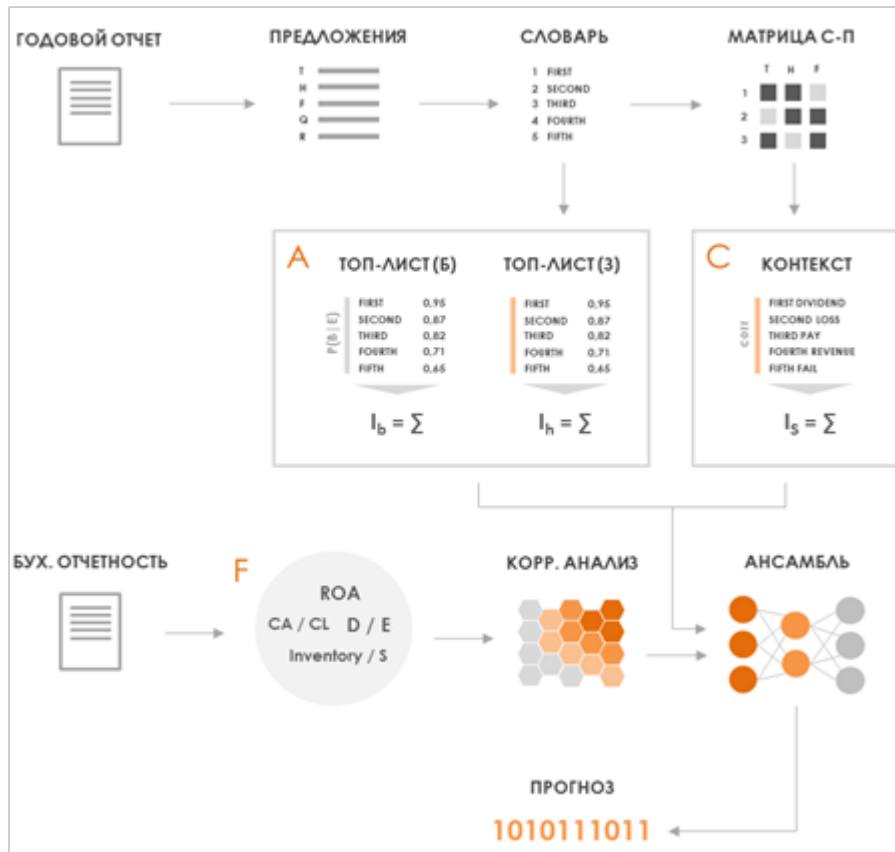


Рисунок 1. Схематическое представление этапов исследования

В ходе исследования выдвигаются три гипотезы.

Гипотеза 1. Использование байесовского ансамбля для прогнозирования несостоятельности компаний оправданно на относительно небольших выборках с высокой вариацией и позволяет классифицировать компании с меньшей ошибкой по сравнению с индивидуальными моделями.

Гипотеза 2. Компании-банкроты и финансово здоровые компании, публикующие годовую отчетность в открытых источниках, отражают в них информацию в разном качестве и количестве, что позволяет выделять слова и фразы, специфичные для каждой из групп компаний.

Гипотеза 3. Включение информации о контекстно зависимых единицах годовых отчетов в модель предсказания неустойчивости компаний позволяет повысить ее предсказательную способность в долгосрочной перспективе, а также уменьшить ошибку неверной классификации банкротов (ошибку первого рода).

Следующие программные продукты и пакеты дополнений были использованы в исследовании: «R 3.1» с пакетами «tm», «SnowballS» и «Rstem»; «MatLab R2013b» с пакетом «Neural Network Toolbox» актуальной версии.

Описание выборки данных

Поскольку исследование предполагает анализ текстовой информации, заключенной в годовой отчетности компаний, изначально выборка ограничена компаниями, предоставляющими отчетность на одном языке или сопоставимых языках.

В данной работе была проанализирована выборка из 84-х нефинансовых российских компаний, предоставляющих годовую отчет и годовую финансовую отчетность на русском языке. Выборка охватывает компании, признанные арбитражным судом банкротами в период с 2010 г. до второго квартала 2014 г., при этом большая часть банкротств приходится на период 2012–2013 гг. Всего выборка включает в себя 44 компании-банкрота и 40 финансово здоровых компаний. Прогнозирование банкротства происходит за один и три года до момента признания компании банкротом. Источником информации послужили базы данных «S&P Capital IQ» и «RUSLANA».

В сравнении со здоровыми компаниями типичный представитель компаний-банкротов в выборке обладает следующими характеристиками:

- Отрицательные и более волатильные показатели рентабельности. Так, например, среднее значение рентабельности активов компаний-банкротов составляет -0,12 при стандартном отклонении 0,2. Эти же характеристики у финансово здоровых компаний составляют 0,02 и 0,06 соответственно.
- Низкая оборачиваемость краткосрочных обязательств, запасов и дебиторской задолженности. Так, средняя оборачиваемость запасов здоровых компаний более чем в два раза превышает аналогичный показатель компаний-банкротов. Максимальное значение оборачиваемости по первой группе компаний также выше. Это может свидетельствовать о накоплении избыточных запасов и неэффективном складском управлении компаниями-банкротами.
- Низкие показатели ликвидности. Соотношение оборотных активов и текущих обязательств компаний-банкротов в среднем в два раза ниже по сравнению с финансово здоровыми компаниями и составляет 1,09.
- Низкие показатели покрытия долга, что вызвано низкой операционной эффективностью.

Для целей исследования исходная выборка была поделена на обучающую и проверочную выборки в пропорции 60% и 40% соответственно. Семантический анализ и обучение ансамбля проводились на обучающей выборке. Проверка модели происходила на тестовой выборке. Таким образом, в обучающую выборку вошло 26 компаний-банкротов и 24 финансовоздоровых компаний, в тестовую выборку – 18 и 16 компаний соответственно.

Этапы проведения исследования

Семантический анализ текстовой информации годовой отчетности

Первая часть исследования связана с анализом текстовой информации в годовой отчетности компании на основе выявления семантических особенностей годового отчета

Путем разбиения предложений на отдельные слова был получен вектор уникальных слов (основ слов), присутствующих в массиве предложений – так называемый «словарь».

Последующий алгоритм семантического анализа состоит из двух блоков: блока «А» и блока «С».

В первом блоке (блоке «А») в каждом годовом отчете происходит учет наличия слов из сформированного словаря. Выделение слов, характерных для группы банкротов, и слов, характерных группы финансово здоровых компаний, происходило на основе сопоставления условных вероятностей.

Данный подход состоит в нахождении условной вероятности банкротства компании при использовании конкретного слова и сопоставлении с общей вероятностью банкротства, равной отношению числа банкротов в выборке к общему размеру выборки. Условная вероятность банкротства при использовании слова «Е» находится по следующей формуле:

$$P(\text{Bankrupt} | E) = E_b / E_a, \quad (1)$$

где – число компаний, упомянувших «Е», – число компаний-банкротов, упомянувших «Е».

Если условная вероятность значительно больше общей вероятности, то данное слово может считаться специфичным для данной группы компаний. Аналогичная процедура применяется и к финансово здоровым компаниям. В результате формируется список слов специфичных для каждой группы. На основе этого списка строится итоговый взвешенный индекс $A^{bankrupt}$ и $A^{healthy}$.

$$A_i^{bankrupt} (A_i^{healthy}) = \sum_{j=1}^N E_{ij} W_j, \quad (2)$$

где равно «1», если j -е слово специфичного списка упоминается в годовом отчете i -й компании

и «0», если не упоминается; – вес, соответствующий отношению условной вероятности j -го слова к сумме условных вероятностей слов специфического списка. Значения индексов находятся в пределах [0; 1].

Второй блок (блок «С») включает в себя контекстный анализ фраз, связанных с ключевым словом. Ключевым может выступать слово, достаточно часто встречающееся как в отчетах компаний-банкротов, так и в отчетах финансово здоровых компаний. Важным является не само слово, а контекст, в котором оно употребляется. Так, например, слово «дивиденд» может встретиться во фразе «дивиденды не выплачивались» и во фразе «выплачены дивиденды в размере». В каждую из этих фраз заложен различный смысл, и, несмотря на частоту употребления ключевого слова, его наличие или отсутствие в конкретном контексте может являться характерным сигналом принадлежности компании к группе банкротов или финансово здоровых компаний.

Для проведения контекстного анализа предварительно была построена матрица, строками которой являются основы слов из словаря, а столбцы соответствуют номеру предложения в исходном массиве данных. В ячейках матрицы отображается число упоминаний слова в конкретном предложении.

Процедура выявления взаимосвязанных слов заключается в поиске слов, наиболее часто упомянутых в связке в одном предложении. Дополняющее слово также проверяется на соответствие контексту. Также фильтруются слова, характерные лишь для конкретных компаний в выборке. В качестве ключевых слов использовался следующий список: «Дивиденд», «Прибыль», «Займ», «Рынок», «Конкурент», «Работник».

В этот список вошли слова, одинаково часто встречающиеся в годовых отчетах как компаний-банкротов, так и здоровых компаний. Данные ключевые слова обычно не несут с собой конкретного негативного или позитивного значения, таким образом, важен именно контекст, в котором они были употреблены. Они также имеют прямое отношение к деятельности компании и в связке с сигнальным словом отражают ее отношение с основными стейкхолдерами: акционерами, кредиторами, работниками и потребителями.

Построение модели предсказания банкротства

В качестве финансовых показателей для построения модели были выбраны популярные среди авторов детерминанты, характеризующие рентабельность, ликвидность, деловую активность, структуру капитала и степень покрытия долга компании. Показатели размера компании также были включены в модель.

Поскольку некоторые показатели строятся на основе схожих метрик и могут являться коррелированными, одним из этапов анализа является избавление от мультиколлинеарности в модели. Для анализа взаимосвязей между независимыми переменными использовалась нейронная сеть, построенная по принципу самоорганизующихся карт (SOM). Коррелированными признавались переменные, обладающие схожей картой весов. Факторы с наименее кластеризованными картами исключались из дальнейшего анализа.

Оставшиеся финансовые показатели вместе с индикаторами, полученными в процессе семантического анализа, составили итоговый массив данных, характеризующий исследуемые компании. Этот массив данных является входным для модели анализа несостоятельности компании. В качестве модели, адекватной устройству данных, в исследовании используется нейросетевой байесовский ансамбль.

Результаты исследования

Блок «А». Слова-маркеры

Основываясь на описанном ранее алгоритме, мы составили список уникальных слов, характерных для каждой из двух групп: компаний-банкротов и финансово здоровых компаний. Для выявления типичных слов из этого списка были исключены слова, встречающиеся менее чем в

15% годовых отчетов. Из списка также были исключены слова, относящиеся к специфике деятельности компаний (например, «фармацевтический», «цемент» и так далее). Итоговый список слов представлен в таблице 2.

Таблица 2

Слова, характерные для каждой из двух групп компаний

Компании-банкроты	научно-технический, погашение, остаток, вынужденный, подрядчик, падение, завершённый
Финансово здоровые компании	интеллектуальный, жизнь, медицинский, интегрированный, доходный, подвержен, лидер, портфель, лучший, ценности, поддерживать, сохранять, потенциальный, население, ориентированный, вклад, статус, преимущественный, роль, прирост

Оба списка слов существенно отличаются друг от друга. Разнятся не только размер списков, но и смысловый окрас входящих в их состав слов. Так, например, в списке слов, характерных для финансово здоровых компаний, присутствуют слова, придающие положительный окрас той или иной фразе, как, например: «лидер», «лучший», «сохранять», «поддерживать». Встречаясь в предложении, данные слова обычно указывают на стремление компании удерживать значимые позиции на рынке, а также сохранять темпы развития. В этом списке также присутствуют слова, указывающие на взаимодействие со стейкхолдерами компании, это: «интеллектуальный», «медицинский», «интегрированный», «портфель», «ценности», «население». Употребленные в таком контексте, как «защита интеллектуальной собственности» или «обязательное медицинское страхование», данные слова указывают на проявление активной позиции компании в отношении своих работников, потребителей. Таким образом, компания демонстрирует укрепление отношений со стейкхолдерами, что в свою очередь указывает на ее стремление к устойчивому долгосрочному развитию и стратегическому росту, ключевым для которого является создание интеллектуального капитала.

В то же время в списке слов, характерных для компаний, обанкротившихся три года спустя, присутствуют слова, придающие негативный окрас фразам, такие как «вынужденный» и «падение». Слово «подрядчик» часто употребляется в контексте выбора компанией подрядчика на основе ценовой конкуренции, а также в таких фразах, как «задолженность перед подрядчиками», «перенос сроков выполнения работ подрядчиками». Таким образом, можно предположить, что компании-банкроты в силу ограниченности бюджета и других возможных причин ведут дела с менее надежными партнерами, что способствует переносу сроков и прочим негативным последствиям.

В таблице 2 представлена суммарная статистика по индивидуальным индексам $A^{bankrupt}$ и $A^{healthy}$. Из таблицы 3 также видно, что средние значения индексов с учетом стандартного отклонения умеренно отдалены друг от друга. Это в свою очередь позволяет говорить о значимости данных индексов для предварительной кластеризации компаний на соответствующие группы.

Таблица 3

Суммарная статистика по индексам $A^{bankrupt}$ и $A^{healthy}$

Метрика	Компании-банкроты					Финансово здоровые компании				
	mean	median	min	max	st.dev.	mean	median	min	max	st.dev.
$A^{bankrupt}$	0.24	0.20	0.00	0.90	0.22	0.15	0.10	0.00	0.61	0.17
$A^{healthy}$	0.09	0.05	0.00	0.55	0.12	0.48	0.45	0.05	0.95	0.26

Блок «С». Ключевые и сигнальные слова

Список взаимосвязанных слов представлен в таблице 4. В него входят сигнальные слова, встречающиеся не менее чем в 15% годовых отчетов. В список не вошли слова, употребляемые в контексте специфики деятельности компании, а также наименования и аббревиатуры компаний.

Связанные ключевые и сигнальные слова

Группа	Ключевое слово					
	«заем»	«дивиденд»	«прибыль»	«работник»	«риск»	«рынки»
Банкроты	-	облигации, процент, деятельность, налог	работа, состав, деятельность, предприятие	среднесписочная численность, на одного, трудов	региональный, платежеспособность, основной	-
Не банкроты	не позднее, долгосрочный, задолженность, показатель	начислен, результат, принят, размер	рентабельный, продажи, продукция, показатель	организация, социальный, развитие, оплата	внутренний, показатель, действие, отсутствие	новый, мировой, позиция, сегмент

Как видно из таблицы 4, сигнальных слов в группе банкротов оказалось меньше, чем в группе финансово здоровых компаний. Трудности, возникающие при выявлении слов, специфичных для компаний-банкротов, связаны с меньшим разнообразием контекста, в котором употребляются ключевые слова. Так, например, в годовых отчетах компаний-банкротов секция, посвященная дивидендам, часто имеет небольшой размер и ограничивается рядом стандартных фраз, вроде «... за указанный период дивиденды не выплачивались...» или «годовым общим собранием акционеров... было принято решение выплатить дивиденды в размере...».

Для устойчивых компаний характерной оказалась связка «работник» и «организация». Часто употребляемая в контексте организации мероприятий по развитию и мотивации персонала, данная связка указывает на стремление компании увеличить отдачу от своих интеллектуальных ресурсов, что в свою очередь характерно для устойчивых компаний. В то же время связка «работник» и «среднесписочная численность» часто употребляется компаниями-банкротами в коротких предложениях с целью перечисления текущих показателей, связанных, например, с оплатой труда. Отсутствие в таких предложениях ориентации на будущую отдачу от интеллектуального капитала и перечисление текущих метрик говорит о преимущественном краткосрочном планировании деятельности компании, что присуще неустойчивым компаниям.

Такие связки, как: «рынки» и «новый», «риск» и «отсутствие», говорят о благоприятных внешних факторах, способствующих устойчивому развитию компании. Логично было бы предположить, что именно такие компании наиболее устойчивы в долгосрочной перспективе.

В итоге были сформированы индексы $C^{bankrupt}$ и $C^{healthy}$ (табл. 5).

Таблица 5

Суммарная статистика по индексам $C^{bankrupt}$ и $C^{healthy}$

Метрика	Компании-банкроты					Финансово здоровые компании				
	mean	median	min	max	st.dev.	mean	median	min	max	st.dev.
$C^{bankrupt}$	0.21	0.20	0.00	0.65	0.15	0.21	0.20	0.00	0.62	0.13
$C^{healthy}$	0.13	0.09	0.00	0.43	0.13	0.38	0.37	0.04	0.80	0.21

В целом, по результатам семантического анализа, можно говорить о различном характере употребления тех или иных слов в годовых отчетах компаний-банкротов и финансово здоровых компаний. Это проявляется как в количественной оценке разнообразия встречаемого контекста, так и в его качественной оценке, что подтверждает Гипотезу 2 исследования.

Финансовые показатели

Основываясь на картах SOM, из дальнейшего анализа были исключены следующие финансовые показатели, показавшие высокую коррелированность с другими индикаторами:

- EBIT / Equity.
- EBIT / CI.

- Коэффициент быстрой ликвидности.
- CL / Sales.
- Debt / TA.
- Debt / Interest.
- Логарифм активов (ln(TA)).

Оставшиеся 18 показателей были объединены с семантическими индикаторами и составили конечный массив входных данных, используемый далее в модели прогнозирования несостоятельности компании.

В соответствии с описанной ранее методологией была построена и обучена модель прогнозирования несостоятельности компании на основе ансамбля искусственных нейронных сетей. Для выделения эффекта, связанного с включением в модель семантических показателей, было проведено три общих тестирования ансамбля:

1. В ходе первого теста ансамбль обучался лишь на отобранных финансовых показателях.
2. Второй тест подразумевает включение в модель индексов $A^{bankrupt}$ и $A^{healthy}$.
3. Третий тест подразумевает включение в модель также индексов $C^{bankrupt}$ и $C^{healthy}$.

Результаты классификации ансамбля, обученного только на финансовых показателях, представлены в таблице 6.

Таблица 6

Результаты тестирования индивидуальных моделей ансамбля нейросетей, обученного на основе только финансовых показателей

№ сети	Доля верно распознанных компаний, %	Ошибка 1-го рода, %	Ошибка 2-го рода, %	Коэффициент К
HCM1	62	17	63	0,35
HCM2	71	17	44	0,28
HCM3	74	6	50	0,23
HCM4	76	17	31	0,23
HCM5	71	50	6	0,32
HCM6	71	6	56	0,26
HCM7	68	44	19	0,34
HCM8	76	28	19	0,24
HCM9	76	28	19	0,24
HCM10	62	6	75	0,33
Среднее по ансамблю	79,4	5,6	37,5	0,18

Как видно из таблицы 6, после процедуры объединения результатов индивидуальных моделей доля верно распознанных компаний составила 79,4% на тестовой выборке. Этот результат превосходит предсказательную способность любой индивидуальной модели в ансамбле. То же самое касается не только общей предсказательной способности, но и ошибки первого рода. Таким образом, можно говорить об эффективности применения байесовского ансамбля для уменьшения общей доли неверно-классифицированных предприятий, что в свою очередь подтверждает Гипотезу 1.

Итоговые результаты 100 циклов тестирования данного ансамбля представлены в таблице 7. В таблице отражены как средние результаты основных показателей эффективности ансамбля, так

и результаты лучшей и худшей системы. Определение лучшей и худшей модели происходило на основании итогового критерия, рассчитанного на обучающей выборке.

Таблица 7

**Результаты тестирования ансамбля нейросетей, обученного на основе только финансовых показателей.
Итоги 100 циклов обучения**

Критерий	Среднее значение		Лучшая система		Худшая система	
	Обучающая выборка	Тестовая выборка	Обучающая выборка	Тестовая выборка	Обучающая выборка	Тестовая выборка
Доля верно распознанных компаний	92,7%	79,9%	98,0%	82,4%	80,0%	73,5%
Ошибка 1-го рода	3,5%	15,1%	0,0%	11,1%	3,8%	5,6%
Ошибка 2-го рода	11,5%	25,8%	4,2%	25,0%	37,5%	50,0%
Итоговый крит. К	0,07	0,19	0,02	0,17	0,18	0,23

Полученные результаты говорят о высокой общей предсказательной способности модели на обучающей выборке и умеренно высокой – на тестовой выборке. Ключевой особенностью моделей является низкая ошибка первого рода, что вызвано характером отбора индивидуальных моделей в ансамбле по критерию. Среднее значение общей доли верно классифицированных компаний на тестовой выборке среди 100 циклов обучения составило 79,9%. Что при значении ошибки первого рода в 15% говорит о сравнимости данной модели с результатами предыдущих исследований в данной области.

Наибольший интерес представляет включение в модель семантических показателей, построенных ранее. Итоги тестирования модели ансамбля с использованием индексов $A^{bankrupt}$ и $A^{healthy}$ представлены в таблице 8.

Таблица 8

**Результаты тестирования ансамбля нейросетей, обученного на основе финансовых показателей и индексов.
Итоги 100 циклов обучения**

Критерий	Среднее значение		Лучшая система		Худшая система	
	Обучающая выборка	Тестовая выборка	Обучающая выборка	Тестовая выборка	Обучающая выборка	Тестовая выборка
Доля верно распознанных компаний	99,1%	87,3%	100,0%	94,1%	98,0%	79,4%
Ошибка 1-го рода	0,4%	8,0%	0,0%	5,6%	0,0%	16,7%
Ошибка 2-го рода	1,3%	17,9%	0,0%	6,3%	4,2%	25,0%
Итоговый крит. К	0,01	0,12	0,00	0,06	0,02	0,21

При сравнении данных результатов с предыдущими видно, что включение в модель показателей, основанных на словах-маркерах, существенно повышает общую предсказательную способность ансамбля как на обучающей, так и на тестовой выборке. Так, доля верно распознанных компаний на тестовой выборке возросла в среднем на 7,4% и составила 87,3% (11,7% и 94,1% соответственно, если сравнивать лучшие системы). Однако основным преимуществом полученной модели ансамбля является существенно меньшее значение ошибки первого рода, снизившееся практически в два раза по сравнению с предыдущей моделью ансамбля им составившее 8%. Также стоит отметить, что значение ошибки второго рода снизилось, хоть и не так существенно.

Далее представлены результаты тестирования модели ансамбля при использовании индексов $A^{bankrupt}$, $A^{healthy}$, $C^{bankrupt}$ и $C^{healthy}$ (см. табл. 9).

Результаты тестирования ансамбля нейросетей, обученного на основе финансовых показателей и всех семантических индексов. Итоги 100 циклов обучения

Критерий	Среднее значение		Лучшая система		Худшая система	
	Обучающая выборка	Тестовая выборка	Обучающая выборка	Тестовая выборка	Обучающая выборка	Тестовая выборка
Доля верно распознанных компаний	99,8%	88,7%	100,0%	94,1%	98,0%	81,5%
Ошибка 1-го рода	0,1%	8,0%	0,0%	5,6%	0,0%	16,7%
Ошибка 2-го рода	0,3%	15,8%	0,0%	6,3%	4,2%	22,3%
Итоговый крит. K	0,00	0,11	0,00	0,06	0,02	0,19

Из таблицы видно, что включение в ансамбль индексов $S^{bankrupt}$ и $S^{healthy}$ приводит к незначительному росту доли верно распознанных предприятий как на обучающей, так и на тестовой выборке. Этот эффект достигнут за счет незначительного снижения среднего значения ошибки второго рода.

Результаты применения данных индексов вместе с финансовыми показателями без включения в модель индексов A представлены в таблице 10.

Таблица 10

Результаты тестирования ансамбля нейросетей, обученного на основе финансовых показателей и индексов. Итоги 100 циклов обучения

Критерий	Среднее значение		Лучшая система		Худшая система	
	Обучающая выборка	Тестовая выборка	Обучающая выборка	Тестовая выборка	Обучающая выборка	Тестовая выборка
Доля верно распознанных компаний	98,9%	83,3%	100,0%	82,4%	98,0%	76,5%
Ошибка 1-го рода	1,4%	15,1%	0,0%	11,1%	0,0%	16,7%
Ошибка 2-го рода	0,9%	24,3%	0,0%	18,8%	4,2%	31,3%
Итоговый крит. K	0,01	0,17	0,00	0,16	0,02	0,33

В целом можно говорить о меньшей значимости индексов $S^{bankrupt}$ и $S^{healthy}$ по сравнению с индексами $A^{bankrupt}$ и $A^{healthy}$ как в случае отдельного их применения, так и в случае совместного использования в модели ансамбля. Отчасти это объясняется наличием связи между индексами и A , что подтверждается схожим узором весов карт SOM.

По результатам трех тестирований можно заключить, что анализ текстовой информации, заключенной в годовых отчетах компаний, позволяет увеличить классификационные возможности модели предсказания несостоятельности компании, основанной исключительно на финансовых данных. Данный эффект проявляется как в увеличении общей доли верно распознанных компаний, так и в значительном снижении ошибки первого рода, что подтверждает Гипотезу 3 исследования.

Список использованной литературы

1. Макеева Е.Ю., Горбатков С.А., Белолипец И.И. Выбор системы экономических показателей для диагностики и прогнозирования банкротств на основе нейросетевого байесовского подхода // Вестник Финансового университета. 2013. № 76. С. 50–62.
2. Bellovary, J.L., Giacomino, D.E., Akers, M.D. (2007), A Review of Bankruptcy Prediction Studies: 1930 to the Present, *Journal of Financial Education*, 33 (2007) 1–42.

3. Clatworthy, M., Jones, M. (2006), Differential patterns of textual characteristics and company performance in the chairman's statement, *Accounting, Auditing & Accountability Journal*, 4(19) (2006) 493–511.
4. Shirata, C., Takeuchi, H., Ogino, S., Watanabe, H. (2011), Extracting key phrases as predictors of corporate bankruptcy. Empirical analysis of annual reports by text mining, *Journal of Emerging Technologies in Accounting*, 8 (2011) 31–44.
5. Smith, M., Taffler, R. (2000), The chairman's statement: A content analysis of discretionary narrative disclosures, *Accounting, Auditing & Accountability Journal*, 5(13) (2000) 624–647.