

УДК 336.7

Е. Ю. Сопельник, Л. В. Федосеева

Донской государственный технический университет,
email: katya_sopelnik@mail.ru; fedoseevvb@gmail.com

СРАВНЕНИЕ ТЕХНИЧЕСКИХ И ФУНДАМЕНТАЛЬНЫХ ПОКАЗАТЕЛЕЙ В ПРОГНОЗИРОВАНИИ ЦЕН НА АКЦИИ

Ключевые слова: контролируемое обучение, технический анализ, фундаментальный анализ, прогнозирование цены акций.

В данной статье даётся сравнение фундаментального и технического анализа при прогнозировании цен на акции с помощью моделей машинного обучения. Рассматривает вопрос о целесообразности совместного использования двух подходов. Тесты, проведённые в 140 компаниях из списка S&P 500 показывают, что модели, использующие индикаторы, основанные на фундаментальном анализе, превосходят те, которые используют индикаторы технического анализа, при этом уровень эффективности варьируется в зависимости от отрасли. Отмечается что, в 95% случаев использование комбинированных показателей приводит к снижению среднеквадратичной ошибки (СКО) по сравнению с использованием только фундаментальных или технических показателей.

E. Y. Sopelnik, L. V. Fedoseeva

Don State Technical University, email: katya_sopelnik@mail.ru;
fedoseevvb@gmail.com

COMPARISON OF TECHNICAL AND FUNDAMENTAL INDICATORS IN STOCK PRICE FORECASTING

Keywords: supervised learning, technical analysis, fundamental analysis, stock price forecasting.

This article compares fundamental and technical analysis in predicting stock prices using machine learning models. Considers the feasibility of using the two approaches together. Tests conducted in 140 companies from the S&P 500 list show that models using indicators based on fundamental analysis are superior to those using technical analysis indicators, with the level of effectiveness varying depending on the industry. It is noted that, in 95% of cases, the use of combined indicators leads to a decrease in the standard error (RMS) compared to the use of only fundamental or technical indicators.

В условиях современной экономики цены на акции определяются индексами спроса и предложения на фондовом рынке, что обусловлено решениями трейдеров в необходимости покупать или продать акции компании [1]. Гипотеза эффективного рынка (ГЭР) утверждает, что цены на акции отражают всю доступную информацию таким образом, что невозможно получить устойчивую прибыль, прогнозируя будущие цены на акции [2, 3]. Таким образом, ГЭР считает, что цены на акции следуют случайному ходу [2, 3]. В отличие от этого адаптивная гипотеза рынка (АГР) утверждает, что цены на акции предсказуемы [4] и могут быть основаны на техническом анализе и фундаментальном анализе – это две основные философские школы, которые используют трейдеры в области финансов при принятии торговых решений и прогнозировании цен на акции [1, 5, 6].

Прогнозирование цены акций на основе методов машинного обучения оказалось популярным и успешным. Тем не менее, опросы показали, что большая часть работ по применению машинного обучения к прогнозированию цен на акции была сосредоточена на использовании только технических индикаторов [8, 9]. Таким образом, существует несоответствие между подходом, применяемым трейдерами в области финансов, и подходом, применяемым исследователями в отношении типа исходных данных, которые следует использовать для составления прогноза. Кроме того, основное внимание исследователей в области машинного обучения в отношении финансового прогнозирования было в значительной степени сфокусировано на прогнозировании фондовых индексов на следующий день, а не отдельных акций.

Цель исследования

Для решения этой проблемы, в статье представлен ряд экспериментов, в которых сравниваются рабочие характеристики моделей с использованием технических, фундаментальных показателей и их комбинации при прогнозировании изменения цены акций отдельных компаний в интервале 126 дней и 252 дней. Результатом является анализ преимуществ совместного использования индикаторов, основанных на техническом и фундаментальном анализе, при прогнозировании цен на акции с помощью моделей машинного обучения.

Методы исследования

Технический анализ использует исторические цены акций компании и информацию об объёме торговли как при принятии решения о том, какой будет цена акции, так и при принятии торгового решения [1, 6].

Фундаментальный анализ рассчитывает ожидаемую цену акций на основе подробного изучения основных бизнес-факторов (прибыльности, операционной эффективности, управленческого опыта и т.д.), связанных с компанией, ее продуктами, отраслью и общей экономикой [1, 7]. С философской точки зрения, «технические аналитики и фундаментальные аналитики диаметрально противоположны друг другу», хотя эти подходы могут быть взаимодополняющими [5, 7]. Исследователи в области финансов выясняли эффективность технического, фундаментального и комбинированного подхода.

При прогнозе однодневной цены технические индикаторы опережали фундаментальные индикаторы (среднеквадратичное значение ошибки 69,9 против 82,5). Однако, при прогнозе доходности на один день вперёд, фундаментальные индикаторы опережали (1,30 против 1,38) технические. Исследования, приведённые в работе [12] показали, что использование фундаментальных показателей (составляющих FSCORE, представляющих собой систему ранжирования, использующую многочисленные финансовые коэффициенты для указания финансовой устойчивости компании) для дополнения технических индикаторов превзошло (информационное

соотношение 0,1845 против 0,1335) импульсную стратегию с использованием только технических индикаторов, когда инвестиционный горизонт составляет 6 месяцев.

Аналогично, в работе [10] проверяется гипотеза о том, даёт ли объединение фундаментального и технического анализа статистически значимый лучший прогноз для стоимости акций по сравнению с использованием только факторов фундаментального анализа (балансовая стоимость и прибыль на акцию) или индикаторов технического анализа (импульсная стратегия). Тесты показали, что фундаментальный анализ (скорректированный R-квадрат 0,7629) и технический анализ (скорректированный R-квадрат 0,7546) были эффективными методами оценки цены акций и что комбинация двух методов дала лучшие предсказания.

Результаты исследования и их обсуждение

Тип исходных данных, используемых при прогнозировании цен на акции, в значительной степени зависит от базового подхода к инвестиционному анализу: фундаментальный или технический анализ. Работа с использованием методов машинного обучения показала явное предпочтение использованию технических индикаторов в качестве входных данных [8, 9]. В большинстве исследований, рассмотренных в обзорах [8, 9], использовались только технические индикаторы, в некоторых – только фундаментальные индикаторы, и лишь немногие исследовали их комбинацию.

Для генерации прогнозов использовались стандартные и улучшенные (оптимизированные с помощью генетических алгоритмов (GA)) нейронные сети (NN) и модели регрессии опорных векторов (SVR). Результаты показали, что улучшенная модель ANN с использованием комбинированного набора индикаторов достигла наивысшего уровня точности (80,51%), за которой последовала простая модель SVM с техническими индикаторами с точностью 79,4%. В исследование были включены только 25 компаний из стран с развивающейся рыночной экономикой. При прогнозировании направления цены акций Apple

на следующий день были использованы прогнозные модели (ANN, SVR, Decision Trees (DT)) с использованием различных источников данных [20]. Источники данных включали рыночные данные (финансовые временные ряды и соотношение P/E), технические индикаторы, трафик Википедии и количество новостей Google.

Проведено сравнение производительности моделей с использованием каждого источника данных по отдельности и комбинированным способом. Модель с наилучшими характеристиками использовала все источники данных и достигла 85% коэффициента попадания при площади под кривой (AUC) 0,874. Хотя исследование было направлено на то, чтобы предсказать изменение курса акций только одной компании на один день вперёд, оно показало, что использование ряда источников данных улучшило точность прогноза. В дополнение к чрезмерной зависимости исключительно от технических индикаторов, в обзорах [8, 9] показано, что в большинстве исследований рассматривалось прогнозирование на один или несколько дней вперёд, и основное внимание уделялось прогнозированию индекса фондового рынка по сравнению с прогнозированием отдельных компаний.

С точки зрения широкой категоризации, большинство подходов к машинному обучению, использующих маркированные данные и контролируемые методы обучения, имеют тенденцию делиться на три основные группы [8]: модели, использующие одну технику машинного обучения, модели, использующие гибридную комбинацию методов машинного обучения с методами оптимизации и моделями, которые представляют собой совокупность различных отдельных моделей ANN и SVR перечислены как успешно развёрнутые методы машинного обучения [8, 9].

Для прогнозирования процентного изменения цены акций выбранной компании в будущем (через 252 и 126 торговых дней) с использованием моделей прогнозирования ANN и SVR с учётом технических индикаторов, фундаментальных индикаторов и их комбинации. Для сгенерированных прогнозов показатель RMSE был рассчитан и получен пу-

тём сравнения с фактическими данными. Эксперименты проводились с использованием языка Python и инструмента извлечения данных с открытым исходным кодом WEKA, и в нем приняли участие 140 компаний, отобранных из индекса S&S P 500, который составляет 80% рынка акций США по размеру капитала. Компании были выбраны на основании наличия достаточных данных за определённый период времени (январь 1996 г. – декабрь 2015 г.), что обеспечило включение периодов, когда на фондовом рынке проявлялись различные состояния (турбулентности, взлёты и падения и т.д.).

Для каждой компании были собраны два набора финансовых данных, технических и фундаментальных, и были созданы ежедневные данные финансовых временных рядов. Наконец, эти два набора были объединены с использованием дат для создания объединённого набора данных.

Для каждой компании получены данные цен акций (цена открытия, цена закрытия, минимальная и максимальная цена) из Quandl, и пакета TTR использовался для генерации 10 технических индикаторов. Чтобы решить проблему с отсутствующими данными, использовалось среднее значение данных за ближайшие доступные торговые дни.

Основные показатели, используемые в экспериментах, можно разделить на группы следующим образом: показатели, относящиеся к эффективности рассматриваемой компании, показатели, относящиеся к прямым конкурентам, показатели, относящиеся к отрасли, к которой принадлежит компания, и макроэкономические показатели. Два главных прямых конкурента каждой компании были определены на основе рыночной капитализации и выбраны из списка, созданного базой данных Thomson One3. База данных IBES4 содержит ежемесячные прогнозные прогнозы финансовых аналитиков компаний, а также их рекомендации по покупке / удержанию / продаже акций. Медиана ежемесячных оценок финансовых аналитиков для прибыли на акцию (EPS) за 1 год и 2 года, а также долгосрочный ожидаемый процент роста в EPS были получены для каждой компании и их конкурентов. Для устранения пропущенных значений использовался

метод «последнего переноса наблюдения». Поскольку эти оценки были доступны только на ежемесячной основе, их частота была преобразована в дневную путём деления ее на данные о ценах закрытия акций предыдущего дня.

Отраслевое обозначение компании было определено с использованием классификации, доступной на веб-сайте Yahoo Finance. Данные по дневным индексам цен для каждой соответствующей отрасли были получены с веб-сайта MSCI6, которые были преобразованы в индикаторы конвергенции и расхождения скользящих средних (MACD) для краткосрочного (с 26 днями и 12 днями) и среднесрочного (с 126 и 12 днями). Помимо этих данных, связанных с компаниями, были использованы макроэкономические показатели⁷, которые были одинаковыми для всех компаний в исследовании. Один из таких индикаторов был основан на данных по иностранной валюте, где ежедневное значение «Торгово-взвешенный индекс доллара США по отношению к основным валютам» было преобразовано в MACD (126 дней и 12 дней). Другим используемым макроэкономическим индикатором являются «фьючерсные данные S&P 500», дневное значение которых было дополнительно преобразовано с использованием математического пакета MACD (26 дней и 12 дней). Окончательный используемый макроэкономический индикатор

был получен из отношения 10-летнего постоянного коэффициента погашения, который был преобразован с использованием MACD (26 дней и 12 дней).

В экспериментах используются два различных горизонта прогнозирования: 126 дней (то есть 6 торговых месяцев) и 252 дня (то есть 1 торговый год). Наборы входных данных, используемые для различных горизонтов, были скорректированы, чтобы обеспечить их актуальность. Например, коэффициент изменения цены (ROC) технических индикаторов рассчитывается на 252/126 дней в зависимости от горизонта прогнозирования. Аналогичным образом, основные показатели будут скорректированы, если формулы для отраслевых и макроэкономических показателей будут соответствовать горизонту прогнозирования.

Данные были разделены на 80% данных обучения и 20% данных тестирования. Стандартным способом обеспечения устойчивости является использование перекрёстной проверки по K-кратному критерию. Однако, поскольку перекрёстная проверка по K-кратной схеме требует случайной выборки для формирования, тестового и обучающего наборов, она не подходит для прогнозирования финансовых временных рядов, в котором важно отделить данные тестирования от обучающих данных, так чтобы хронологический порядок данных был сохранён.

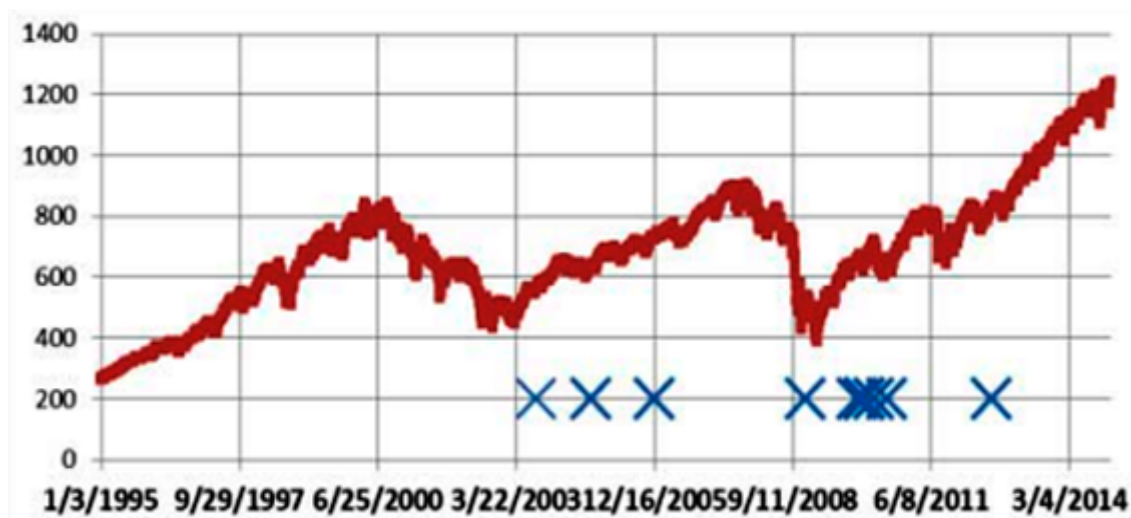


Рис. 1. Случайные контрольные точки, генерируемые на фоне общего рынка

Это гарантирует, что модель не подвергается преждевременному воздействию информации на этапе обучения (смещение вперёд), что потенциально может привести к нереально хорошей производительности. Поэтому во время симуляции данные тестирования использовались хронологически после данных тренировки. Для обеспечения устойчивости было сгенерировано 10 случайных начальных точек, и из каждой начальной точки доступные данные были разбиты на обучающие и тестовые наборы. Каждая случайная начальная точка привела к 1892 наблюдениям на тренировочном комплекте и 473 наблюдениям на испытательном множестве.

На рис. 1 наложено 10 случайных точек на общую производительность фондового рынка; можно видеть, что некоторые тестовые начальные точки (отмеченные x) возникают во время рыночных колебаний вверх, а другие – во время рыночных нисходящих разворотов.

Для экспериментов использовались модели NN и SVR, поскольку обе они были успешно применены к аналогичным задачам [8,9]. Чтобы определить параметры моделей, обучающие наборы были далее разделены на обучающие и проверочные наборы (снова с использованием 80%-20% разбиения и сохраняя хронологический порядок) и параметры модели, которые дали наименьшую частоту ошибок прогнозирования (RMSE) в проверочных наборах, были отложены.

Чтобы настроить модель NN, на основе нескольких тестов было определено идеальное количество скрытых нейронов, скорость обучения и скорость импульса. Определено количество скрытых нейронов, начиная с небольшого числа (квадратный корень из числа объектов) и увеличивая до тех пор, пока производительность модели больше не улучшится. При определении архитектуры были выполнены сценарии, показанные ниже:

– скрытых слоёв: 3,5,7 – коэффициент обучения: 0,05, 0,3, 0,6 – импульс: 0,1 0,3,0,7

Для SVR модели значения C и гамма были протестированы по ряду различных сценариев, как показано в Таблице IV, для определения оптимальной калибровки модели:

– Значения C: 0,125, 0,5, 2 – Значение гаммы: 0,01953, 0,125, 0,5, 1

В ходе эксперимента мы сравнили относительную эффективность технического анализа с фундаментальным анализом с точки зрения машинного обучения. Прогнозирующие характеристики (RMSE) моделей NN и SVR с использованием наборов технических и фундаментальных данных были оценены с использованием данных испытаний и сопоставлены для каждой компании в исследовании. На рис. 2 показано среднее среднеквадратичное значение для каждой отрасли при использовании разных входных данных для прогнозирования на 126 дней.

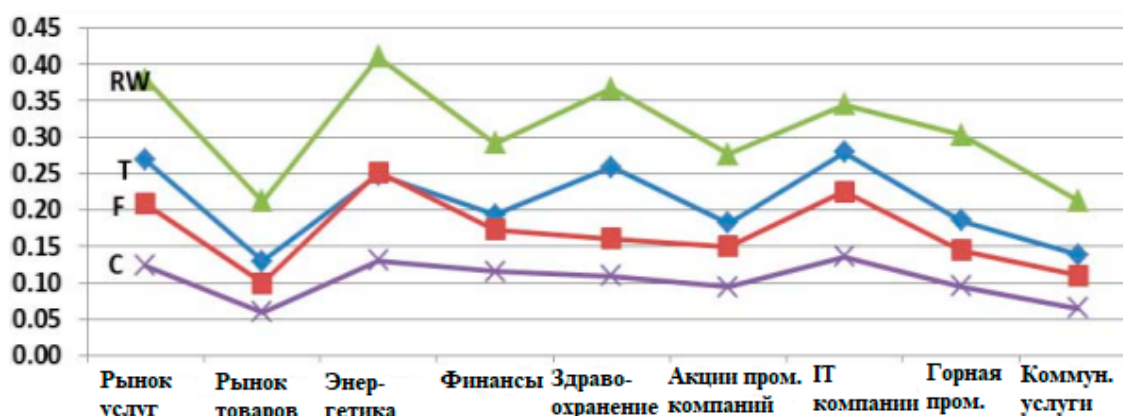


Рис. 2. Среднеквадратическое отклонение моделей, использующих технические (Т), фундаментальные (F) и комбинированные (С) входные данные для прогнозирования данных на 126 дней

Сравнение моделей, основанных (F) и технических (T) индикаторов показывает, что в среднем модели, основанные на фундаментальном анализе (среднеквадратичное значение 0,1464), превосходят модели, основанные на техническом анализе (среднеквадратичное значение 0,1693), независимо от сектора компании. Разрыв между показателями прогнозирования моделей, использующих технические и фундаментальные индикаторы, меньше для фирм в таких секторах, как финансы и энергетика, в то время как разрыв в области здравоохранения больше. На рисунке 3 показано среднее среднеквадратичное среднее значение для каждой отрасли при использовании различных входных данных для прогнозирования на 252 дня, где можно наблюдать те же тенденции, что и при прогнозировании на 126 дней.

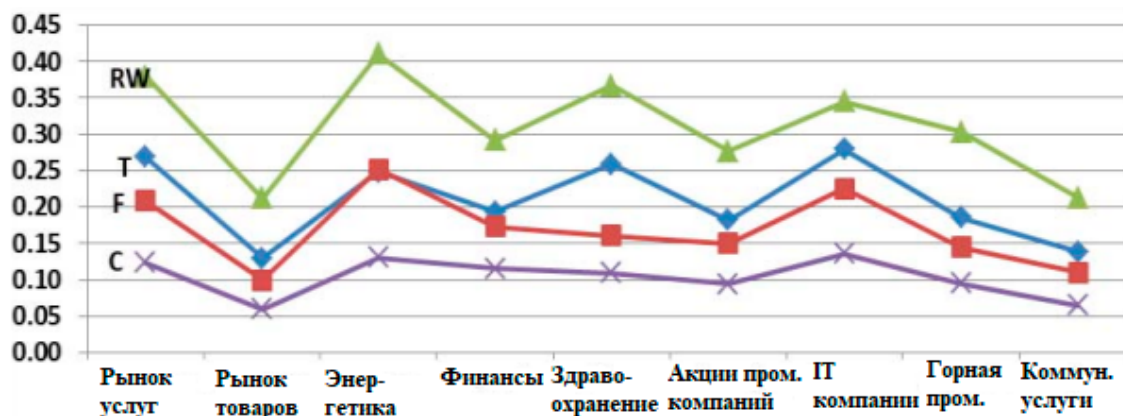


Рис. 3. Среднее среднеквадратичное отклонение моделей с использованием технических (T), фундаментальных (F) и комбинированных (C) входных данных для прогнозирования на 252 дня

Сектор	Кол-во компаний	126 дней				252 дня			
		ANN		SVR		ANN		SVR	
		T	F	T	F	T	F	T	F
Рынок услуг	19	32	68	5	95	26	74	16	84
Рынок товаров	15	40	60	13	87	7	93	13	87
Энергетика	12	50	50	33	67	58	42	50	50
Финансы	14	43	57	14	86	43	57	14	86
Здравоохранение	18	11	89	-	100	-	100	-	100
Акции пром. компаний	27	37	63	22	78	26	74	7	93
IT компании	13	45	55	-	100	36	64	-	100
Горная промышленность	11	45	55	9	91	9	91	18	82
Всего	140	36	64	14	86	24	76	14	86

Рис. 4. Процент удачного прогноза при использовании разных индикаторов

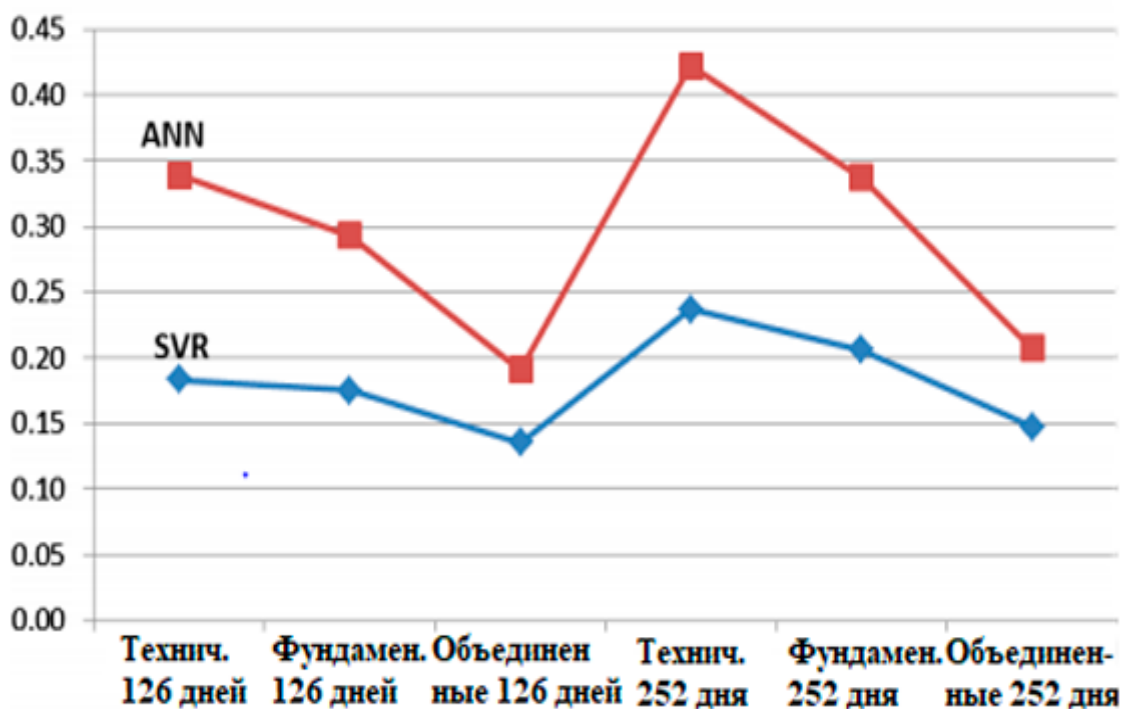


Рис. 5. Сравнение ANN и SVR для прогнозирования задач с различными горизонтами и использованием технических, фундаментальных и комбинированных входных наборов

Независимо от того, составляет ли горизонт прогнозирования 126 дней или 252 дня, в фундаментальные индикаторы (F) модели превосходят технические (T); в определенных отраслях, таких как здравоохранение, это преимущество выше. Кроме того, модели машинного обучения, независимо от горизонта прогнозирования или используемых входных данных, превосходят метод случайного блуждания. Таблица V показывает процент случаев, когда модели, использующие технические (T) или фундаментальные (F) -индикаторы, были лучше, с разбивкой по горизонту прогнозирования и методу машинного обучения (NN против SVR). В большинстве случаев (за исключением моделей NN, применяемых к компаниям в энергетическом секторе), независимо от горизонта прогнозирования и используемого метода машинного обучения, модели, использующие фундаментальные индикаторы, работают лучше, чем модели, использующие технические индикаторы. Это превосходство более очевидно при использовании SVR. При рассмо-

трени модели на основе NN процент случаев, когда модели, использующие фундаментальные индикаторы, превосходят показатели, использующие технические индикаторы, увеличивается, когда горизонт инвестиций увеличивается со 126 до 252 дней.

Рисунок 4 показывает, что модели прогнозирования компаний в секторах здравоохранения и информационных технологий имеют сильное предпочтение использованию фундаментальных показателей независимо от используемого метода машинного обучения или инвестиционного горизонта. Выполнение тестов статистической значимости ($p=0,05$) подтвердило, что во всех случаях, когда основанные на индикаторах модели превосходили технические показатели, основанные на технических показателях, результаты не наблюдались по случайности.

SVR доминировал над NN независимо от горизонта прогнозирования или отрасли компании, и особенно, когда использовались фундаментальные индикаторы, как показано на рисунке 5.

Выводы

Методы машинного обучения были успешно применены, чтобы узнать из прошлых движений цен акций компании и генерировать прогнозы на будущее. Специалисты в области финансов утверждают, что совместное использование технических и фундаментальных показателей может быть полезным, в то время как в центре внимания прогнозирования на основе машинного обучения в основном участвуют только технические показатели.

Эксперименты с участием 140 компаний из списка S&P 500 показали, что при прогнозировании изменения цены акций через 126 или 252 дня модели, использующие фундаментальные индикаторы (среднее значение RMSE 0,1464 и 0,1685 соответственно), превосходят модели, использующие технические индикаторы (среднее значение

RMSE 0,1693 и 0,2110 соответственно). Эксперименты также показали, что изменения стоимости акций компаний в секторах здравоохранения и информационных технологий лучше отражаются в фундаментальных показателях, чем в технических показателях. Тем не менее, в случае финансовых и энергетических компаний лучшие показатели прогнозирования, основанные на использовании фундаментальных показателей, менее выражены. Комбинация показателей привела к улучшению прогнозирования (статистически значимо, значение $p < 0,05$) в более чем 95% случаев. Таким образом, можно сделать вывод, что индикаторы, основанные на техническом и фундаментальном анализе, следует использовать вместе при прогнозировании цен на акции с использованием моделей на основе машинного обучения.

Библиографический список

1. М. С. Томсетт, Начало работы с техническим анализом, Иллюстрированное издание. Хобокен, Сингапур: издательство John Wiley, 2015.
2. Ф. Кампанелла, М. Мустилли и Э. Д'Анджело, «Гипотеза эффективного рынка и фундаментальный анализ: эмпирический тест на европейском рынке ценных бумаг», «Обзор экономики и финансов», с. 27–42, 2016.
3. Б. Г. Малкиель, «Случайный путь по Уолл-стрит»: проверенная временем стратегия успешного инвестирования, 11-е пересмотренное издание. W. W. Norton & Company, 2016.
4. А. В. Ло, Адаптивные рынки Финансовая эволюция со скоростью света. Пресс Принстон, 2017.
5. М. Кранц, Фундаментальный анализ для чайников. Индианаполис, Индиана, США: Wiley Publishing, 2016.
6. Б. Рокфеллер, Технический анализ для чайников. Wiley Publishing, Индианаполис, 2011.
7. С. Р. Кавальканте, С. Р. Бразилейру, Л. Ф. В. Соуза, П. Дж. Нобрега и А. Л. Оливейра, «Вычислительный интеллект и финансовые рынки»
8. Экспертные системы с приложениями. – 2016. – Т. 55. – С. 194–211.
9. Б. Ванстоун и К. Тан, «Обзор применения мягких вычислений в инвестиционной и финансовой торговле», Материалы австралийской и новозеландской конференции по интеллектуальным информационным системам, с. 211-216. Интеллектуальные информационные системы Австралии и Новой Зеландии, Сидней, 2003.
10. Х. Чен, Л. Ченг-Фью и С. Вэй-Кан, «Техническая, фундаментальная и комбинированная информация для отделения победителей от проигравших». Pacific-Basin Finance Journal, том. 39, с. 224-242, 2016.
11. Д. Чандвани и С. М. Салуджа, «Индикаторы рынка в индийском контексте», «Международный журнал компьютерных приложений», том 92 (11), с. 8-17, 2014.
12. Б. Венг, А. М. Ахмед и М. Ф. Мегехед, «Прогноз движения на один день вперед на фондовом рынке с использованием разнородных источников данных», «Экспертные системы с приложениями», вып. 79, с. 153-163, 2017.