

СОВРЕМЕННЫЕ МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ: ПРИМЕНИМОСТЬ НА РЫНКЕ ЦЕННЫХ БУМАГ

Алексей Сергеевич ДМИТРИЕВ

аспирант департамента банковского дела и финансовых рынков,
Финансовый университет при Правительстве Российской Федерации,
Москва, Российская Федерация
alex.inc2015@yandex.ru
ORCID: отсутствует
SPIN-код: 9721-7046

История статьи:

Рег. № 127/2022
Получена 09.03.2022
Получена
в доработанном виде
23.03.2022
Одобрена 06.04.2022
Доступна онлайн
30.05.2022

УДК 336.763

JEL: C45, C53, C88,
G11, G12

Ключевые слова:

интеллектуальный
анализ данных,
рынок ценных бумаг,
эконометрические
модели, управление
портфелем,
машинное обучение

Аннотация

Предмет. Методы интеллектуального анализа данных и их применимость на рынке ценных бумаг.

Цели. Выявить особенности применения на фондовом рынке современных методов интеллектуального анализа данных, обозначить основные постулаты для построения комбинированной модели интеллектуального анализа данных.

Методология. Используются логический и системный подходы, общенаучные методы анализа и синтеза, сравнительного анализа.

Результаты. На основе анализа научных работ и моделей, предложенных авторами этих исследований, выявлены особенности применения на фондовом рынке современных методов интеллектуального анализа данных. В соответствии с результатами, которых добились авторы рассматриваемых работ, обозначены основные постулаты для построения модели интеллектуального анализа данных.

Область применения. Результаты могут быть использованы участниками финансового рынка, органами государственной власти, а также научно-исследовательскими и образовательными организациями.

Выводы. Совмещение нескольких методов, практикуемых в различных моделях интеллектуального анализа данных, и использование при этом инструментария фундаментального и технического анализа и механизма идентификации волатильности активов дает возможность создать принципиально новую комбинированную систему анализа стоимости и доходности активов и портфельного управления.

© Издательский дом ФИНАНСЫ и КРЕДИТ, 2022

Для цитирования: Дмитриев А.С. Современные методы интеллектуального анализа данных: применимость на рынке ценных бумаг // *Финансы и кредит*. — 2022. — Т. 28, № 5. — С. 1178 — 1196.
<https://doi.org/10.24891/fc.28.5.1178>

Интеллектуальный анализ данных можно рассматривать как часть методов анализа данных, который так же имеет совокупность механизмов работы с данными, обнаружения ранее неизвестных, нетривиальных, практически полезных и интерпретируемых знаний в наборах данных, он включает в себя так называемые методы машинного обучения, методы эволюционного программирования и другие. По этим причинам далее иногда будет применяться такой термин, как «машинное обучение» для большей конкретизации сути обсуждения.

Важно указать, что методы интеллектуального анализа данных могут быть использованы и для работы с большими данными, и для сравнительно малых объемов, в то же время в качестве критерия достаточного количества данных применяется сама область исследования или же непосредственно алгоритм анализа. Более того, хотелось бы отметить, что применение интеллектуального анализа данных уже достаточно распространено во многих сферах [1], в том числе в экономике и финансах [2]. Например, согласно исследованию [3], в котором проведен сравнительный анализ частоты применения различных эконометрических моделей и методов для решения задач анализа и прогнозирования финансовой информации, наибольшей популярностью среди ученых пользуются нейронные сети либо различные модификации методов интеллектуального анализа данных.

Если опираться на опыт современной финансовой истории, можно сказать, что возрастает актуальность использования современных технологий интеллектуального анализа данных применительно к сфере инвестиций на фондовом рынке. Их грамотное использование дает возможности для автоматизации рутинных задач, снижения роли человеческого фактора в принятии решений и повышения качества оценки активов. Анализируя литературу, посвященную изучению методов анализа на рынке ценных бумаг [4–6], мы выделили ключевые требования к методам интеллектуального анализа данных на рынке ценных бумаг, предназначенные для построения комплексной модели прогнозирования цен на активы и управления портфелем:

- 1) высокая точность;
- 2) прибыльность;
- 3) сочетания методов технического и фундаментального анализов;
- 4) требуемая частотность операций на рынке ценных бумаг;
- 5) интерпретируемость;

б) фильтры для зашумленности торговых данных;

7) сложность/простота реализации.

Важно, чтобы высокая точность сопровождалась прибыльностью выбранной стратегии. При этом существенной является не столько сама прибыль, сколько ее величина, так как прибыльность стратегии должна превышать доходность по долгосрочным долговым государственным облигациям или доходность безрисковых биржевых индексов.

Технический анализ основан на постулате о том, что все объективные данные о компании уже включены в текущую цену акций. По этой причине предполагается, что изменение рыночной стоимости акции не является прямой реакцией на события, а зависит лишь от оценки событий участниками рынка. Отметим, что низкая точность прогноза характерна для краткосрочного периода, тогда как долгосрочное прогнозирование показывает более высокие результаты. Данная ситуация объясняется временной замедленной реакцией рынка на изменения отдельных финансовых показателей. Фундаментальный анализ основан совершенно на других принципах исследования активов, поэтому при опоре на методы машинного обучения краткосрочные прогнозы являются не совсем подходящей задачей, в отличие от того случая, когда за базис берут технический анализ.

Однако разница построения модели состоит лишь в исходных данных и некоторых настроечных деталях, которые можно подобрать динамически и итеративно. В том случае можно также предположить, что подобная гибкость методов интеллектуального анализа данных, при котором одна и та же модель может использоваться для торговли и в короткой позиции, и в длинной, говорит о другом подходе к анализу финансовой информации, нежели, например, в классических эконометрических моделях. Данная идея позволяет эластично реагировать на сложившуюся конъюнктуру рынка, а также затрачивать значительно меньшее количество усилий на трансформирование модели в новой экономической ситуации, в то время как традиционные идеи имеют свою строгую гранулярность в применении к тому или иному финансово-экономическому явлению. Из описанных выводов можно прийти к заключению, что требование к частоте операций является исключительно прерогативой выбора инвестора и его торговой стратегии.

Интерпретируемость выводов моделей базируется на достаточно простом принципе, а именно: понятно ли пользователю то, как принимает решение

алгоритм, можно ли эти шаги визуализировать и в каком виде представить выводы моделей до того, как они примут решение. Например, С. Рудин [7] согласна с тем, что зачастую высокая степень интерпретируемости моделей интеллектуального анализа данных (особенно это касается сложных архитектур нейронных сетей) ставится под сомнение, хотя существуют такие методы, как градиентный бустинг и различные методы деревьев, которые достаточно хорошо воспринимаются человеком.

Несомненно, эконометрические модели, на которых стоит финансово-математическая основа управления портфелем, в данном ключе высоко интерпретируемы человеком ввиду их идейной составляющей, не предполагавшей алгоритмического решения. Однако стоит упомянуть и тот факт, что модели интеллектуального анализа данных, которые используются для прогнозирования цен на активы и/или для управления портфелем чаще всего являются комбинацией нескольких моделей.

Подобные модели работают и с категориальными признаками, закодировав их особым образом, и с числовыми признаками, то есть учитывают и текстовую информацию, влияние которой невозможно учесть только эконометрическими методами, что дает некоторые преимущества. В то же время асимметричность информации накладывает свои риски на модель машинного обучения, избежать которые может только качественно построенная модель. В то же время подобной проблемы в традиционном подходе не существует. Таким образом, меньшая интерпретируемость современных алгоритмических решений со стороны интеллектуального анализа данных экономическим агентом является обоснованной. Более того, итеративный способ решения задач оптимизации увеличивает проблему интерпретируемости результатов анализа моделью.

Можно заключить, что в таком разрезе классические подходы к анализу финансовой информации и построению на ее основе системы прогнозирования цен на активы и/или управления портфелем и описанные методы разное смотрят на анализ финансовой информации и принятие решений на ее основе.

Фильтрация шумов в торговых данных — это важнейшая часть анализа данных, без которой качество любой модели на фондовом рынке будет низким. И в этом же постулате сталкиваются два различных подхода к измерению данных. Традиционный подход, во-первых, исходит из априорных знаний о рынке или активах, во-вторых, строится зачастую на оценке стационарности временных рядов [8], экспоненциальном сглаживании или регрессионном анализе в своем базисе. Из этого вытекает

несколько проблем, первой из которых — неспособность воспринять конъюнктурные изменения на рынке [9].

Иначе говоря, в определенный момент такие модели либо перестают адекватно работать, требуя кардинальных изменений, либо сильно снижают свое качество. Что же касается фильтрации шумов при стационарной динамике, эти модели будут действительно показывать результаты, прибыльность которых не вызывает сомнений. Однако в то же время рынок является крайне динамичным, имея в себе определенную энтропию множества факторов, которые и отражаются в цене и влияют на портфель в целом, с чем могут не справиться модели классического подхода. Например, лаговый оператор в некоторых классических моделях будет некорректен из-за требований к изменению параметров уравнений авторегрессии. Соответственно, модели управления портфелем, которые основываются на ценах из указанной модели будут автоматически некорректны и будут подталкивать к неверным решениям.

Другим примером разного подхода к данным может послужить тот факт, что при оценке ценных бумаг важную роль играет их реализованная волатильность. Учет реализованной волатильности в анализе ценных бумаг может быть как агрегированный (по портфелю), так и специфицированный, то есть по каждой бумаге. К примеру, в классическом подходе может быть необходим учет реализованной дисперсии, оценки шкал реализованной волатильности с попыткой рассчитать с учетом шума или использовать, например, ядро Барндора — Нильсена [10] для вычисления реализованной дисперсии, а эффективность инвестиционного портфеля мерить, например, моделью Шарпа [11].

В то же время при интеллектуальном анализе данных нам необходим учет реализованной волатильности для каждой бумаги наподобие того, как это делается при градиентном спуске, когда за высокие значения модель штрафуют. И потому нам будет важна величина штрафа, которая поможет определять риск от волатильности бумаги. При этом возможны и другие варианты решений через ансамблевые алгоритмы (даже с привлечением, например, модели Шарпа, которая будет встроена в процесс иным образом) или нейронные сети. Более того, изменить состав портфеля будет несколько проще при использовании методов интеллектуального анализа данных ввиду алгоритмов запоминания весов моделей и комбинированности подходов.

Хотелось бы сказать, что область интеллектуального анализа данных достаточно широка, потому существует множество методов, которые могут

работать в разных условиях, с разными данными и предназначаются для разных задач. Важно отметить, что преимуществом большинства этих подходов, которые раскрыты в *табл. 1, рис. 1–6* является комбинированность относительно данных, что подразумевает возможность обучить модели на множестве вариантов, которые могут произойти на рынке, смоделировав разные варианты развития событий. Объединив разнонаправленные модели, например, на анализе эмоций и предсказании цены актива, мы можем получить модель, которая будет учитывать и то, что происходит в социальной структуре общества, мимикрировав под человеческий анализ, и то, что закладывается в цене на самом рынке, тем самым будет сформирован совершенно иной подход к анализу рыночной информации. Таким образом, это служит дополнительным фильтром для зашумленных данных, помимо множества тренировок, речь о которых шла ранее.

Последним, но не по важности, является требование к сложности или простоте реализации метода. Фактически данное требование стало краеугольным камнем для моделирования системы управления портфелем, оценки активов и принятия рисков. Помимо того, что было раскрыто ранее, существуют и разные подходы к тому, как работать с историческими данными. Стационарные модели пытаются свести к стационарности и грамотно выделить периоды для приведения к тренду. С учетом проблем с изменениями на рынке, о которых уже говорилось, такие модели при «мутации» рынка требуют корректировки.

В то же время модели, основанные на методах машинного обучения, имеют альтернативные способы кластеризации периодов, например, на основе деревьев решений. При достаточном количестве обучения модели возможно ее переустройство, а также внесение изменений в модель управления портфеля и даже принятие решения о реформировании портфеля. Однако в отличие от традиционных методов в таком случае требуется применение более двух моделей, работающих синхронно, интерпретация выводов которых падает. Более того, частота операций, о которых говорилось, также будет вносить свои коррективы в сложность реализации.

Исходя из описанных требований были проанализированы основные методы и модели интеллектуального анализа данных на рынке ценных бумаг и выделены их сильные и слабые стороны для создания оптимальной модели, которая отвечала бы указанным требованиям. В первую очередь хотелось бы отметить, что разбор достоинств моделей, базой которых служит метод опорных векторов (от англ. Support Vector Machine и Support

Vector Regressor) представлены в работе Дж. Чжэн, Ю. Ван, С. Ли, Х. Чен [12] (*рис. 1 и табл. 1*). Другие важные модели и методы были представлены в работах С. Барак, А. Арджманд, С. Ортобелли, М. Вийх, Д. Чандола, В. Ананд, А. Кумар, П. Гош, А. Нойфельд, Дж. Кешари [13–15] (*рис. 2 и табл. 1*). Наиболее часто используемые модели, которые уже реализуют ансамблевые методы и нейронные сети, а также наиболее передовые гибридные модели на основе, например, генетических алгоритмов представлены в работах других авторов, таких как Ю. Чен, К. Лю, Ю. Се, М. Ху, Г.Г. Огнев, Е.Ю. Щетинин¹ (ансамбль деревьев с градиентным спуском и нейронная сеть, *рис. 3 и табл. 1*), А. Кумар, З. Лей, С. Чжан, Дж. Чанг, В. Ту, С. Ю, С. Цинь (нейронная сеть, на основе классификационного алгоритма, *рис. 4 и табл. 1*) и Х. Хунг, Ю. Чуанг, М. Ву, П. Коратамадди, К. Вадхвани, М. Гупта, С. Сандживи, Э. Ахмади, М. Джасеми, Л. Монплеизир, М. Набави, А. Махмуди, П. Джем (комитет нейронных сетей, гибридные модели и обучение с подкреплением от англ. Reinforcement learning, *табл. 1, рис. 5 и 6*) [16–21].

Анализ информации, представленной на рисунках и в таблице, позволяет сформулировать выводы, касающиеся принципов адаптации методов интеллектуального анализа данных к задачам, связанным с оценкой активов и портфельным управлением на рынке ценных бумаг. Сформулируем главные из них.

1. Наиболее качественные результаты показывают модели на основе методов SVM, градиентного бустинга и нейронных сетей.
2. Методы должны быть совмещены в одну модель, с привлечением методов, используемых в нейронной сети, на основе классификационного алгоритма (Sentiment Analysis) и в гибридных моделях.
3. Торговый алгоритм, встроенный в модель, целесообразно создавать при обосновании логики решений в рамках фундаментального анализа в комбинации с техническим либо, наоборот (технический в комбинации с фундаментальным), в зависимости от целей инвестора.
4. Применимость модели зависит от степени стабильности рыночной ситуации, что требует дополнительных критериев индикации стабильности и идентификации критических зон волатильности активов.

¹ Огнев Г.Г., Щетинин Е.Ю. Исследование глубоких нейронных сетей с LSTM архитектурой для прогнозирования финансовых временных рядов // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем. М.: РУДН, 2020. С. 280–283.

Подводя итоги проведенного исследования, отметим главное. Методы интеллектуального анализа данных на сегодняшний день следует рассматривать как альтернативу традиционным методам анализа и управления портфелем и одновременно их закономерное продолжение ввиду возможности работать с большим объемом разнообразной информации и использовать подходы, преодолевающие недостатки и ограничения традиционных методов.

При этом проведенные нами методологические изыскания и изучение опыта создания техник анализа на рынке ценных бумаг приводят нас к выводу о том, что для построения современной модели интеллектуального анализа данных в целях оценки активов и портфельного управления на рынке ценных бумаг требуется комбинирование нескольких методов разных моделей, совместное использование техник и инструментов фундаментального и технического анализа, дополненное разработкой систем идентификации волатильности активов.

Таблица 1
Сравнение базовых методов интеллектуального анализа данных

Table 1
Comparison of basic methods of data mining

Базовый метод	Достоинства базового метода	Недостатки базового метода
Метод опорных векторов (SVM) (SVR)-метод)	<ol style="list-style-type: none"> 1. Возможность интеграции в другие модели. 2. Зачастую эквивалентен двухслойной нейронной сети, что значительно проще в реализации и управлении для управляющего 	<ol style="list-style-type: none"> 1. Неустойчивость к выбросам. 2. Отбор признаков возможен только через перебор. 3. Проблемы спрямляющих пространств, которые могут ухудшить результаты анализа данных
Дерево решений и случайный лес	<ol style="list-style-type: none"> 1. Для решения задачи может быть создано несколько деревьев, которые отличаются условиями классификации. 2. Простота отбора лучшего результата. 3. Высокая интерпретируемость 	<ol style="list-style-type: none"> 1. Средняя устойчивость к шумам, что исправляется ансамблевыми методами. 2. Вероятность переобучения при постоянной «стрижке», что также мешает нахождению лучшего результата. Random forest из множества деревьев исправляет часть этой проблемы. 3. Проблема нахождения оптимума из-за использования эвристики
Ансамбль деревьев и нейронная сеть	<ol style="list-style-type: none"> 1. За счет применения внутреннего бэггинга и градиентным спуском на ошибках, уменьшает минусы деревьев и лесов, и тем самым дает достаточно хорошую точность. 2. Возможность идентифицировать выбросы 	<ol style="list-style-type: none"> 1. Влияние тех методов, которые применяются в ансамбле, на итоговый результат ансамбля все равно существует, а значит, уязвимость к шумам. 2. Сложность регулирования
Нейронная сеть на основе классификации (Sentiment Analysis)	<ol style="list-style-type: none"> 1. Простота интерпретируемости. 2. Возможность интеграции в другие модели. 3. Простота двухклассовой классификации 	<ol style="list-style-type: none"> 1. Возможность переобучения при большом количестве шума, потому необходим порог чувствительности. 2. Отбор признаков возможен только через перебор. 3. Влияние ложной информации
Комитет нейронных сетей	<ol style="list-style-type: none"> 1. За счет большого количества данных и регуляризации сравнительно высокая работа с шумами в данных. 2. Хорошее качество предсказания на однородных данных 	<ol style="list-style-type: none"> 1. Низкая интерпретируемость результатов. 2. Сложность реализации. 3. Требует наибольшего количества данных среди всех методов
Гибридные модели и обучение с подкреплением	<ol style="list-style-type: none"> 1. Поиск нестандартных даже для нейронных сетей подходов. 2. Хорошее качество предсказания на однородных данных 	<ol style="list-style-type: none"> 1. Низкая интерпретируемость результатов. 2. Наивысшая сложность реализации. 3. Проблема количества управляемых переменных. 4. Проблема неполноты информации для дообучения

Источник: авторская разработка

Source: Authoring

Рисунок 1

Анализ модели на основе метода опорных векторов SVM (SVR) метод)

Figure 1

Model analysis based on the SVM (SVR) method

Метод опорных векторов (SVM (SVR)-метод)		
<p>Достоинства модели</p> <ol style="list-style-type: none"> 1. Применение данных биржевых торгов, а также результатов анализа финансовых показателей. 2. Большой объем выборки. 3. Использование отбора методов анализа. 4. Высокие показатели доходности в долгосрочном периоде 	<p>Недостатки модели</p> <ol style="list-style-type: none"> 1. Прогноз цены акции на следующий день демонстрирует низкую точность. 2. Средняя по качеству точность прогнозирования в долгосрочном периоде 	<p>Применимость на фондовом рынке</p> <ol style="list-style-type: none"> 1. Является пригодным для применения на рынке ценных бумаг с некоторыми ограничениями. 2. Применение данных биржевых торгов, финансовых показателей и их отбор доказали свою эффективность

Источник: авторская разработка

Source: Authoring

Рисунок 2

Анализ модели на основе дерева решений или случайного леса

Figure 2

Model analysis based on the Decision Tree or Random forest method

Дерево решений и случайный лес		
<p>Достоинства модели</p> <ol style="list-style-type: none"> 1. Точность прогнозирования достигается наличием разнообразных комбинаций деревьев решений. 2. Применяемость к статичным данным. 3. Возможность кластеризации периодов на рынке 	<p>Недостатки модели</p> <p>Неустойчивость к нестатичным данным</p>	<p>Применимость на фондовом рынке</p> <p>Применима для прогнозирования низковолатильных активов либо только в ансамбле моделей (например, кластеризация методом случайного леса)</p>

Источник: авторская разработка

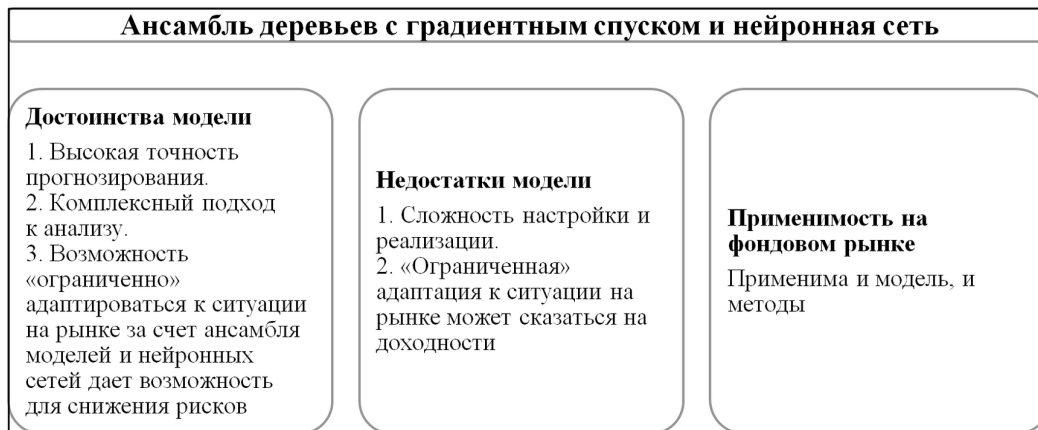
Source: Authoring

Рисунок 3

Анализ модели на основе ансамбля деревьев с градиентным спуском и нейронная сеть

Figure 3

The Gradient Descent Tree Model Analysis and the Neural Network



Источник: авторская разработка

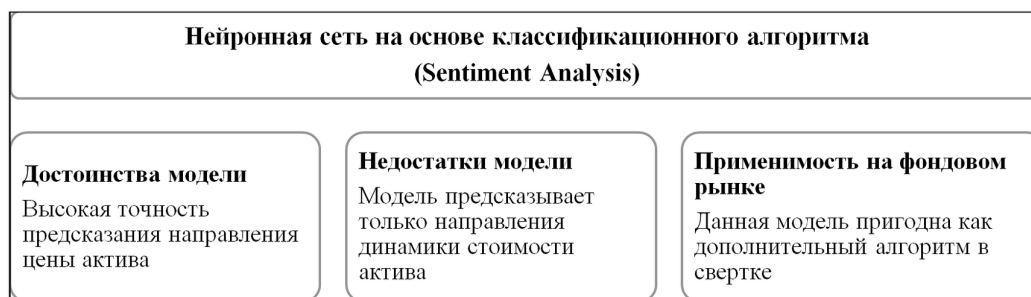
Source: Authoring

Рисунок 4

Анализ модели на основе нейронной сети, которая строится на основе классификационного алгоритма

Figure 4

Model analysis based on a neural network, which is built on the basis of a classification algorithm



Источник: авторская разработка

Source: Authoring

Рисунок 5
Анализ модели на основе комитета нейронных сетей

Figure 5
Model analysis based on a committee of neural networks

Комитет нейронных сетей		
<p>Достоинства модели</p> <ol style="list-style-type: none"> 1. За счет большого количества данных и регуляризации сравнительно высокая работа с шумами в данных. 2. Хорошее качество предсказания на однородных данных 	<p>Недостатки модели</p> <ol style="list-style-type: none"> 1. Высокая частота совершаемых операций моделью. 2. Отсутствие какой-либо адаптации к ситуации из-за скорости совершения операций, что влечет за собой высокие риски 	<p>Применимость на фондовом рынке</p> <p>Подходит только для стратегий внутридневных спекулятивных операций (например, скальпинг)</p>

Источник: авторская разработка

Source: Authoring

Рисунок 6
Анализ модели на основе гибридных моделей и обучения с подкреплением

Figure 6
Model analysis based on hybrid models and reinforcement learning

Гибридные модели и обучение с подкреплением		
<p>Достоинства модели</p> <ol style="list-style-type: none"> 1. Поиск нестандартных даже для нейронных сетей подходов. 2. Хорошее качество предсказания на однородных данных 	<p>Недостатки модели</p> <ol style="list-style-type: none"> 1. Наивысшая сложность реализации. 2. Существование проблем в реализации в ограниченных информацией условиях. 3. Слабая приспособленность к временным данным, используемым на рынке ценных бумаг 	<p>Применимость на фондовом рынке</p> <p>Подходит только как вспомогательный метод оптимизации модели для торговли и обработки специфической информации на рынке</p>

Источник: авторская разработка

Source: Authoring

Список литературы

1. *López-Robles J.R., Rodríguez-Salvador M., Gamboa-Rosales N.K. et al.* The Last Five Years of Big Data Research in Economics. *Econometrics and Finance: Identification and Conceptual Analysis. Procedia Computer Science*, 2019, vol. 162, pp. 729–736.
URL: <https://doi.org/10.1016/j.procs.2019.12.044>
2. *Hayakawa K.* Recent Development of Covariance Structure Analysis in Economics. *Econometrics and Statistics*, 2021.
URL: <https://doi.org/10.1016/j.ecosta.2021.10.002>
3. *Henrique B.M., Sobreiro V.A., Kimura H.* Literature Review: Machine Learning Techniques Applied to Financial Market Prediction. *Expert Systems with Applications*, 2019, vol. 124, pp. 226–251.
URL: <https://doi.org/10.1016/j.eswa.2019.01.012>
4. *Берзон Н.И., Буянова Е.А., Газман В.Д. и др.* Инновации на финансовых рынках. М.: ВШЭ, 2013. 424 с.
5. *Semenkova E.V., Andrianova L.N., Krinichansky K.V.* The Concept of Fair Pricing in the Regulation Framework of the Russian Securities Market. *Journal of Reviews on Global Economics*, 2018, vol. 7, pp. 562–571.
6. *Криничанский К.В., Безруков А.В.* Некоторые практические задачи модели оптимизации портфеля // Журнал экономической теории. 2012. № 3. С. 142—147. URL: <https://cyberleninka.ru/article/n/nekotorye-prakticheskie-zadachi-modeli-optimizatsii-portfelya>
7. *Rudin C.* Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 2019, vol. 1, pp. 206–215. URL: <https://doi.org/10.48550/arXiv.1811.10154>
8. *Zolfaghari M., Gholami S.* A Hybrid Approach of Adaptive Wavelet Transform, Long Short-Term Memory and ARIMA-GARCH Family Models for the Stock Index Prediction. *Expert Systems with Applications*, 2021, vol. 182. URL: <https://doi.org/10.48550/arXiv.1811.10154>
9. *Pagach D.P., Warr R.S.* Analysts Versus Time-Series Forecasts of Quarterly Earnings: A Maintained Hypothesis Revisited. *Advances in Accounting*, 2020, vol. 51, no. 100497. URL: <https://doi.org/10.1016/j.adiac.2020.100497>
10. *Floros C., Gkillas K., Konstantatos C., Tsagkanos A.* Realized Measures to Explain Volatility Changes over Time. *Journal of Risk and Financial Management*, 2020, vol. 13, iss. 6, p. 125.
URL: <https://doi.org/10.3390/jrfm13060125>

11. *Christina Dan Wang, Zhao Chen, Yimin Lian, Min Chen*. Asset Selection Based on High Frequency Sharpe Ratio. *Journal of Econometrics*, 2022, vol. 227, iss. 1, pp. 168–188.
URL: <https://doi.org/10.1016/j.jeconom.2020.05.007>
12. *Zheng J., Wang Y., Li S., Chen H.* The Stock Index Prediction Based on SVR Model with Bat Optimization Algorithm. *Algorithms*, 2021, vol. 14, iss. 10, p. 299. URL: <https://doi.org/10.3390/a14100299>
13. *Barak S., Arjmand A., Ortobelli S.* Fusion of Multiple Diverse Predictors in Stock Market. *Information Fusion*, 2017, vol. 36, pp. 90–102.
URL: <https://doi.org/10.1016/j.inffus.2016.11.006>
14. *Vijh M., Chandola D., Tikkiwal V.A., Kumar A.* Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*, 2020, vol. 167, pp. 599–606. URL: <https://doi.org/10.1016/j.procs.2020.03.326>
15. *Ghosh P., Neufeld A., Sahoo J.K.* Forecasting Directional Movements of Stock Prices for Intraday Trading using LSTM and Random Forests. *Finance Research Letters*, 2022, vol. 46, part A, no. 102280.
URL: <https://doi.org/10.1016/j.frl.2021.102280>
16. *Chen Y., Liu K., Xie Y., Hu M.* Financial Trading Strategy System Based on Machine Learning. *Mathematical Problems in Engineering*, 2020, pp. 10–23.
URL: <https://doi.org/10.1155/2020/3589198>
17. *Kumar A., Lei Z., Zhang C.* Dividend Sentiment, Catering Incentives, and Return Predictability. *Journal of Corporate Finance*, 2022, vol. 72, no. 102128. URL: <https://doi.org/10.1016/j.jcorpfin.2021.102128>
18. *Chang J., Tu W., Yu C., Qin C.* Assessing Dynamic Qualities of Investor Sentiments for Stock Recommendation. *Information Processing & Management*, 2021, vol. 58, iss. 2, no. 102452.
URL: <https://doi.org/10.1016/j.ipm.2020.102452>
19. *Hung H.-C., Chuang Y.-J., Wu M.-C.* Customizable and Committee Data Mining Framework for Stock Trading. *Applied Soft Computing*, 2021, vol. 105, no. 102277. URL: <https://doi.org/10.1016/j.asoc.2021.107277>
20. *Koratamaddi P., Wadhvani K., Gupta M., Sanjeevi S.* Market Sentiment-Aware Deep Reinforcement Learning Approach for Stock Portfolio Allocation. *Engineering Science and Technology, an International Journal*, 2021, vol. 24, iss. 4, pp. 848–859.
URL: <https://doi.org/10.1016/j.jestch.2021.01.007>

21. *Ahmadi E., Jasemi M., Monplaisir L. et al.* New Efficient Hybrid Candlestick Technical Analysis Model for Stock Market Timing on the Basis of the Support Vector Machine and Heuristic Algorithms of Imperialist Competition and Genetic. *Expert Systems with Applications*, 2018, vol. 94, pp. 21–31.
URL: <https://doi.org/10.1016/j.eswa.2017.10.023>

Информация о конфликте интересов

Я, автор данной статьи, со всей ответственностью заявляю о частичном и полном отсутствии фактического или потенциального конфликта интересов с какой бы то ни было третьей стороной, который может возникнуть вследствие публикации данной статьи. Настоящее заявление относится к проведению научной работы, сбору и обработке данных, написанию и подготовке статьи, принятию решения о публикации рукописи.

MODERN METHODS OF DATA MINING: APPLICABILITY IN THE SECURITIES MARKET

Aleksei S. DMITRIEV

Financial University under Government of Russian Federation,
Moscow, Russian Federation
alex.inc2015@yandex.ru
ORCID: not available

Article history:

Article No. 127/2022
Received 9 March 2022
Received in revised
form 23 March 2022
Accepted 6 April 2022
Available online
30 May 2022

JEL classification:

C45, C53, C88, G11,
G12

Keywords: intelligent
data analysis, securities
market, econometric
model, portfolio
management, machine
learning

Abstract

Subject. The article addresses data mining methods, their applicability in the securities market.

Objectives. The aim is to identify the specifics of using modern data mining methods in the stock market, outline basic tenets for building a combined data mining model.

Methods. The study rests on logical and systems approaches, general scientific methods of analysis and synthesis, and comparative analysis.

Results. Based on the analysis of existing works and models, the paper unveils the specifics of using modern methods of data mining in the stock market, defines the main postulates for building a data mining model. The findings can be used by financial market participants, State authorities, and research and educational organizations.

Conclusions. Today, data mining methods are an alternative to the traditional portfolio analysis and management methods, being a logical continuation of them due to the ability to work with a large amount of diverse information and use approaches that overcome the shortcomings and limitations of other methods. To build a model of data mining to evaluate assets and portfolio management in the securities market, it is necessary to combine several methods of different models, combine the fundamental and technical analysis, and create systems for identifying asset volatility.

© Publishing house FINANCE and CREDIT, 2022

Please cite this article as: Dmitriev A.S. Modern Methods of Data Mining: Applicability in the Securities Market. *Finance and Credit*, 2022, vol. 28, iss. 5, pp. 1178–1196.
<https://doi.org/10.24891/fc.28.5.1178>

References

1. López-Robles J.R., Rodríguez-Salvador M., Gamboa-Rosales N.K. et al. The Last Five Years of Big Data Research in Economics, Econometrics and Finance: Identification and Conceptual Analysis. *Procedia Computer Science*, 2019, vol. 162, pp. 729–736.
URL: <https://doi.org/10.1016/j.procs.2019.12.044>

2. Hayakawa K. Recent development of covariance structure analysis in economics. *Econometrics and Statistics*, 2021.
URL: <https://doi.org/10.1016/j.ecosta.2021.10.002>
3. Henrique B.M., Sobreiro V.A., Kimura H. Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 2019, vol. 124, pp. 226–251.
URL: <https://doi.org/10.1016/j.eswa.2019.01.012>
4. Berzon N.I., Buyanova E.A., Gazman V.D. et al. *Innovatsii na finansovykh rynkakh* [Innovations in financial markets]. Moscow, HSE Publ., 2013, 424 p.
5. Semenkova E.V., Andrianova L.N., Krinichansky K.V. The Concept of Fair Pricing in the Regulation Framework of the Russian Securities Market. *Journal of Reviews on Global Economics*, 2018, vol. 7, pp. 562–571.
6. Krinichanskii K.V., Bezrukov A.V. [Some Practical Problems of the Portfolio Optimization Model]. *Zhurnal ekonomicheskoi teorii = Journal of Economic Theory*, 2012, no. 3, pp. 142–147.
URL: <https://cyberleninka.ru/article/n/nekotorye-prakticheskie-zadachi-modeli-optimizatsii-portfelya> (In Russ.)
7. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 2019, vol. 1, pp. 206–215. URL: <https://doi.org/10.48550/arXiv.1811.10154>
8. Zolfaghari M., Gholami S. A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction. *Expert Systems with Applications*, 2021, vol. 182.
URL: <https://doi.org/10.1016/j.eswa.2021.115149>
9. Pagach D.P., Warr R.S. Analysts versus time-series forecasts of quarterly earnings: A maintained hypothesis revisited. *Advances in Accounting*, 2020, vol. 51, no. 100497. URL: <https://doi.org/10.1016/j.adiac.2020.100497>
10. Floros C., Gkillas K., Konstantatos C., Tsagkanos A. Realized Measures to Explain Volatility Changes over Time. *Journal of Risk and Financial Management*, 2020, vol. 13, iss. 6, p. 125.
URL: <https://doi.org/10.3390/jrfm13060125>
11. Christina Dan Wang, Zhao Chen, Yimin Lian, Min Chen. Asset selection based on high frequency Sharpe ratio. *Journal of Econometrics*, 2022, vol. 227, iss. 1, pp. 168–188.
URL: <https://doi.org/10.1016/j.jeconom.2020.05.007>

12. Zheng J., Wang Y., Li S., Chen H. The Stock Index Prediction Based on SVR Model with Bat Optimization Algorithm. *Algorithms*, 2021, vol. 14, iss. 10, p. 299. URL: <https://doi.org/10.3390/a14100299>
13. Barak S., Arjmand A., Ortobelli S. Fusion of multiple diverse predictors in stock market. *Information Fusion*, 2017, vol. 36, pp. 90–102. URL: <https://doi.org/10.1016/j.inffus.2016.11.006>
14. Vijh M., Chandola D., Tikkiwal V.A., Kumar A. Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*, 2020, vol. 167, pp. 599–606. URL: <https://doi.org/10.1016/j.procs.2020.03.326>
15. Ghosh P., Neufeld A., Sahoo J.K. Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Research Letters*, 2022, vol. 46, part A, no. 102280. URL: <https://doi.org/10.1016/j.frl.2021.102280>
16. Chen Y., Liu K., Xie Y., Hu M. Financial Trading Strategy System Based on Machine Learning. *Mathematical Problems in Engineering*, 2020, pp. 10–23. URL: <https://doi.org/10.1155/2020/3589198>
17. Kumar A., Lei Z., Zhang C. Dividend sentiment, catering incentives, and return predictability. *Journal of Corporate Finance*, 2022, vol. 72, no. 102128. URL: <https://doi.org/10.1016/j.jcorpfin.2021.102128>
18. Chang J., Tu W., Yu C., Qin C. Assessing dynamic qualities of investor sentiments for stock recommendation. *Information Processing & Management*, 2021, vol. 58, iss. 2, no. 102452. URL: <https://doi.org/10.1016/j.ipm.2020.102452>
19. Hung H.-C., Chuang Y.-J., Wu M.-C. Customizable and committee data mining framework for stock trading. *Applied Soft Computing*, 2021, vol. 105, no. 102277. URL: <https://doi.org/10.1016/j.asoc.2021.107277>
20. Koratamaddi P., Wadhvani K., Gupta M., Sanjeevi S. Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. *Engineering Science and Technology, an International Journal*, 2021, vol. 24, iss. 4, pp. 848–859. URL: <https://doi.org/10.1016/j.jestch.2021.01.007>
21. Ahmadi E., Jasemi M., Monplaisir L. et al. New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the Support Vector Machine and Heuristic Algorithms of Imperialist Competition and Genetic. *Expert Systems with Applications*, 2018, vol. 94, pp. 21–31. URL: <https://doi.org/10.1016/j.eswa.2017.10.023>

Conflict-of-interest notification

I, the author of this article, bindingly and explicitly declare of the partial and total lack of actual or potential conflict of interest with any other third party whatsoever, which may arise as a result of the publication of this article. This statement relates to the study, data collection and interpretation, writing and preparation of the article, and the decision to submit the manuscript for publication.