

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
Учреждение высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра Теории функций и стохастического анализа

Анализ динамики экономических активов

с помощью методов машинного обучения

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 412 группы

направления 01.03.02 Прикладная математика и информатика

код и наименование направления

механико-математического факультета

наименование факультета, института, колледжа

Ткачевой Дарьи Алексеевны

фамилия, имя, отчество

Научный руководитель:

доцент, к.ф.-м.н., доцент

подпись

А.В.Шаталина

инициалы, фамилия

Заведующий кафедрой:

д-ф.-м.н., доцент

подпись

С.П. Сидоров

инициалы, фамилия

Саратов 2025

СОДЕРЖАНИЕ

Стр.

ВВЕДЕНИЕ	3
1 Основное содержание работы	5
ЗАКЛЮЧЕНИЕ	12

ВВЕДЕНИЕ

Анализ фондового рынка является одним из наиболее перспективных направлений исследований в экономике. Инвесторы все чаще рассматривают вложение капитала в ценные бумаги как наиболее привлекательную возможность получения прибыли. В то же время, принятие решения о покупке или продаже акций считается сложным, так как цены акций колеблются случайным образом и подвержены воздействию любой информации, связанной с акциями.

Сегодня, когда машинное обучение широко применяется во многих сферах жизни, в том числе отраслях экономики, системы искусственного интеллекта ежедневно взаимодействуют с человеком и его социальным окружением. Например, умная станция «Алиса» переключает музыку по голосовой команде человека, биометрия данных, используемая в качестве формы управления идентификаторами доступа и контроля доступа, и т.д. Если говорить об экономике, нейронные сети помогают решить такие задачи, как:

1. Прогнозирование;
2. Аппроксимация;
3. Оптимизация.

Поиск алгоритмического подхода к принятию решения о покупке или продаже акций является актуальной темой. Идея заключается в нахождении решения, которое помогло бы учитывать подмножество факторов, влияющих на фондовый рынок чтобы автоматически предсказывать рост или падение цен на акции через определенный промежуток времени.

Целью данной работы было разработать и сравнить несколько моделей машинного обучения, чтобы определить, какие из них лучше всего подходят для прогнозирования динамики акций. Прогнозы будут делаться на основе акций выбранных компаний. Результатом будет не конкретные цены, а отображение текущей тенденции (движение вверх или вниз).

Задачи, которые будут решены в ходе работы:

1. Обзор материала по фондовым рынкам и машинному обучению.
2. Собрать данные по пяти показателям (Open, Close, High, Low, Volume) акций компаний «Газпром» и «Норникель» и оформить датасет.

3. Изучить построение моделей для сравнения точности предсказаний с использованием следующих методов: метод опорных векторов (SVM), метод деревьев решений (DT), метод случайного леса (RF) и метод искусственной нейронной сети (ANN);
4. На основе подготовленного датасета построить четыре модели прогнозирования, используя следующие методы: SVM, DT, RF и ANN.;
5. Анализ результатов и сравнение эффективности моделей.

Модели машинного обучения, которые будут использованы:

1. Метод опорных векторов (SVM).
2. Деревья решений (DT).
3. Случайный лес (RF).
4. Искусственная нейронная сеть (ANN).

Данная бакалаврская работа состоит из четырех разделов, введения, заключения и приложений.

1 Основное содержание работы

В первом разделе рассматривается теоретическая часть работы, основные сведения такие как: рынок, анализ рынка, экономический рынок, фондовые биржи, торговля акциями, ценная бумага, акции и др. А так же дается определение машинному обучению.

Рынок – это непрерывное экономическое взаимодействие между производителями товаров и потребителями.

Анализ рынка – это изучение показателей этого взаимодействия. Аналитики исследуют потребности клиента и предложения конкурентов. Анализ рынка основан на сборе, обработке и применении этой информации в работе компании.

Экономический рынок - совокупность экономических отношений, возникающих в связи с перемещением свободных денежных средств и иных финансовых инструментов от их владельцев к другим экономическим агентам при активном участии финансовых посредников, для размещения этих средств в целях извлечения максимальной прибыли, но сопровождающиеся высокими рисками.

Фондовые биржи – это объединение покупателей и продавцов различных ценных бумаг. На фондовых биржах торгуют различными ценными бумагами.

Торговля акциями – сделка между двумя физическими лицами, которая не оказывает прямого влияния на корпорацию, чьи акции были первоначально выпущены.

Ценная бумага – это документ, который удостоверяет имущественные права, которые осуществляются при его предъявлении. Ценная бумага выполняет свои функции, если соблюдена форма ее составления и при содержании в ней всех необходимых реквизитов. С экономической точки зрения ценные бумаги представляют собой долю капитала.

Акции - долевые ценные бумаги, выпускаемые с целью аккумуляции капитала для последующей деятельности. При владении которыми акционер получает права как на участие в управлении компанией, так и на получение дивидендов.

Московская биржа – биржевая группа, управляющая единственной в России многофункциональной биржевой площадкой по торговле акциями, облигациями, производными инструментами, валютой, инструментами денежного рынка, драгоценными металлами и товарами. Образована в декабре 2011 года в результате слияния двух основных российских биржевых групп ММВБ и РТС. Основные акционеры: Центральный банк РФ, Сбербанк России, Внешэкономбанк, ЕБРР.

Фундаментальный анализ – это анализ состояния и динамики рынка, основанный на фундаментальных экономических показателях. Целью фундаментального анализа является оценка реальной (внутренней, справедливой) стоимости компании для определения справедливой стоимости ее акций. Задачей фундаментального анализа является выявление недооцененных или переоцененных компаний и, соответственно, бумаг.

Под **техническим анализом** понимается метод прогнозирования рыночной стоимости ценной бумаги на основе информации о ее котировках, основанный на математических вычислениях и статистических данных.

Котировка – установление курса валюты, ценных бумаг, цен на товары на биржах в соответствии с действующим законодательством и сложившейся ситуацией на рынке.

Машинное обучение – это метод анализа данных, который автоматически строит аналитическую модель для определенной задачи, учитывая набор данных, описывающих эту задачу. Алгоритмы машинного обучения могут выполнять задачи как регрессии, так и классификации. Отдельных алгоритмов много, и зачастую они сильно отличаются друг от друга. Однако все алгоритмы машинного обучения используют идею «обучающих данных» для автоматического создания аналитической модели для требуемой задачи. После обучения алгоритмам предоставляются новые образцы данных для выполнения задачи (например, регрессия или классификация). Алгоритмы и составные части, используемые в данной работе, описаны ниже.

Во втором разделе собирается и оформляется датасет для прогнозирования.

Данные, которые использовались для формирования датасета, представляют из себя набор данных о котировках акций компаний "Газпром" и "Норникель" начиная с 01.07.2022 по 21.11.2024. были собраны и получены автоматически с Московской биржи:

- DATE - дата;
- TIME - время (временной шаг был выбран 1 час);
- OPEN - цена открытия;
- HIGH - самая высокая цена до закрытия;
- LOW - самая низкая цена до закрытия;
- CLOSE - цена закрытия;
- VOL - количество сделок за период.

Далее создан датасет, который состоит из 10 рассчитанных технических показателей: SMA10, WMA10, K% и D%, RSI14, Momentum14, MACD, Williams' %R, StoOsc, A/D, CCI.

Данные индикаторы станут основными входными данными для нейронной сети.

Третий раздел посвящен построению моделей машинного обучения.

Метод опорных векторов (Support Vector Machine — SVM) — это очень мощная и универсальная модель машинного обучения, способная выполнять линейную или нелинейную классификацию, регрессию и даже выявление выбросов. Она является одной из самых популярных моделей в машинном обучении.

Основная идея метода — перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с наибольшим зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей классы. Разделяющей гиперплоскостью будет гиперплоскость, создающая наибольшее

расстояние до двух параллельных гиперплоскостей. Алгоритм основан на допущении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.

Деревья решений (DT) — это непараметрический контролируемый метод обучения, используемый для классификации и регрессии. Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной, изучая простые правила принятия решений, выведенные из характеристик данных.

Случайный лес — это набор деревьев решений, где каждое дерево немного отличается от остальных. Идея случайного леса заключается в том, что каждое дерево может довольно хорошо прогнозировать, но скорее всего переобучается на части данных. Если построить много деревьев, которые хорошо работают и переобучаются с разной степенью, можно уменьшить переобучение путем усреднения их результатов.

Для реализации вышеизложенной стратегии необходимо построить большое количество деревьев решений. Каждое дерево должно на приемлемом уровне прогнозировать целевую переменную и должно отличаться от других деревьев. Случайные леса получили свое название из-за того, что в процесс построения деревьев была внесена случайность, призванная обеспечить уникальность каждого дерева. Благодаря своей гибкости, Random Forest применяется для решения практически любых задач в области машинного обучения. Сюда относятся классификации (RandomForestClassifier) и регрессии (RandomForestRegressor), а также более сложные задачи, вроде отбора признаков, поиска выбросов/аномалий и кластеризации.

Нейронные сети — это направление машинного обучения, в котором аналитическая модель строится с использованием многослойной архитектуры, где каждый слой преобразует входные данные и выдает одно или несколько выходных значений. Каждый слой состоит из определенного количества узлов (нейронов). Каждый нейрон принимает входные данные и выполняет линейное преобразование, затем применяет нелинейную функцию

к преобразованному значению и выдает результат. Эта процедура выполняется на входе для каждого нейрона в слое, но с разными весами для каждого нейрона. После этого выходные данные могут быть использованы для классификации или регрессии, либо переданы в другой слой. Процесс настройки весов всей нейронной сетью таким образом, чтобы они давали желаемый выход, является фазой обучения искусственных нейронных сетей.

Сначала к созданному датасету применен метод опорных векторов с полиномиальной гиперплоскостью. Данный способ построения показал наиболее точные результаты в сравнении с другими возможными конфигурациями.

Точность данной модели для акций Газпрома на тестовых данных равна 0.80.

Точность данной модели для акций Норникель на тестовых данных равна 0.79.

Далее строится дерево решений.

Создан массив параметров, который включает в себя все критерии построения модели случайный лес и максимальную глубину от 1 до 7:

С помощью GridSearchCV происходит перебор множеств деревьев на основе заданных параметров, после чего выводится результат лучшего из них:

Точность дерева с данными параметрами составляет 0.76 для акций компании Газпром.

Точность дерева с данными параметрами составляет 0.80 для акций компании "Норникель". Далее применен алгоритм случайный лес.

Точность дерева с данными параметрами составляет 0.77 для акций компании "Газпром" а для акций "Норникель" 0.82 .

Площадь под ROC-кривой AUC (Area Under Curve) является агрегированной характеристикой качества классификации, не зависящей от соотношения цен ошибок. Чем больше значение AUC, тем «лучше» модель классификации. Данный показатель часто используется для сравнительного анализа несколь-

ких моделей классификации. В данном случае сравнивается работа деревьев решений.

Далее спроектирована нейронная сеть ANN с 10 входными нейронами, 3 скрытыми слоями с 10, 5 и 5 нейронами и функциями активацией `relu` и 1 выходным нейроном с сигмоидной функцией активации.

Точность модели для акций "Газпром" составляет 0.81

Точность модели для акций "Норникель" составляет 0.83

В четвертой части представлено сравнение акций. Для сравнения точностей моделей машинного обучения при прогнозировании тренда цен на фондовом рынке построены несколько моделей для каждой выбранной акции и проведено тестирование.

Первая модель, которая представлена в данной работе – метод опорных векторов. В ходе тестирования лучшая точность находится на уровне 80% для прогнозирования акции "Газпром" и 79% .

Вторая модель - дерево решений. В ходе тестирования лучшая точность находится на уровне 80% для прогнозирования акции "Норникель".

Третья модель – случайный лес. В ходе тестирования лучшая точность находится на уровне 82% для прогнозирования акции "Норникель".

Четвертая модель, которая представлена в работе, – искусственная нейронная сеть. В ходе тестирования лучшая точность находится на уровне 83% для акции "Норникель".

Для каждой акции лучшую точность показали разные модели машинного обучения. Это может означать, что для каждой задачи необходимо проводить исследования качества моделей и выбирать лучшую, с наибольшей точностью прогнозирования.

Таким образом, сравнение показало, что в рамках прогнозирования на фондовом рынке, нейронная сеть является наиболее точной моделью с точностью прогнозирования на уровне 83%. Однако, следует учитывать, что каж-

дая модель имеет свои ограничения и может быть более или менее эффективной в зависимости от конкретной задачи и условий прогнозирования.

В приложении представлены исходные программные коды реализации.

ЗАКЛЮЧЕНИЕ

В более ранних исследованиях технические индикаторы применялись напрямую для прогнозирования, но в данном исследовании извлекается информация о тенденции из каждого технического индикатора, а затем используется для прогнозирования. Это позволяет добиться значительного увеличения точности прогнозов.

При повышении точности моделей прогнозирования в будущем, предлагаемая система может быть развернута в режиме реального времени для прогнозирования тренда акций, что делает инвестиции более прибыльными и безопасными. Повышение точности с помощью этого подхода, основанного на распространенных методах инвестирования в акции, также продвигает идею предварительной обработки данных на основе предметной области, в которых алгоритмы машинного обучения используются. Эта идея может быть расширена не только в области акций за счет включения других человеческих подходов к инвестированию, но и в различных других областях, где используются экспертные системы и методы машинного обучения.

В данной работе для построения базы знаний используются десять технических показателей, однако другие макроэкономические переменные, такие как курсы обмена валют, инфляция, государственная политика, процентные ставки и т.д., которые влияют на фондовый рынок, также могут быть использованы в качестве исходных данных для моделей или при построении базы знаний эксперта система. Средний объем акций также является потенциальным кандидатом, который может быть полезен при определении тренда.

Архитектура и правила обучения нейронных сетей, как правило, носят ситуационный характер - разрабатываются для решения определенных (или группы родственных) задач, поэтому важно понимание принципов и особенностей их работы, что позволит профессионально подходить к вопросу выбора готовых или создания новых алгоритмов обучения.

В данной работе решены следующие задачи:

1. Сделан обзор материала по фондовым рынкам и машинному обучению и оформлен датасет;
2. Построение моделей машинного обучения для сравнения точности предсказанных значений с использованием: метода опорных векторов

(SVM), метода дерева решений (DT), метода случайного леса (RF) и метода искусственной нейронной сети (ANN).;

3. Анализ получившихся результатов и сравнение моделей.

Для каждого временного периода лучшую точность показали разные модели машинного обучения. Это может означать, что для каждой задачи необходимо проводить исследования качества моделей и выбирать лучшую, с наибольшей точностью прогнозирования.