

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники
 Направление подготовки (специальность) 09.04.01 «Информатика и вычислительная техника»
 Отделение информационных технологий

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

Тема работы
Формирование стратегии инвестирования на фондовой бирже при помощи методов машинного обучения.

УДК 004.5:330.322.013:336.761

Студент

Группа	ФИО	Подпись	Дата
8ВМ03	Мнацакянн Рафаэль Эдуардович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ	Спицин В.Г.	д.т.н, профессор		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Былкова Т.В.	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ООД ШБИП	Федоренко О.Ю.	д.м.н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП,	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ	Спицин В.Г.	д.т.н, профессор		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	УК(У)-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	УК(У)-2. Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	УК(У)-3. Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	УК(У)-4. Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	УК(У)-5. Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определить и реализовать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
Общепрофессиональные компетенции	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований
ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования
ОПК(У)-7	Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
Профессиональные компетенции	
ПК(У)-1	Способен к созданию программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов
ПК(У)-2	Способен проектировать сложные пользовательские интерфейсы
ПК (У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК (У)-4	Способен осуществлять руководство разработкой комплексных систем
ПК(У)-5	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники
 Направление подготовки (специальность) 09.04.01 «Информатика и вычислительная техника»
 Отделение информационных технологий

УТВЕРЖДАЮ:
 Руководитель ООП
 _____ Спицын В.Г.
 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

Студенту:

Группа	ФИО
8ВМ03	Мнацаканян Рафаэль Эдуардович

Тема работы:

Формирование стратегии инвестирования на фондовой бирже при помощи методов машинного обучения.
--

Утверждена приказом директора (дата, номер)	03.02.2022 № 34-63/с
---	----------------------

Срок сдачи студентом выполненной работы:	
--	--

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе</p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<p>Набор научных статей о биржевой торговле. Данные о финансовых показателях компаний за различные периоды времени.</p>
---	---

<p>Перечень подлежащих исследованию, проектированию и разработке вопросов</p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<p>Реферат; Введение;</p> <ol style="list-style-type: none"> 1. Обзор предметной области; 2. Подтверждение гипотез; 3. Разработка анализирующей системы; 4. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение; 5. Социальная ответственность; 6. Выводы; 7. Список использованных источников.
<p>Перечень графического материала</p> <p><i>(с точным указанием обязательных чертежей)</i></p>	

Консультанты по разделам выпускной квалификационной работы

(с указанием разделов)

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Былкова Татьяна Васильевна
Социальная ответственность	Федоренко Ольга Юрьевна

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	24.01.2022
---	------------

Задание выдал руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ	Спицин В.Г.	д.т.н, профессор		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ03	Мнацаканян Рафаэль Эдуардович		

Реферат

Выпускная квалификационная работа содержит 73 страницы, 20 рисунков, 18 таблиц, 29 литературных источника, 1 приложение.

Ключевые слова: Классификация, Фондовая биржа, Дерево решений, Бустинг, Автоматизированный анализ.

Объектом исследования является система автоматизированного анализа акций рынка ценных бумаг. Данная система выбирает акции, которые недооценены рынком и таким образом формирует стратегию инвестирования.

Цель работы – разработка системы для автоматического анализа финансовых показателей.

В исследовании представлен обзор существующих стратегий инвестирования и принципы, на которых они основываются.

Область применения: научно-исследовательские и конструкторские работы, непосредственное использование по назначению.

Экономическая эффективность проекта достигается за счёт выбора наиболее простых для реализации и эффективных алгоритмов.

Разработаны меры по снижению влияния вредных и предотвращения воздействия опасных факторов. Разработка оказывает вред окружающей среде только с точки зрения образования твёрдых отходов из люминесцентных ламп и комплектующих оргтехники. Разработаны меры предотвращения пожара и правил поведения во время пожара.

В дальнейшем возможно развитие проекта при помощи анализа новостного фона и внешних факторов.

Оглавление

Реферат	6
1. Введение.....	10
2. Аналитический обзор предметной области.....	11
2.1. Принципы разделения торговых систем	11
2.2. Информация, на основании которой можно делать прогнозы...	13
2.2.1. Фундаментальный анализ	14
2.2.2. Технический анализ.....	15
2.2.3. Важные неучтенные параметры	16
3. Проверка гипотез.....	17
3.1. Полученные для работы данные.....	17
3.2. Анализ ключевых функций решения.....	19
3.3. Статистическое решение	20
4. Создание анализирующей системы.....	23
4.1. Используемые алгоритмы и инструменты	23
4.1.1. Классификация как задача машинного обучения.....	23
4.1.2. Обзор используемых классификаторов	24
4.1.2.1. Искусственные нейронные сети.....	24
4.1.2.2. Градиентный бустинг	24
4.1.2.3. Случайный лес	25
4.1.3. Используемые метрики для анализа результата.....	25
4.1.3.1. Матрица несоответствий.....	25
4.1.3.2. Метрики	26
4.1.3.3. Используемые инструменты.....	26
4.2. Создание моделей машинного обучения.....	27

4.3.	Интерпретируемость решения.....	29
4.3.1.	Проблема интерпретируемости.....	29
5.	Вывод.....	31
6.	Список литературы	32
7.	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.....	35
7.1.	Предпроектный анализ	35
7.1.1.	Потенциальные потребители результатов исследования	35
7.1.7.	Оценка сравнительной эффективности исследования.....	46
8.	Социальная ответственность.....	53
8.1.	Правовые и организационные вопросы обеспечения безопасности.....	53
8.1.1.	Правовые нормы трудового законодательства	53
8.1.1.1.	Режим рабочего времени	54
8.1.1.2.	Защита персональных данных работника	54
8.1.1.3.	Оплата и нормирования труда.....	54
8.1.1.4.	Виды компенсаций при работе во вредных условиях труда... ..	55
8.1.2.	Эргономические требования к правильному расположению и компоновки рабочей зоны	56
8.1.3.	Производственная безопасность	57
8.1.3.1.	Отклонение показателей микроклимата.....	57
8.1.3.2.	Превышение уровня шума.....	57
8.1.3.3.	Недостаточная освещенность рабочей зоны.....	58
8.1.3.4.	Повышенное значение напряжение в электрической цепи, замыкание которой может произойти через тело человека.....	61

8.1.4.	Экологическая безопасность	62
8.1.5.	Безопасность в чрезвычайных ситуациях	62
8.2.	Выводы по разделу	64
8.3.	Законодательные акты	65
ПРИЛОЖЕНИЕ А.....		66

1. Введение

Современная экономическая система плотно связана с фондовой биржей, которая перераспределяет финансовые средства, питающие развитие участвующих экономических агентов. Каждый участник фондовой биржи должен принимать множество решений о том, когда и какие действия нужно совершать и какие активы необходимо использовать. Так как каждое удачное вложение и действие приносит прибыль, а каждое неудачное приносит убытки, все экономические агенты стараются максимизировать свою информированность для принятия наилучших решений [1].

Так как фондовой бирже уже сотни лет, за это время выработалось огромное количество различных стратегий и тактик, некоторое из которых основаны на глубоком анализе технических параметров, некоторые на анализе психологии поведения инвесторов, другие на статистическом подборе самых успешных компаний, а иные просто на персональной приязни к какому-то активу. Большое количество институциональных инвесторов изобретают свои методы, которые являются коммерческой тайной и не раскрываются общественности. С момента развития вычислительных машин, их сразу применили в финансовом секторе, в том числе и для анализа ценных бумаг.

В отличие от множества сфер науки, невозможно получить state-of-art систему по биржевой аналитике или торговле, в отличие, например, от систем распознавания образов или систем перевода из одного натурального языка в другой. Высокая стоимость таких систем и крайне высокая ценность удачных экземпляров значительно тормозит развитие в этом направлении, потому что каждому новому исследователю не на что опираться, а существующие подходы, опубликованные в научных журналах, являются лишь небольшими исследовательскими экспериментами, а не глубоко проработанными фреймворками [2].

Тем не менее, в рамках данной работы была поставлена задача произвести систему проводящую аналитику различных биржевых

инструментов и формирующую стратегию инвестирования с минимальным участием пользователя.

2. Аналитический обзор предметной области

Несмотря на то, что фондовый рынок содержит большое количество различных агентов, он тщательно стандартизирован. В результате нескольких крупных кризисов были созданы государственные комиссии, одна из которых SEC, была создана в США в 1934 году для контроля над рынком ценных бумаг и увеличения доверия инвесторов в рыночные инструменты. Благодаря строгому контролю каждая компания, которая присутствует на рынке, обязана раскрывать всю свою финансовую отчетность. Например, для листинга на Нью-Йоркской бирже компания обязана иметь оборот не менее \$50 млн и подготовить все данные о своей работе за несколько лет. За время подготовки к листингу компания так же должна уточнить свою структуру, систематизировать свои активы и схему корпоративного управления. После того как оперативные управляющие компании поймут стоимость бизнеса, можно рассчитать будущую стоимость акции и соответственно общую капитализацию компании [1][2].

Такой строгий контроль не только обеспечивает относительную стабильность финансовых рынков, но и генерирует большие объемы отчетности, которые можно автоматически анализировать.

2.1. Принципы разделения торговых систем

Множество исторических данных, их доступность и высокая мотивация участников рынка получить конкурентное преимущество подтолкнуло многие исследовательские группы, индивидуальных ученых и инвесторов к разработке систем прогнозирования различной природы [1].

Системы разделяются по типу ценных бумаг, которые они обрабатывают:

- **Акции** – фактически являются правом владения частью компании. При покупке достаточно большого объема инвестор может получить право голосовать на собраниях или даже начать владеть

компанией. Так же каждый держатель акции может претендовать на прибыль компании.

- Облигации – являются долговыми расписками, которые компания или государство обязуются погасить с процентами. Чем надежнее заемщик, тем меньших процент он предлагает держателям своего долга. Бывают различных сроков погашения.
- Фьючерсные контракты – обязательство приобрести или продать некоторый актив в будущем. Удобный инструмент для стабилизации рынка и фиксации цены на нужную покупку (например нефть или газ) на несколько месяцев вперед.
- Ценные металлы – Драгоценные металлы. Золото, платина и другие.
- Фиатные валюты – Валюты различных стран, выпускаемые центральными банками. Доллары, евро, рубли и другие.
- Цифровые валюты – криптовалюты и стейблкоины. Различные виртуальные активы, которые поддерживаются вычислительной мощностью системы или особым алгоритмом.
- Паи – Смесь различных ценных бумаг, доля владения которыми предоставляется за небольшую сумму денег за минимальную единицу.

Некоторые активы являются более волатильными, некоторые являются более стабильными. В рамках дальнейшей работы будем рассматривать фонд акций, как достаточно динамичный, но все еще предсказуемый в отличие от фонда цифровых валют, которые не предоставляют никакую отчетность контролирующим органам [2].

Так же системы разделяются по длительности между открытием и закрытием позиции в активе. Открытой позицией считается покупка или короткая продажа ценной бумаги.

- Ультракоткие позиции – Частое открытие и закрытие позиций (так же известно как скальпинг). Время удержания позиции может не превышать нескольких секунд. Подобные спекуляции часто проводятся при помощи автоматизированных торговых систем [1][6].
- Внутрдневные позиции – Открытые позиции будут закрыты к концу дня. Некоторые брокеры берут комиссию за перенос некоторых позиций на следующий день, поэтому внутрдневная торговля довольно популярна у частных инвесторов.
- Среднесрочные позиции – Инвестиции на 1-3 месяцев. Удержание некоторой позиции. Такой срок позволяет получить прибыль в случае удачной конъюнктуры рынка, хорошего отчета компании или удачной новости.
- Долгосрочные позиции – Инвестиции на года. Позволяют заработать на общем росте рынка акций (порядка 2% в год в среднем). При достаточном времени инвестирования появляется значительный шанс сохранить накопленные средства. Считается эффективным средством борьбы с инфляцией.

В рамках данной работы будет рассмотрен вариант с среднесрочными позициями, так как для краткосрочных позиций не существует необходимости в анализе показателей компаний, а долгосрочные инвестиции тяжело обрабатывать и тестировать, из-за геометрически увеличивающейся сложности. Остановимся на инвестициях временным промежутком в 3 месяца.

2.2. Информация, на основании которой можно делать прогнозы

Существует два основных подхода прогнозирования стоимости акций. Фундаментальный анализ и технический анализ. Так же существует различные эвристические подходы, при помощи которых участники рынка надеются обнаружить паттерн в динамике цены. Такие подходы, например,

как волновой подход, не нашли серьезного научного и статистического подтверждения. [1]

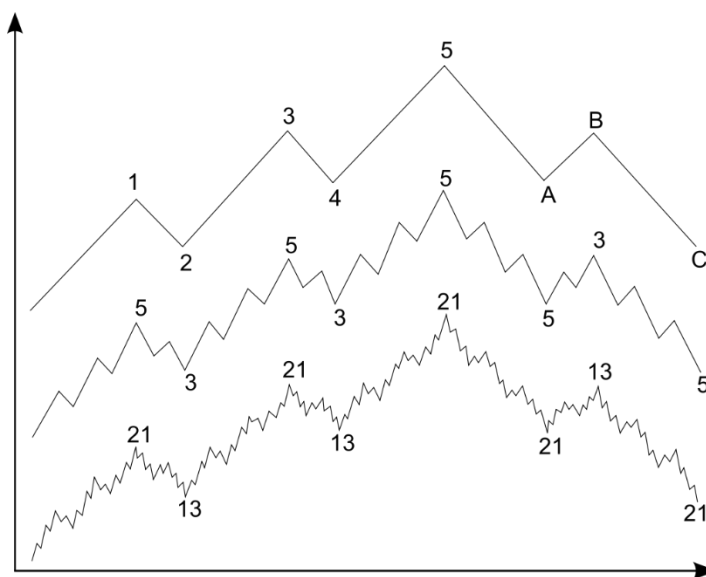


Рисунок 1 – Визуализация волновой теории Эллиота, которая гласит, что все движения рынка можно представить как 8 чередующихся в произвольном порядке волн (5 по тренду и 3 против тренда)

1. Фундаментальный анализ

Фундаментальным анализом ценных бумаг называют метод извлечения информации о внутренней ценности ценных бумаг в отрыве от цены, которая на данный момент установлена рынком, как равновесная. Эксперты изучают различные параметры, такие как финансовая отчетность компании, макроэкономические факторы, состояние определенной отрасли экономики и темпы экономического роста [1].

Целью такого анализа является получение для бумаги экспертного заключения о том, является ли она недооценена или переоценена рынком. Соответственно в таком случае не оптимальности рынка, трейдер может заработать, открыв длинную или короткую позицию соответственно [1][2].

Данный метод анализа использует информацию другой природы в отличие от технического анализа. Технический анализ рассматривает движение на бирже, изменение объема торгов и сдвиг кривой спроса и

предложения, фундаментальный анализ не использует подобную информацию в своей работе.

Фундаментальный анализ возможен благодаря требованиям к прозрачности компаний. Для каждой компании рассчитываются бесчисленные коэффициенты, при помощи которых проводится сравнительный анализ смежных компаний и определяется их перспективность.

У теории фундаментального анализа есть критики, основной аргумент которых является то, что финансовый рынок настолько сложная система, которая зависит от настолько большого количества факторов, что фактически невозможно проанализировать всё, а без этой информации точность предсказания будет крайне низкой. Далее в ходе работы будут рассмотрены информационные ограничения системы.

2. Технический анализ

Техническим анализом называется оценка ценных бумаг посредством анализа текущего баланса спроса и предложения. Так же анализируются исторические данные и по этим параметрам пытаются прогнозироваться тренды [2].

Существует множество инструментов технического анализа, которые, однако, имеют скорее эвристическую доказуемость, чем какую-то доказанную статистическую закономерность.

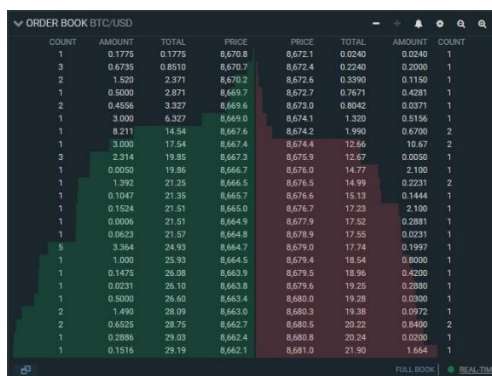


Рисунок 2 – Биржевой стакан – инструмент, позволяющий получить информацию о настоящих покупателях в данный момент времени, что они готовы купить и что продать и по какой цене

Биржевой стакан используется для получения информации о колебаниях спроса и предложения. Данная информация может быть крайне полезна в краткосрочной перспективе, потому что повышение спроса может спровоцировать рост ценной бумаги.



Рисунок 3 – График стоимости акции в виде японских свечей.

Технический анализ фактически использует левую половину графика для того, чтобы предсказать значения на правой половине. Многими учеными данный подход считается не обоснованным из-за недетерминированности рынка. В свою очередь сторонники технического анализа утверждают, что инструмент основан не на предсказании графика, а на прогнозировании поведения людей на рынке, которые в схожих ситуациях будут действовать

одинаково

3. Важные неучтенные параметры

Были рассмотрены предпосылки фундаментального и технического анализа, однако ценные бумаги зависят не только от информации, которая обрабатывается этими подходами. Так же важной для фондового рынка является новостная информация. Эпидемии, новые законы и политические кризисы значительно влияют на стоимость ценных бумаг. Иногда влияние

таких событий на стоимость многократно превышает влияние параметров, рассматриваемых в фундаментальном и техническом анализе [3][4].

Внешние события и новостной фон не учитываются в рамках данной работы, однако, очевидно, что для промышленно готового решения необходимо рассмотреть способы работы и с подобными данными тоже. Существуют методы извлечения новостного фона по заголовкам при помощи анализа естественных языков, которое может быть применено в будущем поверх существующей модели.

3. Проверка гипотез

3.1. Полученные для работы данные

В рамках дальнейшей работы сосредоточимся на решении задачи при помощи анализа фундаментальных показателей. Для работы был собран специальный датасет, состоящий из множества значений отчетов компаний за 25 лет [17].

Каждая запись состояла из:

Значения, которые необходимо прогнозировать:

- E_{t+1} Earnings (ib - spi) in year $t + 1$ Прибыль за год
- EPS_{t+1} Earnings (ib - spi) in year $t + 1$ scaled by shares outstanding (csho) Прибыль за год, деленная на количество акций

Входные параметры для регрессионных моделей:

- E_t Earnings (ib - spi) in year t
- A_t Total assets (at)
- D_t Dividend payment (dvc)
- DD_t Dummy variable indicating dividend payers
- $NegE_t$ Dummy variable indicating negative earnings
- AC_t суммированные неденежные активы, включая краткосрочные долги и налоги

- EPS_t + Earnings per share when earnings are positive, and zero otherwise
- AC_t – Accruals per share when accruals are negative, and zero otherwise
- AG_t Percentage change in total assets
- NDD_t Dummy variable indicating zero dividend per share
- DIV_t Dividend per share (dvpsx_f)
- BTM_t Закредитованность к рынку. Отношение.
- $Price_t$ Stock price as of three months after
- BVE_t Book value of equity (ceq)
- $TACC_t$ Total accruals

Входные параметры моделей машинного обучения:

- Sale - sales (продажи)
- COGS - cost of goods sold (себестоимость проданных товаров)
- XSGA - selling, general and admin. exp. (общие траты)
- XAD - Advertising expense (траты на маркетинг)
- XRD - Research and development (траты на исследования)
- DP_t - Depreciation and amortization (dp)
- $XINT_t$ Interest and related expense (xint)
- $NOPIO_t$ non-operating income (expense) – other (nopio)
- TXT - Income taxes (налог на прибыль)
- $XIDO_t$ Extraordinary items and discontinued operations (xido)
- E - Earnings (прибыль)
- Common dividend (dvc)
- CHE_t Средства и краткосрочные инвестиции (che)
- $INVT_t$ Суммарная инвентаризация (invt)
- $RECT_t$ Суммарная прибыль (rect)
- ACS - (стоимость всего имущества)

- $PPENT_t$ Стоимость недвижимости и физических объектов
- $IVAO_t$ Инвестиции
- $INTAN_t$ Нематериальные активы (intan)
- AT_t Общие активы (at)
- AP_t Балансовый счет (ap)
- DLC_t Долги в текущий обязательствах(dlc)
- TXP_t Налоги на прибыль (txp)
- LCT_t Все пассивы (lct)
- $DLTT_t$ Долгосрочные долги (dltt)
- LT_t Все долги (lt)
- SEQ_t Обычный капитал (seq)

всего 28 параметров для 1000 компаний за 25 лет

Датасет был дополнен информацией о стоимости акций на конец торгового дня с шагом в неделю. Таким образом любая запись имела дополнительно поле – стоимость на момент отчета и стоимость через три месяца. На основании именно этих данных в последствии и осуществлялось обучение всех моделей [17].

3.2. Анализ ключевых функций решения

Ключевыми функциями решения будут показатели мультиклассовой точности. Ограничим нашу задачу как задачу прогнозирования, вырастет ли рассматриваемый актив в течении трех месяцев, упадет в цене или не изменится [5][6].

Для данной задачи классическим является регрессионное решение с получением численного значения стоимости, однако из-за проблем с интерпретируемости, конечному пользователю сложно доверять такой модели. Одним из способов повысить как точность, так и степень доверия к модели – переопределить задачу подобным образом.

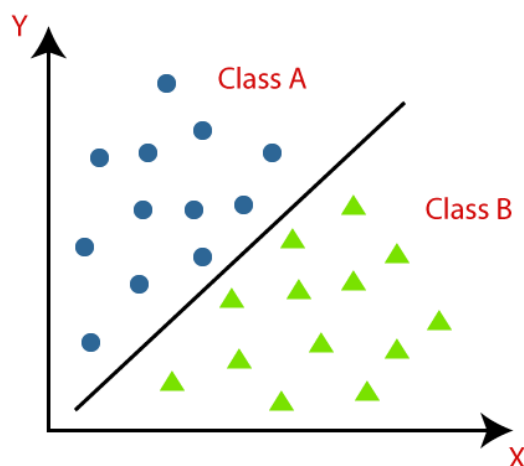


Рисунок 4 – Изображение классификации. Доверие результатам расчета больше из-за более низкой точности предсказания. Пользователю достаточно легко поверить, что через три месяца акция скорее всего вырастет и намного сложнее поверить, что ровно через три месяца акция будет стоить конкретное значение, предсказанное моделью

Так же важной функцией будет являться способность модели обосновать свой прогноз. Это добавит модели достоверности и значительно повысит потребительские качества всей системы [6][8][9].

3.3. Статистическое решение

Перед применением моделей глубокого обучения необходимо проверить принципиальную возможность получения значительных результатов, основываясь на методе подбора ценных бумаг с определенными параметрами фундаментальных параметров. Это позволит будущим пользователям больше доверять системе, потому что доказанность принципиального решения этой задачи неочевидна [10].

Проработаем особые мультипликаторы:

- $\frac{P}{E}$. Цена акции к приведенной прибыли.
- $\frac{P}{B}$. Цена акции к приведенной балансовой стоимости.
- $\frac{P}{CF}$. Цена акции к приведенному свободному денежному потоку.

Приведенные значения являются популярными показателями для сравнения нескольких ценных бумаг между собой. Показатели позволяют понять, насколько акции более оценены по отношению к другой акции при прочих равных [3][15].

Проработаем две стратегии:

- Quantitative Momentum Strategy

Инвестирование в наиболее растущие акции

Отсортированные акции по данной стратегии:

	Ticker	Price	Number of Shares to Buy
0	A	98.19	199
1	AAL	13.89	1411
2	AAP	161.13	121
3	AAPL	467.65	41
4	ABBV	97.29	201
5	ABC	104.97	186
6	ABMD	305.78	64
7	ABT	102.94	190
8	ACN	237.88	82
9	ADBE	453.91	43
10	ADI	120.52	162

Рисунок 5 – Полученные акции с наибольшим мультипликатором $\frac{P}{E}$ и скоростью роста за предыдущий период

- Quantitative Value Strategy

Инвестирование в наиболее недооцененные акции

Отсортированные акции по данной стратегии:

	Ticker	Price	Number of Shares to Buy
0	HPE	9.830	1994
1	FTI	8.870	2210
2	AAL	13.360	1467
3	CCL	15.610	1256
4	HFC	25.660	764
5	HPQ	19.140	1024
6	XRX	17.800	1101
7	TPR	16.010	1224
8	LB	27.330	717
9	SYF	25.300	775
10	TAP	39.340	498

Рисунок 6 – Полученные акции с наименьшим мультипликатором $\frac{P}{B}$

Так как акции растут в среднем на 2% в год, то не имеет смысла рассматривать эффективность, как результат того, что отобранные акции подорожали, а не подешевели. В качестве базового показателя, с которым будем сравнивать приведенные решение будем использовать индекс S&P500 [1].

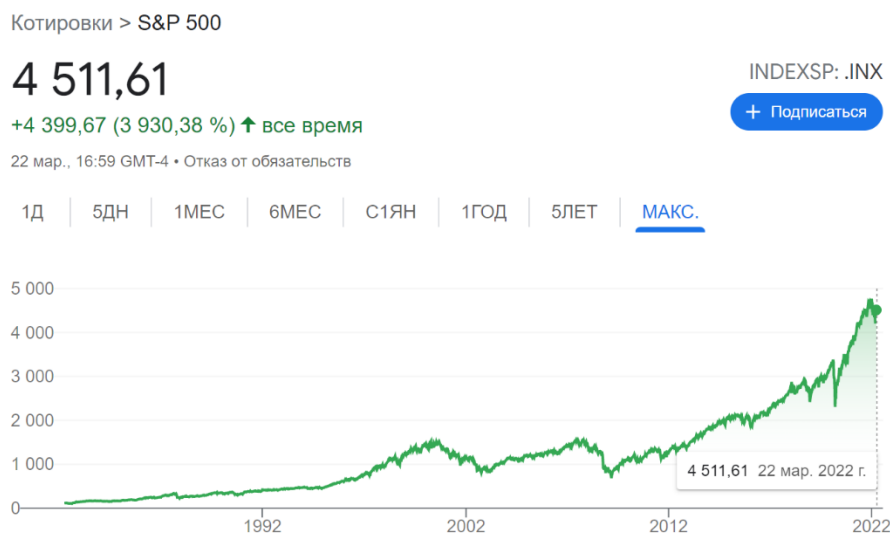


Рисунок 7 – Биржевой индекс S&P500 является взвешенной комбинацией 500 самых крупных торгуемых компаний. Исторически считается хорошей пассивной стратегией инвестиции, потому что он растет, когда весь рынок растет и не сильно падает при падении, потому что содержит в своем составе лишь крупнейшие компании, обесценивание которых достаточно нечасто [2]

В сравнении с S&P500 Quantitative Momentum Strategy за 2020 год росла быстрее на 31,2%.

В сравнении с S&P500 Quantitative Value Strategy за 2020 год росла быстрее на 5,1%.

По итогам проведенного эксперимента можно заключить, что модификаторы, основанные на фундаментальных показателях, позволяют неплохо сортировать акции компаний. Далее рассмотрим более глубокую проработку данного метода, основанного на моделях искусственного интеллекта.

4. Создание анализирующей системы

4.1.Используемые алгоритмы и инструменты

1. Классификация как задача машинного обучения

Классификация в рамках машинного обучения описывает проблему разделения объектов на основании специфических признаков, результатом решения задачи является правило разделение объектов, которое можно применять даже для на данных, которые не использовались для обучения классификатора. В рамках данной работы каждая запись имеет корректный лейбл, что является подразделом машинного обучения – обучение с экспертом [15].

Для обучения данные делятся на три части: обучающую тестовую и валидационную выборки. Обучающая выборка предоставляется модели машинного обучения, которая постепенно рассчитывает ошибку, которую она совершает и изменяет свои веса [15].

Классификатор обучается посредством постепенной сходимости с целью минимизации размера ошибки. Это процедура контролируется разработчиком для того, чтобы избежать переобучения модели на тестовых данных [5][10].

Проблема переобучения является причиной по которому бессмысленно переучивать модель для достижения запредельной точности на тестовой выборке. Модель начинает терять предсказательную способность на внешних данных и перестает быть промышленно применимой.

Валидационная используется для контроля функции потерь. Если модель продолжает обучаться, а на валидационных данных ошибка уже начинает расти, то это показатель начала переобучения модели. Дальнейшая подстройка весов приведёт лишь к ухудшению предсказательных характеристик модели.

После окончания обучения используется тестовая выборка – данные, которые не были доступны сети тем или иным образом. Именно на этих данных производится финальная верификация и расчет результатов и метрик.

Существуют разные топологии моделей и разные алгоритмы, которые применяются для их сходимости. Они используются в разных сферах деятельности и для разных форматов данных, однако нет универсального способа определить, без экспериментов самую точную топологию модели на конкретной задаче, поэтому в рамках проведенной работы были исследованы несколько различных классификаторов.

2. Обзор используемых классификаторов

4.1.2.1. Искусственные нейронные сети

Искусственной нейронной сетью называют модель, состоящую из множества соединенных вычислительных ядер, подстраивающие свои веса для максимального распространения нужного сигнала. Архитектура была изобретена под влиянием изучения биологического строения нервной системы живых существ [7][8].

Разные виды нейронных сетей имеют схожий принцип работы. Сеть является связанным графом, который аккумулирует в своих узлах веса, которые использует для модификации сигналов, проходящих через определенный узел. Проходя от истока сети к стоку сигнал преобразуется в ожидаемый от данной модели ответ.

В дальнейшем для модели хранятся веса всех связей, которые и решают задачу преобразования сигнала[10]. Данная модель является крайне гибкой, поэтому при помощи ИНС решают широкий спектр различных задач.

4.1.2.2. Градиентный бустинг

Бустинг является способом обучения слабых моделей, когда каждая новая модель старается исправить ошибку предыдущих моделей и обучается с большим весом на данных, для которых совершаются больше всего ошибок.

Главным различием бустингов являются сами модели, которые лежат в его основе. В работе будет использован градиентный буст (англ, adaboost) и АДА буст(англ, adaboost) [13].

4.1.2.3. Случайный лес

Случайный лес — это модель, которая является ансамблем над обученными деревьями решений. Дерево является самостоятельным классификатором. На каждой ветке такого дерева происходит сравнение с некоторым условием и спускаясь по условиям от корня дерева к листьям дерево и совершает классификацию. Несмотря на то, что единичное дерево обладает отличной способностью к обоснованию совершенных выборов, качество классификации оставляет желать лучшего. Поэтому множество таких деревьев объединяются при помощи алгоритма случайного леса, который обучает множество различных деревьев, а после использует ответ, который получается при помощи голосования всего ансамбля.

3. Используемые метрики для анализа результата

4.1.3.1. Матрица несоответствий

Матрица несоответствия — это метод демонстрации межклассовой ошибки в задаче классификации. Иногда высокая точность работы достигается за счёт сильного дисбаланса класса, или любого другого ‘нечестного’ приема, которому обучится модель. Поэтому необходимо проконтролировать какие классы путаются чаще всего для того, чтобы потом принять решение о том, как модифицировать рабочий пайплайн для исправления проблемы [16]. Сама матрица хранит на главной диагонали все верно угаданные классы, а все остальные значения — это ошибки классификатора.

		<i>P</i>	<i>N</i>
True label	<i>P</i>	True Positives (TP)	False Negatives (FN)
	<i>N</i>	False Positives (FP)	True Negatives (TN)
		Predicted label	

Рисунок 8 – Общий вид матрицы несоответствий

Наличие большинства значений на главной диагонали говорит о хорошей предсказательной способности модели.

4.1.3.2. Метрики

Accuracy – количество верно предсказанных классов

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision мера способности модели отличать один класс от другого

$$precision = \frac{TP}{TP + FP} \quad (2)$$

Recall есть мера демонстрирующая как модель реагирует на отдельный класс

$$recall = \frac{TP}{TP + FN} \quad (3)$$

F-мера – рассчитывается как среднее гармоническое precision и recall. Универсальный параметр, который принимает высокие значения только когда большая часть предсказаний правдива без перекосов и без забывания малочисленных классов [6].

$$f1 = 2 * \frac{precision * recall}{precision + recall} \quad (4)$$

Выбор метрик зависит от решаемой задачи.

4.1.3.3. Используемые инструменты

В работе использовались изученные в рамках пройденного обучения инструменты по анализу данных и обучению моделей.

Основным языком программирования был выбран Python, который является крайне популярным в задачах анализа данных и машинного обучения.

Некоторые использованные библиотеки:

- `scikit-learn` – Библиотека, содержащая классификаторы, описанные в предыдущих пунктах и методы расчёта метрик для обученных моделей;
- `Tensorflow` с использованием `Keras`. Фреймворк для работы с нейронными сетями различных топологий.
- `Pandas` – библиотека для обработки структурированных данных.
- `Numpy` – библиотека для ускоренных матричных вычислений.

4.2. Создание моделей машинного обучения

Обучение проводится при помощи созданного ранее датасета. Выходы модели:

- Компания недооценена
- Компания переоценена
- Компания оценена верно

На вход модели подаются примеры, состоящие из 28 параметров.

Обучение проводилось для моделей случайного леса и градиентного бустинга. Дополнительно проводились примеры обучения модели, основанной на полносвязанной нейронной сети.

Результаты обучения сведены в таблицу:

Таблица 1 – метрики лучших моделей в своих топологиях

Модель	Random forest	Gradient boosting	Dense NN1	Dense NN2
Accuracy	0.6815	0.6712	0.6577	0.6134
Recall	0.6433	0.6311	0.5933	0.6314
Precision	0.621	0.6811	0.5814	0.6211

Были проведены множественные эксперименты с подбором оптимальных коэффициентов и макропараметров обучения. Лучшие результаты обучения представлены ниже [6].

Параметры обучения случайного леса:

- `n_estimators=500`,
- `criterion='mse'`,
- `oob_score= True`,
- `n_jobs=-1`,
- `random_state=10`,
- `cv=5`,
- `n_jobs=-1`,
- `scoring='neg_mean_squared_error'`

Параметры обучения градиентного бустинга:

- `learning_rate=0.1`,
- `n_estimators=500`,
- `loss=' huber'`,
- `alpha=0.7`,
- `subsample=0.9`,
- `random_state=10`,
- `cv=5`,
- `scoring='neg_mean_squared_error'`

Топология Dense NN1:

Optimizer = ADAM

- 28 relu
- Dropout 0,4
- 256 relu
- Dropout 0,4
- 256 relu
- 3 softmax

Топология Dense NN2:

Optimizer = ADAM

- 28 relu
- 128 relu
- Dropout 0,4
- 512 relu
- Dropout 0,4
- 512 relu
- 128 relu
- 3 softmax

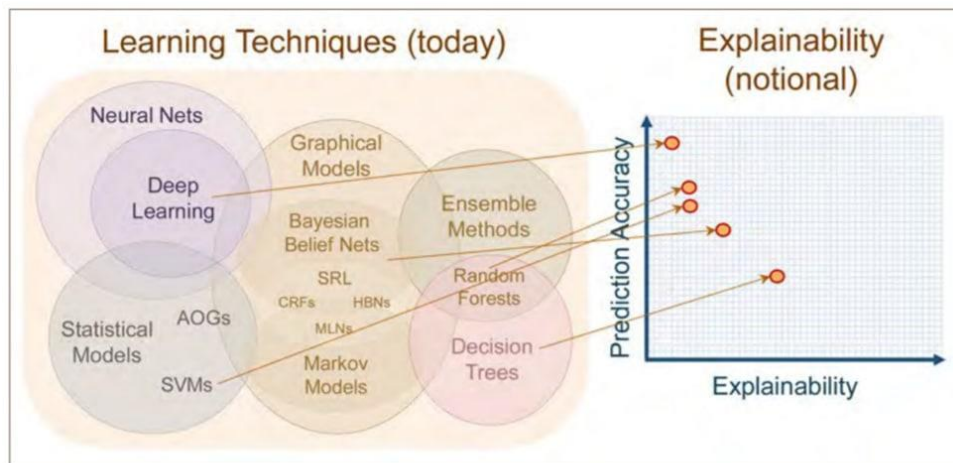
4.3.Интерпретируемость решения

1. Проблема интерпретируемости

Пользователи считают объяснение решения важнее самого решения. Особенно сильно эта проблема влияет, когда система принимает ответственные решения, заменяя собой эксперта.

Существует консенсус о объясняемых моделях:

- Наиболее ценными являются такие рекомендации, которые могут быть объяснены и понятны пользователю без необходимости слепого доверия алгоритму, работающему по принципу черного ящика.
- Чем важнее сфера применения, и чем дороже стоимость ошибки, тем больше контроля над анализатором хочет пользователь.



Source: DARPA

Рисунок 9 – Ранжирование топологий моделей по степени их объяснимости

К сожалению, наиболее точные результаты были получены слабо объяснимыми моделями, так что для продолжения работы необходимо использовать внешнюю функциональность, а не внутренние параметры обученной модели [17].

Для понимания природы предсказания был проведен анализ важности параметров и их ранжирование [15].

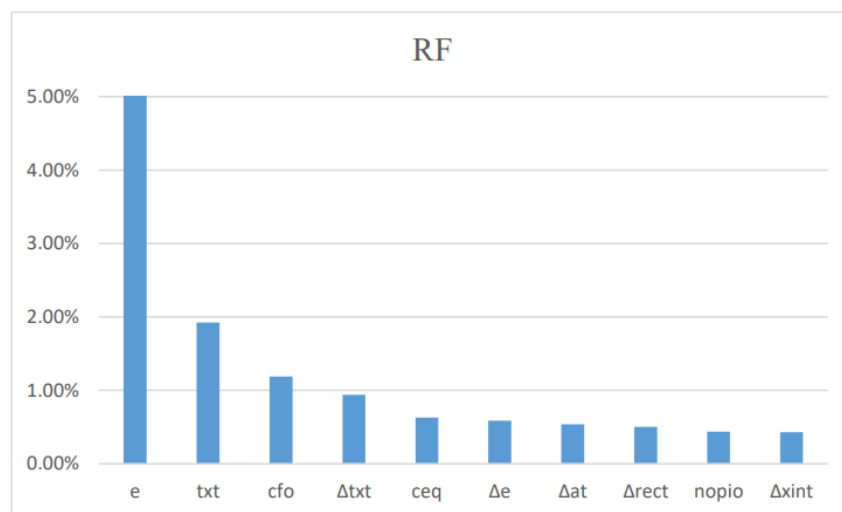


Рисунок 10 – степень влияния параметра на результат

Самым значимым параметром является размер прибыли компании и плотно коррелирующий с ним размер налоговой нагрузки. Базовая модель коррелирует с существующим представлением реального мира. Попробуем проинтерпретировать отдельные тесты при помощи библиотеки `shap`.

Векторы Шепли основаны на идеи проведения серии игр с незначительно изменяемыми входными параметрами. По размеру изменения выхода модели становится понятно, какой параметр является наиболее важным, что даже при незначительном его изменении результат работы модели серьезно изменяется [18].



Рисунок 11 – Предсказание библиотекой `shap` разбивает самые значительные и влиятельные параметры и показывает какой параметр повлиял в какую сторону. Например, в этом примере компания имела годовую прибыль ниже среднего для отрасли, поэтому мы видим значительный перевес синего (визуализация негативного влияния параметра)

5. Вывод

В результате работы проведён глубокий анализ предметной отрасли, изучены существующие схемы стратегии инвестирования.

Сформированы гипотезы о применимости стратегий инвестирования на основе анализа биржевых показателей.

Создана программа, предлагающая необходимые ценные бумаги, согласно стратегии инвестирования, проведен анализ с существующими пассивными инвестиционными стратегиями.

Создана рекомендательная модель на основе методов машинного обучения, лучшей моделью оказалась модель, основанная на случайном лесе с точностью 0.685.

Разработан и протестирован способ обоснования принятого системой решения.

6. Список литературы

1. Troy Segal, Fundamental Analysis. [Электронный ресурс]. URL: <https://www.investopedia.com/terms/f/fundamentalanalysis.asp>
2. Adam Hayes, Technical Analysis. [Электронный ресурс]. URL: <https://www.investopedia.com/terms/t/technicalanalysis.asp>
3. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989.
4. Спицын В.Г. Применение искусственных нейронных сетей для обработки информации: Методические указания к лабораторным работам / Спицын В.Г., Цой Ю.Р. – Томск: Издательство ТПУ, 2007 – 32 с.
5. Применение нейронных сетей для задач классификации // basegroup [Электронный ресурс]. URL: <https://basegroup.ru/community/articles/classification>
6. Метод деревьев решений для задачи классификации // edu.kpfu [Электронный ресурс]. URL: https://edu.kpfu.ru/pluginfile.php/91556/mod_resource/content/3/Decision%20tree%20s_1.pdf
7. Breiman L. Bagging predictors // Machine Learning, 1996, vol. 24, no. 2, pp. 123–140.
8. Leo Breiman. Random Forests // Machine Learning, October 2001, Volume 45, Issue 1, pp 5-32. 71
9. Caruana R., Niculescu-Mizil A. An Empirical Comparison of Supervised Learning Algorithms // Department of Computer Science, Cornell University, Ithaca, NY 14853 USA.
10. Mark R. Segal. Machine Learning Benchmarks and Random Forest Regression // Division of Biostatistics, University of California, San Francisco, CA 94143-0560, April 14, 2003.

11. L. Mason , J. Baxter , P. Bartlett , M. Frea. Boosting Algorithms as Gradient Descent // 2000
12. Рашка С. Python и машинное обучение / пер. с англ. А. В. Логунова. М.: ДМК Пресс, 2017. – 418 с.: ил. 13. Открытый курс машинного обучения. Тема 5. Композиции: бэггинг, случайный лес. [Электронный ресурс]. URL: <https://habr.com/ru/company/ods/blog/324402/#3-ocenka-vazhnosti-priznakov>
13. Открытый курс машинного обучения. Тема 10. Градиентный бустинг. [Электронный ресурс]. URL: <https://habr.com/ru/company/ods/blog/327250/#4-itog-pro-teoriyu-gbm>
14. Метрики в задачах машинного обучения. [Электронный ресурс]. URL: <https://habr.com/ru/company/ods/blog/328372/>
15. Davy Cielen, Arno D.B. Meysman, Mohamed Ali Introducing Data Sciences – Manning Publications Co., 2016.
16. Конспекты курса МФТИ «Введение в машинное обучение», специализация – «Машинное обучение и анализ данных»
17. Kai Cao, Haifeng You. Fundamental Analysis via Machine Learning. School of Business and Management Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong., 2021
18. Мирошниченко Лена, Вектор Шепли. Лекция НИУ ВШЭ 2020. [Электронный ресурс]. URL: https://math.hse.ru/data/2020/12/15/1358168643/лекция_23_Вектор_Шепли.pdf

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСООБЪЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
8ВМ03	Мнацаканян Рафаэль Эдуардович

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	магистратура	Направление/специальность	09.04.01 «Информатика и вычислительная техника»

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Стоимость ресурсов определялась по средней рыночной стоимости, и в соответствии с окладами сотрудников организации.
2. <i>Нормы и нормативы расходования ресурсов</i>	30% районный коэффициент, коэффициент дополнительной заработной платы 12%;
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	30% отчисления во внебюджетные фонды

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого и инновационного потенциала НТИ</i>	Провести предпроектный анализ
2. <i>Разработка устава научно-технического проекта</i>	Представить Устав научного проекта магистерской работы
3. <i>Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок</i>	Разработать план управления НТИ
4. <i>Определение ресурсной, финансовой, экономической эффективности</i>	Рассчитать сравнительную эффективность исследования

Перечень графического материала (с точным указанием обязательных чертежей):

1. *Сегментирование рынка*
2. *Оценка конкурентоспособности технических решений*
3. *Матрица SWOT*
4. *График проведения и бюджет НТИ*
5. *Оценка ресурсной, финансовой эффективности НТИ*

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Былкова Т.В.	канд.экон.наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ03	Мнацаканян Р.Э.		

7. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

7.1. Предпроектный анализ

7.1.1. Потенциальные потребители результатов исследования

Целью проекта является разработка рекомендательной системы, осуществляющей автоматизированный анализ инвестиционных инструментов. Данная система может использоваться участниками торгов на фондовой бирже, институциональными и частными инвесторами, а также различными исследователями и аналитиками. Система может использоваться как напрямую, так и для формирования второго мнения, если уже есть заключение другого эксперта.

Сегментировать рынок услуг можно по степени потребности использования возможностей системы. Результат сегментирования представлен на таблице 2.

Таблица 2 – Карта сегментирования рынка

		Вид		
		Анализ инвестиционных инструментов	Автоматическое формирование стратегии инвестирования	Проверка существующего экспертного мнения
<i>Размер</i>	Крупные			
	Средние			
	Мелкие			

	Институциональные инвесторы		Исследователи и аналитики		Частные инвесторы
--	-----------------------------	--	---------------------------	--	-------------------

7.1.2. Анализ конкурентных решений

Задача имеет большое количество решений, большинство из которых предоставляются платно, как сервис. Выделим 3 популярных конкурентных решения.

ТС2000 – платформа, предоставляющая автоматические торговые стратегии и комбинированную аналитику по различным инструментам фондовой биржи.

StockRover – платформа, предоставляющая аналитические отчеты по 7000 тысячам компаний в Северной Америке.

Wallstreetzen – платформа агрегирует прогнозы различных аналитиков.

Оценка созданного продукта и сравнение с конкурентами приводится в таблице 3.

Таблица 3 – Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы				Конкурентоспособность			
		Б _ф	Б _{к1}	Б _{к2}	Б _{к3}	К _ф	К _{к1}	К _{к2}	К _{к3}
1	2	3	4	5	6	7	8	9	10
Технические критерии оценки ресурсоэффективности									
1. Повышение производительности труда	0,078	5	5	3	2	0,391	0,391	0,234	0,156
2. Простота эксплуатации	0,055	5	4	3	5	0,273	0,219	0,164	0,273
3. Интерпретируемость	0,070	4	2	5	5	0,281	0,141	0,352	0,352
4. Точность	0,078	3	3	4	3	0,234	0,234	0,313	0,234
5. Удобство эксплуатации	0,055	3	3	2	4	0,164	0,164	0,109	0,219
6. Степень автоматизации	0,055	5	3	1	1	0,273	0,164	0,055	0,055
7. Энергоэкономичность	0,016	4	5	5	5	0,063	0,078	0,078	0,078
8. Требования к оборудованию	0,039	2	5	5	5	0,078	0,195	0,195	0,195
9. Работа по сети Интернет	0,008	5	5	5	5	0,039	0,039	0,039	0,039
10. Дополнительные функциональные возможности	0,055	1	3	1	1	0,055	0,164	0,055	0,055
Экономические показатели эффективности									
1. Конкурентоспособность продукта	0,078	4	3	2	3	0,313	0,234	0,156	0,234
2. Цена	0,070	5	3	2	3	0,352	0,211	0,141	0,211
3. Срок выхода на рынок	0,063	2	4	5	3	0,125	0,250	0,313	0,188
4. Скорость обучения использованию	0,063	4	2	1	3	0,250	0,125	0,063	0,188
5. Стоимость масштабирования	0,063	3	3	2	1	0,188	0,188	0,125	0,063

6. Стоимость модификации	0,063	4	2	1	1	0,250	0,125	0,063	0,063
7. Стоимость внедрения	0,078	5	3	2	3	0,391	0,234	0,156	0,234
8. Уровень проникновения на рынок	0,016	1	3	2	4	0,016	0,047	0,031	0,063
Итого	1	65	61	51	57	3,734	3,203	2,641	2,898

Согласно проведенному сравнительному анализу, можно сделать вывод о высокой конкуренции в рамках решаемой задачи. Конкурирующие системы доступны по подписке, что позволяет достаточно быстро начать использование той или иной платформы. Однозначным плюсом рассматриваемого в данной работе решения является доступный исходный код, что позволяет модифицировать систему для удовлетворения новых потребностей пользователей без необходимости поиска нового инструментария в стороннем сервисе. Так же конкурентным преимуществом является полностью автоматизированная аналитика с готовым результатом, который при желании можно интерпретировать. Конкуренты предоставляют обширную аналитику, для успешного использования которой, необходимы специальные знания фондовой биржи. Таким образом можно сделать вывод о том, что в отличии от конкурентов продукт сильнее нацелен на частных инвесторов без профильного образования и глубоких познаний в инвестировании.

7.1.3. SWOT-анализ

В ходе SWOT-анализа были определены приоритетные направления развития разработки, а также способы для нивелирования слабых сторон разработки и противодействия возможным угрозам. Это позволяет перейти ко второму этапу анализа.

Для второго этапа анализа необходима интерактивная матрица проекта, которая представлена в таблице 4.

Таблица 4 – Интерактивная матрица проекта

Сильные стороны проекта				
Возможности		Си1	Си2	Си3
	В1	+	-	-

	B2	-	-	+
	B3	-	+	-
Слабые стороны проекта				
Возможности		Сл1	Сл2	Сл3
	B1	+	-	-
	B2	-	+	-
	B3	-	-	+
Сильные стороны проекта				
Угрозы		Си1	Си2	Си3
	У1	+	-	-
	У2	-	+	+
Слабые стороны проекта				
Угрозы		Сл1	Сл2	Сл3
	У1	+	-	+
	У2	-	+	-

Результат третьего этапа анализа приведен в таблице 5.

Таблица 5 – SWOT-анализ

	Сильные стороны	Слабые стороны
	Си1. Минимальные требования навыков от пользователя.	Сл1. Низкая способность предсказывать значительные новостные изменения.
	Си2. Высокая степень автоматизации.	Сл2. Высокая зависимость от качества данных.
	Си3. Работа с ценными бумагами на всех рынках.	Сл3. Глубина предсказаний ограничена 3 месяцами.
Возможности		
V1. Доработка функционала (учитывать новостной фон, общие тенденции на финансовом рынке)	V1Си1. Возможно добавить усложненные конфигурационные настройки для продвинутых пользователей для расширения кастомизации предсказаний.	V1Сл1. Фундаментальная проблема, которая может быть решена значительными исследованиями и расширением функционала. Существуют теоретические исследования по NLP которые могут быть использованы.
V2. Улучшение качества работы в целях улучшения точности	V2Си3. Увеличение количества данных для обучения значительно повысит качество предсказаний. К тому же информация с нескольких рынков позволять	V2Сл2. Выбранные модели не могут быть получены в достаточном качестве без проведения серьезных подготовительных работ с данными. Проблема решается, однако

	отслеживать глобальные тенденции.	необходимо закладывать адекватное время на её решение.
В3. Улучшение производительности в целях увеличения скорости	В3Си2. Улучшение скорости принятия решений могут позволить расширить систему для автоматического более высокочастотного трейдинга.	В2Сл3. Для улучшения быстродействия был выбран один временной отрезок. Если переобучить модель, то можно добиться любых временных отрезков, однако работа над несколькими отрезками одновременно приведет к замедлению скорости работы. Необходимо понять, что важнее в использовании перед тем, как модифицировать систему подобным образом.
Угрозы		
У1. Непредвиденные обстоятельства, не учитываемые созданной моделью	У1Си1 Упрощение работы может создать у пользователя впечатление хорошего понимания экономической ситуации. Естественная недетерминированность рынков в свою очередь всегда может оказать влияние за пределами любых предсказательных способностей как текущей системы, так и любого эксперта.	У1Сл1. Проблема неразрешима при текущей архитектуре. Однако конкуренты так же не способны её решить. У1Сл3. Можно развернуть проблему, если значительно ограничить область видимости модели. Если она будет предсказывать на неделю, то вероятность значительных новостей, которые кардинально развернут рынок уменьшится.
У2. Появление более качественных аналогов	У2Си2 (У2Си3) любые доработки конкурирующий систем будут развиваться в каком-то определенном направлении. Из-за особенностей предметной области, всегда можно найти особенную черту, которая позволит конкурировать с любыми аналогами.	У2Сл2. Более качественные аналоги скорее всего будут иметь доступ к дополнительной информации. Можно будет провести дополнительные исследования, чтобы понять, как можно догнать улучшить показатели модели.

Самой сильной стороной проекта является отличная модифицируемость и хорошая автоматизация, которая не требует значительных познаний от пользователей. Слабыми сторонами проекта является слабая прогнозируемость некоторых событий и большая зависимость от использованных для обучения данных.

7.1.4. Инициация проекта

Устав научного проекта магистерской работы:

1. Цели и результат проекта. Информация по заинтересованным сторонам представлена в таблице 6:

Таблица 6 – Заинтересованные стороны проекта

Заинтересованные стороны	Ожидания
Пользователь	Получение автоматизированной аналитики, позволяющей информировано создавать свою стратегию инвестирования.
Разработчик	Получение заработной платы
Компания разработчика	Получение прибыли с полноценного комплекса продукта
Научный руководитель, студент	Выполненная выпускная квалификационная работа

Цели и результат проекта представлены в таблице 7.

Таблица 7 – Цели и результат проекта

Цели проекта	<ul style="list-style-type: none">– Изучить предметную область– Проверить работоспособность отбора ценных бумаг при помощи анализа фундаментальных показателей– Создание системы, основанной на методах машинного обучения– Создание методов обоснования принятого моделью решения
Ожидаемые результаты	<ul style="list-style-type: none">– Создание системы с ощутимой предсказательной способностью– Сдана выпускная квалификационная работа

Критерии приемки	– Успешное тестирование на специально отобранной тестовой выборке данных
Требования к результату проекта	– Выполнены все пункты функционального требования. – Разработанный функционал полностью соответствует ожиданиям заинтересованных сторон.

7.1.5. Планирование управления научно-техническим проектом

В таблице 8 приведен порядок работ, выполняемых в ходе разработки, и исполнитель каждой работы.

Таблица 8 – Перечень работ и исполнителей при разработке модуля.



№	Наименование работы	Исполнители работы
1	Выбор научного руководителя работы	Мнацаканян Рафаэль Эдуардович
2	Составление и утверждение темы работы	Мнацаканян Рафаэль Эдуардович Спицын Владимир Григорьевич
3	Составление календарного плана-графика выполнения работы	Мнацаканян Рафаэль Эдуардович Спицын Владимир Григорьевич
4	Подбор и изучение литературы по теме работы	Мнацаканян Рафаэль Эдуардович
5	Анализ предметной области	Мнацаканян Рафаэль Эдуардович
6	Проверка гипотезы прогнозирования по фундаментальным показателям	Мнацаканян Рафаэль Эдуардович
7	Разработка модели машинного обучения	Мнацаканян Рафаэль Эдуардович
8	Создание модуля обоснования принятия решения	Мнацаканян Рафаэль Эдуардович
9	Согласование выполненной работы с научным руководителем	Спицын Владимир Григорьевич
10	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Мнацаканян Рафаэль Эдуардович
11	Подведение итогов, оформление работы	Мнацаканян Рафаэль Эдуардович Спицын Владимир Григорьевич

График Ганта построим в виде таблицы с разбивкой по декадам (10 дней) за период времени выполнения научного проекта. При этом работы на графике следует выделить различным цветом, в зависимости от исполнителей, ответственных за ту или иную работу.

Таблица 9 – Календарный план-график проведения НИОКР по теме

№	Вид	Исп	Т раб, дн.	Продолжительность														
				Янв.			Февр.			Март			Апр.			Май		
				1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	Выбор научного руководителя работы	С	8	■														
2	Составление и утверждение темы работы	С	4		■													
		Р	2		■													
3	Составление календарного плана-графика выполнения работы	С	5			■												
		Р	1			■												
4	Подбор и изучение литературы по теме работы	С	18				■	■										
5	Анализ предметной области	С	14					■	■									
6	Проверка гипотезы прогнозирования по фундаментальным показателям	С	8							■								
7	Разработка модели машинного обучения	С	26								■	■	■					
8	Создание модуля обоснования принятия решения	С	26											■	■			
9	Согласование выполненной работы с научным руководителем	Р	1														■	
10	Выполнение других частей работы	С	7														■	

	(финансовый менеджмент, социальная ответственность)																	
11	Подведение итогов, оформление работы	С	5															
		Р	1															

-  Руководитель (Р)
-  Студент (С)

Итого руководитель потратит на проект около 5 дней, а студент – 121 день.

7.1.6. Бюджет научного исследования

В материальные затраты научного исследования вошли затраты на электроэнергию в сумме 2000 рублей.

Поскольку необходимое оборудование и лицензии на программное обеспечение, которое было использовано при реализации разработки, уже были приобретены, затраты включают в себя только затраты на амортизацию оборудования студента.

Норма амортизации для использованного оборудования составляет

$$A_n = \frac{1}{n} * 100\% = \frac{1}{3} * 100\% = 33.33\%, \quad (5)$$

где n – срок полезного использования оборудования.

Тогда с учетом того, что продолжительность работы составляет четыре месяца, формула для расчета итоговой суммы амортизации основных средств выглядит следующим образом:

$$A = \frac{C * A_n * 4}{12 * 100\%}, \quad (6)$$

где C – стоимость оборудования.

Во время проведения научного исследования потребовался ПК с современной графической картой для ускорения расчетов. Общая стоимость ПК – 150000. Исследования проводились исключительно с использованием свободного ПО стоимостью 0 рублей. Расчет затрат на амортизацию представлен в таблице 10.

Таблица 10 – Расчет затрат на амортизацию

Наименование	Стоимость, руб	Затраты, руб
Амортизация ПК	180000	15000
Итого	180000	15000

Должность руководителя – профессор, д.т.н. – 37300 рублей в месяц.

Должность инженера – студент – 23800 рублей в месяц.

Месячная заработная плата рассчитывается по формуле:

$$Z_{\text{осн}} = Z_{\text{дн}} * T_p, \quad (7)$$

где $Z_{\text{осн}}$ – среднедневная зарплата, руб;

T_p – продолжительность работ, выполняемых работником, раб. дни.

Среднемесячная зарплата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_m * k_p * M}{F_d} \quad (8)$$

где Z_m – месячный оклад работника, руб;

M – количество месяцев работы без отпуска в течении года: (10.4);

F_d – действительный годовой фонд рабочего времени персонала (представлен в таблице 11).

Таблица 11 – Баланс рабочего времени участников разработки

Показатели рабочего времени	Руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней		
– выходные дни	52	82
– праздничные дни	11	14
Потери рабочего времени		
– отпуск	56	24
– невыходы по болезни	-	-
	254	217

Среднемесячная зарплата руководителя составляет:

$$Z_{\text{дн}} = \frac{Z_m * k_p * M}{F_d} = \frac{37300 * 1.3 * 10.4}{254} = 1985.41$$

Среднемесячная зарплата студента составляет:

$$Z_{\text{дн}} = \frac{Z_{\text{м}} * k_p * M}{F_{\text{д}}} = \frac{23800 * 1.3 * 10.4}{217} = 1482.83$$

Расчет основной зарплаты представлен в таблице 12.

Таблица 12 – Расчет основной заработной платы

Исполнители	$Z_{\text{ок}}$, руб.	k_p	$Z_{\text{м}}$, руб.	$Z_{\text{дн}}$, руб.	T_p , раб. дн.	$Z_{\text{осн}}$, руб.
Руководитель	37300	1.3	48490	1985.41	5	9927.05
Инженер	23800	1.3	30940	1482.83	121	179422.43
Итого по статье $Z_{\text{осн}}$						189349.48

Запланируем коэффициент дополнительной зарплаты в размере 10% от основной заработной платы. В таблице 13 приведен расчет дополнительной заработной платы.

Таблица 13 – Заработная плата исполнителей ВКР, руб.

Заработная плата	Руководитель	Инженер
Дополнительная зарплата	992.705	17942.24
Итого	18934,94	

Коэффициент отчислений на уплату во внебюджетные фонды: 0.3.

Расчет отчислений на социальные нужды предоставлен в таблице 14.

Таблица 14 – отчисления на социальные нужды, руб

Заработная плата	Руководитель	Инженер
Зарплата исполнителя	10919.755	197364.673
Отчисления на социальные нужды	3275.9265	59209.4019
Итого	62485.3284	

Реализованное исполнение не является единственным возможным. Предлагается сравнить существующее исполнение (Исп. 1) с двумя другими. Исполнение 2 рассчитано при помощи облачного сервиса, что значительно снизит траты на амортизацию оборудования (из-за уменьшение требований к

рабочей машине), однако время разработки не изменится, так как все проводимые исследования останутся теми же.

Стоимость аренды облака – 38370 рублей в месяц. Необходимо два месяца непрерывных вычислений.

В рамках исполнения 3 рассмотрим вариант с использованием готовых фреймворков, которые значительно снизят время разработки ценой ухудшения качества работы системы и её меньшей специализации на решаемой задаче. Затраченные часы для руководителя не изменятся. Затраченные часы для студента – 102 рабочих дня.

В таблице 15 представлена группировка затрат по статьям.

Таблица 15 – Бюджет затрат НТИ

Наименование	Сумма, руб.		
	Исп. 1	Исп. 2	Исп. 3
Материальные затраты	2000	76740	2000
Амортизация оборудования	15000	2000	15000
Заработная плата исполнителей	208284.42	208284.42	177293,28
Отчисления на социальные нужды	62485.32	62485.32	53187,98
Итого	287769,74	349509,74	247481,26

7.1.7. Оценка сравнительной эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный показатель финансовой эффективности научного исследования получают в ходе оценки бюджета затрат трех или более вариантов исполнения научного исследования. Для этого наибольший интегральный показатель реализации технической задачи принимается за базу

расчета как знаменатель, с которым соотносятся финансовые значения по всем вариантам исполнения.

Интегральный финансовый показатель разработки определяется как:

$$I_{\Phi}^p = \frac{\Phi_{pi}}{\Phi_{max}} \quad (9)$$

Где I_{Φ}^p – интегральный финансовый показатель разработки,

Φ_{pi} – стоимость i -го варианта исполнения,

Φ_{max} – максимальная стоимость исполнения разработки (в т.ч. аналоги).

$$I_{\Phi}^p = \frac{\Phi_{pi}}{\Phi_{max}} = \frac{287769,74}{349509,74} = 0.82$$

$$I_{\Phi}^a = \frac{\Phi_{pi}}{\Phi_{max}} = \frac{349509,74}{349509,74} = 1$$

$$I_{\Phi}^b = \frac{\Phi_{pi}}{\Phi_{max}} = \frac{247481,26}{349509,74} = 0.70$$

Полученная величина интегрального финансового показателя разработки отражает соответствующее численное увеличение бюджета затрат разработки в разгах (значение больше единицы), либо соответствующее численное удешевление стоимости разработки в разгах (значение меньше единицы, но больше нуля).

Определенная величина показывает, что для реализации был выбран относительно недорогой метод. Необходимо понять, изменение потребительских характеристик различных методов, чтобы окончательно понять, насколько выбранное исполнение эффективно.

Можно определить следующим образом:

$$I_{pi} = \sum_{i=1}^n a_i * b_i^p, \quad (10)$$

где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i^p – бальная оценка i -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания;

Расчет приведен в таблице 16.

Таблица 16 – Сравнение характеристик вариантов исполнения проекта

Критерии \ ПО	Вес	Исп. 1	Исп. 2	Исп. 3
Удобство использования	0,19	5	5	3
Масштабируемость	0,15	4	4	3
Требуемые ресурсы	0,15	4	2	4
Функциональность	0,15	4	4	4
Удобство обслуживания	0,07	5	5	3
Срок разработки	0,07	3	3	5
Надёжность	0,19	5	5	3
Итого (сумма)	1	25	28	22

Сравнение интегрального показателя эффективности текущего проекта и аналогов позволит определить сравнительную эффективность проекта.

Таблица 17 – Сравнительные показатели эффективности разработки

№	Показатели	Исп. 1	Исп. 2	Исп. 3
1	Интегральный финансовый показатель разработки	0.82	1	0.70
2	Интегральный показатель ресурсоэффективности разработки	4,38	4,07	3,46
3	Интегральный показатель эффективности	5,34	4,07	4,94
4	Сравнительная эффективность аналогов и разработки		1,31	1,08

Сравнение значений интегральных показателей эффективности позволяет судить о приемлемости существующего варианта решения поставленной в магистерской диссертации технической задачи с позиции финансовой и ресурсной эффективности. По результатам проведенных расчетов первый вариант оказался наиболее эффективным в сравнении с другими исполнениями. Качество работы сопоставимо со вторым исполнением, которое, однако, значительно дороже. В свою очередь третье исполнение несмотря на то, что оно дешевле, проигрывает в качестве почти по всем пунктам.

Итак, сравнение конкурентных технических решений показало, что разрабатываемая система имеет лучшие качества.

Далее был сформирован SWOT-анализ, в котором балы описаны сильные и слабые стороны проекта, в выявлении возможностей и угроз для реализации проекта, для выявления соответствия и несоответствия была составлена интерактивная матрица проекта. Анализ показал существование уникальной ниши для проекта из-за высокой степени секретности исходного кода конкурентных проектов. Основной риск можно ограничить границами применимости проекта, так же проект может дорабатываться для нивелирования негативного влияния случайных событий.

Проведена оценка готовности проекта к коммерциализации, которая показала, что перспективность разработки средняя.

В рамках процессов инициации определены внутренние и внешние заинтересованные стороны проекта с их ожиданиями от проекта, цели и результат проекта.

План проекта представлен на диаграмме Ганта, из которого видно какой исполнитель (студент или руководитель) какой вид работ осуществлял и в течении какого количества дней.

Рассчитан интегральный финансовый показатель в ходе оценки бюджета затрат трех вариантов исполнения. Интегральный показатель реализации разработки равен 5,34. Значение меньше единицы, что говорит о численном

удешевлении стоимости в разах. Рассчитан интегральный показатель ресурсоэффективности для трех вариантов исполнения: 4,07, 4,94. Расчет интегрального показателя эффективности для разработки и аналога позволил рассчитать сравнительную эффективность разработки, которая говорит о приемлемости существующего варианта решения поставленной в магистерской диссертации технической задачи с позиции финансовой и ресурсной эффективности.

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа		ФИО	
8ВМ03		Мнацаканян Рафаэль Эдуардович	
Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	магистратура	Направление/специальность	09.04.01 «Информатика и вычислительная техника»

Тема ВКР:

Формирование стратегии инвестирования на фондовой бирже при помощи методов машинного обучения.	
Исходные данные к разделу «Социальная ответственность»:	
<p>Введение</p> <ul style="list-style-type: none"> – Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика) и области его применения. – Описание рабочей зоны (рабочего места) при разработке проектного решения/при эксплуатации 	<p><i>Объект исследования Программный модуль анализа финансовых показателей ценных бумаг</i> <i>Область применения: фондовая биржа</i> <i>Рабочая зона: офис</i> <i>Размеры помещения: Ширина комнаты составляет $b=3$ м, длина $a=6$ м, высота $H=3$ м. Площадь помещения будет составлять $S=ab=18$ м², объем $V=abh=54$ м³;</i> <i>Количество и наименование оборудования рабочей зоны: персональный компьютер, устройства ввода и вывода информации, ЖК монитор.</i> <i>Рабочие процессы, связанные с объектом исследования, осуществляющиеся в рабочей зоне: присутствует окно, через которое может производиться вентиляция помещения, принудительная вентиляция отсутствует; в зимнее время помещение отапливается; в помещении используется комбинированное освещение.</i></p>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<p>1. Правовые и организационные вопросы обеспечения безопасности при разработке проектного решения:</p> <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. 	<p>Работа над системой регулируется следующими нормативно-правовыми актами:</p> <ul style="list-style-type: none"> – Трудовой кодекс Российской Федерации от 30.12.2001 № 197-ФЗ (ред. От 1.3.2022). – Федеральный закон № 152 О персональных данных (с изменениями на 2 июля 2021 года). – Кодекс Российской Федерации об административных правонарушениях (с изменениями на 28 мая 2022 года). – Федеральный закон 173-ФЗ О трудовых пенсиях в Российской Федерации (с изменениями на 8 декабря 2020 года). – ГОСТ 12.2.032-78 Система стандартов безопасности труда рабочее место при выполнении работ сидя. – СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания. – ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности. – ГОСТ 12.1.003-83 Система стандартов безопасности труда (ССБТ). Шум. Общие требования безопасности – ГОСТ 12.1.038-82 ССБТ Электробезопасность. Предельно

	<p>допустимые значения напряжений прикосновения и токов.</p> <ul style="list-style-type: none"> – ГОСТ Р 53692-2009 Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов. – ГОСТ 12.1.004-91 Пожарная безопасность. Общие требования.
<p>2. Производственная безопасность <u>при разработке проектного решения:</u></p> <ul style="list-style-type: none"> – Анализ выявленных вредных и опасных производственных факторов – Расчет уровня опасного или вредного производственного фактора 	<p>Вредные факторы:</p> <ul style="list-style-type: none"> – отклонение показателей микроклимата; – превышение уровня шума на рабочем месте; – недостаточная освещенность рабочей зоны; – психофизиологические факторы (монотонность труда, нервнопсихические перегрузки, перенапряжение зрительных анализаторов). <p>Был проведен расчет освещенности рабочей зоны.</p> <p>Опасные факторы:</p> <ul style="list-style-type: none"> – нарушение предельно допустимых значений напряжений прикосновения и токов; – короткое замыкание; – статическое электричество
<p>3. Экологическая безопасность <u>при разработке проектного решения</u></p>	<p>Воздействие на селитебную зону происходит из-за повышенного электромагнитного излучения.</p> <p>Воздействие на литосферу, гидросферу и атмосферу происходит при утилизации компьютера и периферийных устройств (принтеры, МФУ, веб-камеры, наушники, колонки, телефоны), аккумуляторных батарей, люминесцентных ламп, макулатуры.</p>
<p>4. Безопасность в чрезвычайных ситуациях <u>при разработке проектного решения</u></p>	<p>Возможные ЧС: пожары, грозы, ураганы, оползни.</p> <p>Наиболее типичная ЧС пожар.</p>
Дата выдачи задания для раздела по линейному графику	
01.03.2022	

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ООД ШБИП	Федоренко О.Ю.	Д.М.Н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ03	Мнацаканян Рафаэль Эдуардович		

8. Социальная ответственность

Результатом выполнения дипломной работы является рекомендательная система для автоматизированного создания инвестиционных стратегий на фондовой бирже. Разработанные в результате работы алгоритмы могут быть применены отдельно или как часть большей системы. Использование данной системы позволяет единственному человеку при помощи ЭВМ обрабатывать большие объемы информации без значительных временных затрат при помощи предварительно обученных моделей. Потенциальными пользователями системы могут быть частные и институциональные инвесторы и исследователи рынка ценных бумаг. География использования системы не ограничена.

В данном разделе рассматриваются опасные и вредные факторы, оказывающие негативное влияние на разработчика системы, пользователя, окружающей среды и общества в целом. Так же рассматриваются правовые вопросы и организационные мероприятия при чрезвычайных ситуациях.

Система во время разработки и работы не должна наносить вред персоналу, пользователям, окружающей среде и как-либо нарушать действующее законодательство.

Разработка программных средств происходит при помощи ПЭВМ (персональная электронная вычислительная машина) в офисном помещении с использованием сети Интернет.

8.1. Правовые и организационные вопросы обеспечения безопасности

8.1.1. Правовые нормы трудового законодательства

Регулирование отношений между организацией и сотрудниками, в том числе требований к продолжительности рабочего времени и оплаты труда осуществляется законодательством РФ, а именно трудовым кодексом РФ. Согласно ТК РФ[1] работнику обеспечены гарантии в недопустимости дискриминации в сфере труда, запрещение принудительного труда и социальные гарантии.

8.1.1.1. Режим рабочего времени

Режим рабочего дня является предметом рассмотрения в главе 16 ТК РФ [1]. Продолжительность рабочего дня и размер оплаты труда для каждого работника устанавливается в трудовом договоре. Работник обязуется отработать количество времени, оговоренного в договоре, с учетом существующего графика. Работа может выполняться по схеме 5+2, суточными дежурствами, свободными графиком и другими способами.

8.1.1.2. Защита персональных данных работника

Работа с персональными данными сотрудника регламентируется федеральным законом РФ от 27 июля 2006 года № 152-ФЗ «О персональных данных» [2]. Защита персональных данных работника от неправомерного использования или утраты должна быть обеспечена работодателем за счет его средств в порядке установленном ТК РФ [1] и иными федеральными законами.

Согласно статье 10 152-ФЗ [2] так же существуют категории персональных данных, которые не подлежат обработке. Такие как, например данные, касающиеся расовой, национальной принадлежности, политических взглядов, религиозных или философских убеждений, состояния здоровья и интимной жизни.

Лица, виновные в нарушении положений законодательства Российской Федерации в области персональных данных при обработке персональных данных работника, привлекаются к дисциплинарной и материальной ответственности в порядке, установленном ТК РФ [1] и иными федеральными законами, а также привлекаются к гражданско-правовой, административной (ст. 5.39, 13.11, 13.14 КоАП РФ) [2] и уголовной ответственности (ст. 137, 140, 272 УК РФ) в порядке, установленном федеральными законами.

8.1.1.3. Оплата и нормирования труда

Оплата и нормирование труда рассматривается в 6 разделе ТК РФ [1]. Работники имеют окладную часть и стимулирующие выплаты, а статья 133 ТК РФ [1] фиксирует минимальный размер оплаты труда. Месячная зарплата

работника полностью отработавшего норму рабочего времени не может быть меньше установленного законом МРОТ.

Сам размер заработной платы регламентируется существующим договором между работником и компанией. В большинстве случаев компания так же выступает налоговым агентом и выполняет отчисление в социальные фонды и налоговую службу в размерах, установленных законодательством для конкретного работника.

Нормальная продолжительность рабочего дня не должна превышать 40 часов в неделю. Согласно статье 152 ТК РФ [1] сверхурочная работа оплачивается в первые два часа работы не менее чем полуторную ставку, а за последующие часы – не менее чем двойную ставку.

8.1.1.4. Виды компенсаций при работе во вредных условиях труда

Согласно статье 92 ТК РФ [1], продолжительность рабочего времени должна быть сокращена для работников, рабочее место которых относится к 3 или 4 степени опасности труда. Так же подобным работникам доступен ежегодный дополнительный оплачиваемый отпуск сроком не менее 7 календарных дней.

Так же согласно статье 147 ТК РФ [1], зарплата работников, занятых на работах с вредными и (или) опасными условиями труда, устанавливается в повышенном размере. Размер повышения составляет не менее 4%.

Дополнительно к вышеперечисленному, работники подобных категорий получают возможность досрочного назначения трудовой пенсии. Согласно статье 27 и 27.1 Федерального закона «О трудовых пенсиях в Российской Федерации» [4] существует возможность раннего выхода на пенсию в зависимости от категории вредности рабочего места и рабочего стажа.

8.1.2. Эргономические требования к правильному расположению и компоновки рабочей зоны

Разработка, тестирование и дальнейшее использование происходит за сидячим рабочим местом, оборудованном ПЭВМ. Общие эргономические требования описаны в ГОСТ 12.2.032-78 [5]. Все манипуляторы и другие используемые в работе объекты должны располагаться в зоне моторной доступности (изображение 12)

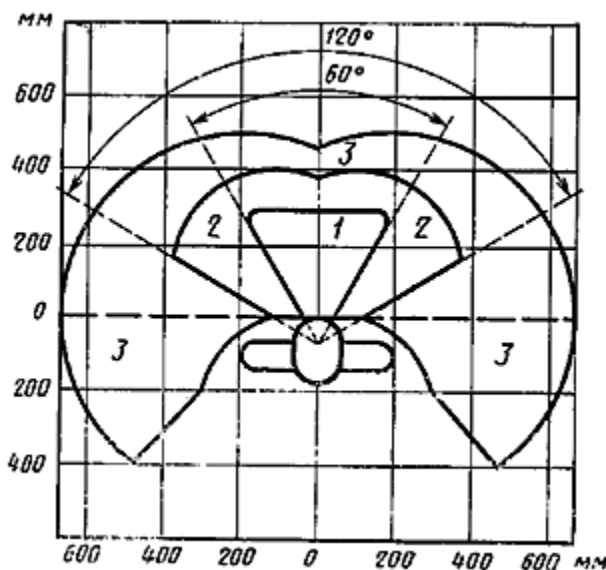


Рисунок 12 – Зона досягаемости моторного поля в горизонтальной плоскости. Так в зоне 1 рекомендуется размещать самые важные органы управления. Зона 2 для часто используемых органов управления и зона 3 для всех остальных органов управления.

Так же в стандарте описаны требования к освещению, эргономике сидения и рабочего стола и пространства для ног. Все эти требования существуют для обеспечения удобной и продуктивной работы без дискомфорта и вреда для здоровья. Вредные факторы и требования безопасности будут рассмотрены ниже.

8.1.3. Производственная безопасность

8.1.3.1. Отклонение показателей микроклимата

Комфортные условия для работы создаются оптимальным сочетанием температуры, относительной влажности и скорости движения воздуха. На рабочих местах пользователей ПЭВМ должны обеспечиваться оптимальные параметры микроклимата в соответствии с СанПиН 1.2.3685-21 [6]

Согласно этому документу, должны быть соблюдены требования, описанные в таблице 17

Таблица 17 – Оптимальные параметры микроклимата в производственном помещении пользователя ПК

Период года	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность, %	Скорость движения воздуха, м/с
Холодный	22-24	21-25	40-60	0,1
Теплый	23-25	22-26	40-60	0,1

Для поддержания оптимальных значений микроклимата используется системы отопления и кондиционирования воздуха, тепловая изоляция нагретых поверхностей оборудования. При исследовании микроклимата было выявлено, что в кабинетах, где выполнялась разработка, параметры микроклимата соответствуют требованиям СанПиН.

8.1.3.2. Превышение уровня шума

Шум является одним из распространенных в производстве вредных факторов. Его создают работающее оборудование, преобразователи напряжения, работающие осветительные приборы дневного света и другие источники шума. Шум может стать причиной снижения работоспособности и повышенной утомляемости. Значительные превышения уровня шума на рабочем месте вызывают необратимые изменения в органах слуха человека, также оказывают неблагоприятное влияние на весь организм человека через

нервную систему. В результате ослабляется внимание, ухудшается память, снижается реакция, что вызывает увеличение числа ошибок при работе

Требования к допустимому уровню шума были описаны в ГОСТ 12.1.003-83 [11] и СанПиН 1.2.3685-21 [6]. Согласно данному документу, допустимый уровень шума составляет не более 50 дБа.

В процессе выполнения проекта на улице около производственного помещения проводились непродолжительные ремонтные работы, во время которых шумовое загрязнение значительно превышало норму. В рамках улучшения условий труда офис был временно изменен на аналогичное помещение с другой стороны здания, в котором уровень шума не превышал допустимый уровень.

Все источники громкого шума (принтеры, сервера и другое оборудование) были изолированы в отдельном помещении.

8.1.3.3. Недостаточная освещенность рабочей зоны

Плохое естественное и искусственное освещение рабочего места оказывает влияние на физическое и психологическое состояние пользователя, что неблагоприятно сказывается на его работе. Не надлежащее качество освещения может привести к ухудшению зрения.

Согласно СП 52.13330.2016 [7] при работах III зрительного разряда и подразряда г (работы высокой точности) освещенность при системе общего освещения должна быть не ниже 200 Лк.

Расчет общего равномерного искусственного освещения горизонтальной рабочей поверхности выполняется методом коэффициента использования светового потока, учитывающим световой поток, отраженный от потолка и стен. Длина помещения $A = 6\text{ м}$, ширина $B = 3\text{ м}$, высота $H = 3\text{ м}$. Высота рабочей поверхности над полом $h_p = 0.8\text{ м}$.

Площадь помещения:

$$S = A * B = 6 * 3 = 18\text{ м}^2 \quad (11)$$

Коэффициент отражения стен, оклеенных светлыми обоями с окнами, без штор $p_c = 30\%$, потолка светлой поверхности $p_{\text{п}} = 50\%$. Коэффициент запаса, учитывающий загрязнение светильника, для помещений с малым выделением пыли равен $K_3 = 1,5$. Коэффициент неравномерности для люминесцентных ламп $Z = 1,1$.

Согласно нормативным документам, принятым в предприятии, помещение освещено тремя светильниками типа ОДОР-2-40 с двумя лампами ЛД-40 со световым потоком $\Phi_{\text{п}} = 2300$ Лм. По паспорту длина светильника $A_{\text{св}} = 1227$ мм, ширина $B_{\text{св}} = 265$ мм. Мощность лампы – 40 Вт.

Интегральным критерием оптимальности расположения светильников является величина λ , которая для люминесцентных светильников с защитной решеткой лежит в диапазоне 1,1-1,3. Примем $\lambda = 1,2$.

Расстояние светильников от перекрытия (свес): $h_c = 0,4$ м.

Высота светильника над рабочей поверхностью определяется по формуле:

$$h = H - h_p - h_c = 3 - 0,8 - 0,4 = 1,8\text{м} \quad (12)$$

Индекс помещения определяется по формуле:

$$i = \frac{A*B}{h*(A+B)} = \frac{6*3}{1,8*(6+3)} = 1,11 \quad (13)$$

Из методической таблицы коэффициент использования светового потока, показывающий какая часть светового потока ламп попадает на рабочую поверхность, для светильников типа ОДОР с люминесцентными лампами при $p_c = 30\%$, $p_{\text{п}} = 50\%$ и индексе помещения $i = 1,11$ равен $\eta = 39\% = 0,39$.

План помещения и размещения светильников с люминесцентными лампами представлен на рисунке 13.

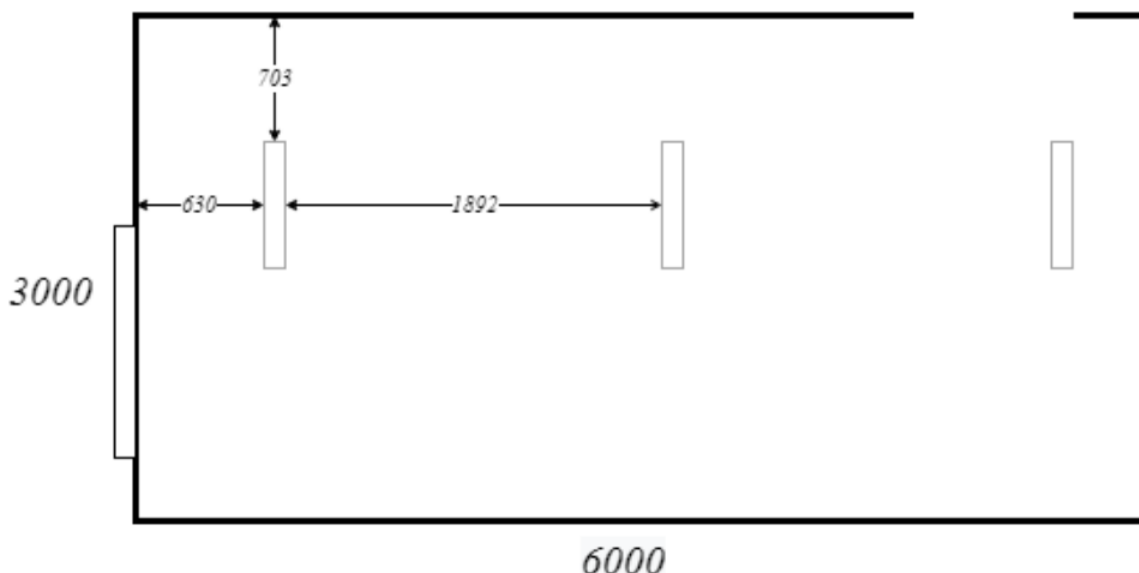


Рисунок 13 – схема расположения светильников с люминесцентными лампами в производственном помещении

Общее число светильников: $N_{\text{св}} = 3$. Соответственно количество люминесцентных ламп: $N_{\text{лл}} = 6$.

Расчет светового потока группы люминесцентных ламп светильника определяется по формуле:

$$\Phi_{\text{рас}} = \frac{E \cdot A \cdot B \cdot K_3 \cdot Z}{N_{\text{лл}} \cdot \eta} = \frac{200 \cdot 6 \cdot 3 \cdot 1,5 \cdot 1,1}{6 \cdot 0,39} = 2538 \text{ лм} \quad (14)$$

Делаем проверку выполнения условия:

$$-10\% \leq \frac{\Phi_{\text{п}} - \Phi_{\text{рас}}}{\Phi_{\text{лд}}} \cdot 100\% \leq 20\% \quad (15)$$

$$\frac{\Phi_{\text{п}} - \Phi_{\text{рас}}}{\Phi_{\text{лд}}} \cdot 100\% = \frac{2300 - 2538}{2300} \cdot 100\% = -10\% \quad (16)$$

Таким образом: $-10\% \leq -10\% \leq 20\%$, необходимый световой поток светильника не выходит за пределы требуемого диапазона, хотя и находится на грани.

Мощность осветительной установки рассчитывается по формуле:

$$P_{\text{уст}} = N \cdot P_{\text{л}} = 6 \cdot 40 = 240 \text{ Вт}. \quad (17)$$

Рассчитаем удельную мощность осветительной установки по формуле:

$$P_{\text{уд}} = \frac{N \cdot P_{\text{л}}}{S} = \frac{6 \cdot 40}{18} = 13,3 \text{ Вт/м}^2 \quad (18)$$

Далее перечислены общие требования и рекомендации к организации

освещения на рабочем месте.

- рабочие места следует размещать таким образом, чтобы естественный свет падал преимущественно слева, а дисплеи мониторов были ориентированы боковой стороной к световым проемам;
- система общего равномерного освещения должна регулировать искусственное освещение в помещениях для эксплуатации ПЭВМ.

Вышеперечисленные меры полностью соблюдаются, что позволяет сохранить зрение и избежать пагубного воздействия на глаза во время разработки и эксплуатации результатов выпускной квалификационной работы.

8.1.3.4. Повышенное значение напряжение в электрической цепи, замыкание которой может произойти через тело человека

Напряжение прикосновения – напряжение, появляющееся на теле человека при одновременном прикосновении к двум точкам проводников или проводящих частей, в том числе при повреждении изоляции. Напряжение может возникнуть на корпусах и каркасах оборудования, в частности ПЭВМ, в случае полного или частичного повреждения электрической изоляции оборудования или питающих оборудование кабелей.

Воздействие напряжения отрицательно влияет на здоровье человека. Стандарт ГОСТ 12.1.038-82 ССБТ [8] устанавливает предельно допустимые значения напряжений прикосновений и токов, протекающих через тело человека.

Напряжения прикосновения и токи, протекающие через тело человека, при нормальном (неаварийном) режиме работы, не должны превышать значений, указанных в таблице 18.

Таблица 18 – Напряжения прикосновения и токи, протекающие через тело человека, при нормальном режиме работы

Вид тока	U, В	I, мА
	Не более	
Переменный, 50 Гц	2,0	0,3
Переменный, 400 Гц	3,0	0,4
Постоянный	8,0	1,0

В помещении, в котором проводились работы происходил необходимый контроль оборудования, повышение значений напряжения не фиксировалось.

8.1.4. Экологическая безопасность

Непосредственно разработка и эксплуатация платформы, в состав которого входит модуль, не оказывают воздействие на окружающую среду. Негативное влияние на атмосферу, гидросферу и литосферу могут оказывать используемые для разработки и эксплуатации вещества. Требования к обращению с отходами описаны в ГОСТ Р 53692-2009 [9].

ПЭВМ, батарейки, оргтехнику в случае выхода из строя, либо по причине устаревания необходимо утилизировать через специализированные организации, имеющие лицензию на данную деятельность.

Отходы от канцелярских товаров, продуктов питания и личной гигиены и другие твердые отходы в случае их неправильной утилизации могут оказывать негативное воздействие на почвенный покров. Защита почвенного покрова от данного воздействия реализуется за счет сбора, сортировки и утилизации отходов и их организованного захоронения.

Все отходы, образующиеся в ходе разработки и эксплуатации результатов выпускной квалификационной работы, утилизируются без оказания негативного влияния на окружающую среду.

8.1.5. Безопасность в чрезвычайных ситуациях

К чрезвычайным ситуациям, которые могут возникнуть во время разработки или эксплуатации, можно отнести возможность возникновения и распространения пожара, так как работа происходит на ПЭВМ. В современных ПЭВМ очень высокая плотность размещения элементов электронных схем, также в непосредственной близости друг от друга

располагаются соединительные провода и кабели. При протекании по ним электрического тока выделяется значительное количество теплоты, при этом возможно оплавление изоляции и возникновение возгорания. Возникновение других видов чрезвычайных ситуаций маловероятно и нецелесообразно для рассмотрения в контексте данной квалификационной работе.

Необходимо соблюдение техники пожарной безопасности для предотвращения возникновения и распространения пожара. Общие требования к пожарной безопасности описаны в ГОСТ 12.1.004-91 [10].

Далее перечислены правила поведения в случае возникновения данной чрезвычайной ситуации.

- В случае возникновения пожара люди должны покинуть помещение в течение минимального времени согласно плану эвакуации (рисунок 3).
- В помещениях с компьютерной техникой, недопустимо применение воды и пены ввиду опасности повреждения или полного выхода из строя дорогостоящего электронного оборудования.
- Для тушения пожаров необходимо применять углекислотные и порошковые огнетушители, которые обладают высокой скоростью тушения, большим временем действия, возможностью тушения электроустановок, высокой эффективностью борьбы с огнем. Воду разрешено применять только во вспомогательных помещениях.
- Помещение должно быть оборудовано пожарными извещателями, которые позволяют оповестить дежурный персонал о пожаре.



Рисунок 14 – План эвакуации

В помещении, где происходила разработка и происходит эксплуатация программных средств, были соблюдены все правила техники безопасности. Сотрудникам были объяснены правила поведения в случае возникновения и распространения пожара.

8.2. Выводы по разделу

В рамках раздела «Социальная ответственность» процесс выполнения и результаты дипломной работы были рассмотрены с точки зрения социальной ответственности на моральные, общественные, экономические, экологические последствия и ущерб здоровью человека. Дополнительно был выполнен анализ на предмет выявления основных опасных и вредных факторов и оценена степень их воздействия на человека, общество и природную среду. Были предложены методы для защиты и минимизации воздействий выявленных факторов, а также методы предотвращения и устранения возможных чрезвычайных ситуаций.

8.3.Законодательные акты

1. Трудовой кодекс Российской Федерации от 30.12.2001 № 197-ФЗ (ред. От 1.3.2022).
2. Федеральный закон № 152 О персональных данных (с изменениями на 2 июля 2021 года).
3. Кодекс Российской Федерации об административных правонарушениях (с изменениями на 28 мая 2022 года).
4. Федеральный закон 173-ФЗ О трудовых пенсиях в Российской Федерации (с изменениями на 8 декабря 2020 года).
5. ГОСТ 12.2.032-78 Система стандартов безопасности труда рабочее место при выполнении работ сидя.
6. СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания.
7. СП 52.13330.2016 Естественное и искусственное освещение.
8. ГОСТ 12.1.038-82 ССБТ Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов.
9. ГОСТ Р 53692-2009 Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов.
- 10.ГОСТ 12.1.004-91 Пожарная безопасность. Общие требования.
- 11.ГОСТ 12.1.003-83 Система стандартов безопасности труда (ССБТ). Шум. Общие требования безопасности

ПРИЛОЖЕНИЕ А

Formation of an investment strategy on the stock exchange using machine learning methods

Обучающийся:

Группа	ФИО	Подпись	Дата
8ВМ03	Мнацаканян Рафаэль Эдуардович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ	Спицын Владимир Григорьевич	Д.Т.Н.		

Консультант – лингвист отделения иностранных языков, школы ИШИТР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Ануфриева Татьяна Николаевна			

1. Literature Review

1.1. Forecasting problem

First of all, it's important to understand nature of problem. In this work algorithm have to predict data, based on historical and some additional values. The problem can be handled differently.

It is important to review what time series is, as well as time series analysis and forecasting. Time Series pertains to the sequence of observations collected in constant time intervals: daily, monthly, quarterly or yearly. Time Series Analysis can be described as understanding internal correlation in data that helps predicts what is going to happened.

To be predictable data have to be specific. It should have:

- Trends (to describe increasing or decreasing behavior of the time series frequently presented in linear modes).
- Seasonality (to highlight the repeating pattern of cycles of behavior over time)
- Cyclicity (to identify the repetitive changes in the time series and define their placement in the cycle).

Without those components no model can be good enough to work on that prediction task.

There are different appliances to forecasting problems.

1.2. Existing forecasting methods

Forecasting problem can be targeted with classical naive approach.

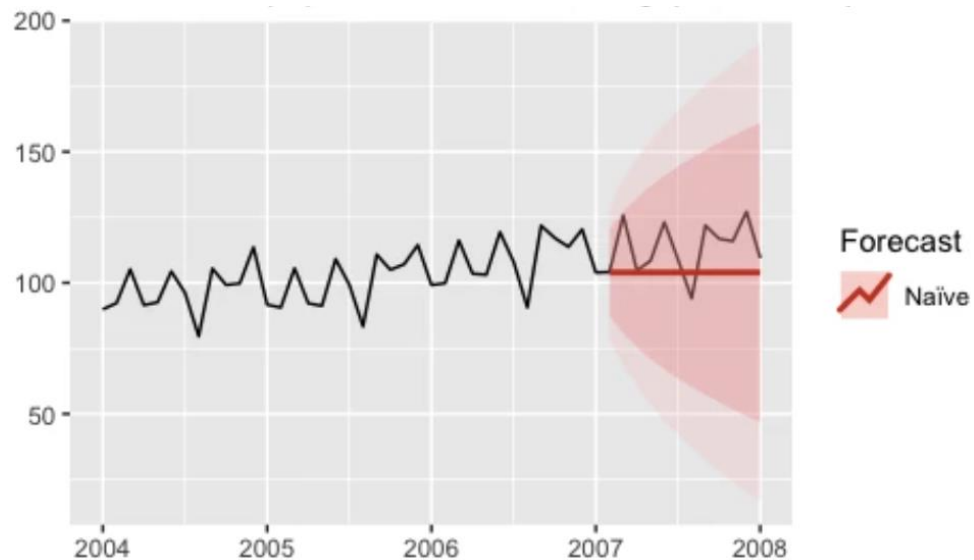


Image 1 – Visualization of naive forecast approach.

In most cases, naive models are applied as a random walk (with the last observed value used as a unit for the next period forecast) and a seasonal random walk (with a value from the same period of the last observed time span used as a unit of the forecast).

Machine learning approaches is much more sophisticated.

For example, modern step-of-the-art solution is LSTM networks.

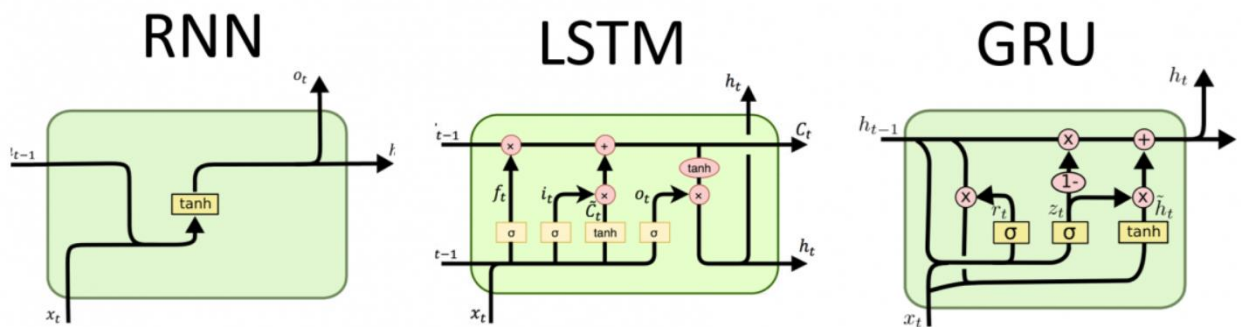


Image 2 – Different recurrent artificial neurons with memory.

LSTM cells (special RNN cells) were developed to find the solution to the issue with gradients by presenting several gates to help the model make a decision on what information to mark as significant and what information to ignore. GRU is another type of gated recurrent network.

That model can learn and remember data and provide forecast by knowing all history data point simultaneously.

However, to create advertise model problem can be targeted from another angle. In provided research data was split to test and train batches and for every test sample we calculate answer – whether company successful in 3 month or not. That give opportunity to solve classification problem instead of regression problem.

Now there are several specific data points that should be distinct one from another.

2. Quality metrics

The following are the metrics used to assess the quality of the detection algorithm.

2.1. Class Recall

The Class Recall metric describes the ability of the detection algorithm to detect objects of a certain class. This metric can be calculated using the following formula:

$$\text{Class Recall} = \frac{TP}{TP+TN},$$

where TP is the number of correct predictions of a certain class, Ground truths is the number of true detections for a certain class. The higher the Class Recall value, the better the algorithm's ability to identify all objects of a certain class

2.2. Class Precision

The Class Precision metric describes the proportion of objects correctly identified as belonging to a certain class relative to all objects that the detection algorithm attributed to this class. This metric can be calculated using the following formula:

$$\text{Class Precision} = \frac{TP}{\text{Detections}},$$

where TP is the number of correct predictions of a certain class, Detections is the number of all detections for this class. The higher the Class Precision value, the better the detector's ability to distinguish objects of a certain class from others.

2.3. Class AP and MAP

The AP (Average Precision) metric is defined as the area under the precision-recall curve for objects of a certain class. The higher the AP value, the better the balance between Class Precision and Class Recall, and therefore the more confidence in the detector's performance.

The MAP (Mean Average Precision) metric is the average AP value for all classes

2.4. Class AR and MAR

The AR (Average Recall) metric is defined as the doubled area under the recall-IOU curve for objects of a certain class. The MAR (Mean Average Recall) metric is the average AR for all classes

2.5. F1 score

In the F1 score, we compute the average of precision and recall. To mix its meanings to one universal value its common to use the harmonic mean. The F1 score formula is shown here:

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Since the F1 score is an average of Precision and Recall, it means that the F1 score gives equal weight to Precision and Recall.

It means that:

- A model will obtain a high F1 score if both Precision and Recall are high
- A model will obtain a low F1 score if both Precision and Recall are low
- A model will obtain a medium F1 score if one of Precision and Recall is low and the other is high

That metric is great for brief understanding of model robustness. It generally much more representable than precision and recall metrics separately.

2.6.ROC

ROC represents for receiver operating characteristic curve. It's a graph showing the performance of a classification model at all classification thresholds.

This curve plots two parameters:

- Recall (true positive rate)
- False Positive Rate

On ROC is specific representation of those two parameters.

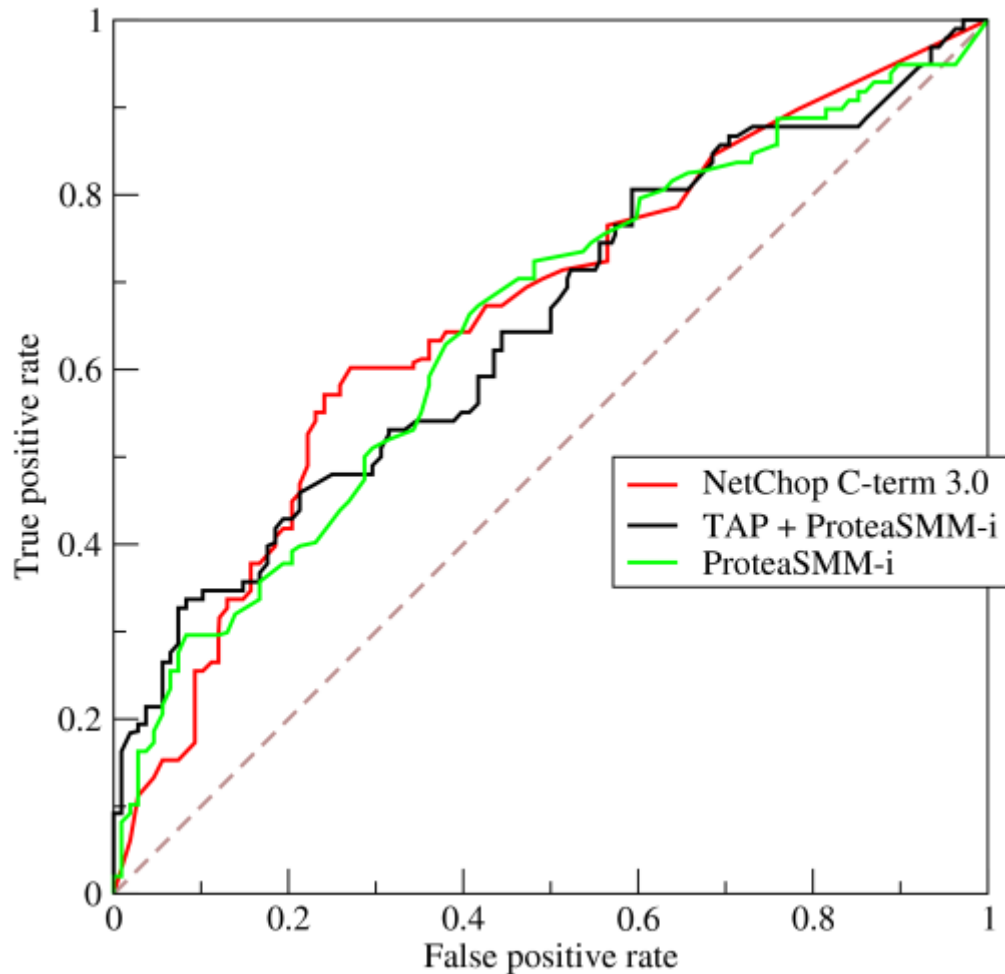


Image 3 – ROC curve.

The most useful part of that graph is called AUC (area under curve). Size of that area represents how good model perform. If area is lower than 50% of all graph area it means that the solution is worse than just random guessing. On the other hand, if curve is above main diagonal, it means that model is robust and have some predictive capability.

It is great tool to roughly estimate whether the model is good or not.

3. Solution algorithm

In that research a lot of different classification algorithms was used. Best results were achieved with random forest and gradient boosting.

A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.

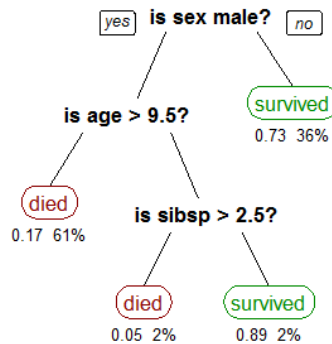


Image 4 – Decision tree.

A random forest algorithm consists of many decision trees. The ‘forest’ generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

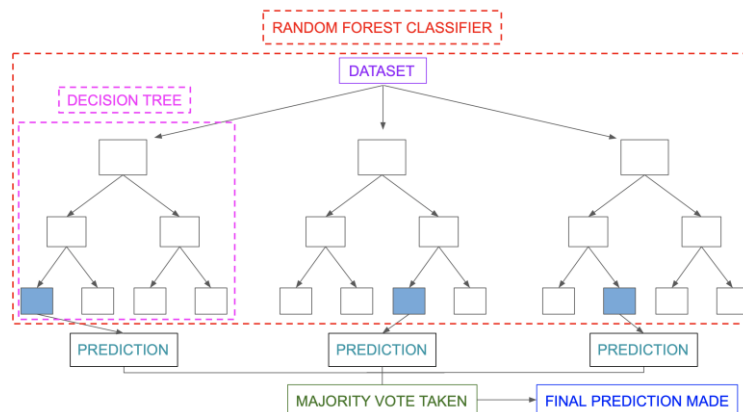


Image 5 – Random Forest.

Another algorithm that provided great forecasting result was gradient boosting. That is approach of combining lots of weak classifiers to achieve better accuracy than any of them can produce separately.

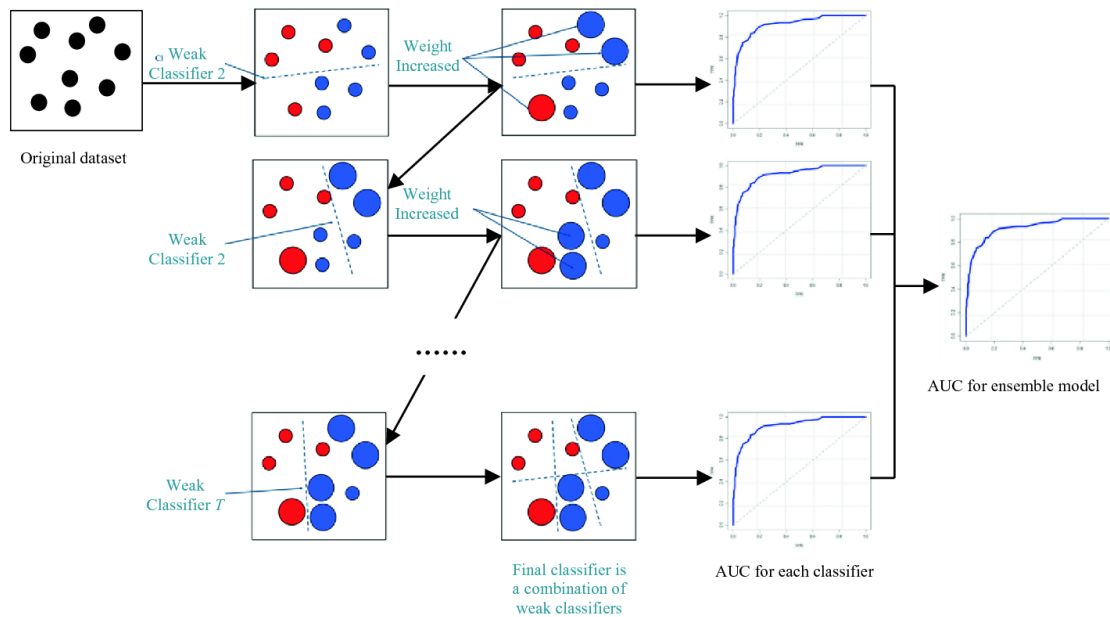


Image 6 – ensemble of weak classifiers with gradient boosting.

In a result random forest get AP 0,6815 and gradient boosting gets AP: 0,6712.