

Абу Сувейлим Мухаммед Мунифович

Маслова Анастасия Сергеевна

Коротун Илья Игоревич

ул. Орджоникидзе, 3

+79004852795

**Предоставление услуг анализа тональности финансовых  
новостей**

Научное направление: NLP

Москва 2024 г.

## Оглавление

<b>Резюме бизнес-плана .....</b>	<b>4</b>
<b>Цель и сущность предлагаемого проекта .....</b>	<b>4</b>
<b>Стадия развития проекта на момент составления бизнес-плана.....</b>	<b>4</b>
<b>Доказательства выгоды .....</b>	<b>5</b>
<b>Состав конкретных мероприятий, которые необходимо реализовать в рамках проекта .....</b>	<b>5</b>
<b>Средства, необходимые для реализации проекта .....</b>	<b>5</b>
<b>Ключевые факторы успеха.....</b>	<b>6</b>
<b>Краткое описание инновационной идеи, положенной в основу проекта .....</b>	<b>7</b>
<b>Характеристика проблемы и обоснование необходимости ее решения .....</b>	<b>7</b>
<b>Основные цели и задачи .....</b>	<b>7</b>
<b>Описание инновационной идеи .....</b>	<b>7</b>
<b>Механизм реализации инновационной идеи.....</b>	<b>8</b>
<b>Требования к ресурсному обеспечению инновационной идеи .....</b>	<b>8</b>
<b>Оценка экономической эффективности инновационной идеи.....</b>	<b>9</b>
<b>Описание планируемой к производству продукции.....</b>	<b>9</b>
<b>Наименование продукции.....</b>	<b>9</b>
<b>Назначение и область применения .....</b>	<b>9</b>
<b>Перспективы выпуска продукции на конкретном рынке .....</b>	<b>9</b>
<b>Возможность замещения, в том числе импортозамещения.....</b>	<b>10</b>
<b>Краткое описание и основные технико-экономические и потребительские характеристики.....</b>	<b>10</b>
<b>Конкурентоспособность продукции.....</b>	<b>10</b>
<b>Возможности повышения конкурентоспособности .....</b>	<b>10</b>
<b>Степень готовности.....</b>	<b>10</b>
<b>Анализ рынка.....</b>	<b>11</b>
<b>Размер рынка .....</b>	<b>11</b>
<b>Темпы роста рынка, фаза развития спроса.....</b>	<b>11</b>
<b>Уровень удовлетворения спроса.....</b>	<b>11</b>

<b>Характер отрасли .....</b>	<b>12</b>
<b>Динамика продаж аналогов за последние 5 лет по России, СНГ, в мире .....</b>	<b>12</b>
<b>Прогнозы развития отрасли.....</b>	<b>12</b>
<b>Основные и потенциальные конкуренты (наименования и адреса фирм - основных производителей товара, их сильные и слабые стороны) .....</b>	<b>12</b>
<b>Уровень рентабельности отрасли .....</b>	<b>13</b>
<b>Маркетинговый план .....</b>	<b>13</b>
<b>Конечные потребители.....</b>	<b>13</b>
<b>Уровень удовлетворения спроса, его характер.....</b>	<b>13</b>
<b>Особенности сегмента рынка.....</b>	<b>14</b>
<b>Конкурентные преимущества услуги конкурентов и предприятия, реализующего проект.....</b>	<b>14</b>
<b>Планируемая доля рынка .....</b>	<b>14</b>
<b>Патентная ситуация, возможность для конкурентов производить соответствующую продукцию без нарушения патентных прав претендента .....</b>	<b>14</b>
<b>Обоснование цены на продукцию .....</b>	<b>14</b>
<b>Организация сбыта .....</b>	<b>15</b>

## Резюме бизнес-плана

### Цель и сущность предлагаемого проекта

Цель проекта – разработка и внедрение автоматизированной модели для анализа тональности финансовых новостей, что позволит компаниям, занимающимся управлением портфелем акций, эффективно собирать и анализировать данные рынка.

Анализ тональности финансовых новостей — это двухэтапная задача. Первая — это задача выделения и классификации именованных сущностей в тексте, таких как имена людей, названия организаций, даты, местоположения, суммы денег и другие типы специфических объектов. Вторая задача – это анализ и классификация тональности текста. Наш проект представляет собой модель на основе фреймворка ruABSA с использованием модели BERT, которая решает задачи распознавания сущностей и анализа тональностей.

### Стадия развития проекта на момент составления бизнес-плана

Сейчас модель обучается на англоязычных датасетах и выводит двухэлементный кортеж, включающий только сущность и ее тональность.

Пример:

Ввод: `Apple stocks went up by 10% on Monday, according to TASS."`

Вывод: `("Apple", Positive), ("TASS", Neutral)`

В будущем планируется обучать модель на русскоязычных данных для распространения на российском рынке. Кроме того, планируется, что программа будет определять пятиэлементный кортеж (`sentiment-holder`, `sentiment-target`, `sentiment-target-aspect`, `sentiment`, `sentiment-time`).

Пример:

**Ввод:** "Apple stocks went up by 10% on Monday, according to TASS."

**Вывод:** ("Apple", "Apple", "Apple stocks", Positive, Monday),  
("TASS", "Apple", "Apple stocks", Neutral, Monday)

### **Доказательства выгоды**

Если в качестве взять набор данных финансовых новостей, состоящий из 10000 строк, то чтобы вручную извлечь сущности и тональности текста относительно них, потребуется минимум 30 часов. Наша модель способна выполнить эту задачу меньше чем за минуту, что позволяет сократить временные и ресурсные расходы.

### **Состав конкретных мероприятий, которые необходимо реализовать в рамках проекта**

- приобрести оборудование, мощности которого хватит для качественного обучения модели, для начала – сервер с 16 видеокарт NVIDIA A100;
- улучшить модель, чтобы она соответствовала требованиям точности вывода. Вместо двухэлементного кортежа (объект тональности, тональность) планируется увлечения количество элементов, таких как атрибут объекта тональности и время, когда мнение было выражено;
- запатентовать модель;
- найти компании, готовые приобрести наши услуги и заключить с нами договор;
- заключить договоры с компаниями.

### **Средства, необходимые для реализации проекта**

Для реализации проекта нам потребуются инвестиции в оборудование и программное обеспечение, что потребует стартового капитала в размере минимум 200 000 рублей (одна видеокарта NVIDIA RTX 4090). В среднем реализация данного проекта требует 15-20 миллионов рублей. Ниже

представлена таблица с расходами на все необходимое оборудование для функционирования стартапа (табл. 1).

<b>Количество</b>	<b>Оборудование</b>	<b>Стоимость</b>
16	NVIDIA A100	16 000 000 руб.
6	Intel XEON gold 6258r	1 953 600 руб.
8	Samsung 970 PRO 1TB NVMe M.2 SSD	160 000 руб.
4	Samsung 860 EVO 1TB SATA SSD	40 000 руб.
10	Western Digital 10TB Ultrastar DC HC520	300 000 руб.
1	Broadcom MegaRAID	80 000 руб.
1	Synology DS1819+ 8 x 10TB HDDs	340 000 руб.
6	Corsair Hydro Series H150i Pro	90 000 руб.
16	EKWB Custom Loop Kit	640 000 руб.

Таблица 1. Необходимое оборудование и его стоимость

## Ключевые факторы успеха

Наш проект будет считаться успешным, если ежемесячный доход составит более 400000 рублей с учетом расходов на электричество, аренду офиса/места для сервера, хранилища и пр.

## **Краткое описание инновационной идеи, положенной в основу проекта**

### **Характеристика проблемы и обоснование необходимости ее решения**

Существует много рабочих моделей для определения тональности в различных сферах, но нет моделей для определения тональности в финансовых новостях. Существующие модели нельзя применять к финансовым новостям, иначе вывод будет некорректным. Кроме того, на данный момент в РФ недостаточно компаний, предоставляющих услуги анализа для дальнейшего финансового консалтинга.

### **Основные цели и задачи**

Основной целью проекта является разработка и внедрение автоматизированной модели для анализа тональности финансовых новостей, что позволит компаниям, занимающимся управлением портфелем акций, эффективно собирать и анализировать данные рынка. Задачи заключаются в следующем: представить проект компаниям, занимающимся управлением портфелем акций, и заключить договор с одной или несколько компаний. В случае, если заключить договор не получится, будет необходимо улучшить модель и подать заявку на грант.

### **Описание инновационной идеи**

Наша инновационная идея заключается в создании специализированной модели машинного обучения, способной анализировать тональность финансовых новостей с высокой точностью. В основе модели лежат современные методы обработки естественного языка (NLP), включая использование архитектуры BERT, адаптированной для анализа финансовых текстов. Это позволит не только определять общий тон новостей (положительный, отрицательный, нейтральный), но и учитывать специфические финансовые термины и контексты, что обеспечит более точный и релевантный анализ.

## **Механизм реализации инновационной идеи**

Реализация проекта будет проходить в несколько этапов:

1. Исследование и сбор данных: Сбор и аннотирование большого объема финансовых новостей для создания обучающей выборки.
2. Разработка модели: Адаптация архитектуры BERT и других современных NLP моделей для специфики финансовых текстов.
3. Обучение и тестирование: Обучение модели на собранных данных и её тестирование для обеспечения высокой точности.
4. Внедрение и интеграция: Внедрение модели в рабочие процессы клиентов, интеграция с их системами управления данными и аналитики.
5. Поддержка и улучшение: Постоянное обновление модели на основе новых данных и обратной связи от пользователей.

## **Требования к ресурсному обеспечению инновационной идеи**

Для реализации проекта потребуются:

- Финансовые ресурсы: Инвестиции в размере 15-20 миллионов рублей, включающие затраты на оборудование, программное обеспечение и оплату труда специалистов.
- Человеческие ресурсы: Команда из экспертов в области машинного обучения, аналитиков данных, финансовых консультантов и разработчиков программного обеспечения.
- Технические ресурсы: Высокопроизводительные серверы для обучения моделей и хранения данных, лицензии на необходимое программное обеспечение и доступ к специализированным базам данных финансовых новостей.



## **Оценка экономической эффективности инновационной идеи**

Экономическая эффективность проекта оценивается на основе нескольких факторов:

- Снижение затрат: автоматизация анализа финансовых новостей позволит клиентам сократить расходы на ручную обработку и анализ данных.
- Увеличение доходов: более точный и своевременный анализ новостей поможет клиентам принимать обоснованные инвестиционные решения, что повысит их доходы.
- Рентабельность инвестиций: ожидается, что вложенные в проект средства окупятся в течение первых двух лет за счет подписок на услугу и продажи лицензий на использование модели.

## **Описание планируемой к производству продукции**

### **Наименование продукции**

«Moscow Analytica»

### **Назначение и область применения**

Программное обеспечение предназначено для автоматизации анализа тональности финансовых новостей, что позволяет компаниям, занимающимся управлением акционерными портфелями, эффективно собирать и анализировать данные для принятия обоснованных инвестиционных решений.

### **Перспективы выпуска продукции на конкретном рынке**

С учетом растущего спроса на инструменты анализа больших данных и автоматизации процессов в финансовом секторе продукт имеет значительные перспективы, особенно на рынках с высокой концентрацией инвестиционных компаний и бирж.

## **Возможность замещения, в том числе импортозамещения**

Продукт может заменить импортные аналоги, так как основан на открытых технологиях и моделях, таких как BERT, и может быть адаптирован под специфику локального рынка.

## **Краткое описание и основные технико-экономические и потребительские характеристики**

Модель использует передовые методы машинного обучения для анализа тональности текста, обеспечивая высокую точность и скорость обработки данных. Точность работы модели составляет 88%.

## **Конкурентоспособность продукции**

Благодаря использованию современных алгоритмов и возможности кастомизации, «Moscow Analytica» выделяется на фоне конкурентов и предлагает уникальное решение для анализа финансовых данных. На данный момент на рынке не существует ПО или модель способна определять тональность финансовых новостей и объекты тональности одновременно с точностью нашей модели.

## **Возможности повышения конкурентоспособности**

Дальнейшее развитие продукта может включать интеграцию с дополнительными источниками данных, улучшение алгоритмов обработки естественного языка и расширение функционала для анализа различных типов финансовых инструментов. Также рассматривается возможность включения финансовых данных на других языках в работы модели.

## **Степень готовности**

На данный момент продукт готов, но было принято решение дальнейшего развития для более точных расчетов и более подробного вывода. Сейчас модель определяет тональность текста и сущность, к которому тональность обращена,

но в дальнейшем планируется улучшить модель. Вместо двухэлементного картежа (объект тональности, тональность) планируется увлечения количество элементов, таких как атрибут объекта тональности и время, когда мнение было выражено.

## **Анализ рынка**

### **Размер рынка**

Рынок больших данных (БД) в России растет. Исследование АБД показало, что в 2018 году объем индустрии составил ₽89 млрд, а спустя три года, в 2021-м — ₽170 млрд.

### **Темпы роста рынка, фаза развития спроса**

Российский рынок ВІ находится в активной стадии трансформации. На конец 2023 года российским компаниям пока так и не удалось потеснить зарубежные решения, которые продолжают обеспечивать потребность в аналитике приблизительно для половины российских заказчиков. Несмотря на то, что 2022 год оказался одним из самых сложных периодов для сектора БД, в этом же году открылись новые уникальные возможности для отечественных компаний в связи с уходом с российского рынка иностранных гигантов. По оценке АБД рынок больших данных в России составил 170 млрд. руб. в 2021 году, при консолидации действий бизнеса и государства будет обеспечен рост на 90% до 319 млрд руб. к концу 2024 года.

### **Уровень удовлетворения спроса**

Существующие решения часто ограничены и не полностью удовлетворяют спрос на точный анализ, что создает возможности для «Moscow Analytica».

## **Характер отрасли**

Отрасль динамична, с быстрым внедрением новых технологий и постоянным изменением требований рынка. В 2021 г. В России утвердили первый национальный стандарт в области больших данных. BI-рынка (Business Intelligence) в России в 2020-2022 гг. составил примерно 13%, в то время как средний прирост мирового рынка – 11%. После ухода с российского рынка иностранных ПО ситуация на рынке не стабильная. «В настоящий момент доля отечественных BI-систем находится в диапазоне 35-45% при выраженном тренде на увеличение. Ряд экспертов, при условии неизменной конъюнктуры рынка, прогнозируют рост отечественной доли рынка до 85%-90% в ближайшие три года. Мои прогнозы более осторожные – около 63-65% до конца 2026 года» – делится экспертизой директор по развитию и цифровой трансформации РДТЕХ Евгений Осьминин.

## **Динамика продаж аналогов за последние 5 лет по России, СНГ, в мире**

Продажи аналитических инструментов показали стабильный рост, особенно в сегментах, связанных с финансовыми технологиями и искусственным интеллектом. 70% популярных в России BI-систем разработаны зарубежными вендорами

## **Прогнозы развития отрасли**

Аналитики предсказывают продолжение роста спроса на интеллектуальный анализ данных, особенно в финансовом секторе. По данным Ассоциации больших данных, дополнительный эффект от использования больших данных для отраслей экономики составит 1,6 трлн. рублей операционной прибыли.

## **Основные и потенциальные конкуренты (наименования и адреса фирм - основных производителей товара, их сильные и слабые стороны)**

Конкуренты включают как крупные технологические компании, так и специализированные стартапы. Ключевыми игроками являются те, кто

предлагает аналогичные аналитические инструменты, но «Moscow Analytica» может выделиться за счет точности анализа.

Основной конкурент - компания Neoflex (АО "Неофлекс Консалтинг", 127015, Москва, ул. Вятская, д. 35, стр. 4, 1 подъезд, 2 этаж), потенциально способная оказывать данную услугу, имеет похожую модель, но ее точность составляет всего 52%.

### **Уровень рентабельности отрасли**

Отрасль обладает высоким потенциалом рентабельности, особенно для продуктов, которые могут предложить уникальные и эффективные решения для анализа данных. По данным TadViser российские вендоры BI, которые ранее занимали по различным оценкам не более 10-15% рынка, оказались просто не готовы предложить полноценную замену ставших недоступными платформам.

### **Маркетинговый план**

#### **Конечные потребители**

Существующие потребители - компании, занимающиеся консалтингом в сфере управлением портфелем акций.

Потенциальные потребители - финансовые аналитики, инвестиционные банки, хедж-фонды и частные инвесторы, стремящиеся к автоматизации анализа рыночных данных.

#### **Уровень удовлетворения спроса, его характер**

Спрос на автоматизированный анализ данных является равномерным, так как потребность в точной и оперативной информации о рынке актуальна вне зависимости от времени года.

## **Особенности сегмента рынка**

Рынок финансовых технологий высококонкурентен и быстро развивается. Компании постоянно ищут способы улучшения аналитических инструментов для принятия решений. 70% инструментов анализа финансовых данных было иностранным ПО.

## **Конкурентные преимущества услуги конкурентов и предприятия, реализующего проект**

Пока в России нет компаний, которые занимались бы предоставлением услуги анализа тональности новостей в финансовой сфере. Кроме того, качество подобной услуги сравнительно невелико.

## **Планируемая доля рынка**

Цель – захватить до 5% рынка в течение 1,5 года после запуска продукта, с последующим увеличением доли благодаря расширению функционала, увеличению числа сотрудников и улучшению сервиса.

## **Патентная ситуация, возможность для конкурентов производить соответствующую продукцию без нарушения патентных прав претендента**

Необходимо провести патентный поиск и оценить возможность патентования ключевых технологий и методик, что снизит риск копирования продукта конкурентами.

## **Обоснование цены на продукцию**

Цены на услуги будут определяться после исследования рынка и средней цены на подобную услугу. Для того, чтобы удостовериться в окупаемости, будут учитываться также затраты на оборудование, аренду и выплату зарплат сотрудникам.

## **Организация сбыта**

Продажи будут осуществляться через прямые продажи крупным клиентам и через онлайн-платформу для мелких и средних компаний. Также планируется сотрудничество с консультационными агентствами и интеграция с популярными платформами для трейдинга.