

# Разработка торговой стратегии автоследования

**Куратор:**

Кирилл Долматов

**Участники:**

Владимир Лыткин

Вишняков Александр

# Описание изначальной задачи

## Идея:

В проекте предлагается создать торговую систему, основанную на мониторинге сделок и сообщений в социальной сети инвесторов Тинькофф Пульс

## Техническая часть:

**Инженерная часть** - сбор, хранение, потоковая обработка данных и построение дашбордов, мониторинг, получение альтернативных данных, дообучение моделей в проде

**Пул NLP направлений:** анализ тональности, NER, кластеризация текстов

**ML:** генерация признаков, кластеризация пользователей по профилю торговли, построение предиктивной модели, перебор параметров

# Идея проекта

В нашем проекте предлагается сделать телеграмм бота, который будет выдавать некую аналитику по ценным бумагам московской биржи.

**Идеальный конечный результат нашего проекта выглядит следующим образом:**

на вход мы подаем тикер компании (тикер - это название ценной бумаги) – на выходе получаем аналитику по бумаге, которая позволит ответить на следующие вопросы:

- Переоценена ли бумага? Какая хорошая цена для покупки бумаги? Стоит ли её покупать?
- Какие есть интересные бумаги для приобретения на рынке?
- Какой новостной фон у компании? (хороший/нормальный/плохой)
- Каким потенциалом обладает компания? Сколько она может стоять через некоторый промежуток времени?

# Описание данных их сбора

## Finam

- Новости
- Исторические данные котировок
- Тикеры акций мосбиржи



**ФИНАМ**

## МОЕХ

- Финансовые агрегаты



## Тинькофф Пульс

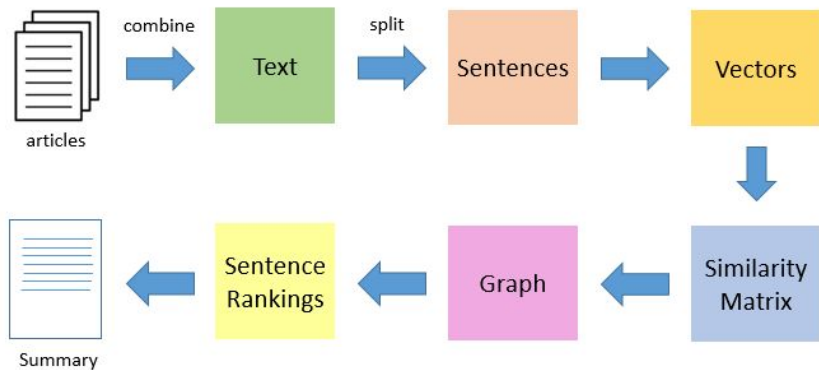
- Данные пользователей
- Публикации пользователей по бумаге



**ТИНЬКОФФ**

# Описание модели

В качестве первой модели была использована модель **TextRank** для суммаризации текста.



Мы испытали два вида этой модели: базовую и с применением трансформера. Модель с применением трансформера выдаёт визуально лучший результат, но имеет большее время работы.

В нашем боте по итогу работает базовая модель, тк модель с применением трансформера работает слишком долго для нашей задачи (примерно 50 секунд на новость)

# Описание модели

В качестве второй задачи решается задача классификации новостей.

Таргет: Рост акции в течение 10 торговых дней после публикации.

Для этой задачи пробуются подходы с использованием предобученной **Word2Vec** и **BERT** с логистической регрессией, но работы по данной модели ещё не закончены.

# Пример реализованного функционала



бот поднимается локально

