

Высшая школа кибертехнологий, математики и статистики (факультет)
Кафедра информатики

Выпускная квалификационная работа

Направление 38.03.05 «Бизнес-информатика»
профиль «Цифровая трансформация бизнеса»

тема:

Разработка и внедрение системы предиктивной аналитики в условиях импортозамещения

Выполнил обучающийся Крабу Кира Сергеевна

Группа 15.11Д-БИЦТ09/216

Научный руководитель выпускной
квалификационной работы
Батищев А.В., к.э.н, доцент,
доцент кафедры информатики



Актуальность, цель и задачи работы

Актуальность:

1. Недоступность западных систем
2. Недостаточная функциональность и доступность отечественных решений
3. Повышенный интерес к автоматизации аналитики

Объект: ООО «ПИКС Роботикс» (PIX Robotics)

Предмет: процесс предсказательного анализа фондового рынка

Цель:

Разработка системы прогнозирования котировок обыкновенных акций российских компаний (Сбер и ВТБ) на горизонте 20 рабочих дней с учетом тональности новостей и формированием торгового сигнала

Задачи:

1. Анализ существующих решений
2. Проектирование БД
3. Автоматизация сбора данных
4. Построение модели прогнозирования
5. Визуализация результатов в BI
6. Оценка экономической эффективности

Компания PIX Robotics



РЭУ.РФ
РОССИЙСКИЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ
ИМЕНИ Г.В. ПЛЕХАНОВА



Российский разработчик
RPA, BI и BPM-систем

PIX Robotics

Цели

Выход на рынок инвестиционной
аналитики

Увеличение прибыли и снижение убытков

Интеграция платформы аналитики в
экосистему компании

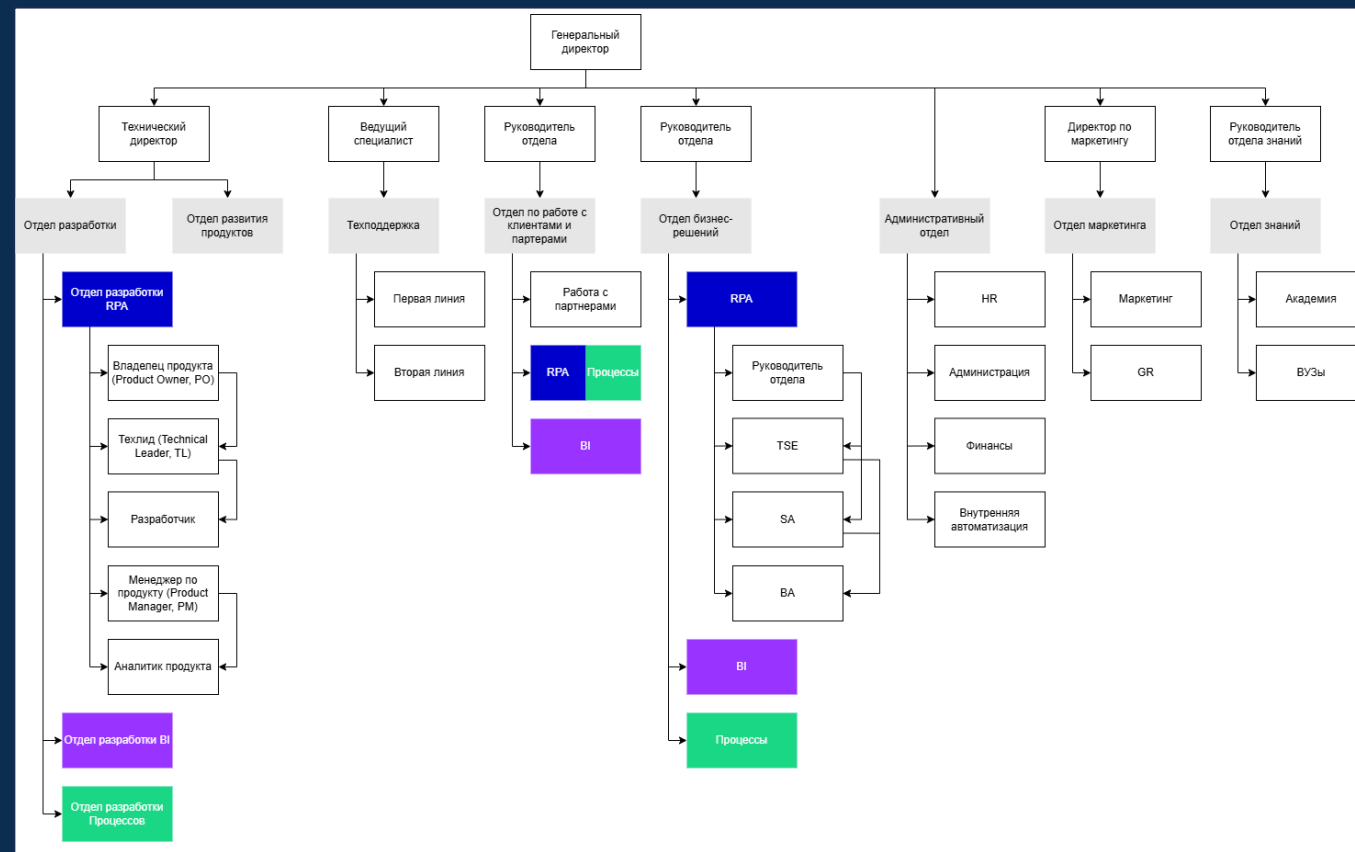
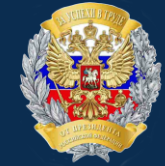


Рис.1 – Организационная структура PIX Robotics

Используемые продукты: PIX RPA, PIX BI

Процесс «как есть»



Узкие места:

1. Ручной сбор новостей и технических индикаторов
2. Субъективный анализ новостей
3. Формирование таблиц и отчетов вручную

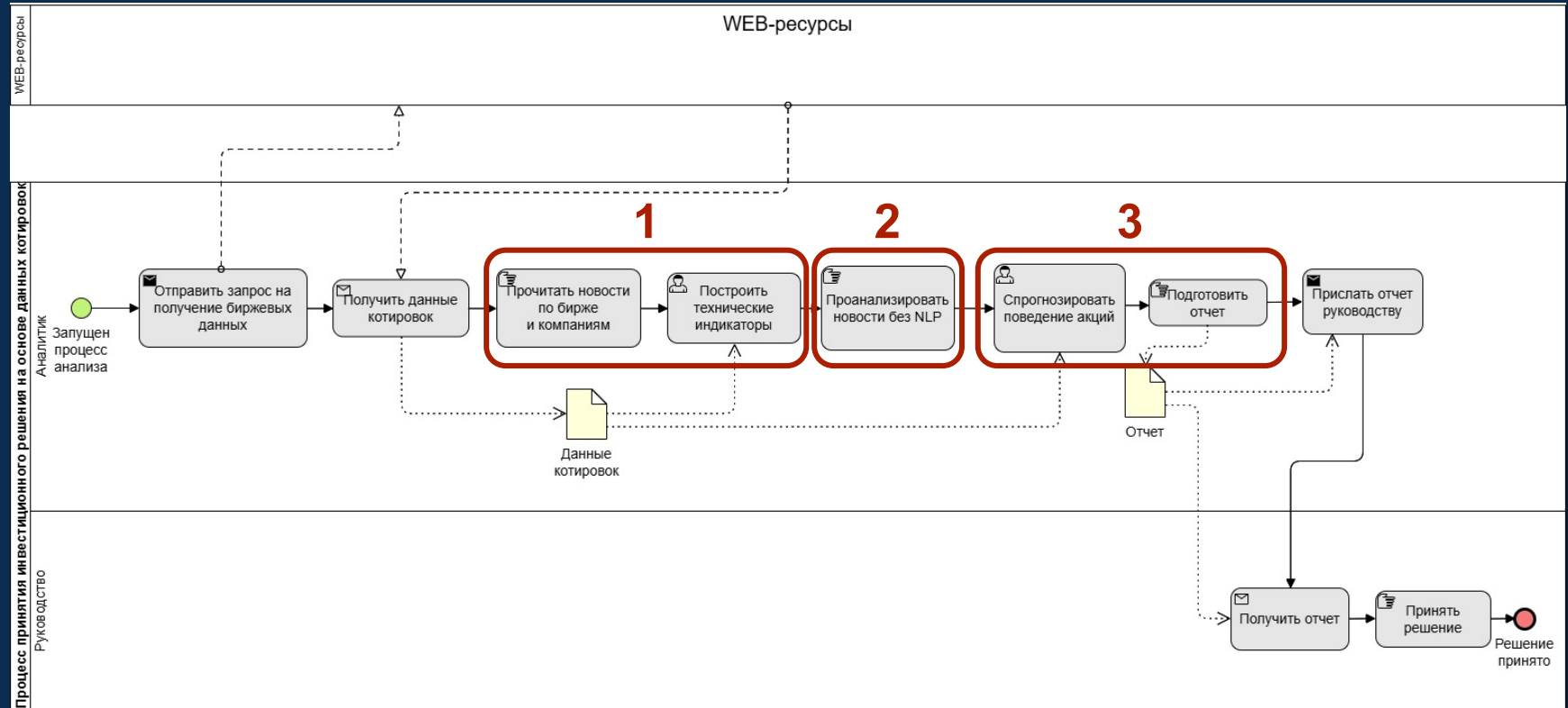


Рис.2 – Процесс «как есть»

Постановка задачи:

Разработать систему прогнозирования цен акций Сбера и ВТБ на 20 дней вперед, которая:

Автоматически собирает и анализирует данные

Формирует торговые сигналы (BUY/SELL/HOLD)

Визуализирует результаты на BI-дашборде

Процесс «как должно быть»

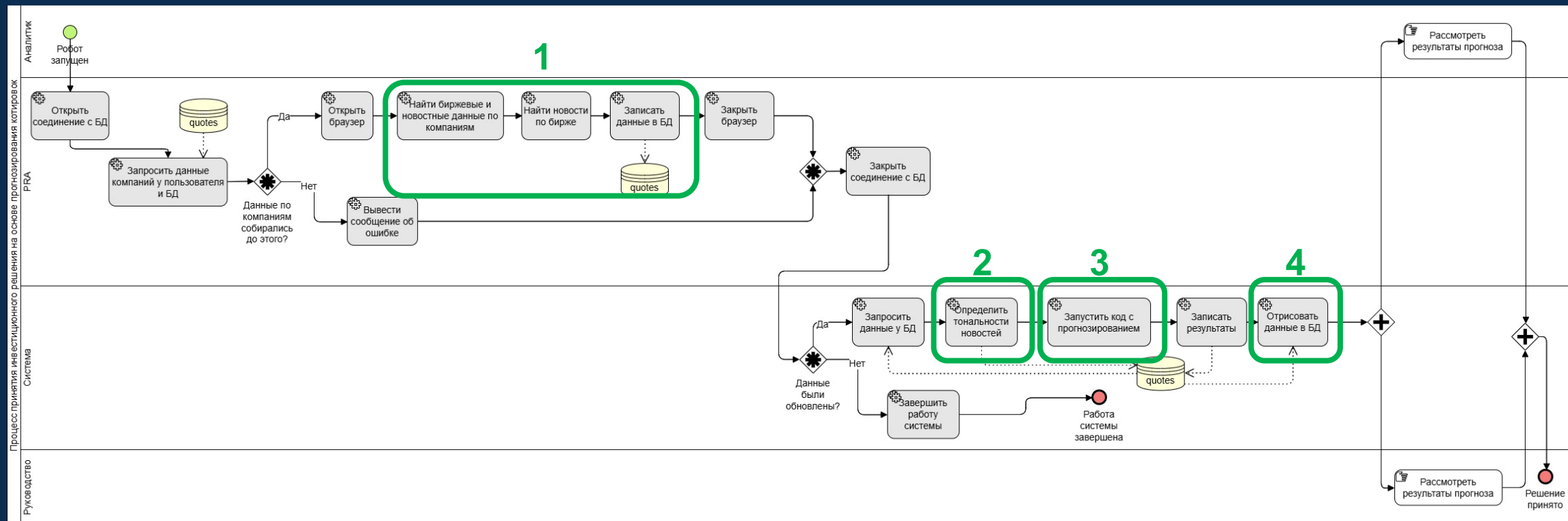


Рис.3 – Процесс «как должно быть»

После автоматизации:

1 Автоматизированный сбор данных

2 Анализ тональности новостей

3 Формирование прогноза и торгового сигнала

4 Визуализация в BI

Техническое задание



Функциональные требования

1. Автоматический сбор биржевых данных и новостей по российским компаниям – ПАО «Сбербанк» и ПАО «ВТБ» – с помощью RPA в базу данных PostgreSQL
2. Анализ тональности новостных текстов с использованием предобученной нейросетевой модели RuBERT
3. Формирование прогноза стоимости закрытия акции на горизонте 20 рабочих дней вперед
4. Генерация торговых сигналов «BUY / HOLD / SELL»
5. Отображение результатов анализа в виде дашборда с прогнозами, метриками и показателями портфеля в системе PIX BI

Нефункциональные требования

1. Время формирования прогноза по одной компании – не более 60 секунд; сбор данных – не более 240 секунд
2. R^2 (коэффициент детерминации) должен быть не ниже 0,96 на валидационной выборке
3. Система должна позволять расширение на новые компании без доработки архитектуры
4. Доступ к данным и прогнозам осуществляется через защищенное соединение (авторизация по логину и паролю)

Календарно-ресурсный план проекта

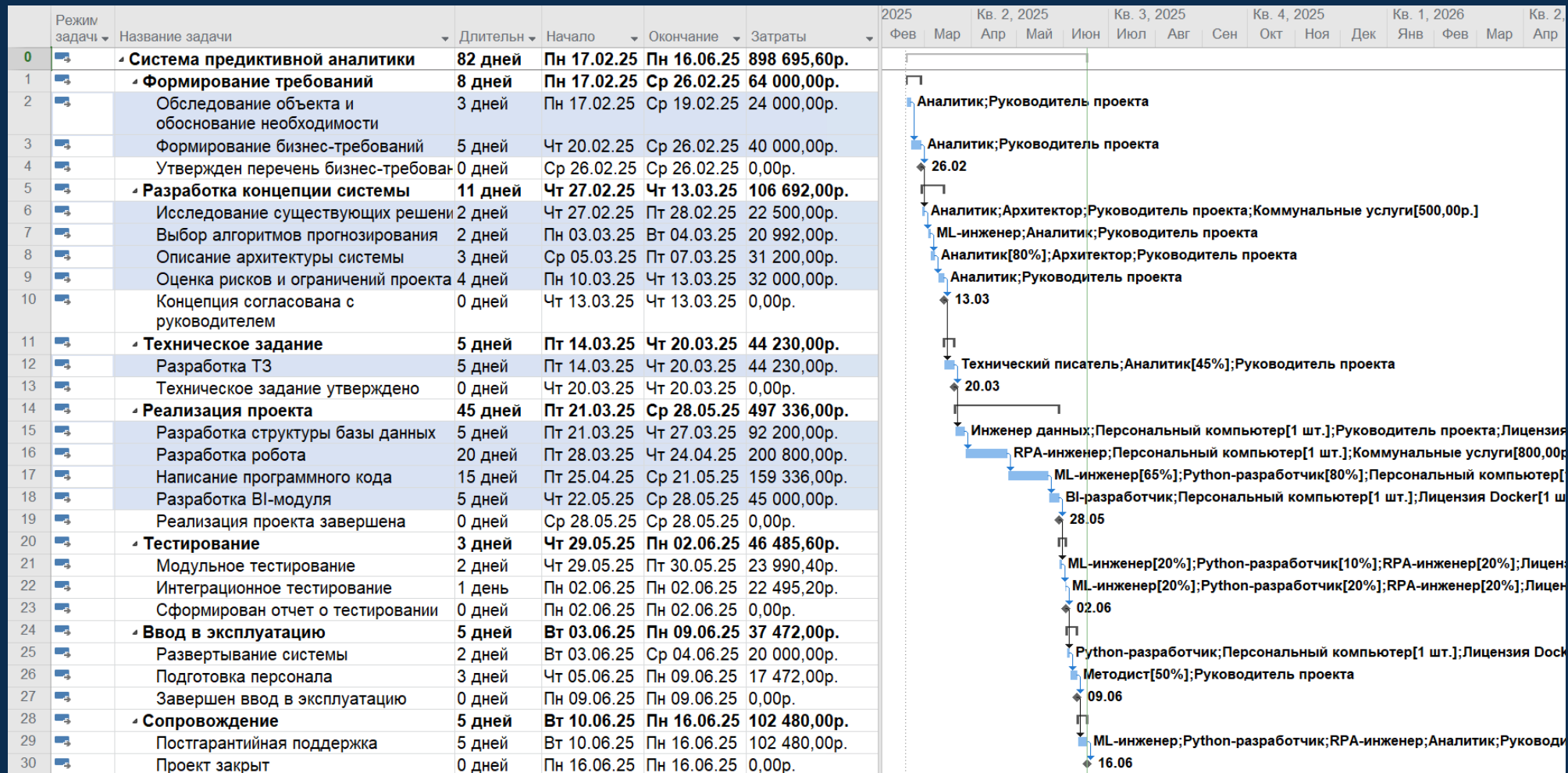


Рис.4 – Календарно-ресурсный план

Проект занимает 82 дня

Роль в проекте – аналитик и RPA-инженер

Архитектура решения



РЭУ.РФ
РОССИЙСКИЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ
ИМЕНИ Г.В. ПЛЕХАНОВА



Биржевые данные собираются для:
ПАО «Сбер»
ПАО «ВТБ»

Новости собираются для:
ПАО «Сбер»
ПАО «ВТБ»
Биржи

Все компоненты связаны и работают автоматически

Работа с данными

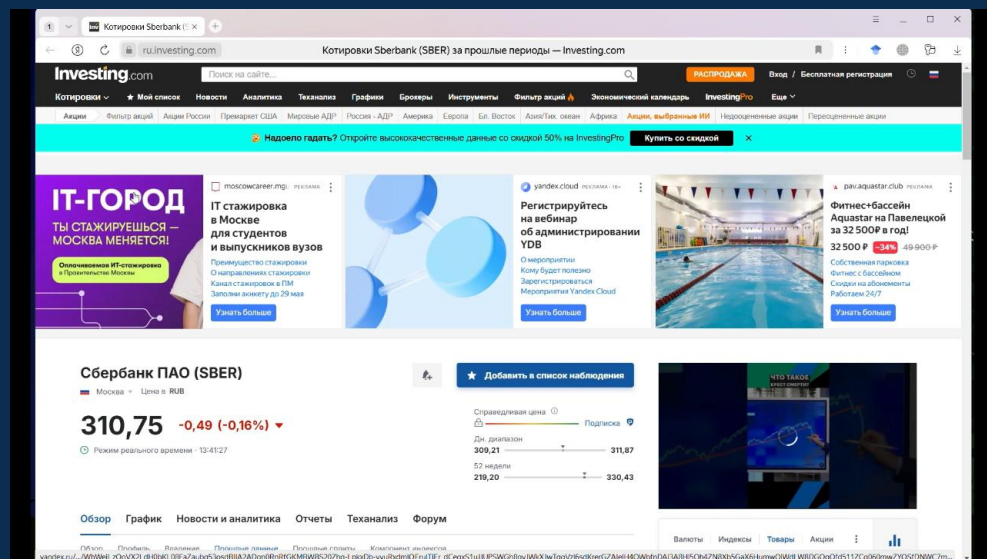
PIX RPA

Робот проходит по следующим сайтам:



LENTA.RU

1. Дата
2. Цены открытия/закрытия
3. Минимальная/максимальная цены
4. Объем торгов
5. Среднее скользящее за 5, 10, 20 дней
6. Текст новости



Видео работы RPA (часть сбора данных Сбера, ускоренно в 4 раза)

PostgreSQL

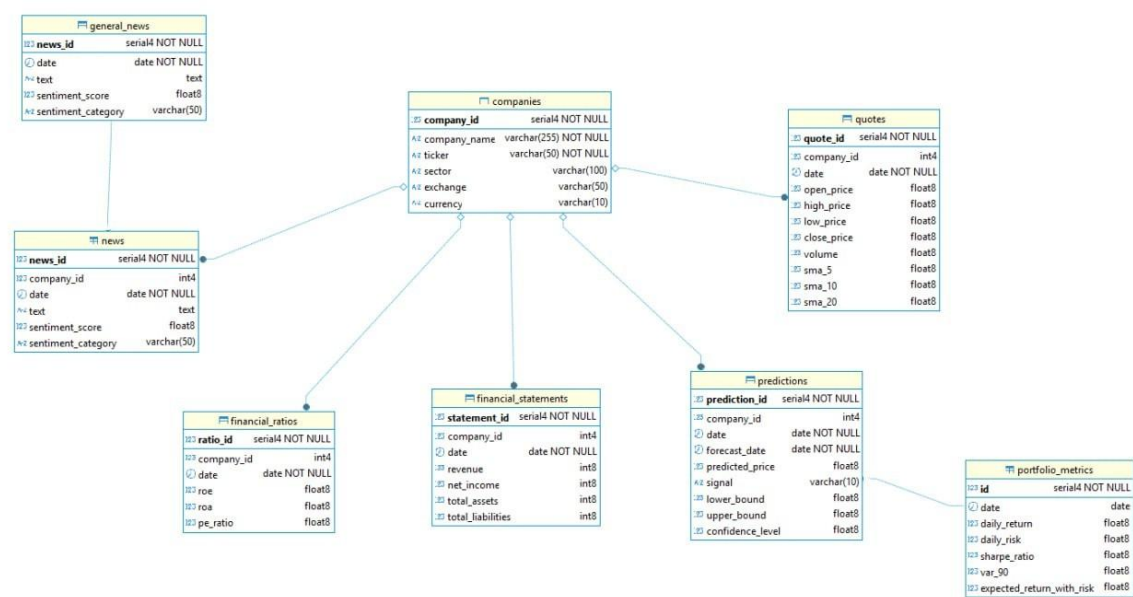


Рис.5 – ERD БД

8 таблиц для хранения информации о компаниях, котировках их акций, новостях и сформированных прогнозах

Созданы «триггеры» для автоматизации запуска модуля аналитики



Модель машинного обучения: тональность новостей

Задача МО: построить модель регрессии для предсказания цены закрытия акции на горизонте 20 рабочих дней на основе признаков: исторических котировок, скользящих средних, финансовых показателей компании и тональности новостей.

Классификация тональности текста новостей реализуется с помощью модели RuBERT. Дополнительно считаются показатели сформированного инвестиционного портфеля.

Таблицы «news» и «general_news»
(новости по компаниям и новости по бирже)

news_id [PK] integer	company_id integer	date date	text text
806	1	2025-05-20	Сбер с начала 2025 года помог своим клиентам
911	1	2025-05-29	На XIII Международной встрече высоких предс

Рис.6 – Пример данных в таблице «news»

Подключение
предобученной модели
RuBERT

Тональность текста
эмоциональное отношение автора
высказывания к некоторому
объекту, выраженное в тексте

Ансамбль моделей МО

sentiment_score double precision
0.05055155311352932
-0.22296574920238493

Рис.7 – Пример
сформированной оценки

Интервал оценки: [-1;1]

Оценка < -0,33: отрицательная
Оценка > 0,33: положительная
Между: нейтральная



Модель машинного обучения: ансамбль моделей МО

На вход:

720 строк

Целевой признак
– «close_price»

Ансамбль
моделей МО

	company_id integer	date date	open_price double precision	close_price double precision	volume double precision	sma_10 double precision	net_income bigint	pe_ratio double precision	company_sentiment double precision	market_sentiment double precision
1	1	2025-03-03	309.8	305.5	69.49	311.28	1560000000000	3.99	0.2690725735448397	0.06295125336580067
2	2	2025-03-03	9	90.77	111.8	91.96	6099700000000	0.8	-0.0022893364173228073	0.06295125336580067
3	1	2025-03-04	306.9	316.4	88.13	311.41	15600000000000	3.99	0.07181773682694055	-0.16889173051611583

Рис.8 – Пример входных данных

1. Линейная регрессия
2. Деревья решений
3. Случайные леса
4. Метод k-ближайших соседей
5. Алгоритм градиентного бустинга
6. Алгоритм ансамблевого обучения «XGBoost»

метод объединения

Стекинг:
градиентный бустинг

Метрики качества:

Коэффициент детерминации: 0,997

Среднеквадратичная ошибка (MSE): 35,6

Средняя абсолютная ошибка (MAE): 4,05

Средняя абсолютная ошибка в процентах (MAPE): 2,06%

На выходе:

40 строк
(20 для Сбера, 20 для ВТБ)

для таблицы «predictions»

prediction_id [PK] integer	company_id integer	date date	forecast_date date	predicted_price double precision	signal character varying (10)
5	1	2025-05-05	2025-05-13	303.9402511171787	HOLD
25	2	2025-05-05	2025-05-13	98.1957000210819	HOLD
6	1	2025-05-05	2025-05-14	303.94228634759935	HOLD
26	2	2025-05-05	2025-05-14	98.21678331205217	HOLD
27	2	2025-05-05	2025-05-15	98.16797889935762	HOLD

Сигнал формируется
с помощью метода
Монте-Карло

Рис.9 – Пример сформированных данных в таблице «predictions»

1 строка

для таблицы «portfolio_metrics»

id [PK] integer	date date	daily_return double precision	daily_risk double precision	sharpe_ratio double precision	var_90 double precision	expected_return_with_risk double precision
1	2025-05-12	0.0009	0.0214	0.03	2.74	-2.05

Рис.10 – Пример сформированных данных в таблице «portfolio_metrics»

BI-визуализация

PIX BI



РЭУ.РФ
РОССИЙСКИЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ
ИМЕНИ Г.В. ПЛЕХАНОВА

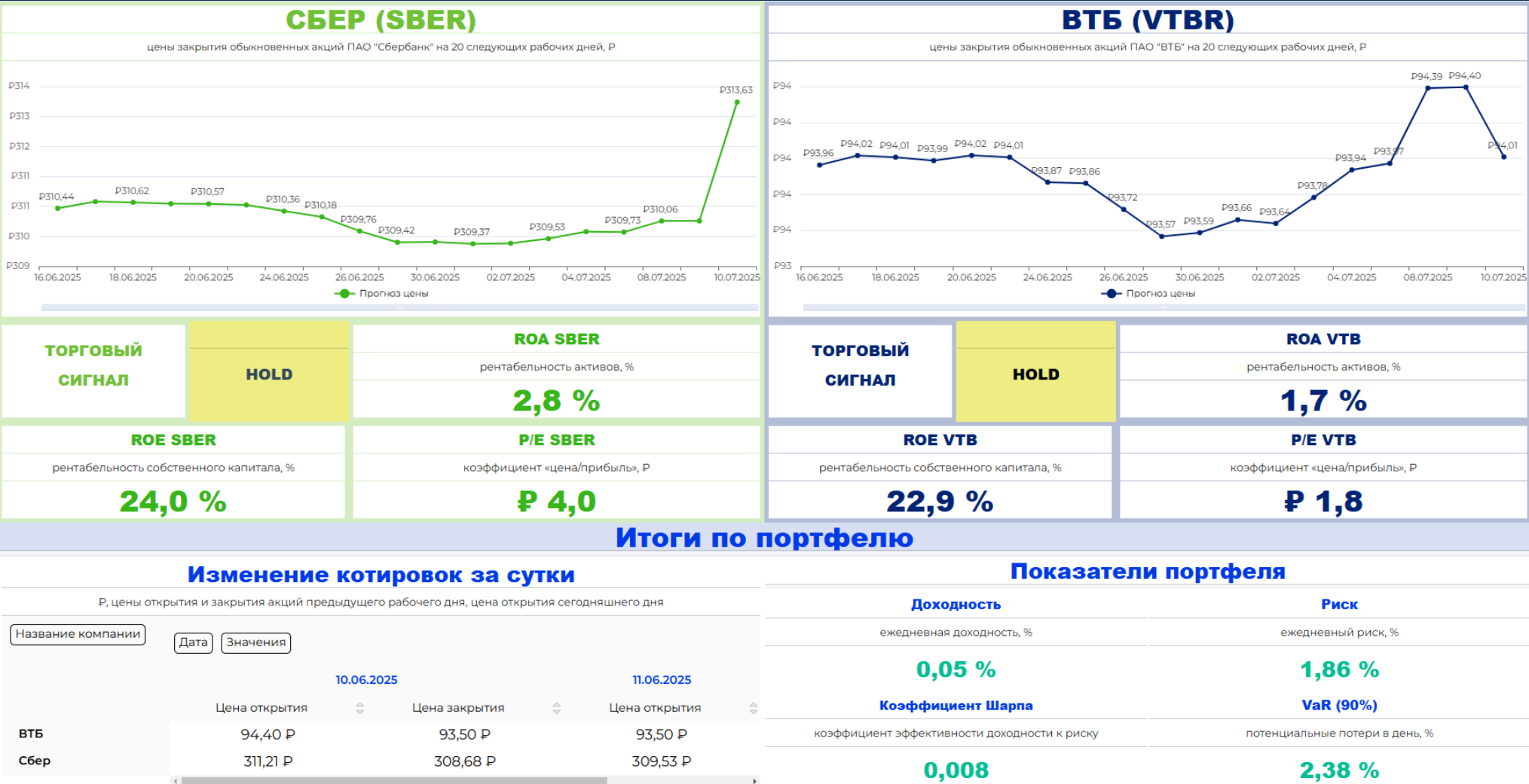
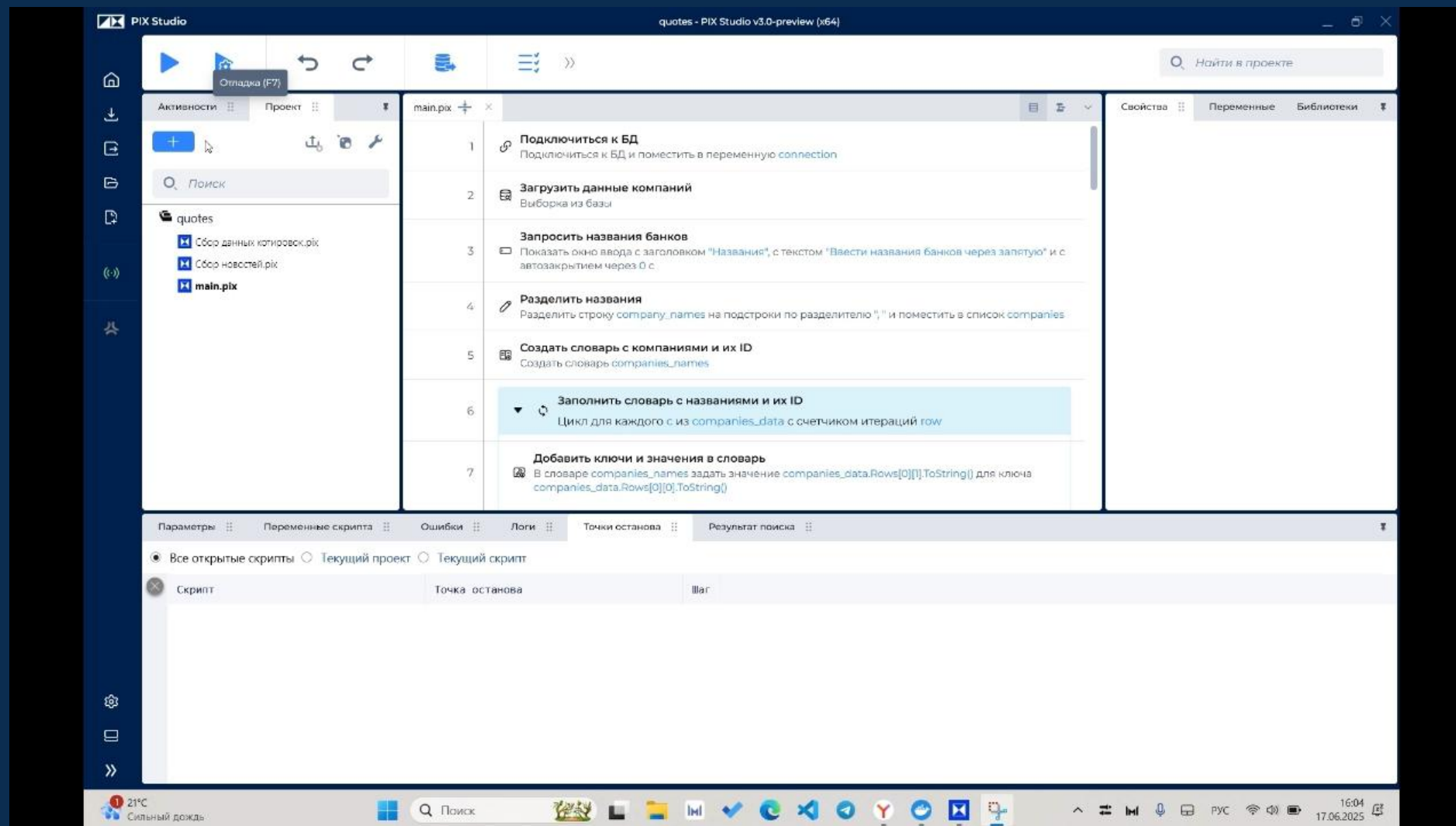


Рис.11 – Итоговый дашборд в PIX BI

Видео работы системы



Видео работы системы, фрагменты ускорены, изначальная длина видео – 9:01

Экономическая эффективность



Статья затрат	Сумма, руб.
Капитальные затраты	
Фонд оплаты труда	707 384,60
Настройка инфраструктуры (сервер/ПО/доступ)	115 374
Внедрение, тестирование, отладка системы	46 437
Итого:	869 195,60
Операционные затраты	
Поддержка, мониторинг, отчетность	22 000
Технические корректировки	7 500
Итого в год:	29 500
Всего:	898 695,60

До внедрения		После внедрения
Затраченное время	40 ч/мес (480 ч/год)	2 ч/мес (24 ч/год)
Оплата труда (только по данной задаче аналитика)	32 000 руб./мес (384 000 руб./год)	1 600 руб./мес (19 200 руб./год)

Прямая экономия

Снижение трудозатрат аналитика

экономия 95%

Снижение стоимостных затрат

364 800 Р ежегодно

Ставка дисконтирования при всех расчетах – 25%

Срок окупаемости

$$DPP = \min \left\{ t: \sum_{k=1}^t \frac{364\,800}{(1 + 0,25)^k} \geq 898\,695,60 \right\}$$

4 года и 4 месяца

NPV

$$NPV = \sum_{t=1}^5 \frac{364\,800 - 29\,500}{(1 + 0,25)^t} - 898\,695,60$$

3 019,9Р

ROI

$$ROI = \frac{3\,019,984}{898\,695,60} * 100\% \approx 0,3$$

0,3%

Результаты и перспективы



РЭУ.РФ
РОССИЙСКИЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ
ИМЕНИ Г.В. ПЛЕХАНОВА

- | | | | |
|----------|---|----------|--|
| 1 | Все поставленные задачи выполнены | 4 | Интеграция с продуктами компании PIX Robotics |
| 2 | Разработана и протестирована рабочая система прогнозирования котировок | 5 | Потенциал масштабирования на другие компании без изменения архитектуры |
| 3 | Точность модели соответствует требованиям | 6 | Потенциал масштабирования на другие источники данных |



тема:

**Разработка и внедрение системы предиктивной аналитики в
условиях импортозамещения**

Спасибо за внимание!

Выполнила обучающаяся Крабу Кира Сергеевна

Группа 15.11Д-БИЦТ09/216

Научный руководитель выпускной
квалификационной работы
Батищев А.В., к.э.н, доцент,
доцент кафедры информатики