

Содержание

Содержание.....	1
Глава 2 Обзор источников.....	2
2.1 Введение в анализ валютных пар и финансовых временных рядов.....	2
2.2 Источники данных и способы сбора финансовых данных.....	4
2.2.1 Основные категории источников данных.....	4
2.2.2 Учет санкций и практические рекомендации для РФ.....	6
2.3 Подходы к построению конвейеров данных (Data Pipeline, ETL).....	7
2.3.1 Классические ETL-подходы.....	7
2.3.2 Потоковая обработка данных (streaming pipelines).....	9
2.3.3 Архитектуры конвейеров данных в финансовых системах.....	10
2.4 Методы анализа и моделирования валютных курсов.....	11
2.4.1 Статистические методы анализа временных рядов.....	11
2.4.2 Базовые модели машинного обучения.....	12
2.4.3 Современные и SOTA-модели для финансовых временных рядов...	13
2.5 Метрики качества и способы оценки моделей.....	15
2.6 Аналоги и существующие решения для мониторинга валютных пар.....	19
2.7 Выводы по обзору источников.....	22
Источники.....	24

Глава 2 Обзор источников

2.1 Введение в анализ валютных пар и финансовых временных рядов

Валютный рынок играет ключевую роль в глобальной финансовой системе, обеспечивая обмен валютами, хеджирование рисков и поддержку международной торговли. Хотя в России основной организованный сегмент — валютный рынок Московской биржи (МОЕХ FX) — предоставляет данные по парам вроде USD/RUB, EUR/RUB и CNY/RUB, значительный интерес для анализа представляют зарубежные рынки, включая глобальный внебиржевой рынок Forex и биржи вроде СМЕ или ЕUREX. Это особенно актуально для мониторинга кросс-курсов (например, EUR/USD, GBP/JPY) и экзотических пар, где объёмы торгов достигают триллионов долларов ежедневно по данным Банка международных расчётов (BIS).

Однако в контексте российской реальности сбор данных по зарубежным финансовым инструментам, включая валютные пары и акции (stocks), сталкивается с вызовами из-за санкций, введённых после 2022 года. Эти ограничения затрагивают доступ к некоторым API и платформам (например, прямой доступ к NYSE или NASDAQ для акций может требовать VPN или альтернативных провайдеров), а также усложняют интеграцию с международными брокерами. Несмотря на это, использование зарубежных данных остаётся критически важным: оно позволяет анализировать глобальные тренды, корреляции с российским рублём и влияние макроэкономических событий (например, решений ФРС США или ЕЦБ). В частности, данные по зарубежным акциям (например, AAPL, GOOGL с NASDAQ) могут служить дополнением к анализу валютных пар, поскольку волатильность акций часто коррелирует с валютными курсами (эффект "риск-он/риск-офф").

Финансовые временные ряды валютных курсов и акций обладают общими стилизованными фактами (stylized facts), которые усложняют моделирование:

1. Нестационарность — уровни цен (например, курс EUR/USD или цена акции TSLA) обычно имеют тренды, в то время как логарифмические доходности стационарны.
2. Слабая автокорреляция доходностей — изменения близки к белому шуму, согласуясь с гипотезой эффективного рынка.
3. Кластеризация волатильности — периоды высокой волатильности группируются (например, во время кризисов или earning reports для акций).
4. Тяжёлые хвосты и лептокуртичность — экстремальные события (крахи, как в марте 2020) происходят чаще, чем при нормальном распределении.
5. Асимметрия волатильности — отрицательные шоки усиливают волатильность сильнее положительных.
6. Долгая память и мульти尺度ность — свойства зависят от временного масштаба.

Для зарубежных акций сбор данных проще через открытые API (Yahoo Finance, Alpha Vantage, Polygon.io), в отличие от российских акций (например, GAZP или SBER на МОЕХ). Тем не менее, интеграция таких данных в конвейер требует учёта юридических аспектов (compliance с OFAC/EU санкциями) и технических решений (proxy, зеркала данных). Важность использования зарубежных данных подчеркивается их объёмом и качеством: например, исторические данные по S&P 500 акциям позволяют тестировать модели на больших датасетах, что повышает робастность анализа валютных пар.

Анализ сочетает фундаментальный (экономические индикаторы, корпоративные отчёты), технический (индикаторы вроде MA, RSI) и количественный подходы (статистика, ML). В работе акцент на разработке data

`pipeline` для мониторинга, с учетом санкций: приоритет зарубежным источникам для валютных пар и акций.

2.2 Источники данных и способы сбора финансовых данных

Сбор качественных и своевременных финансовых данных является фундаментальной частью любого конвейера данных для мониторинга валютных пар и связанных активов (включая зарубежные акции). В условиях 2025–2026 годов доступ к данным определяется несколькими факторами: типом данных (реального времени, исторические, EOD — end-of-day), задержкой (*latency*), объёмом, стоимостью и ограничениями, связанными с санкциями для пользователей из РФ.

Для анализа валютных пар (FX) и зарубежных акций предпочтительны источники, предоставляющие:

- Реал-тайм данные (tick-by-tick или с частотой 1–60 секунд) — для мониторинга и высокоскоростных стратегий.
- Исторические данные (от минутных до ежедневных баров за 10–20+ лет) — для обучения моделей и бэктестинга.
- Доступные API (REST, WebSocket) с JSON/CSV-ответами.

2.2.1 Основные категории источников данных

1. Бесплатные API (с ограничениями по запросам/объёму):

Alpha Vantage — один из наиболее популярных бесплатных сервисов. Предоставляет реал-тайм и исторические данные по валютным парам (FX rates, intraday, daily), акциям (stocks), криптовалютам и техническим индикаторам. Подходит для прототипирования и академических работ.

Finnhub — бесплатный API с реал-тайм котировками по forex, stocks, crypto. Хороший охват валютных пар (включая majors и crosses), WebSocket для

streaming. Бесплатный tier позволяет до нескольких тысяч запросов в месяц, но с лимитом на одновременные соединения.

Financial Modeling Prep (FMP) — бесплатный доступ к историческим и EOD данным по акциям и forex. Подходит для фундаментальных данных (балансы, отчёты компаний), но реал-тайм ограничен.

Yahoo Finance (через неофициальные библиотеки вроде `yfinance` в Python) — исторические данные по акциям и валютам (через тикеры вроде EURUSD=X). Официальный API прекратил работу в 2017 году, но `yfinance` остаётся популярным (хотя нестабильным из-за изменений в scraping). Альтернативы: EODHD, Polygon.io.

2. Платные/профессиональные API с высоким качеством (рекомендуются для production):

Polygon.io — один из лидеров 2025 года. Полные данные по forex (более 1750 пар, quotes, aggregates), stocks (US markets), WebSocket для реал-тайм. Высокая точность, низкая задержка, исторические данные с 2000-х.

EODHD (End-of-Day Historical Data) — часто признаётся лучшим для forex в 2025 году. Более 1100 валютных пар, реал-тайм через WebSocket, исторические данные (tick, 1-min, daily). Хороший баланс цены/качества, есть бесплатный тестовый доступ.

Twelve Data — реал-тайм и исторические данные по stocks, forex, crypto. Поддержка WebSocket, технические индикаторы. Freemium-модель.

TraderMade, FCS API, ExchangeRatesAPI — специализированы на forex, обновление каждые 60 секунд, исторические данные. Подходят для простых задач мониторинга.

3. Альтернативные способы сбора (если API ограничены санкциями или лимитами):

Web scraping сайтов брокеров/агрегаторов (Investing.com, TradingView, Forex Factory) — с использованием библиотек

BeautifulSoup/Selenium/Playwright. Риски: блокировка IP, нестабильность формата данных, юридические аспекты (ToS запрещают scraping в большинстве случаев).

Открытые датасеты — Kaggle, Quandl (теперь Nasdaq Data Link), HistData.com (бесплатные исторические forex данные в формате CSV/MT4).

Библиотеки-обёртки — pandas_datareader, ccxt (для crypto, но частично forex), openbb (открытый стек с множеством источников).

2.2.2 Учет санкций и практические рекомендации для РФ

В 2025–2026 годах санкции (OFAC, EU) ограничивают прямой доступ к некоторым US-based провайдерам (например, Bloomberg, Refinitiv) для резидентов РФ, но большинство retail-API (Alpha Vantage, Polygon, Finnhub, EODHD) остаются доступны через VPN/proxy или без ограничений (поскольку они не подпадают под прямые запреты). Рекомендуется:

- Использовать несколько источников (multi-source redundancy) для fallback.
- Хранить исторические данные локально после первого скачивания.
- Для реал-тайм — предпочтать WebSocket над polling, чтобы минимизировать запросы.
- Тестировать на compliance: избегать данных, напрямую связанных с санкционными сущностями.

2.3 Подходы к построению конвейеров данных (Data Pipeline, ETL)

Конвейер данных (data pipeline) представляет собой последовательность процессов, обеспечивающих автоматизированный сбор, обработку, хранение и доставку данных от источников до потребителей (моделей анализа, дашбордов, систем мониторинга). В контексте мониторинга валютных пар и зарубежных

акций конвейер должен учитывать специфику финансовых временных рядов: высокую частоту обновлений (от тиковых до минутных данных), объём (миллионы записей в день), требования к низкой задержке (latency) для реал-тайм анализа и надёжности (отказоустойчивость, поскольку пропуск данных может привести к неверным торговым решениям или сигналам).

Различают два основных класса конвейеров: классические пакетные (batch) ETL и потоковые (streaming). В финансовых системах часто применяют гибридные архитектуры (lambda или карпа), сочетающие оба подхода для баланса между скоростью и надёжностью.

2.3.1 Классические ETL-подходы

Классический ETL (Extract — Transform — Load) — это пакетная обработка данных, где данные извлекаются из источников периодически (ежедневно, ежечасно или по расписанию), преобразуются (очистка, агрегация, feature engineering) и загружаются в целевое хранилище (data warehouse, lake или базу данных).

Основные характеристики и инструменты (2025–2026 гг.)

Преимущества: Простота реализации, высокая надёжность (легко отлаживать, повторять), хорошая обработка больших исторических объёмов, низкие затраты на инфраструктуру для нереал-тайм задач.

Недостатки: Задержка (latency) от минут до часов, не подходит для мониторинга в реальном времени (например, детекции аномалий в валютных курсах во время новостей).

Типичные этапы в финансовом контексте:

Извлечение: Поллинг API (Alpha Vantage, Polygon.io, EODHD) по cron-задачам.

Трансформирование: Обработка OHLCV (Open-High-Low-Close-Volume), расчёт лог-доходностей, удаление выбросов, ресэмплинг, feature extraction (технические индикаторы: MA, RSI, Bollinger Bands).

Отправка: Загрузка в PostgreSQL/TimescaleDB (для time-series), Snowflake, BigQuery или S3/Parquet для lake.

Популярные инструменты и фреймворки (2025):

- Airbyte / Fivetran — open-source / managed ELT (Extract-Load-Transform, где трансформация происходит после загрузки).
- Apache Airflow — оркестратор для DAG (Directed Acyclic Graphs), широко используется в финансах для сложных зависимостей.
- dbt (data build tool) — для трансформаций в SQL после загрузки.
- AWS Glue, Azure Data Factory, Google Cloud Dataflow — облачные managed ETL.

В финансовых применениях классический ETL часто используется для nightly/historical обновлений (бэктестинг моделей, расчёт долгосрочных метрик) и compliance-отчётов.

2.3.2 Потоковая обработка данных (streaming pipelines)

Потоковая (streaming) обработка позволяет обрабатывать данные в реальном времени по мере их поступления, без ожидания полного батча. Это критично для мониторинга валютных пар, где задержка в секунды может быть фатальной (например, во время всплеска волатильности на новостях ФРС или геополитике).

Ключевые характеристики:

- Low-latency (от миллисекунд до секунд).

- Exactly-once / at-least-once семантика для гарантии доставки.
- Stateful processing (агрегация по окнам времени, join потоков).
- Масштабируемость горизонтальная (на кластере).

Основные технологии (2025–2026)

- Apache Kafka — базовый транспорт (message broker) для финансовых данных; Confluent Platform добавляет enterprise-функции (Schema Registry, ksqlDB).
- Apache Flink и Spark Streaming — движки обработки (Flink лидирует в low-latency и stateful streaming).
- Kafka Streams и ksqlDB — лёгкие для простых задач.
- Облачные альтернативы: AWS Kinesis + Lambda/Flink, Google Pub/Sub + Dataflow, Azure Event Hubs + Stream Analytics.
- Для time-series: InfluxDB, TimescaleDB с continuous aggregates; ClickHouse для аналитики в реальном времени.

Лучшей практикой по моему мнению было бы использовать именно streaming для high-frequency данных (tick/minute), но комбинировать с batch для тяжёлых трансформаций (например, retraining моделей раз в день).

2.3.3 Архитектуры конвейеров данных в финансовых системах

В количественных финансах и high-frequency trading (HFT) архитектуры эволюционировали от монолитных к микросервисным и event-driven.

Основные паттерны (2024–2025):

- Lambda Architecture (классическая гибридная):
 - Batch layer (Hadoop/Spark для исторических данных).
 - Speed layer (Kafka + Flink/Spark Streaming для реал-тайм).
 - Serving layer (Cassandra/Redis для запросов).
- Kappa Architecture (современный тренд):
 - Всё как stream: исторические данные replay'ятся из Kafka.

- Преимущества: единая логика, проще масштабировать.
 - Применяется в hedge funds и prop-trading для factor modeling, HFT.
- Event-Driven и Microservices:
 - Kafka как backbone.
 - Микросервисы: ingestion service → processing → ML inference → alerting/visualization.
 - Примеры: AWS serverless (Kinesis + Lambda + SageMaker) для GenAI в factor pipelines.
- Специфические для финансов:
 - High-Frequency Trading (HFT): FPGA/ASIC для ультранизкой latency, co-location у бирж, custom C++/Rust движки.
 - Quantitative Finance или Algo Trading: Kafka + Flink для streaming + TimescaleDB для хранения + MLflow/Vertex AI для моделей.
 - Мониторинг валютных пар: WebSocket ingestion → Kafka topic → Flink job (feature extraction, anomaly detection) → PostgreSQL и Grafana для дашбордов.

В рамках ВКР мной рекомендуется Карра-подобная архитектура на базе Kafka + Flink/Python (для простоты) или Airflow + Kafka для гибкости. Это позволит собирать данные из API в реальном времени, обрабатывать потоково и хранить для последующего анализа/моделирования.

2.4 Методы анализа и моделирования валютных курсов

Моделирование валютных курсов (FX rates) — одна из наиболее сложных задач в финансовом анализе из-за нестационарности, высокой шумности, тяжёлых хвостов распределения доходностей и влияния внешних факторов (макроэкономика, geopolitika, monetarная политика). Методы эволюционировали от классической статистики к современным глубоким нейронным сетям и foundation-моделям.

В контексте разработки конвейера данных для мониторинга валютных пар (включая majors вроде EUR/USD, USD/JPY и кросс-курсы) важно выбрать подходы, которые хорошо работают с временными рядами высокой частоты (minute/daily) и позволяют интегрировать в streaming-пайплайн для реал-тайм предсказаний или детекции аномалий.

2.4.1 Статистические методы анализа временных рядов

Статистические методы остаются базовыми и часто используются как benchmark или в hybrid-моделях благодаря интерпретируемости и низким требованиям к вычислениям.

Ключевые подходы:

1. ARIMA/SARIMA (AutoRegressive Integrated Moving Average) — классика для стационарных или дифференцированных рядов. Хорошо работает на лог-доходностях валютных пар при отсутствии сильных структурных сдвигов. Расширения: ARIMAX с экзогенными переменными (процентные ставки, нефть).
2. GARCH семейство (Generalized Autoregressive Conditional Heteroskedasticity): ARCH (Engle, 1982), GARCH (Bollerslev, 1986), EGARCH (асимметрия), TGARCH, IGARCH. Критически важны для моделирования кластеризации волатильности — ключевого stylized fact FX-рядов. Применяются для прогнозирования волатильности (VaR, опционы) и как компонента в hybrid-моделях.
3. Векторные авторегрессии (VAR/VEC) — для многомерного анализа (например, USD/RUB и нефть Brent).
4. Exponential Smoothing (Holt-Winters, ETS) — простые, но эффективные для коротких горизонтов.
5. Cointegration и Error Correction Models (ECM) — для пар с долгосрочным равновесием (например, EUR/USD и GBP/USD).

Преимущества: Интерпретируемость, статистическая значимость, низкий риск переобучения.

Недостатки: Предполагают линейность и стационарность, плохо справляются с нелинейностями и regime shifts (например, после 2022 года в рублёвых парах).

В 2025 году эти методы часто комбинируют с ML (например, AR-GARCH + LSTM) для улучшения точности на 10–20 % по сравнению с чистой статистикой.

2.4.2 Базовые модели машинного обучения.

Базовые ML-модели превосходят чистую статистику за счёт нелинейности и способности захватывать взаимодействия фич.

Основные модели:

Линейные регрессоры (Linear Regression, Ridge/Lasso) — baseline с lagged фичами и техническими индикаторами.

Деревья решений и ансамбли:

- Random Forest, Extra Trees — robust к шуму, feature importance.
- Gradient Boosting: XGBoost, LightGBM, CatBoost — state-of-the-art среди tree-based для табличных данных. LightGBM особенно эффективен для больших датасетов FX (минутные бары).
- Support Vector Regression (SVR) — с RBF-ядром для нелинейных зависимостей.
- K-Nearest Neighbors (KNN) — для pattern matching (редко, но полезно в комбинации).

Feature engineering (ключевой этап): lagged returns, rolling statistics (mean, std, skew), технические индикаторы (MA, EMA, RSI, MACD, ATR), sentiment proxies (если доступны), внешние фичи (VIX, нефть, ставки).

Преимущества: Хорошая обобщаемость на out-of-sample, интерпретируемость (SHAP values).

Применение в FX: XGBoost/LightGBM часто дают лучшие результаты на daily/hourly данных по сравнению с ARIMA/GARCH.

2.4.3 Современные и SOTA-модели для финансовых временных рядов

С 2020–2025 годов доминируют глубокие модели, особенно Transformer-based и hybrid-архитектуры, которые лучше захватывают долгосрочные зависимости и мульти尺度ность.

В современных задачах прогнозирования финансовых временных рядов наблюдается переход от классических рекуррентных моделей к более сложным attention-based и гибридным архитектурам. Модели на основе LSTM и GRU по-прежнему используются в качестве базовых решений благодаря их способности учитывать временные зависимости, однако их эффективность снижается при работе с длинными временными горизонтами из-за проблемы затухания градиентов.

Существенный прогресс был достигнут с адаптацией Transformer-архитектур для временных рядов. Модификации классического Transformer, ориентированные на работу с длинными и многомерными последовательностями, демонстрируют высокие результаты на современных бенчмарках. Дополнительное улучшение качества прогнозирования обеспечивают модели, использующие предварительную декомпозицию временных рядов на трендовые и сезонные компоненты, что позволяет частично учитывать нестационарность финансовых данных.

Особое распространение в последние годы получили гибридные модели, сочетающие рекуррентные сети и Transformer-механизмы внимания. Такой

подход позволяет эффективно захватывать как локальные, так и глобальные временные зависимости и оказался особенно востребованным в финансовых приложениях. Параллельно развивается направление foundation-моделей, предварительно обученных на больших объемах финансовых временных рядов, которые демонстрируют заметное превосходство над традиционными архитектурами за счет переноса знаний.

Несмотря на высокую точность современных state-of-the-art моделей, их применение связано с повышенными требованиями к данным и вычислительным ресурсам, а также с ограниченной интерпретируемостью. Согласно актуальным результатам, наилучшее качество прогнозирования валютных курсов обеспечивают гибридные Transformer-архитектуры и foundation-модели, снижая ошибку по сравнению с LSTM-подходами на 15–30 %.

В рамках данной выпускной квалификационной работы целесообразно использовать поэтапный подход: от простых базовых моделей к более сложным гибридным архитектурам, интегрированным в полный цикл машинного обучения с регулярным переобучением.

2.5 Метрики качества и способы оценки моделей

Оценка качества моделей прогнозирования валютных курсов и других финансовых временных рядов требует использования специализированных метрик, поскольку стандартные показатели регрессии не всегда адекватно отражают особенности финансовых данных, такие как нестационарность, асимметричная волатильность и наличие экстремальных значений. В рамках данной работы используются метрики, оценивающие как точность прогнозирования уровня временного ряда, так и корректность предсказания

направления его изменения, а также статистические методы сравнения альтернативных моделей.

Для оценки точности восстановления уровня прогнозируемого значения широко применяется средняя абсолютная ошибка, обладающая высокой интерпретируемостью и устойчивостью к выбросам. Данная метрика определяется следующим образом:

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

где фактическое значение временного ряда в момент времени t обозначает реальное наблюдение, а прогнозное значение временного ряда в момент времени t — соответствующее предсказанное значение. В контексте валютных курсов абсолютная ошибка имеет наглядную экономическую интерпретацию, выражаясь в пунктах (pips).

Наряду с абсолютными метриками используются квадратичные показатели ошибки, такие как среднеквадратичная ошибка и её корень:

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2, \quad \text{RMSE} = \sqrt{\text{MSE}}$$

Данные метрики более чувствительны к крупным отклонениям прогнозов, что делает их особенно полезными в задачах, где важно penalize большие ошибки. По этой причине RMSE часто используется в качестве основной метрики при сравнении моделей на современных бенчмарках.

Для сопоставления качества прогнозов на различных финансовых инструментах применяются относительные показатели ошибки, в частности средняя абсолютная процентная ошибка:

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

Несмотря на удобство интерпретации, данная метрика может быть нестабильной при значениях временного ряда, близких к нулю. В связи с этим в современных исследованиях предпочтение часто отдаётся симметричной модификации процентной ошибки, более устойчивой к малым значениям:

$$\text{sMAPE} = \frac{200\%}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|}$$

В задачах прогнозирования валютных курсов традиционно используется сравнение моделей с наивным предположением о случайному блуждании, при котором прогноз на следующий момент времени равен последнему наблюдению. Для этой цели применяется коэффициент Тейла:

$$U = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - y_{t-1})^2}}$$

Значение коэффициента менее единицы указывает на превосходство модели над наивным прогнозом, тогда как значение, превышающее единицу, свидетельствует о худшем качестве прогнозирования.

Помимо точности восстановления уровня временного ряда, в финансовых приложениях особое значение имеет корректность предсказания направления изменения цены. Для оценки данного аспекта используется метрика направленной точности, определяемая как доля случаев, в которых модель правильно предсказывает знак изменения временного ряда:

$$\text{DA} = \frac{1}{n} \sum_{t=1}^n \mathbb{I} (\text{sign}(y_t - y_{t-1}) = \text{sign}(\hat{y}_t - y_{t-1}))$$

где индикаторная функция используется для фиксации совпадения направлений. Значения данной метрики, превышающие 50 %, указывают на наличие прогностической способности, превосходящей случайное угадывание. В качестве обобщения может использоваться средняя направленная точность, усреднённая по нескольким горизонтам прогнозирования.

Для более сбалансированной оценки бинарных прогнозов направления применяется коэффициент корреляции Мэттьюса:

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

где true positive и true negative соответствуют корректно предсказанным случаям роста и падения, а false positive и false negative — ошибочным классификациям. Данная метрика учитывает все элементы матрицы ошибок и рекомендуется для использования при несбалансированных данных.

В задачах прогнозирования волатильности применяются специализированные показатели, характерные для моделей условной дисперсии, а также методы сравнения прогнозируемой и реализованной волатильности. Дополнительно могут использоваться метрики, основанные на результатах симулированных торговых стратегий, такие как коэффициент Шарпа, совокупная доходность и максимальная просадка.

Для подтверждения статистической значимости различий в качестве прогнозов альтернативных моделей используются специальные тесты

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{f}_d(0)}{T}}}$$

сравнения. Наиболее распространённым является тест Диболда—Мариано, предназначенный для проверки гипотезы о равной точности прогнозов:

где

$$d_t = L(e_{1t}) - L(e_{2t}),$$

среднее значение разности потерь обозначает усреднённую разницу значений функции потерь двух моделей, а оценка спектральной плотности на нулевой частоте используется для нормализации статистики теста. Для сравнения вложенных моделей применяется тест Кларка—Уэста, а для отбора множества статистически неразличимых лучших моделей используется процедура Model Confidence Set.

2.6 Аналоги и существующие решения для мониторинга валютных пар.

Существующие решения для мониторинга валютных пар (currency pairs) можно разделить на несколько категорий: коммерческие торговые платформы и брокерские приложения, специализированные API и data providers, открытые/самостоятельные data pipelines (в основном на GitHub), а также AI-ориентированные инструменты и дашборды. Эти аналоги помогают понять лучшие практики интеграции данных, обработки в реальном времени и визуализации, а также выявить пробелы, которые можно закрыть в рамках ВКР (например, кастомный end-to-end пайплайн с акцентом на ML-модели и обход ограничений санкций).

Большинство трейдеров и институционалов используют готовые платформы с встроенным мониторингом:

- MetaTrader 4/5 (MT4/MT5) — индустриальный стандарт для forex. Поддерживает реал-тайм котировки, кастомные индикаторы, Expert Advisors (автоматизация), графики и алERTы. Интеграция с большинством брокеров (OANDA, FOREX.com, tastyfx). Преимущества: огромная экосистема, MQL-язык для скриптов. Недостатки: устаревший UI, зависимость от брокера.
- TradingView — веб/мобильная платформа с мощными чартами, социальными функциями, Pine Script для индикаторов. Реал-тайм данные по 100+ валютным парам, интеграция с брокерами. Популярна для технического анализа и алERTов.
- OANDA, FOREX.com, tastyfx — брокерские платформы с собственными дашбордами: реал-тайм rates, экономический календарь, новости, мобильные apps. Поддерживают 80+ пар, low spreads.
- Instinct FX (Bank of America) — институциональное решение для e-trading с алгоритмами и liquidity access.

Для корпоративного/институционального уровня: Bloomberg Terminal, Refinitiv Eikon — полный мониторинг с новостями, аналитикой, но дорого и ограничено санкциями для РФ.

Много примеров end-to-end пайплайнов для образовательных/исследовательских целей:

- The-Forex-Data-Pipeline (Sabareh / jaiswalanshul) — полный пайpline: сбор данных по валютным курсам, обработка (Airflow DAGs), подготовка для ML. Включает ingestion, cleaning, feature engineering.
- Forex-Currency-Processing-Airflow-Hdfs-Hive-Spark (slatawa) — big data подход: Airflow для оркестрации, HDFS/Hive для хранения, Spark для трансформаций. Загружает rates из внешнего API.

- ForexDataPipeline (emrhnksck) — простой Airflow-проект: fetch → process → store forex rates.
- Forex_Data_Pipeline_Project (M-Abdullah-Jutt) — end-to-end с Power BI дашбордом: нормализация, визуализация цен.
- Другие: проекты с Binance/Bybit/Yahoo Finance интеграцией, но адаптируемые под FX (lokhiufung/trading-data).

Эти репозитории демонстрируют типичные архитектуры: Airflow + Python + PostgreSQL/TimescaleDB, иногда Spark/Kafka для масштаба.

AI-ориентированные и современные решения (2025–2026)

- Jenova.ai — AI Forex Market Agent: реал-тайм анализ с веб-поиском, rate differentials, COT data, technical signals, risk management.
- GoMoon.ai — AI economic calendar с анализом влияния событий на рынки.
- Другие AI-tools: для sentiment analysis, predictions, automated monitoring.

Таблица 1. Сравнение аналогов

Решение / Тип	Реал-тайм	ML/Прогнозы	Open-source	Ключевые особенности
MetaTrader 5	Да	Частично (EA)	Нет	Трейдинг + индикаторы
TradingView	Да	Нет	Нет	Чарты, алERTы, скрипты
Polygon.io / EODHD API	Да (WS)	Нет	Нет	Источник для пайплайна
The-Forex-Data-Pipeline (GitHub)	Нет (batch)	Да (подготовка)	Да	Airflow + ML-ready
Jenova.ai AI Agent	Да	Да (AI)	Нет	Реал-тайм intelligence

В отличие от аналогов, проект ВКР фокусируется на кастомном streaming пайплайне (Kafka/Flink-подобном) с интеграцией SOTA-моделей (Transformer hybrids), периодическим retraining и визуализацией (Grafana/Power BI), с учётом доступности данных в условиях ограничений.

2.7 Выводы по обзору источников.

Проведённый обзор научных публикаций, технической документации и практических решений, включая работы из arXiv и Google Scholar, открытые репозитории GitHub, а также отчёты регулирующих и инфраструктурных организаций и актуальную API-документацию за 2025–2026 годы, позволяет сформулировать ряд обобщённых выводов. Финансовые временные ряды валютных курсов характеризуются выраженной нестационарностью, кластеризацией волатильности и наличием тяжёлых хвостов распределений, что делает неэффективным использование единственного класса моделей. В связи с этим в работе обоснован комбинированный подход, предполагающий использование статистических моделей в качестве базового уровня, методов машинного обучения для учёта нелинейных зависимостей и современных нейросетевых архитектур для достижения максимальной точности прогнозирования.

Анализ источников данных показывает, что на практике наиболее надёжным является сбор информации через специализированные внешние API, обеспечивающие стабильность и полноту исторических и текущих котировок. При этом ориентация на глобальные валютные пары и зарубежные финансовые инструменты позволяет упростить доступ к данным и увеличить объём обучающих выборок. Для обработки таких данных целесообразно применение гибридной архитектуры конвейера, сочетающей пакетную обработку

исторических данных и потоковую обработку для мониторинга в реальном времени, при этом stream-ориентированный подход рассматривается как наиболее перспективный для финансовых приложений.

В ходе обзора также установлено, что корректная оценка качества прогнозных моделей требует использования набора взаимодополняющих метрик, отражающих как точность предсказания значений, так и их практическую применимость в торговых сценариях, с обязательной проверкой статистической значимости и out-of-sample тестированием. Анализ существующих программных решений и открытых реализаций показывает, что они, как правило, охватывают лишь отдельные этапы обработки данных и не предоставляют полностью настраиваемого сквозного решения с интеграцией современных моделей и учётом ограничений доступа к данным.

На основании проведённого обзора в данной выпускной квалификационной работе обоснован выбор технологического стека и методов для практической реализации, включающий использование Python-экосистемы, потоковой обработки данных, градиентного бустинга и Transformer-ориентированных моделей, а также инструментов визуализации для мониторинга результатов. Совокупность проанализированных источников подтверждает актуальность выбранной темы и наличие достаточной теоретической и практической базы для реализации предложенного решения с возможностью дальнейшего расширения функциональности.

Источники

Литература:

1. Cont, R. Empirical properties of asset returns: stylized facts and statistical issues // Quantitative Finance. – 2001. – Vol. 1. – P. 223–236.
2. Tsay, R. S. Analysis of Financial Time Series. – 3rd ed. – Wiley, 2010. – 720 p.
3. Engle, R. F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation // Econometrica. – 1982. – Vol. 50. – No. 4. – P. 987–1007.
4. Bollerslev, T. Generalized Autoregressive Conditional Heteroskedasticity // Journal of Econometrics. – 1986. – Vol. 31. – P. 307–327.
5. Diebold, F. X., Mariano, R. S. Comparing Predictive Accuracy // Journal of Business & Economic Statistics. – 1995. – Vol. 13. – No. 3. – P. 253–263.
6. Theil, H. Applied Economic Forecasting. – Amsterdam: North-Holland Publishing Company, 1966.
7. Hyndman, R. J., Koehler, A. B. Another look at measures of forecast accuracy // International Journal of Forecasting. – 2006. – Vol. 22. – P. 679–688.
8. Hansen, P. R., Lunde, A., Nason, J. M. The Model Confidence Set // Econometrica. – 2011. – Vol. 79. – No. 2. – P. 453–497.
9. Clark, T. E., West, K. D. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models // Journal of Econometrics. – 2007. – Vol. 138. – P. 291–311.

Электронные источники и документация:

10. Bank for International Settlements (BIS). Triennial Central Bank Survey of foreign exchange and OTC derivatives markets in 2022 (и обновления 2025).
– Режим доступа: <https://www.bis.org/statistics/rpfx22.htm>
11. Московская биржа (МОЕХ). Данные по объёмам торгов на валютном рынке. – Режим доступа: <https://www.moex.com/s868> (данные за 2024–2025 гг.)
12. Alpha Vantage API Documentation. – Режим доступа:
<https://www.alphavantage.co/documentation/> (2025–2026)
13. Finnhub API Documentation. – Режим доступа: <https://finnhub.io/docs/api> (2025)
14. Polygon.io Forex & Stocks API Documentation. – Режим доступа:
<https://polygon.io/docs> (2025)
15. EODHD (End-of-Day Historical Data) API Documentation. – Режим доступа:
<https://eodhd.com/financial-apis> (2025)
16. Twelve Data API Documentation. – Режим доступа:
<https://twelvedata.com/docs> (2025)
17. Apache Airflow Documentation. – Режим доступа:
<https://airflow.apache.org/docs/>
18. Apache Kafka Documentation. – Режим доступа:
<https://kafka.apache.org/documentation/>
19. Apache Flink Documentation. – Режим доступа: <https://flink.apache.org/docs/>
20. MetaTrader 5 (MT5) Documentation. – Режим доступа:
<https://www.metatrader5.com/ru/terminal/help>
21. TradingView Pine Script Documentation. – Режим доступа:
<https://www.tradingview.com/pine-script-docs/>
22. GitHub: The-Forex-Data-Pipeline (автор Sabareh / jaiswalanshul). – Режим доступа: [https://github.com/...](https://github.com/)

23.GitHub: Forex-Currency-Processing-Airflow-Hdfs-Hive-Spark (автор slatawa). – Режим доступа: [https://github.com/...](https://github.com/)

24.GitHub: ForexDataPipeline (автор emrhnsck). – Режим доступа: [https://github.com/...](https://github.com/)

25.GitHub: Forex_Data_Pipeline_Project (автор M-Abdullah-Jutt). – Режим доступа: [https://github.com/...](https://github.com/)

Научные работы и модели (2023–2025):

26.Zhou, H. et al. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting // arXiv preprint arXiv:2310.06625, 2023.

27.Nie, Y. et al. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers // arXiv preprint arXiv:2211.14730 (PatchTST), 2022–2023.

28.Liu, M. et al. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting // NeurIPS 2021.

29.ФинCast: A Foundation Model for Financial Time Series Forecasting // arXiv, 2025.

30.LENS: Large Pre-trained Transformer for Exploring Financial Time Series Regularities // arXiv, 2025.

31.LSTM–Transformer hybrid model for financial time series forecasting // MDPI Journal (или аналог), 2025.

32.STL-Transformer: Seasonal-Trend Decomposition with Transformer // ACM proceedings, 2025.