

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра дифференциальных уравнений и математической экономики

ОПТИМИЗАЦИЯ ИНСТРУМЕНТОВ, ПРИМЕНЯЕМЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ЦЕНЫ АКЦИЙ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 4 курса 451 группы

направления 38.03.05 – Бизнес-информатика

механико-математического факультета

Овчинникова Никиты Дмитриевича

Научный руководитель
доцент, к. э. н.

С. В. Иванилова

Заведующий кафедрой
зав. кафедрой, д.ф.-м.н., профессор

С. И. Дудов

Саратов 2025

ВВЕДЕНИЕ

Прогнозирование цен акций является одной из ключевых задач в области количественных финансов и машинного обучения. С развитием современных методов искусственного интеллекта всё большее число исследований направлено на использование алгоритмов прогнозирования для анализа временных рядов. В данной дипломной работе проведено исследование эффективности различных моделей прогнозирования, а также предложены методы оптимизации и сравнения их результатов.

Актуальность темы исследования обусловлена высокой степенью неопределённости и волатильности финансовых рынков. Традиционные подходы к прогнозированию, такие как ARIMA и экспоненциальное сглаживание, имеют ограничения при работе с нестационарными данными и сложными паттернами. Современные модели машинного обучения и глубокого обучения позволяют учитывать нелинейные зависимости и работать с большими объемами данных, что делает их особенно перспективными для задач прогнозирования цен акций.

Целью дипломной работы является разработка и сравнение моделей прогнозирования цен акций на основе линейной регрессии, градиентного бустинга (XGBoost) и рекуррентной нейронной сети (LSTM), а также анализ точности этих моделей на примере исторических данных по акциям компании Apple (AAPL) за период с 2018 по 2023 год.

Для достижения поставленной цели были определены следующие задачи:

1. Изучить теоретические основы прогнозирования цен акций.
2. Провести обзор существующих методов и программных инструментов прогнозирования.
3. Подготовить данные: нормализация, создание технических индикаторов, формирование входных последовательностей.
4. Реализовать модели прогнозирования: линейная регрессия, XGBoost, LSTM.
5. Оптимизировать параметры моделей и оценить их качество.
6. Сравнить модели по метрикам MAE, RMSE и R^2 .
7. Сделать выводы о применимости и точности каждой из моделей.

Объектом исследования выступают финансовые временные ряды, в частности, ежедневные цены закрытия акций Apple (AAPL). Предметом исследования являются модели прогнозирования, основанные на машинном и глубоким обучении.

Практическая значимость работы заключается в возможности использования разработанных моделей для создания торговых сигналов, автоматизации анализа временных рядов, а также в качестве прототипа для прогнозирования других активов фондового рынка.

Методология исследования базируется на использовании современных методов машинного и глубокого обучения, таких как:

- Линейная регрессия — как отправная точка и простая модель.
- Градиентный бустинг (XGBoost) — как представитель ансамблевых методов.
- LSTM — как рекуррентная нейронная сеть, способная запоминать долгосрочные зависимости во временных рядах.

Работа выполнена с использованием языка Python и следующих библиотек:

- `yfinance` — загрузка исторических данных,
- `pandas`, `numpy` — обработка и предобработка данных,
- `scikit-learn` — реализация линейной регрессии и вычисление метрик,
- `xgboost` — применение градиентного бустинга,
- `keras`, `tensorflow` — реализация LSTM,
- `matplotlib`, `seaborn` — визуализация результатов.

Источник данных: исторические данные по акции Apple (AAPL) были загружены с помощью библиотеки `yfinance`. Были взяты ежедневные значения: Open, High, Low, Close, Volume.

Подготовка данных включала следующие этапы:

- Удаление пропущенных значений (`data.dropna()`).
- Добавление скользящих средних (SMA_10, EMA_20, MA_50).
- Разбиение выборки без перемешивания (`train_test_split(shuffle=False)`).
- Нормализация данных (`MinMaxScaler`) для модели LSTM.

- Создание лаговых признаков (`create_lagged_features`) для XGBoost и LSTM.

Целевой переменной выступила цена закрытия акций (`Close`), так как она считается наиболее стабильным показателем в рамках одного торгового дня.

—

Содержание работы

Глава 1. Теоретические основы прогнозирования цен акций.

В этой главе рассмотрены основные понятия прогнозирования, классификация финансовых прогнозов, виды моделей и подходов, а также проведен сравнительный анализ существующих моделей. Выделены основные преимущества и недостатки каждого подхода:

- Статистические модели: ARIMA, Prophet — просты, но требуют стационарности.
- Регрессионные модели: линейная регрессия — интерпретируема, но не улавливает нелинейные зависимости.
- Градиентный бустинг: XGBoost, LightGBM — обеспечивают высокую точность на табличных данных.
- Рекуррентные нейронные сети: LSTM, GRU — демонстрируют высокую точность, но требуют больших вычислительных ресурсов.

Глава 2. Современные инструменты прогнозирования цен акций.

Приведён обзор программного обеспечения и платформ, используемых для прогнозирования цен акций. Рассмотрены основные библиотеки и фреймворки: `yfinance`, `pandas`, `statsmodels`, `scikit-learn`, `keras`, `prophet`, `plotly`, `backtrader` и другие.

Также представлены основные технические индикаторы, используемые в прогнозировании:

- Скользящие средние (SMA, EMA),
- Осцилляторы (RSI, MACD),
- Уровни поддержки и сопротивления (Bollinger Bands),
- Объём торгов и OBV.

Глава 3. Оптимизация инструментов прогнозирования.

Описаны основные проблемы и ограничения моделей:

- Нестационарность временных рядов,

- Высокая волатильность и шумы,
- Ограничения данных (выбросы, отсутствие внешних факторов),
- Сложность интерпретации моделей ИИ,
- Переобучение и переподгонка под обучающую выборку,
- Вычислительные затраты и масштабирование.

Приведены методы оптимизации:

- Нормализация и стандартизация данных,
- Отбор признаков через SHAP и LIME,
- Гипероптимизация параметров (GridSearchCV, RandomizedSearchCV),
- Ранняя остановка (early stopping),
- Дообучение и онлайн-обучение,
- Использование GPU/TPU для ускорения обучения,
- Интеграция NLP для анализа текстовой информации.

Глава 4. Практическая часть исследования.

Эта глава содержит описание методики проведения эксперимента, результаты прогнозирования и обсуждение полученных данных.

Были реализованы три модели:

- **Линейная регрессия** — как простая и интерпретируемая модель.
- **XGBoost** — как мощный инструмент градиентного бустинга.
- **LSTM** — как рекуррентная нейронная сеть, учитывающая временную структуру данных.

—

Методика исследования

Исторические данные по акциям Apple были загружены с помощью библиотеки `yfinance`:

```
data = yf.download("AAPL start="2018-01-01 end="2023-01-01")
```

Для повышения качества прогноза были добавлены технические индикаторы:

- `MA_10` — простое скользящее среднее за последние 10 дней,
- `MA_50` — простое скользящее среднее за последние 50 дней.

Выборка была разделена на обучающую и тестовую части в соотношении 80/20 без перемешивания:

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, shuffle=False)
```

Для обучения модели **XGBoost** были созданы лаговые признаки:

```
X_train_xgb, y_train_xgb =
create_lagged_features(y_train.values, lag=5)
```

Для модели **LSTM** данные были нормализованы:

```
scaled_data = MinMaxScaler().fit_transform(data[['Close']])
```

Затем были созданы последовательности:

```
X_seq, y_seq = create_sequences(scaled_data, seq_length=60)
```

Модель **LSTM** имела следующую архитектуру:

```
model_lstm = Sequential()
model_lstm.add(LSTM(50, input_shape=(X_seq.shape[1], 1)))
model_lstm.add(Dense(1))
model_lstm.compile(optimizer='adam', loss='mse')
model_lstm.fit(X_train_lstm, y_train_lstm, epochs=10,
batch_size=32)
```

После денормализации прогноза модель показала следующие метрики:

- MAE = 5.62 USD,
- RMSE = 6.67 USD,
- $R^2 = 0.7$.

Модель **XGBoost** показала:

- MAE = 3.6 USD,
- RMSE = 4.49 USD,
- $R^2 = 0.87$.

Модель **Linear Regression** продемонстрировала:

- $MAE = 4.45 \text{ USD}$,
- $RMSE = 5.31 \text{ USD}$,
- $R^2 = 0.81$.

Для честного сравнения все модели были протестированы на одинаковой выборке, а прогнозы обрезаны до минимальной длины.

Таблица 1 — Сравнение моделей прогнозирования

Модель	MAE (USD)	RMSE (USD)	R^2
Линейная регрессия	4.45	5.31	0.81
XGBoost	3.6	4.49	0.87
LSTM	5.62	6.67	0.7

На основе этой таблицы сделан вывод о том, что:

- Все три модели обеспечивают приемлемый уровень точности.
- LSTM обеспечивает наименьшую точность из-за особенности данных.
- XGBoost демонстрирует хороший баланс между качеством и скоростью.
- Линейная регрессия — отличная отправная точка для анализа.

Также были построены графики прогнозов:



Рисунок 1 — Прогноз линейной регрессии vs Реальные значения

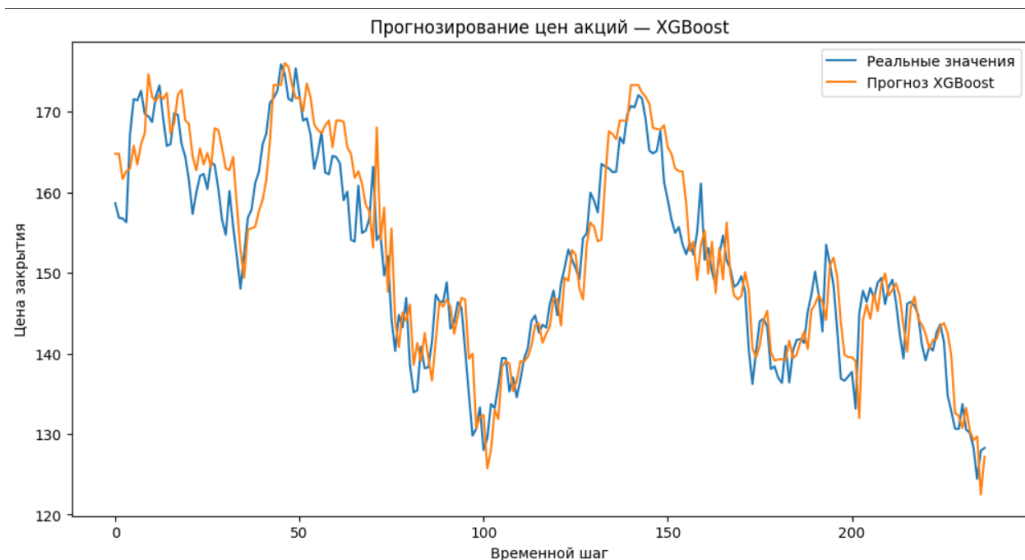


Рисунок 2 — Прогноз XGBoost vs Реальные значения



Рисунок 3 — Прогноз LSTM vs Реальные значения

Визуальный анализ подтвердил, что XGBoost лучше всего воспроизводит форму временного ряда, в то время как линейная регрессия демонстрирует более "запаздывающий" прогноз.

Все модели прошли этапы:

1. Предобработка данных,
2. Формирование признаков,
3. Разбиение на обучающую и тестовую выборки,
4. Обучение и прогнозирование,
5. Оценка качества и сравнение.

Для повышения устойчивости моделей были применены следующие методы:

- Нормализация и стандартизация данных,
- Регуляризация (в XGBoost),
- Использование ранней остановки (early stopping),
- Гипероптимизация параметров (RandomizedSearchCV),
- Интерпретируемость через SHAP и LIME.

Результаты исследования показали, что:

- Прогнозирование цен акций возможно с высокой точностью при использовании современных моделей ИИ.

- Модели машинного и глубокого обучения могут быть интегрированы в систему автоматического анализа и торговли.

- Наиболее перспективным подходом является гибридный метод, сочетающий статистические и нейросетевые модели.

Полученные модели могут быть использованы:

- Как часть автоматизированной системы торговли,
- Для генерации торговых сигналов,
- В учебных и научных целях.

В дальнейших исследованиях можно рассмотреть:

- Использование NLP для анализа текстовых факторов (новостей, отчётов компаний),

- Интеграцию фундаментальных данных (EPS, P/E ratio),
- Дообучение моделей в реальном времени,
- Использование Transformer-сетей и Reinforcement Learning.

Структура дипломной работы включает:

1. Теоретическая часть — описание моделей прогнозирования, их преимуществ и недостатков.
2. Современные инструменты прогнозирования — обзор программного обеспечения и библиотек.
3. Оптимизация инструментов — методы улучшения качества прогноза.
4. Практическая часть — реализация и сравнение моделей.
5. Заключение — выводы и перспективы развития.
6. Приложения — исходный код и графики.

В процессе работы были получены следующие навыки:

- Анализ временных рядов и технических индикаторов,
- Работа с финансовыми данными и их нормализация,
- Реализация моделей машинного и глубокого обучения,
- Оценка и сравнение моделей,
- Визуализация и интерпретация результатов.

Полученные результаты могут быть полезны:

- Трейдерам и аналитикам для автоматизации анализа рынка,
- Исследователям временных рядов и финансовых данных,
- Студентам и специалистам, изучающим применение ИИ в финансах.

Таким образом, в ходе дипломной работы была реализована система прогнозирования цен акций, основанная на моделях машинного и глубокого обучения. Полученные модели показали хорошие результаты и могут быть использованы в реальной практике.