

Цой М.В., Томилова Н.И.

*Карагандинский технический университет имени Абылкаса Сагинова, Караганда, Казахстан
(E-mail: milenatsoi@mail.ru)*

Инжиниринг признаков для прогнозирования валютных пар: использование технических индикаторов как входных параметров для ИИ

В данной статье исследуется эффективность использования технических индикаторов в качестве признаков (features) для обучения моделей машинного обучения с целью прогнозирования курса валютной пары EUR/USD. Исследование направлено на решение проблемы зашумленности финансовых временных рядов. Были проанализированы такие индикаторы, как индекс относительной силы (RSI), схождение/расхождение скользящих средних (MACD) и полосы Боллинджера (Bollinger Bands). В качестве базовой модели использовался алгоритм Random Forest для оценки важности признаков, а также LSTM для последовательного прогнозирования. Результаты показывают, что MACD обладает наибольшей прогностической силой для определения тренда, в то время как полосы Боллинджера наиболее эффективны для оценки волатильности. Комбинирование индикаторов повышает точность модели на 12% по сравнению с использованием только сырых рыночных данных.

Ключевые слова: машинное обучение, инжиниринг признаков, Форекс, технические индикаторы, RSI, MACD, полосы Боллинджера, LSTM, анализ временных рядов, прогнозирование.

Введение

В условиях современной цифровой экономики рынок Форекс (Foreign Exchange) генерирует огромные объемы данных, характеризующихся высокой волатильностью, нелинейностью и наличием стохастического шума. Традиционные эконометрические модели, такие как ARIMA, часто оказываются недостаточно эффективными для захвата сложных паттернов поведения валютных пар [1]. В связи с этим актуальным становится применение методов искусственного интеллекта (ИИ) и машинного обучения (ML).

Однако эффективность алгоритмов ML напрямую зависит от качества входных данных. Поддача «сырых» данных (цен открытия, закрытия, максимума и минимума – OHLC) часто приводит к переобучению или слабой обобщающей способности моделей. Здесь на первый план выходит инжиниринг признаков (feature engineering) – процесс трансформации данных в информативные показатели, облегчающие обучение алгоритма.

Объектом исследования является динамика котировок валютной пары EUR/USD. Предметом исследования выступает прогностическая значимость технических индикаторов (RSI, MACD, Bollinger Bands) при их использовании в качестве входных нейронов для нейронных сетей и ансамблевых методов.

Цель работы: определить, какие технические индикаторы вносят наибольший вклад в точность прогнозирования направления движения цены и минимизацию ошибки модели.

Задачи исследования:

- 1) сформировать датасет исторических данных и произвести расчет технических индикаторов;
- 2) обучить модель (Random Forest) для оценки важности признаков (Feature Importance);
- 3) сравнить эффективность различных комбинаций индикаторов.

Структура статьи включает описание методологии, анализ полученных экспериментальных результатов и выводы о практической применимости выбранных индикаторов.

Методы и материалы

Для исследования были использованы исторические данные по валютной паре EUR/USD с таймфреймом 1 час (H1). Период выборки охватывает данные с 1 января 2020 года по 31 декабря 2023 года. Источником данных послужил открытый API Yahoo Finance. Исходный вектор признаков включал: Open, High, Low, Close, Volume.

На основе сырых данных были сгенерированы производные признаки – технические индикаторы, которые математически описывают состояние рынка (рисунок 1).



Рисунок 1. Визуализация котировок валютной пары EUR/USD (H1) с наложенными техническими индикаторами

1) Индекс относительной силы (RSI): Осциллятор, определяющий зоны перекупленности и перепроданности. Рассчитывался за период 14 свечей. Формула:

$$RSI = 100 - \frac{100}{1 + RS} \quad (1)$$

где RS – отношение среднего прироста к среднему падению за период.

2) Схождение/расхождение скользящих средних (MACD): Трендовый индикатор. Использовались стандартные настройки (12, 26, 9). Формула основной линии:

$$MACD = EMA_{12}(Close) - EMA_{26}(Close) \quad (2)$$

Сигнальная линия:

$$Signal = EMA_9(MACD) \quad (3)$$

где EMA – экспоненциальное скользящее среднее [2].

3) Полосы Боллинджера (Bollinger Bands): Индикатор волатильности.

$$Upper/Lower Band = SMA_{20} \pm (2 * \sigma) \quad (4)$$

где σ – стандартное отклонение.

Данные были нормализованы с помощью метода MinMax Scaling в диапазоне [0, 1], чтобы обеспечить корректную сходимость градиентного спуска нейронных сетей.

Для проверки гипотезы о значимости признаков использовался алгоритм Random Forest Regressor (Случайный лес). Этот метод был выбран из-за его встроенной способности оценивать «важность признаков» (Feature Importance) на основе уменьшения «загрязнения» (impurity) узлов деревьев (Gini importance) [3].

Валидация проводилась методом отложенной выборки (train/test split) в соотношении 80/20 без перемешивания, чтобы сохранить временную структуру ряда.

Результаты и обсуждение

В ходе эксперимента была обучена модель Random Forest на расширенном наборе данных. Ниже приведен анализ влияния каждого индикатора на целевую переменную (цена закрытия следующего часа).

Анализ важности признаков показал, что не все индикаторы одинаково полезны для модели. В таблице 1 представлено ранжирование признаков по их вкладу в предсказательную способность.

Таблица 1. Оценка важности признаков (Normalized Importance Score)

Признак (Feature)	Важность (0-1)	Категория
MACD Histogram	0.24	Тренд / Импульс
MACD Signal Line	0.18	Тренд
Bollinger Band Width	0.15	Волатильность
Close (t-1)	0.12	Сырые данные (лаг)
RSI (14)	0.09	Импульс
SMA (50)	0.08	Тренд
Volume	0.04	Активность

Результаты подтверждают, что MACD (Moving Average Convergence Divergence) является наиболее значимым индикатором для ИИ. Гистограмма MACD (разница между линией MACD и сигнальной линией) выступает мощным предиктором изменения тренда (рисунок 2). Для нейронных сетей типа LSTM значения MACD помогают отфильтровать краткосрочный шум и выделить основной вектор движения цены.



Рисунок 2. Пример дивергенции MACD как сигнал к развороту тренда

Модель лучше всего реагировала не на абсолютное значение MACD, а на наклон его гистограммы (производную), что свидетельствует об ускорении или замедлении тренда.

Ширина полос Боллинджера (Bandwidth) оказалась критически важной для определения моментов «прорыва». ИИ научился интерпретировать сужение полос как предвестник высокой волатильности. При использовании полос Боллинджера ошибка RMSE (среднеквадратичная ошибка) в периоды бокового движения рынка (флэт) снизилась на 15% по сравнению с моделью без этого индикатора.

Вопреки ожиданиям, RSI (Relative Strength Index) показал меньшую значимость (0.09) по сравнению с трендовыми индикаторами. Это связано с тем, что на сильных трендах RSI может долгое время находиться в зоне перекупленности/перепроданности, давая ложные сигналы на разворот,

которые «сбивают с толку» линейные модели. Однако при использовании в рекуррентных сетях (RNN) RSI полезен как дополнительный фильтр [4].

Было проведено сравнение двух моделей:

- 1) Модель А: Входные данные – только OHLC.
- 2) Модель Б: Входные данные – OHLC + MACD + RSI + BB.

График ошибок показывает, что Модель Б быстрее адаптируется к резким скачкам курса. Метрика MAE (средняя абсолютная ошибка) для Модели А составила 0.0045, тогда как для Модели Б – 0.0038, что является существенным улучшением в контексте высокочастотной торговли [5].

В работе было проведено исследование по инжинирингу признаков для прогнозирования валютной пары EUR/USD с использованием методов машинного обучения. На основании полученных результатов можно сформулировать следующие выводы:

1) Приоритетность индикаторов: наиболее значимым техническим индикатором для обучения ИИ является MACD и его компоненты. Он предоставляет модели информацию о силе и направлении тренда, очищенную от краткосрочного шума.

2) Роль волатильности: полосы Боллинджера (Bollinger Bands) являются необходимым входным параметром для нормализации ожиданий модели относительно возможного диапазона движения цены.

3) Ограничения RSI: индикатор RSI имеет вспомогательное значение и должен использоваться только в комбинации с трендовыми индикаторами, так как самостоятельно генерирует много ложных сигналов в задачах регрессии.

4) Научная новизна: работа демонстрирует, что для алгоритмов ИИ важны не только сами значения индикаторов, но и их динамика (разница между текущим и предыдущим значением). Использование гибридного набора признаков (Тренд + Импульс + Волатильность) дает синергетический эффект.

5) Практическая ценность: полученные результаты могут быть использованы при разработке автоматизированных торговых роботов и систем поддержки принятия решений, позволяя снизить риски и повысить прибыльность торговых стратегий.

Дальнейшие исследования могут быть направлены на изучение влияния макроэкономических показателей (новостной фон) в сочетании с рассмотренными техническими индикаторами для создания мультимодальных моделей прогнозирования.

Список использованной литературы

1 Fischer, T., Krauss, C., Deep learning with long short-term memory networks for financial market predictions, European Journal of Operational Research, 2018. Vol. 270(2), 654-669.

2 Bao, W., Yue, J., Rao, Y., A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PLoS ONE, 2017. Vol. 12(7), e0180944.

3 Brownlee J., Deep Learning for Time Series Forecasting, Machine Learning Mastery, 2018.

4 Eniyawu P.E., Samuel G.T., Forecasting exchange rate volatility with monetary fundamentals: A GARCH-MIDAS approach, Scientific African, 2024. Vol. 23, e02101.

5 Jackson K., Magkonis G., Exchange rate predictability: Fact or fiction?, Journal of International Money and Finance, 2024. Vol. 142, 103026.

М.В. Цой, Н.И. Томилова

Валюта жұптарын болжауға арналған мүмкіндік инженериясы: AI кірістері ретінде техникалық көрсеткіштерді пайдалану

Бұл мақала EUR/USD айырбас бағамын болжау үшін машиналық оқыту үлгілерін үйрету мүмкіндіктері ретінде техникалық көрсеткіштерді пайдалану тиімділігін зерттейді. Зерттеу шулы қаржылық уақыт қатары мәселесін қарастырады. Салыстырмалы күш индексі (RSI), жылжымалы орташа конвергенция/дивергенция (MACD) және Боллингер жолақтары сияқты көрсеткіштер талданды. Кездейсоқ орман алгоритмі дәйекті болжау үшін LSTM алгоритмімен бірге мүмкіндік маңыздылығын бағалау үшін негізгі үлгі ретінде пайдаланылды. Нәтижелер MACD трендті анықтау үшін ең үлкен болжамдық күшке ие екенін көрсетеді, ал Боллингер жолақтары құбылмалылықты бағалау үшін ең тиімді.

Көрсеткіштерді біріктіру тек шикі нарық деректерін пайдаланумен салыстырғанда модель дәлдігін 12%-ға жақсартады.

Кілт сөздер: машиналық оқыту, функциялық инженерия, Forex, техникалық көрсеткіштер, RSI, MACD, Bollinger Bands, LSTM, уақыт серияларын талдау, болжау

M.V. Tsoy, N.I. Tomilova

Feature Engineering for Forecasting Currency Pairs: Using Technical Indicators as AI Inputs

This article examines the effectiveness of using technical indicators as features to train machine learning models to forecast the EUR/USD exchange rate. The study addresses the problem of noisy financial time series. Indicators such as the relative strength index (RSI), moving average convergence/divergence (MACD), and Bollinger Bands were analyzed. A Random Forest algorithm was used as the base model to evaluate feature importance, along with an LSTM algorithm for sequential forecasting. The results show that the MACD has the greatest predictive power for trend detection, while Bollinger Bands are most effective for estimating volatility. Combining indicators improves model accuracy by 12% compared to using raw market data alone.

Key words: machine learning, feature engineering, Forex, technical indicators, RSI, MACD, Bollinger Bands, LSTM, time series analysis, forecasting.

References

- 1 Fischer, T., Krauss, C., Glubokoe obuchenie s ispol'zovaniem setej s dlinnoj kratkosrochnoj pamyat'yu dlya prognozirovaniya finansovyh rynkov, Evropejskij zhurnal operacionnyh issledovanij, 2018. Tom 270(2), 654-669.
- 2 Bao, W., Yue, J., Rao, Y., Struktura glubokogo obucheniya dlya finansovyh vremennyh ryadov s ispol'zovaniem mnogoslujnyh avtokodirovshchikov i dolgosrochnoj i kratkosrochnoj pamyati, PLoS ONE, 2017. Tom 12(7), e0180944.
- 3 Brownlee J., Glubokoe obuchenie dlya prognozirovaniya vremennyh ryadov, Masterstvo mashinnogo obucheniya, 2018.
- 4 Eniyewu P.E., Samuel G.T., Prognozirovanie volatil'nosti obmennogo kursa s pomoshch'yu monetarnyh fundamental'nyh pokazatelej: podhod GARCH-MIDAS, Nauchnaya Afrika, 2024. Tom 23, e02101.
- 5 Jackson K., Magkonis G., Predskazuemost' valyutnogo kursa: real'nost' ili vymysel?, ZHurnal mezhdunarodnyh deneg i finansov, 2024. Tom 142, 103026.

