

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра веб-технологий и компьютерного моделирования

БЫЧКОВ Владислав Витальевич

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Аннотация к магистерской диссертации

Научный руководитель:
К.В. Лыков
доктор физ.-мат. наук

Минск, 2025

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Магистерская диссертация «Машинное обучение для анализа временных рядов» посвящена актуальной задаче повышения эффективности методов анализа и прогнозирования временных рядов. Работа соотносится с приоритетными направлениями кафедры, связанными с разработкой интеллектуальных систем обработки и анализа данных, а также внедрением современных методов машинного обучения в практику анализа сложных динамических процессов.

Цель исследования — разработка, реализация и практическое применение моделей для прогнозирования временных рядов на примере курса валют, с использованием как классических статистических методов, так и современных нейросетевых архитектур, с последующим сравнительным анализом их эффективности на основе backtest-моделирования.

Задачи исследования:

Провести обзор и анализ классических статистических моделей временных рядов (AR, MA, ARIMA, экспоненциальное сглаживание) и выявить их сильные и слабые стороны при работе с реальными данными.

Изучить различные нейросетевые архитектуры (RNN, LSTM), методы их обучения и выявить их сильные и слабые стороны при работе с реальными данными.

Реализовать автоматическое извлечение дополнительных признаков из исходного ряда, что позволяет расширить описание входных данных и повысить информативность модели.

Применить гибридный подход, интегрировав в модель закодированные прогнозы финансовых экспертов, что даёт возможность учитывать как алгоритмически извлечённые признаки, так и экспертное мнение.

Сравнить практическую эффективность классических статистических моделей и современных нейросетевых архитектур при прогнозировании временных рядов валютного курса, используя результаты бэктестирования для оценки их реального торгового потенциала.

Объектом исследования являются временные ряды, отражающие поведение динамических систем во времени.

Предметом исследования являются методы анализа и прогнозирования временных рядов на основе статистических моделей и нейронных сетей.

Хронологические рамки исследования охватывают период с 2023 по 2025 годы и обусловлены доступностью достоверных, полноценных временных рядов за указанный период, представляющих практический интерес для анализа.

Методология исследования включает методы математической статистики, регрессионный анализ, а также алгоритмы машинного обучения, в первую очередь рекуррентные нейронные сети (RNN, LSTM).

Научная новизна работы заключается в комплексном применении современных методов машинного обучения для анализа и прогнозирования

временных рядов. В работе использованы актуальные нейросетевые архитектуры (RNN, LSTM) и современные методы оптимизации, что позволило в ряде случаев повысить качество прогнозирования по сравнению с традиционными статистическими моделями. В работе используется расширенный набор признаков, включающий как автоматически извлечённые статистические характеристики временных рядов, так и дополнительные факторы — экспертные оценки и новостную информацию. Построена гибридная модель, объединяющая количественные и качественные признаки, что позволяет более полно учитывать влияние различных факторов на динамику временного ряда.

Положения, выносимые на защиту:

1. Обоснована практическая эффективность применения современных нейросетевых архитектур (RNN, LSTM) для анализа и прогнозирования временных рядов в задачах с выраженной нелинейностью и нестационарностью.

2. Разработан комплексный подход к построению моделей прогнозирования, предусматривающий автоматическое извлечение статистических признаков, а также включение внешних источников информации — финансовых индикаторов и экспертных оценок — в качестве дополнительных количественных признаков.

3. Предложена методика преобразования экспертных мнений и новостных текстов в формат, пригодный для включения в процесс машинного обучения временных рядов.

Апробация результатов магистерской диссертации осуществлялась в рамках научно-исследовательской практики в ЗАО «Банк ВТБ (Беларусь)». Также результаты исследования опубликованы в виде тезисов доклада на конференции «Экономика Беларуси: рост, инновации, безопасность» - 2025, Минск.

Структура магистерской диссертации отражает логику проведённого исследования и состоит из введения, трёх глав, заключения, списка использованных источников.

В первой главе представлены теоретические основы анализа временных. Во второй главе рассматриваются модели временных рядов: классические статистические подходы и нейросетевые архитектуры. Третья глава посвящена практическим экспериментам: подготовке данных, реализации моделей, проведению экспериментов и анализу результатов. В заключении сформулированы основные выводы исследования и обозначены перспективы дальнейшей работы в этой области.

АГУЛЬНАЯ ХАРАКТЕРЫСТЫКА ПРАЦЫ

Магістарская дысертация "Машыннае навучанне для аналізу часовых шэрагаў" прысвечана актуальнай задачы павышэння эфектыўнасці метадаў аналізу і прагназавання часовых шэрагаў. Праца суадносіцца з прыярытэтнымі напрамкамі кафедры, звязанымі з распрацоўкай інтэлектуальных сістэм апрацоўкі і аналізу даных, а таксама ўкараненнем сучасных метадаў машыннага навучання ў практыку аналізу складаных дынамічных працэсаў.

Мэта даследавання - распрацоўка, рэалізацыя і практычнае прымяненне мадэляў для прагназавання часовых шэрагаў на прыкладзе курса валют, з выкарыстаннем як класічных статыстычных метадаў, так і сучасных нейросетевых архітэктур, з наступным параўнальным аналізам іх эфектыўнасці на аснове backtest-мадэлявання.

Задачы даследавання:

Правесці агляд і аналіз класічных статыстычных мадэляў часавых шэрагаў (AR, MA, ARIMA, экспанентнае згладжванне) і выявіць іх моцныя і слабыя бакі пры працы з рэальнымі дадзенымі.

Вывучыць розныя нейрасеткавыя архітэктурны (RNN, LSTM), метады іх навучання і выявіць іх моцныя і слабыя бакі пры працы з рэальнымі дадзенымі.

Рэалізаваць аўтаматычнае выманне дадатковых прыкмет з зыходнага шэрагу, што дазваляе пашырыць апісанне ўваходных дадзеных і павысіць інфарматыўнасць мадэлі.

Ужыць гібрыдны падыход, інтэграваўшы ў мадэль закадавання прагнозы фінансавых экспертаў, што дае магчымасць улічваць як алгарытмічна вынятыя прыкметы, так і экспертнае меркаванне.

Параўнаць практычную эфектыўнасць класічных статыстычных мадэляў і сучасных нейросетевых архітэктур пры прагназаванні часовых шэрагаў валютнага курса, выкарыстоўваючы вынікі бэктэставання для ацэнкі іх рэальнага гандлёвага патэнцыялу.

Аб'ектам даследавання з'яўляюцца часовыя шэрагі, якія адлюстроўваюць паводзіны дынамічных сістэм у часе.

Прадметам даследавання з'яўляюцца метады аналізу і прагназавання часовых шэрагаў на аснове статыстычных мадэляў і нейронавых сетак.

Храналагічныя рамкі даследавання ахопліваюць перыяд з 2023 па 2025 гады і абумоўлены даступнасцю дакладных, паўнацэнных часовых шэрагаў за названы перыяд, якія ўяўляюць практычную цікавасць для аналізу.

Метадалогія даследавання ўключае метады матэматычнай статыстыкі, рэгрэсійны аналіз, а таксама алгарытмы машыннага навучання, у першую чаргу рэкурэнтныя нейронавыя сеткі (RNN, LSTM).

Навуковая навізна працы заключаецца ў комплексным ужыванні сучасных метадаў машыннага навучання для аналізу і прагназавання часовых шэрагаў. У працы выкарыстаны актуальныя нейрасеткавыя архітэктурны (RNN, LSTM) і сучасныя метады аптымізацыі, што дазволіла ў шэрагу выпадкаў павысіць якасць

прагнавання ў параўнанні з традыцыйнымі статыстычнымі мадэлямі. У працы выкарыстоўваецца пашыраны набор прыкмет, які ўключае як аўтаматычна вынятыя статыстычныя характарыстыкі часовых шэрагаў, так і дадатковыя фактары – экспертныя адзнакі і навінную інфармацыю. Пабудавана гібрыдная мадэль, якая аб'ядноўвае колькасныя і якасныя прыкметы, што дазваляе больш поўна ўлічваць уплыў розных фактараў на дынаміку часовага шэрагу.

Палажэнні, якія выносяцца на абарону:

1. Абгрунтавана практычная эфектыўнасць прымянення сучасных нейросетевых архітэктур (RNN, LSTM) для аналізу і прагнавання часовых шэрагаў у задачах з выяўленай нелінейнасцю і нестацыянарнасцю.

2. Распрацаваны комплексны падыход да пабудовы мадэляў прагнавання, які прадугледжвае аўтаматычнае выманне статыстычных прыкмет, а таксама ўключэнне знешніх крыніц інфармацыі – фінансавых індыхатараў і экспертных адзнак – у якасці дадатковых колькасных прыкмет.

3. Предложена методыка пераўтварэнні экспертных меркаванняў і навінавых тэкстаў у фармат, прыдатны для ўключэння ў працэс машыннага навучання часовых шэрагаў.

Апрацацыя вынікаў магістарскай дысертацыі ажыццяўлялася ў рамках навукова-даследчай практыкі ў ЗАТ "Банк ЗГБ (Беларусь)". Таксама вынікі даследавання апублікаваны ў выглядзе тэзісаў даклада на канферэнцыі "Эканоміка Беларусі: рост, інавацыі, бяспека" - 2025, Мінск.

Структура магістарскай дысертацыі адлюстроўвае логіку праведзенага даследавання і складаецца з увядзення, трох частак, заключэння, спісу выкарыстаных крыніц.

У першым раздзеле прадстаўлены тэарэтычныя асновы аналізу часовых. У другім раздзеле разглядаюцца мадэлі часавых шэрагаў: класічныя статыстычныя падыходы і нейросетевыя архітэктury. Трэці раздзел прысвечана практычным эксперыментам: падрыхтоўцы дадзеных, рэалізацыі мадэляў, правядзенню эксперыментаў і аналізу вынікаў. У заключэнні сфармуляваны асноўныя высновы даследавання і абазначаны перспектывы далейшай працы ў гэтай галіне.

GENERAL DESCRIPTION OF WORK

The master's thesis 'Machine Learning for Time Series Analysis' is devoted to the topical task of improving the efficiency of methods for analysing and forecasting time series. The work is related to the priority areas of the department, which are connected with the development of intelligent data processing and analysis systems, as well as the implementation of modern machine learning methods in the analysis of complex dynamic processes.

The aim of the research is to develop, implement and apply models for forecasting time series using the example of exchange rates, using both classical statistical methods and modern neural network architectures, followed by a comparative analysis of their effectiveness based on backtesting.

Research objectives:

Conduct a review and analysis of classical statistical time series models (AR, MA, ARIMA, exponential smoothing) and identify their strengths and weaknesses when working with real data.

Study various neural network architectures (RNN, LSTM), methods of training them, and identify their strengths and weaknesses when working with real data.

Implement automatic extraction of additional features from the original series, which allows you to expand the description of the input data and increase the informativeness of the model.

Apply a hybrid approach by integrating coded forecasts from financial experts into the model, which makes it possible to take into account both algorithmically extracted features and expert opinion.

Compare the practical effectiveness of classical statistical models and modern neural network architectures in forecasting currency exchange rate time series, using backtesting results to assess their real trading potential.

The object of the study is time series reflecting the behaviour of dynamic systems over time.

The subject of the study is methods for analysing and forecasting time series based on statistical models and neural networks.

The chronological framework of the study covers the period from 2023 to 2025 and is determined by the availability of reliable, complete time series for the specified period that are of practical interest for analysis.

The research methodology includes methods of mathematical statistics, regression analysis, and machine learning algorithms, primarily recurrent neural networks (RNN, LSTM).

The scientific novelty of the work lies in the comprehensive application of modern machine learning methods for the analysis and forecasting of time series. The work uses current neural network architectures (RNN, LSTM) and modern optimisation methods, which in a number of cases has made it possible to improve the quality of forecasting compared to traditional statistical models. The work uses an extended set of features, including both automatically extracted statistical

characteristics of time series and additional factors — expert assessments and news information. A hybrid model has been constructed that combines quantitative and qualitative features, which allows for a more complete consideration of the influence of various factors on the dynamics of the time series.

Theses defended:

1. The practical effectiveness of using modern neural network architectures (RNN, LSTM) for analysing and forecasting time series in tasks with pronounced non-linearity and non-stationarity is substantiated.

2. A comprehensive approach to building forecasting models has been developed, providing for the automatic extraction of statistical features, as well as the inclusion of external sources of information — financial indicators and expert assessments — as additional quantitative features.

3. A methodology has been proposed for converting expert opinions and news texts into a format suitable for inclusion in the machine learning process for time series.

The results of the master's thesis were tested as part of a research internship at VTB Bank (Belarus). The research results were also published in the form of abstracts at the conference 'Economy of Belarus: Growth, Innovation, Security' - 2025, Minsk.

The structure of the master's thesis reflects the logic of the research and consists of an introduction, three chapters, a conclusion, and a list of sources used.

The first chapter presents the theoretical foundations of time series analysis. The second chapter examines time series models: classical statistical approaches and neural network architectures. The third chapter is devoted to practical experiments: data preparation, model implementation, conducting experiments, and analysing results. The conclusion formulates the main findings of the research, provides recommendations on the application of the developed methods, and outlines prospects for further work in this area.