

ПРИКЛАДНАЯ МАТЕМАТИКА И МЕХАНИКА

APPLIED MATHEMATICS AND MECHANICS

Вестник Сыктывкарского университета.

Серия 1: Математика. Механика. Информатика. 2024.

Выпуск 4 (53)

Bulletin of Syktovkar University.

Series 1: Mathematics. Mechanics. Informatics. 2024; 4 (53)

Научная статья

УДК 539.3

https://doi.org/10.34130/1992-2752_2024_4_4

ОЦЕНКА ИНВЕСТИЦИОННОЙ АКТИВНОСТИ НА ОСНОВЕ НОВОСТНОГО ФОНА

Петр Викторович Борков¹,
Ольга Амнджановна Мальцева¹,
Ирина Владимировна Полякова¹,
Евгения Николаевна Старцева²

¹Банк России, maltseva.rs@yandex.ru

²Сыктывкарский государственный университет
имени Питирима Сорокина

Аннотация. Целью исследования является построение опережающего показателя инвестиционной активности на основе анализа новостей Telegram-каналов. Имеющиеся данные Росстата по инвестиционной активности публикуются с временным лагом и подвергаются корректировке, что затрудняет их использование при оперативной оценке текущей экономической ситуации. В работе рассматриваются два подхода: первый — на основе фильтра по ключевым словам, второй — на основе языковой модели «BERT». Оба подхода демонстрируют статистически значимую корреляцию с данными Росстата.

Ключевые слова: текстовый анализ, машинное обучение, BERT, rubert-tiny2, инвестиционная активность, Python

Благодарности. Авторы выражают благодарность Полехиной Алисе Александровне, Гусевой Анне Викторовне за предоставленную базу данных новостей Telegram-каналов, а также коллегам Банка России и Сыктывкарского государственного университета имени Питирима Сорокина, участвовавших в обсуждениях работы, за полезные замечания и предложения.

Для цитирования: Борков П. В., Мальцева О. А., Полякова И. В., Старцева Е. Н. Оценка инвестиционной активности на основе новостного фона // *Вестник Сыктывкарского университета. Сер. 1: Математика. Механика. Информатика.* 2024. Вып. 4 (53). С. 4–20. https://doi.org/10.34130/1992-2752_2024_4_4

Настоящая статья отражает личную позицию авторов. Содержание и результаты данного исследования не следует рассматривать, в том числе цитировать в каких-либо изданиях, как официальную позицию Банка России или указание на официальную политику или решения регулятора. Любые ошибки в данном материале являются исключительно авторскими.

Article

ESTIMATION OF INVESTMENT ACTIVITY BASED ON THE NEWS BACKGROUND

Petr V. Borkov¹, Olga A. Maltseva¹,
Irina V. Polyakova¹, Evgenija N. Startseva²

¹Bank of Russia, maltseva.rs@yandex.ru

²Pitirim Sorokin Syktyvkar State University

Abstract. The purpose of our research is to build a leading indicator of investment activity based on the analysis of Telegram-channel news. The available Rosstat data on investment activity are published with a time lag and are subject to adjustment, which makes it difficult to use them in an operational assessment of the current economic situation. The paper considers two approaches: the first is based on a keyword filter, the second is based on the „BERT“ language model. Both approaches demonstrate a statistically significant correlation with Rosstat data.

Keywords: NLP, machine learning, BERT, rubert-tiny2, investment activity, Python

Acknowledgments. The authors express their gratitude to Alisa A. Polekhina, Anna V. Guseva for providing the database of news of Telegram channels, as well as colleagues of the Bank of Russia and Pitirim Sorokin Syktyvkar State University, who participated in discussions of the work, for useful comments and suggestions.

For citation: Borkov P. V., Maltseva O. A., Polyakova I. V., Startseva E. N. Estimation of investment activity based on the news background. *Vestnik Syktyvkarского университета. Seriya 1: Matematika. Mekhanika. Informatika* [Bulletin of Syktyvkar University, Series 1: Mathematics. Mechanics. Informatics], 2024, no 4 (53), pp. 4–20. (In Russ.) https://doi.org/10.34130/1992-2752_2024_4_4

The views expressed herein are solely those of the authors. The content and results of this research should not be considered or referred to in any publications as the Bank of Russia's official position, official policy, or decisions. Any errors in this document are the responsibility of the authors.

1. Введение

Динамика инвестиционной активности является одним из важных аспектов экономического развития и потенциала предприятия, отрасли, региона, страны, что объясняет внимание исследователей к этой теме. Основными показателями инвестиционной активности в России считаются такие данные Росстата, как динамика инвестиций в основной капитал и валовое накопление основного капитала. После перехода в 2016 году с ежемесячного наблюдения на квартальное оперативность официальных показателей Росстата по инвестициям в основной капитал ухудшилась. Кроме того, еще одной из проблем официальной статистики является значительный временной лаг¹, который возникает с момента сбора информации до публикации агрегированных данных [1].

В литературе для оценки экономической ситуации достаточно широко применяется анализ новостного фона [2–7], при этом в качестве источника информации используются статьи крупных изданий. Новизной исследования является выбранный источник — сообщения из Telegram-каналов. Возможность получать информацию оперативно, практически в любых разрезах и с любой частотой — неоспоримые преимущества новостной базы Telegram-каналов. Еще одним преимуществом базы дан-

¹1-я оперативная оценка публикуется на 40-й рабочий день, после отчетного квартала, окончательная оценка — февраль четвертого года, после отчетного года.

ных является наличие информации не только профильных специалистов, а еще настроений и ожиданий в обществе в целом.

Исследование строится на предположении о том, что сигналы об изменениях инвестиционной активности в сообщениях появляются гораздо раньше, чем эти изменения находят отражение в данных Росстата. Целью работы является построение опережающего индекса инвестиционной активности в России, рассчитанного на основе оценки динамики новостей, который статистически значимо коррелирует с данными Росстата.

2. Материалы и методы

В качестве источника информации использована база данных, которая содержит более 11,4 миллионов сообщений из 691 российского ТГ-канала, с историей существования не менее 12 кварталов. Анализ охватывает период с сентября 2015 по сентябрь 2023 года. Пул каналов сформирован на основе каталога tgstat.ru². Характеристики базы данных приведены в табл. 1.

Таблица 1

Структура базы данных Telegram-сообщений за весь период наблюдения (с сентября 2015 по сентябрь 2023)

Категории (по tgstat.ru)	Количество каналов	Общее количество сообщений в канале (тыс.)			
		Среднее	Медиана	Минимум	Максимум
Экономика (ecn)	386	8,6	3,8	0,4	253,8
Новости и СМИ (nws)	131	40,7	23,6	1,1	216,5
Политика (plt)	89	21,6	15,8	2,5	160,1
Криптовалюты (cpr)	44	11,4	5,7	0,9	62,3
Бизнес и стартапы (bsn)	29	7,9	4,3	0,8	82,9
Блоги (blg)	12	12,9	4,1	0,5	99,1
ВСЕГО	691	16,6	6,9	0,4	253,8

Для предварительной оценки «информационного наполнения» базы данных проанализировано окружение слова «инвест»³. Облако слов

²Полное описание базы данных в работе А. Гусевой, А. Полехиной «Построение индекса восприятия Банка России: анализ данных телеграмм-каналов с использованием методов машинного обучения» готовится к публикации. Список каналов, подготовленный указанными авторами, размещен на ресурсе http://github.com/Alisa-llm/tg_channels_cbr (дата обращения: 14.10.2024).

³Проанализированы и размечены тексты, содержащие в себе строку «инвест». Для анализа контекста слова «инвест» выбрано ± 5 слов, окружающих ключевое слово. Для построения «Облака слов» была проведена лемматизация.

(рис. 1) продемонстрировало, что в сообщениях речь идет как о реальных инвестициях (регион, отрасль, программа, проект, инфраструктура, строительство, новостройка, недвижимость, стартап, производство), так и о финансовых (фонд, облигация, биржа, ценная бумага и т. д.). Следовательно, выбранная база данных может служить источником информации об инвестиционной активности предприятий.

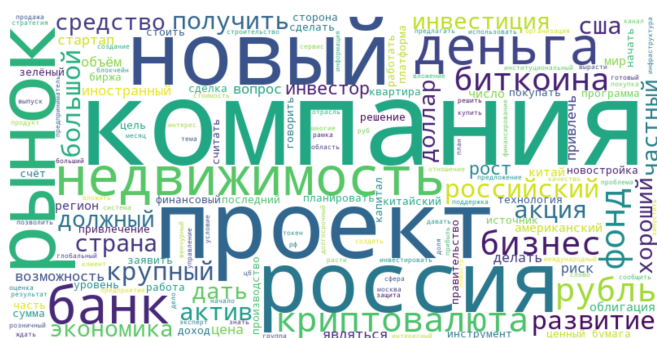


Рис. 1. Облако лемматизированных слов, построенное из окружения «инвест»
+/- 5 слов

Для поиска текстов о фактах или факторах инвестиционной активности применялись два подхода, первый основан на использовании фильтра по ключевым словам, второй — на языковой модели машинного обучения BERT [8]. Оба подхода использовали один и тот же набор данных.

В качестве целевого показателя в работе используется динамика инвестиций в основной капитал (ИОК) за период с 2002 по 2023 год по данным Росстата. Данные после сезонного сглаживания методом X-13 ARIMA/SEATS приведены к базисному виду к 4 кв. 2019 года.

Для построения индексов инвестиционной активности по новостному фону и оценки их качества используются методы корреляционно-регрессионного анализа, метод главных компонент [9; 10]. Для построения моделей, оценивания параметров, проверки гипотез используется период с I кв. 2018 года по IV кв. 2022 года и уровень значимости $p = 0,05$, если не оговорено иное. Оценка прогнозных свойств индексов осуществляется на периоде I – IV кв. 2023 года с помощью средней абсолютной ошибки MAE и среднеквадратической ошибки RMSE.

Исследование проводилось при помощи статистического пакета Eviews 11, библиотек Python PyTorch, Transformers, Scikit-learn, Wordcloud.

3. Результаты исследования и обсуждение

Фильтр по ключевым словам. Наивный индекс инвестиционной активности

Фильтр по ключевым словам широко применяется в экономической литературе для оценки экономической ситуации. Так авторами в работе [2] был предложен индекс неопределенности (IPU) как доля статей, содержащих отдельные выражения, из категорий «экономика», «неопределенность» и «политика» в общем количестве статей. Расчет индекса (IPU) осуществляется по множеству развитых и развивающихся стран [3–5].

В работе при построении наивного индекса инвестиционной активности применялся фильтр по ключевому слову «инвест». Как показало «облако слов», фильтруя таким способом сообщения, отбираются данные, касающиеся инвестиций как в основной капитал (реальных инвестиций), так и финансовых инвестиций. В данном подходе мы полагаемся на тесную связь реальных инвестиций с финансовыми вложениями.

Для каждого Telegram-канала с выбранной периодичностью (день, месяц или квартал) подсчитывается доля постов, содержащих ключевое слово. В рамках исследования была выбрана квартальная частота данных. Далее, полученные доли для каждого источника нормализуются к нулевому среднему и на одно стандартное отклонение на всем периоде. На последнем шаге конструируется сводный TG-index как среднее значение нормированных временных рядов⁴:

$$\text{TG-index}_t = \frac{1}{N} \sum_{j=1}^N \frac{\nu_{j,t} - m_j}{s_j}, \quad (1)$$

где $\nu_{j,t}$ — доля постов в j -ом канале за период t с ключевым словом, m_j — средняя доля постов на j -ом канале на всем периоде, s_j — стандартное отклонение доли постов на j -ом канале на всем периоде.

Построенный сводный индекс (TG-index) и базисный индекс инвестиций в основной капитал (ИОК) представлены на рис. 2. Согласно расширенному тесту Дики – Фуллера (ADF), оба временных ряда являются интегрированными первого порядка $I(1)$. Тест Энгла – Грэнджера (Engle – Granger test) не обнаружил статистически значимой коинтеграции рядов с линейным трендом.

⁴Исключая выбросы в данных.

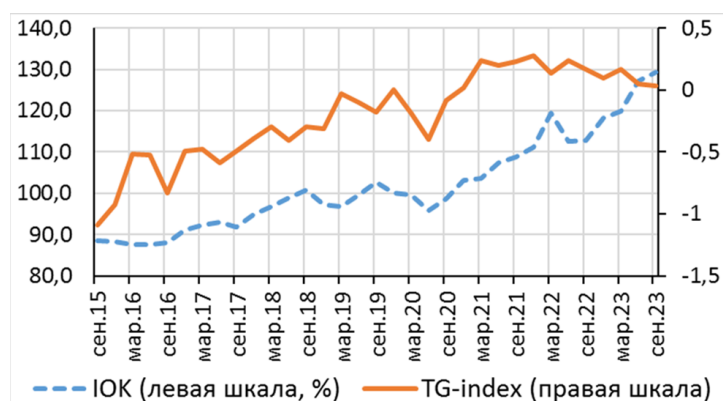
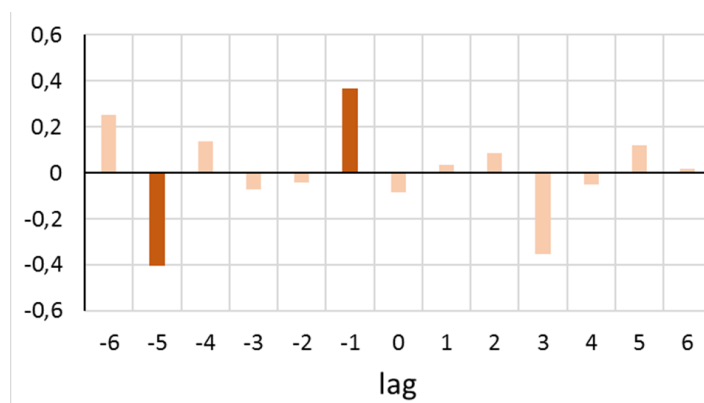


Рис. 2. Динамика TG-index и ИОК

На основе анализа кросс-коррелограммы рядов в первых разностях (рис. 3)⁵ для опережающей оценки инвестиционной активности предлагаем использовать следующий индекс:

$$\text{NAIVE} = \Delta[\text{TG-index}(-1) - \text{TG-index}(-5)]. \quad (2)$$

Далее мы будем называть его наивным индексом инвестиционной активности (NAIVE). Индекс NAIVE статистически значимо коррелирует с динамикой инвестиций ($\Delta \log(\text{ИОК})$) на уровне 0,48 ($p\text{-value} = 0,03$) и по построению является опережающим с лагом в I квартал.

Рис. 3. Кросс-коррелограмма $\Delta\text{TG-index}$ и $\Delta \log(\text{ИОК})$

⁵Для ИОК рассматривается первая разность логарифмов базисного индекса, которая интерпретируется как темп прироста инвестиций за квартал.

Проанализировав динамику наивного индекса по различным типам каналов и применив метод главных компонент, мы достигли более высокой корреляции на выборке из политических (plt), новостных (nws) каналов и блогов (blg) для следующей модификации индекса:

$$\begin{aligned} \text{NAIVE}_{\text{md}} = & 0,5 \cdot \text{NAIVE}(\text{plt}, \text{nws}, \text{blg}) - \\ & - 0,4 \cdot \text{NAIVE}(\text{nws}) - 0,07 \cdot \text{NAIVE}(\text{plt}), \end{aligned} \quad (3)$$

где NAIVE (пул каналов) — наивный индекс, построенный по соответствующей выборке каналов. Корреляция модифицированного наивного индекса с динамикой инвестиций составила 0,72 (p-value < 0,001).

Для оценки прогнозных свойств модифицированного индекса построена линейная регрессия и проверено ее качество на тестовой выборке с I кв. 2023 по IV кв. 2023 года. Сравнение прогнозируемых результатов с фактическими данными (рис. 4) демонстрирует, что модифицированный наивный индекс может выступать в качестве опережающего прогнозного индекса инвестиционной активности. Так, значения ошибок прогнозирования модели MAE и RMSE составили 0,7 п.п. и 0,9 п.п. соответственно, что ниже, чем при использовании авторегрессионной модели AR(1) (1,7 п.п. и 2,6 п.п.)⁶.



Рис. 4. Фактические и модельные значения темпа прироста ИОК (квартал к кварталу)

⁶Была использована модель AR(1) со структурным сдвигом в I кв. 2009 г.

В рамках исследования возник вопрос, насколько влияет выбранный пул каналов на качество индекса. Для этого случайно выбирали 80 % каналов каждой категории и строили по ним индекс. Результаты показали, что для наивного индекса интервал изменения корреляций небольшой, что говорит об устойчивости к ротации каналов. Корреляция модифицированного индекса демонстрирует его более сильную чувствительность к выбранному пулу каналов, что требует дальнейшего исследования (табл. 2, рис. 5).

Таблица 2

Характеристики распределения коэффициентов корреляции наивных индексов и $\Delta \log(\text{ИОК})$

	mean	std	10 %	25 %	50 %	75%	90 %
NAIVE	0,47	0,03	0,43	0,45	0,47	0,49	0,51
NAIVE _{md}	0,61	0,09	0,49	0,55	0,62	0,67	0,72

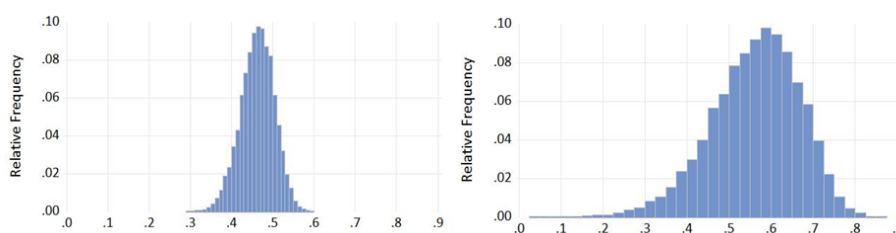


Рис. 5. Распределение коэффициентов корреляции индексов с $\Delta \log(\text{ИОК})$: слева — наивный индекс, справа — модифицированный

Применение языковой модели «BERT»

Поскольку фильтр не может учесть контекст сообщений, было решено провести эксперимент с языковыми моделями, которые позволяют извлекать информацию из неструктурированных данных. Анализ базы данных и разметка части сообщений показали, что потребуются языковая модель, учитывающая сложный контекст сообщений: при осуществлении разметки специалисты по-разному оценивали сообщения как с точки зрения отнесения их к реальным инвестициям, так и с точки зрения определения тональности. Мы остановили свой выбор на языковой модели BERT, разработанной в Google AI Language [8]. Архитектура модели BERT позволяет изучать контекст слова исходя из его окружения в целом и достигает более глубокого понимания текста.

В рамках этого подхода была проведена разметка новостных сообщений в двух разрезах — страна / реальные инвестиции (табл. 3). Первый — относится ли сообщение к России. Второй — отражает ли сообщение факторы и факты осуществления инвестиционной деятельности. Разметка для второй категории (реальные инвестиции) имеет сильный дисбаланс классов, который при обучении корректировался подбором весов для каждого класса.

Таблица 3

Данные по разметке текста (количество сообщений)

	Россия			Реальные инвестиции		
	TRUE	FALSE	Всего	TRUE	FALSE	Всего
Размер обучающей выборки (Train)	2188	940	3128	565	2624	3189
Размер тестовой выборки (Test)	384	176	560	113	453	566
Всего	2572	1116	3688	678	3077	3755

Для классификации была выбрана языковая модель «rubert-tiny2» (Hugging Face cointegrated/rubert-tiny2) [11; 12]. Работает только с русским языком, поддерживает последовательности до 2048 токенов и имеет большой словарный запас 83 тыс. токенов. Модель «rubert-tiny2» имеет классификационный слой нейронов, который генерирует 312 выходных параметров для классификации. Для выполнения задач бинарной классификации в разрезе страна / реальные инвестиции был добавлен выходной слой с двумя нейронами (листинг 1)⁷.

Листинг 1

Подготовка модели

```
# Импорт библиотек
import torch
from transformers import BertTokenizer, BertForSequenceClassification
from torch.utils.data import Dataset, DataLoader
from transformers import AdamW, get_linear_schedule_with_warmup

# Загрузка модели
model_type = 'cointegrated/rubert-tiny2'
model = BertForSequenceClassification.from_pretrained(model_type)
```

⁷Исходный код размещен: https://github.com/PetrBorkov/invest_activity (дата обращения: 14.10.2024).

Окончание листинга 1

```
# загрузка токенизатора
tokenizer = BertTokenizer.from_pretrained(model_type)

# Замена классификационного слоя
out_features = model.bert.encoder.layer[1].output.dense.out_features
model.classifier = torch.nn.Linear(out_features, n_classes)

# Перемещение модели на выбранное устройство (GPU или CPU)
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
model.to(device)
```

Использовался оптимизатор AdamW, в котором задавался относительно низкий learning rate (проводились эксперименты с весами от $7e-6$ до $2e-5$), так как модель необходимо дообучить. Наилучший результат показало обучение на 15 эпохах с использованием линейного планировщика обучения (линейное снижение learning rate), заданием весов класса (для второй модели — это $[0,15; 1]$). Learning rate на уровне $7e-6$ в сочетании с планировщиком позволяет не допускать быстрого переобучения (листинг 2). Результаты обучения приведены в табл. 4 и на рис. 6.

*Листинг 2***Подготовка данных и настройка параметров**

```
# Создание обучающего/валидационного DataLoader
train_set = CustomDataset(X_train, y_train, tokenizer)
valid_set = CustomDataset(X_valid, y_valid, tokenizer)
train_loader = DataLoader(train_set, batch_size=2, shuffle=True)
valid_loader = DataLoader(valid_set, batch_size=2, shuffle=True)

# Инициализация оптимизатора AdamW и планировщика обучения
epochs = 15 # количество эпох обучения
optimizer = AdamW(model.parameters(), lr=7e-6, correct_bias=False, \
                  no_deprecation_warning=True)
scheduler = get_linear_schedule_with_warmup(
    optimizer,
    num_warmup_steps=len(train_loader) * 2,
    num_training_steps=len(train_loader) * epochs
)
class_weights = torch.tensor([0.15, 1.0]) # веса классов
#инициализация функции потерь
loss_fn = torch.nn.CrossEntropyLoss(weight=class_weights).to(device)
```

Таблица 4

Результаты обучения моделей

	Россия (1, 0)	Реальные инвестиции (1, 0)
F1-мера	0,900	0,731
Precision	0,938	0,800
Recall	0,940	0,673

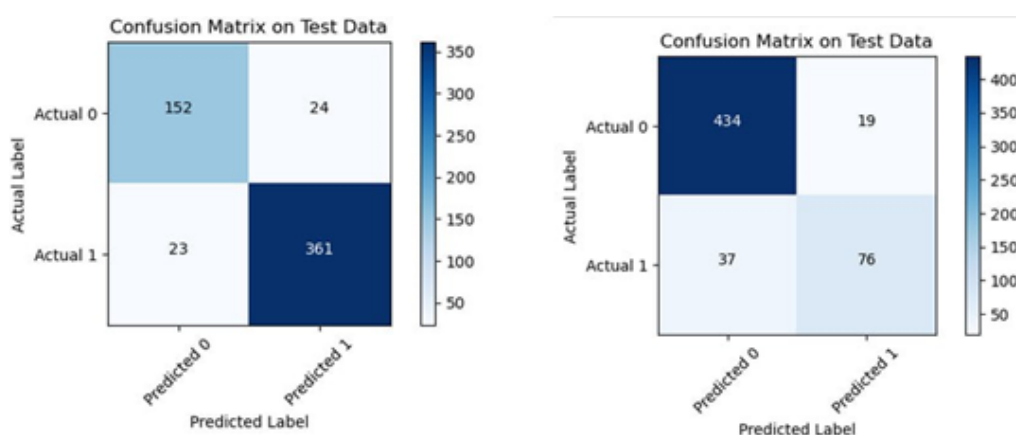


Рис. 6. Матрица ошибок классификации новостей по принадлежности к России (слева), реальным инвестициям (справа)

С помощью обученной модели произведена разметка базы и для каждого канала подсчитана доля сообщений, относящихся к теме «Инвестиционная активность в России». Далее был рассчитан сводный индекс TGbert по всей базе аналогично тому, как это было сделано при построении наивного индекса. Анализ TGbert-index, рассчитанный по всем каналам, не выявил в его динамике статистически значимых корреляций с динамикой инвестиций.

При анализе динамики построенного индекса в разрезе категорий каналов и применении метода главных компонент была выявлена статистически значимая корреляция на выборке из политических (plt), новостных (nws), экономических (ecn) и криптовалютных (crp) каналов для следующего показателя:

$$\text{BERT-index} = \Delta[2 \times \text{TGbert}(\text{plt}) + \text{TGbert}(\text{crp}) - \text{TGbert}(\text{ecn}) - 1,5 \times \text{TGbert}(\text{nws})](-1). \tag{4}$$

Как и наивный индекс, BERT-index является опережающим с лагом в I квартал и коррелирует с инвестициями ($\Delta \log(\text{IOK})$) на уровне 0,68 ($p\text{-value} = 0,001$). Полученный результат сопоставим с результатами модифицированного индекса, но уступает ему по качеству прогноза: ошибки прогнозирования MAE и RMSE составили 1,4 п.п. и 1,9 п.п. При проверке на устойчивость к ротации каналов BERT-index продемонстрировал чувствительность, аналогичную той, что наблюдалось для модифицированного индекса (табл. 5).

Таблица 5

Характеристики распределения коэффициентов корреляции Bert-index и $\Delta \log(\text{IOK})$

	mean	std	10 %	25 %	50 %	75%	90 %
BERT-index	0,61	0,08	0,51	0,56	0,62	0,67	0,71

4. Заключение

В исследовании подтверждается предположение о том, что в информационном поле сигналы об изменениях инвестиционной активности проходят раньше, чем отражаются в данных официальной статистики. Модифицированный вариант индекса (NAIVEmd) может использоваться в качестве опережающего индекса при построении краткосрочного прогноза динамики инвестиций в основной капитал. Наивный индекс инвестиционной активности (NAIVE) показал устойчивость к набору каналов и может рассматриваться в качестве своего рода «benchmark» при оценке ситуации. Результаты, основанные на архитектуре BERT и на том же наборе данных, что и для наивного индекса, не продемонстрировали существенного улучшения качества индексов. Возможно, это обусловлено высокой чувствительностью к качеству разметки. При разметке текста понятие «инвестиционная активность» трактовалось достаточно широко: в сообщениях учитывались не только факты, а еще факторы, оказывающие влияние на инвестиционную активность.

Список источников

1. Регламент оценки, корректировки и публикации данных статистического наблюдения за строительством и инвестициями в основной капитал. Утвержден приказом Росстата от 26 сен-

- тября 2016 г. № 544 // СПС «КонсультантПлюс». URL: <https://www.consultant.ru> (дата обращения: 14.10.2024).
2. **Baker S. R., Bloom N., Davis S. J.** Measuring Economic Policy Uncertainty // *Quarterly Journal of Economics*. 2016. Vol. 131 (4). Pp. 1593–1636. DOI: 10.1093/qje/qjw/024.
 3. **Cerda R., Silva A., Valente J. T.** Impact of economic uncertainty in a small open economy: the case of Chile // *Applied Economics*. 2018. Vol. 50. No 26. Pp. 2894–2908. URL: <https://arxiv.org/pdf/1810.04805> (дата обращения: 14.10.2024).
 4. **Zalla R.** Economic Policy Uncertainty in Ireland // *Atlantic Economic Journal*. 2017. Vol. 45 (2). Pp. 267–271. DOI: 10.1007/s11293-017-9536-8.
 5. **Петрова Д., Трунин П.** Оценка уровня неопределенности экономической политики // *Деньги и кредит*. 2023. № 82 (3). С. 48–61.
 6. **Яковлева К.** Оценка экономической активности на основе текстового анализа // *Деньги и кредит*. 2018. № 77 (4). С. 26–41. DOI: 10.31477/rjmf.201804.26.
 7. **Колужнов Д. В., Колужнов Е. Д., Ляхнова М. В.** Учет информационного фона в DSGE-модели экономики России с адаптивным обучением // *Мир экономики и управления*. 2023. Т. 23 (4). С. 60–82. DOI: 10.25205/2542-0429-2023-23-4-60-82.
 8. **Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova.** BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2019. Vol. 1. Long and Short Papers. Pp. 4171–4186, Minneapolis, Minnesota. URL: <https://aclanthology.org/N19-1423/> (дата обращения: 14.10.2024).
 9. **Носко В. П.** Эконометрика : в 2 кн. М.: Издательский дом «Дело» РАНХиГС, 2021. Кн. 1. 704 с. (Академический учебник).
 10. **Мхитарян В. С. и др.** Эконометрика : учебник / под ред. д-ра экон. наук, проф. В. С. Мхитаряна. М.: Проспект, 2009. 384 с.

11. **Dale David.** Маленький и быстрый BERT для русского языка. June, 2021. URL: <https://habr.com/ru/post/562064> (дата обращения: 14.10.2024).
12. **Dale David.** Рейтинг русскоязычных энкодеров предложений. June, 2022. URL: <https://habr.com/ru/articles/669674> (дата обращения: 14.10.2024).

References

1. Regulations for assessments, adjustments and publication of statistical observation data on construction and investments in fixed capital. The order of Rosstat dated 26 September 2016 No 544. *SPS „Konsul'tant Plyus“* [SPS „ConsultantPlus“ was approved]. Available at: <https://www.consultant.ru> (accessed: 10/14/2024). (In Russ.)
2. **Baker S. R., Bloom N., Davis S. J.** Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics*. 2016. Vol. 131 (4). Pp. 1593–1636. DOI:10.1093/qje/qjw/024.
3. **Cerda R., Silva A., Valente J. T.** Impact of economic uncertainty in a small open economy: the case of Chile. *Applied Economics*. 2018. Vol. 50. No 26. Pp. 2894–2908. Available at: <https://arxiv.org/pdf/1810.04805> (accessed: 14.10.2024).
4. **Zalla R.** Economic Policy Uncertainty in Ireland. *Atlantic Economic Journal*. 2017. Vol. 45 (2). Pp. 267–271. DOI: 10.1007/s11293-017-9536-8.
5. **Petrova D., Trunin P.** Assessment of the level of uncertainty of economic policy. *Den'gi i kredit* [Money and credit]. 2023. No 82 (3). Pp. 48–61. (In Russ.)
6. **Yakovleva K.** Assessment of economic activity based on text analysis. *Den'gi i kredit* [Money and credit]. 2018. No 77 (4). Pp. 26–41. DOI: 10.31477/rjmf.201804.26. (In Russ.)
7. **Kolyuzhnov D. V., Kolyuzhnov E. D., Lyakhnova M. V.** Taking into account the information background in the DSGE model of the

Russian economy with adaptive learning. *Mir ekonomiki i upravleniya* [World of Economics and Management]. 2023. Vol. 23 (4). Pp. 60–82. DOI: 10.25205/2542-0429-2023-23-4-60-82. (In Russ.)

8. **Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova.** BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2019. Vol. 1. Long and Short Papers. Pp. 4171–4186, Minneapolis, Minnesota. Available at: <https://aclanthology.org/N19-1423/> (accessed: 14.10.2024).
9. **Nosko V. P.** *Ekonometrika : v 2 kn.* [Econometrics : in 2 books]. Moscow: Publishing House „Delo“ RANEPА, 2021. Book 1. 704 p. (Academic textbook). (In Russ.)
10. **Mkhitaryan V. S. et al.** *Ekonometrika : uchebnik* [Econometrics : textbook]. Ed. Doctor of Economics sciences, prof. V. S. Mkhitaryan. Moscow: Prospekt, 2009. 384 p. (In Russ.)
11. **Dale David.** *Malen'kiy i bystryy BERT dlya russkogo yazyka* [Small and fast BERT for the Russian language]. June, 2021. Available at: <https://habr.com/ru/post/562064> (accessed: 14.10.2024).
12. **Dale David.** *Reyting russkoyazychnykh enkoderov predlozheniy* [Rating of Russian-language sentence encoders]. June, 2022. Available at: <https://habr.com/ru/articles/669674> (accessed: 14.10.2024).

Сведения об авторах / Information about authors

Борков Петр Викторович / Petr V. Borkov

руководитель направления / Head of Workstream

Отделение — Национальный банк по Республике Коми Северо-Западного главного управления Центрального банка Российской Федерации / Division — National Bank of the Republic of Komi of the North-Western Main Branch of the Central Bank of the Russian Federation

167000, Россия, г. Сыктывкар, ул. Ленина, 53 / 53, Lenin Street, Syktvkar, 167000, Russia

Мальцева Ольга Аминджановна / Olga A. Maltseva

ведущий экономист / Lead Economist

Отделение — Национальный банк по Республике Коми Северо-Западного главного управления Центрального банка Российской Федерации / Division — National Bank of the Republic of Komi of the North-Western Main Branch of the Central Bank of the Russian Federation

167000, Россия, г. Сыктывкар, ул. Ленина, 53 / 53, Lenin Street, Syktyvkar, 167000, Russia

Полякова Ирина Владимировна / Irina V. Polyakova

главный экономист / Chief Economist

Отделение — Национальный банк по Республике Коми Северо-Западного главного управления Центрального банка Российской Федерации / Division — National Bank of the Republic of Komi of the North-Western Main Branch of the Central Bank of the Russian Federation

167000, Россия, г. Сыктывкар, ул. Ленина, 53 / 53, Lenin Street, Syktyvkar, 167000, Russia

Старцева Евгения Николаевна / Evgenija N. Startseva

старший преподаватель кафедры прикладной математики и компьютерных наук / senior lecturer Department of Applied Mathematics and Computer Science

Сыктывкарский государственный университет имени Питирима Сорокина / Pitirim Sorokin Syktyvkar State University

167001, Россия, г. Сыктывкар, Октябрьский пр., 55 / 55, Oktyabrsky Ave., Syktyvkar, 167001, Russia

Статья поступила в редакцию / The article was submitted 06.11.2024

Одобрена после рецензирования / Approved after reviewing 13.11.2024

Принята к публикации / Accepted for publication 20.11.2024