

# Прогнозирование инфляции с использованием новостных индексов

Волгина Елизавета, э-401

### Актуальность:

1. Высокая инфляция по всему миру (Европа: октябрь 2022 г. 11,5 % г/г, Россия: апрель 2022 г. 17,83 %). (народнохозяйственная цель)
2. Уровень инфляции важен для домохозяйств (контракты в номинальных величинах ) и является одним из основных таргетов при принятии решений центральными банками.
3. Публикуя инфляционные прогнозы, центральные банки могут формировать инфляционные ожидания населения
4. Два способа учёта инфляционных ожиданий при прогнозировании инфляции (через опросы и фондовый рынок) имеют недостатки

### Мотивация:

1. Новостные индексы могут повысить предсказательную силу прогнозных моделей за счёт учета «инсайтов» экономического развития, которые сложно уловить в других экономических переменных (Kalamara et al., 2022, Drinth, 2023), что будет способствовать получению более точного прогноза инфляции и проведению более эффективной денежно-кредитной политики.
2. Новости отражают информацию, важную при учёте инфляции, но которую сложно учесть через другие переменные. Например, на цены продуктов питания, в частности томатов, лука и картофеля, влияют экстремальные погодные условия (засуха, сильные дожди, холод), которые отражаются в новостях и влияют на инфляцию (Bhanu and Abhishek, 2022)
3. Новостные индексы позволят строить оперативные прогнозы внутри периода, так как новости обновляются каждый день, в то время как стандартные модели опираются на показатели квартальной или месячной частотности (Barbaglia et al., 2023).
4. Модели, основанные на новостных индексах, могут давать более точные прогнозы в период стресса и кризиса экономики (Kalamara et al., 2022, Drinth, 2023), что особенно важно в последние волатильные годы.
5. Использование новостных индексов может решить проблему неровного края, на которую указывал Bayuza (2018), возникающую при прогнозировании в реальном времени из-за разницы в публикации макроэкономических данных

Цель работы — определить, улучшает ли использование новостных индексов в моделях машинного обучения точность прогнозов инфляции.

Объект исследования — новостные индексы для целей прогнозирования инфляции.

Предмет исследования — предсказательная сила моделей прогнозирования инфляции методами машинного обучения с использованием новостных индексов.

Исследовательский вопрос — помогают ли новостные индексы в моделях машинного обучения повысить точность прогнозов инфляции?

Задачи:

1. На основе анализа литературы выявить методы и модели машинного обучения, позволяющие наилучшим образом прогнозировать инфляцию.
2. Построить новостные индексы для целей прогнозирования инфляции.
3. Осуществить прогноз инфляции с использованием построенных новостных индексов.
4. Сопоставить точность прогнозов моделей, использующих и не использующих новостные индексы. На основе полученных результатов сформулировать рекомендации по методике прогнозирования инфляции

Выдвигаемая гипотеза: использование новостных индексов в моделях машинного обучения повышает качество прогнозов инфляции.

Новизна работы заключается в том, что я использую новостные индексы для непосредственного прогнозирования инфляции, а также применяю новостные индексы в моделях машинного обучения

# Обзор методов определения тональностей новостей

- Для составления лексикона можно основываться на словаре Loughran and McDonald (2011), адаптируя его к своему корпусу текстов (Dorinth et al., 2023, Bhanu and Abhishek, 2022, Rambaccussing and Kwiatkowski, 2020)
- Эффективным способом агрегирования тональности отдельных слов в тональность всего документа является подход, использованный в Thorsrud (2020), основанный на взвешенной сумме скоров и словах-переключателях.
- Наилучшей предобученной моделью классификации текстов для целей определения тональности новостей оказывается модель BERT. (Shapiro et al., 2022)
- Практически эффективным методом векторизации текстов является TF-IDF. (Rambaccussing and Kwiatkowski, 2020, Kalamara, 2022, Larsen and Thorsrud, 2022, Евстигнеева и Карпов, 2023)
- Практически используемой моделью ML для определения тональности новостей на основе размеченного корпуса текстов является LSVM (Linear Support Vector Machine) (Rambaccussing and Kwiatkowski, 2020)

# Обзор методов построения новостных индексов и моделей прогнозирования инфляции методами ML

## Построение новостных индексов

- Распространенная модель тематического моделирования — LDA (Евстигнеевой и Карпова, 2023, Dorinth and Winter, 2023, Thorsrud, 2020)
- Хорошие результаты показывает подход, комбинирующий тематическое моделирование и определение тональности новостей (Thorsrud, 2020)

## Модели прогнозирования инфляции с использованием новостных индексов

- Сложно построить стандартные эконометрические модели, которые бы улучшали предсказательную силу простых моделей в краткосрочном периоде, таких как случайное блуждание (Atkeson- Ohanian, 2001) и модель изменяющихся во времени главных компонент (МА с изменяющимися параметрами (Stock and Watson, 2007) ) (Stock and Watson, 2010)
- Основные модели, показавшие высокую точность при прогнозировании инфляции, — случайный лес (RF) (Medeiros et al., 2021) , модель долгосрочной краткосрочной памяти (LSTM) (Barkan et al., 2023, Almosova and Andresen, 2019)

# Данные

- Данные по новостям были получены с сайта РИА-Новости за период с 01.01.2015 по 30.01.2024 (по данным Медиалогии являлся лидером по цитируемости в СМИ и соцсетях в 2023 году). Всего было собрано 190 978 новостных статей. (Можно брать один наиболее популярный местный источник Drinth (2023))
- Лемматизация, токенизация и удаления стоп-слов. Лист стоп-слов был взят из библиотеки NLTK, согласно работам Евстигнеевой и Карпова (2023), Kalamara et al. (2022), а также дополнен рядом слов, характерных для получившегося корпуса ( «год», «gia», "тысяча" , "миллион" , "миллиард", некоторые имена и фамилии)
- Индекс потребительских цен (ИПЦ) (Росстат, в процентах на конец месяца к концу предыдущего месяца)
- Дополнительные регрессоры (?) (Pavlov, 2020):
  - индекс физического объема ВВП (реальный ВВП), производительность труда (отношение реального ВВП к числу занятых), агрегат денежной массы M2, объем выданных кредитов в реальном выражении, норма безработицы, экспорт в реальном выражении, цена на нефть в долларах США, реальный располагаемый доход, процентная ставка денежного рынка, ИПЦ
  - помесячные процентные изменения

# Определение тональности новостей на основе лексикона

- Наиболее гибкий и вычислительно удобный, по сравнению со всеми другими способами (Algaba et al., 2020).
- Базовый словарь -- модификация словаря Loughran and McDonald (2011), построенная в работе Dorinth and Winter (2023), перевод с нидерландского + добавление важных слов («санкция», «сво») (после обработки из 1 620 слов получилось 1323)
- Алгоритм определения тональности всего документа из Bhanu and Abhishek (2022)
- Файл с словами-переключателями был получен из пакета для R "sentometrics» (Ardia et al., 2021). Файл содержал 132 слова с соответствующим каждому слову коэффициенту. Я перевела полученный файл на русский язык и скорректировала, получила файл с 89 словами и соответствующими им коэффициентами.

# Алгоритм определения тональности новости

1) В день  $t$  для каждого слова  $i$  из новости  $d_n$ , которое встречается в словаре, определяется соответствующая тональность (+1 или -1):

$$S_{int}$$

2) Для каждого слова со скором выделяется его контекст: 4 слова до и 2 слова после. Если какое-то из контекстных слов содержится в списке со словами-переключателями, то тональность слова домножалась на соответствующий коэффициент:

$$v_i * S_{int}$$

3) Скор для всей новости определялся как сумма скоров, определенных на первых шагах, скорректированных на общее количество слов в новости.

$$S_{nt} = \frac{1}{w_d} \sum_{i=1}^{Q_d} v_i * S_{int}$$

4) На последнем шаге скоры каждой новости агрегируются в скор для конкретного дня ( $NSS_t$ ), просто суммируя скоры каждой новости.

$$NSS_t = \sum_{n=1}^{N_t} S_{nt}$$

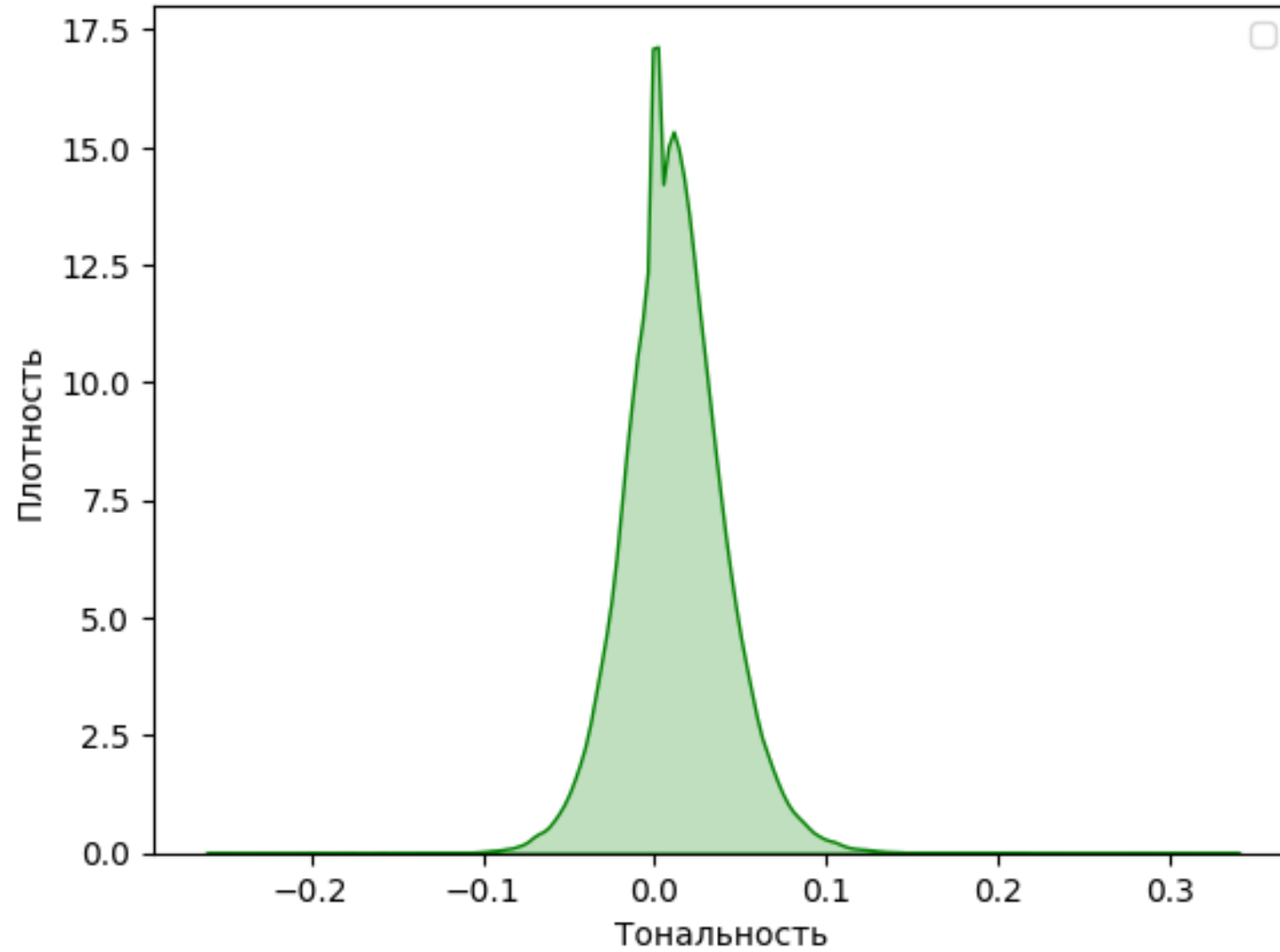
$N_t$  — общее количество новостей в день  $t$

$Q_d$  -- общее количество слов из словаря в новости  $d_n$

$w_d$  — общее количество слов всего в новости

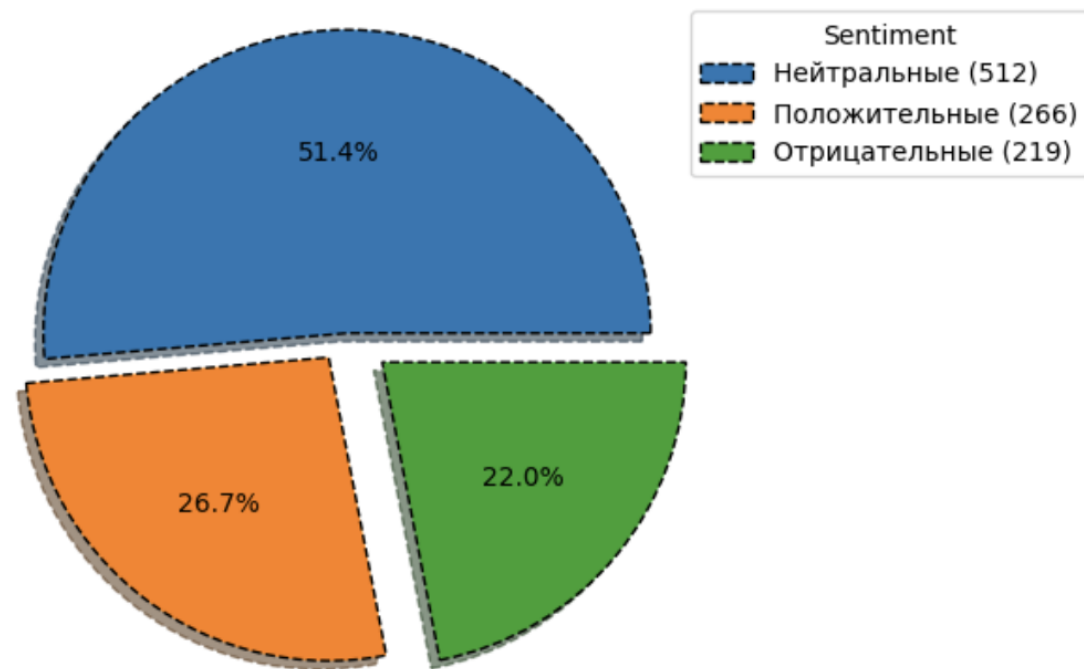


Кривая плотности распределения тональностей по лексикону



# Определение тональности новостей на основе моделей ML (1)

- Была размечена 1 000 новостей, которая использовалась в качестве обучающей выборке для моделей ML
- Негативной новостью считалась та, в которой описывались проинфляционные события (введение санкций, снижение курса рубля, рост социальных выплат, снижение ключевой ставки и т. д.)
- 700 новостей было использовано в качестве обучающей выборки и 300 — в качестве тестовой.



## Определение тональности новостей на основе моделей ML (2)

- Было построено 4 модели наиболее популярные для целей классификации — логистическая регрессия (Logistic Regression) и линейный метод опорных векторов (Linear Support Vector Machine)
- Для представления текстов в числовой вид был использован метод TF-IDF и предобученная модель нейронных сетей RuBert.

Модель	Метрика	Отрицательная	Нейтральная	Положительная
TF-IDF + Logreg	Precision	0.36	0.56	0.50
	Recall	0.14	0.84	0.33
	F1-score	0.20	0.67	0.39
	Accuracy	0.53	0.53	0.53
TF-IDF + LSVM	Precision	0.33	0.56	0.45
	Recall	0.18	0.75	0.35
	F1-score	0.23	0.64	0.39
	Accuracy	0.50	0.50	0.50
RuBert + Logreg	Precision	0.53	0.61	0.45
	Recall	0.46	0.73	0.35
	F1-score	0.50	0.66	0.40
	Accuracy	0.56	0.56	0.56
RuBert + LSVM	Precision	0.42	0.56	0.35
	Recall	0.39	0.62	0.31
	F1-score	0.41	0.59	0.33
	Accuracy	0.48	0.48	0.48

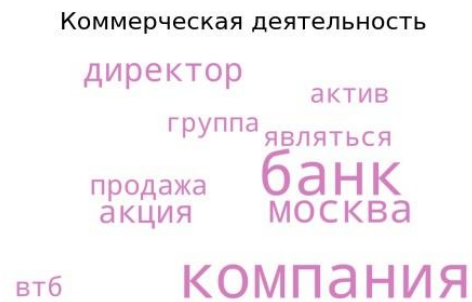
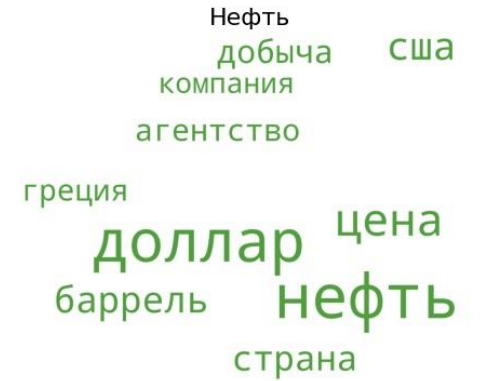
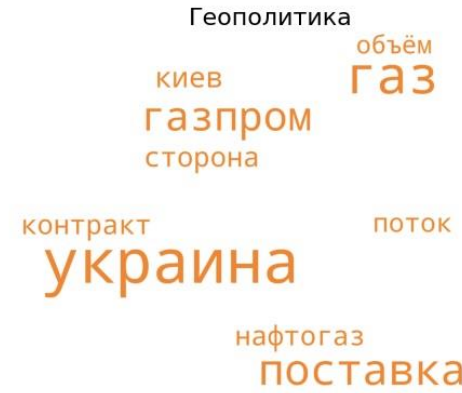
Таблица 1: Метрики качества моделей обучения с учителем для определения тональности новостей

*in progress....*

# Тематическое моделирование (TF-IDF + LDA)

Критерии:  
осмысленность и  
когерентность

№ темы	Название темы
Тема 1	Санкции
Тема 2	Геополитика
Тема 3	Нефть
Тема 4	Инфляция
Тема 5	Курс
Тема 6	Форумы/встречи
Тема 7	Коммерческая деятельность
Тема 8	Региональное развитие
Тема 9	Федеральные программы



# Дальше

- Определить тональность тем (9 временных рядов)
- Построить модель (точно LSTM и .....
- Сравнить с .....

# Ограничения исследования

- Только один новостной источник
- Много спорных моментов при ручной разметке (например, инвестиции в строительство нового предприятия или развитие региона)
- Будет дополняться

# Литература (1)

- 1) Евстигнеева А., Карпов Д. (2023) Влияние негативных новостей на восприятие инфляции населением // Банк России: серия докладов об экономических исследованиях. № 111.
- 2) Andres-Escayola, E., Ghirelli, C., Molina, L. *et al.* (2023) Using Newspapers for Textual Indicators: Guidance Based on Spanish- and Portuguese-Speaking Countries. *Comput Econ.*
- 3) van Dijk, Dorinth and de Winter, Jasper, (February 17, 2023). Nowcasting GDP using tone-adjusted time varying news topics: Evidence from the financial press .De Nederlandsche Bank Working Paper No. 766
- 4) Dooruj Rambaccussing, Andrzej Kwiatkowski (2020), Forecasting with news sentiment: Evidence with UK newspapers, *International Journal of Forecasting* , 36 (4), pp. 1501-1516, ISSN 0169-2070
- 5) Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S.(2022). Making text count: Economic forecasting using newspaper text. *Journal of Applied Econometrics*, 37(5), 896–919.
- 6) Pratap, Bhanu and Ranjan, Abhishek (June 1, 2022), Forecasting Food Inflation using News-based Sentiment Indicators . RBI-Occasional Papers-Vol. 42, No.2, 2021
- 7) Shiller, Robert J. 2017. "Narrative Economics." *American Economic Review*, 107 (4): 967-1004.
- 8) Carroll, C.D., 2003. Macroeconomic expectations of households and professional forecasters. *the Quarterly Journal of economics*, 118(1), pp. 269–298.
- 9) Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. "Rational Inattention: A Review." *Journal of Economic Literature*, 61 (1): 226-73.
- 10) Harris, Z. S., 1954. Distributional structure. *Word*, 10(2-3), pp. 146–162.
- 11) Salton, G. & Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), pp. 513–523.
- 12) Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022.
- 13) Luca Barbaglia, Sergio Consoli & Sebastiano Manzan (2023) Forecasting with Economic News, *Journal of Business & Economic Statistics*, 41:3, 708-719
- 14) Tetlock, P. C. (2007), “Giving Content to Investor Sentiment: The Role of Media in the Stock Market,” *Journal of Finance*, 62, 1139–1168.
- 15) Algaba, A., Ardia, D., Bluteau, K., Borms, S., and Boudt, K. (2020), “Econometrics Meets Sentiment: An Overview of Methodology and Applications,” *Journal of Economic Surveys*, 34, 512–547.
- 16) Petropoulos Petalas, D., van Schie, H., & Hendriks Vettehen, P. (2017). Forecasted economic change and the self-fulfilling prophecy in economic decision-making. *PLoS ONE*, 12(3)

# Литература (2)

- 17) Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics* 38(2), 393–409.
- 18) Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and SocialMedia*, 8(1), 216-225.
- 19) Larsen, V.H. and Thorsrud, L.A. (2022), Asset returns, news topics, and media effects†. *Scand. J. of Economics*, 124: 838-868.
- 20) Stock, J. H. & Watson, M. W. (2010), Modeling inflation after the crisis, Technical report, National Bureau of Economic Research
- 21) STOCK, J.H. and WATSON, M.W. (2007), Why Has U.S. Inflation Become Harder to Forecast?. *Journal of Money, Credit and Banking*, 39: 3-33.
- 22) Atkeson, A. & Ohanian, L. (2001), ‘Are phillips curves useful for forecasting inflation?’, *Federal Reserve bank of Minneapolis Quarterly Review* 25, 2–11.
- 23) Fraiberger, S. P., Lee, D., Puy, D., & Ranciere, R. (2021). Media sentiment and international asset prices. *Journal of International Economics*, 133, 103526.
- 24) Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, 228(2), 221–243.
- 25) Hutto, C., and E. Gilbert (2014): “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” in Eighth International Conference on Weblogs and Social Media (ICWSM-14).
- 26) Isabelle Salle & Yuriy Gorodnichenko & Olivier Coibion, 2023. "[Lifetime Memories of Inflation: Evidence from Surveys and the Lab](#)," [NBER Working Papers](#) 31996, National Bureau of Economic Research, Inc.
- 27) Coulombe, P. G., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5), 920–964. <https://doi.org/10.1002/jae.2910>
- 28) Pavlov, E. (2020). Forecasting Inflation in Russia Using Neural Networks. *Russian Journal of Money and Finance*, 79(1), pp. 57–73.
- 29) Baybuza, I. (2018). Inflation Forecasting Using Machine Learning Methods. *Russian Journal of Money and Finance*, 77(4), pp. 42–59.
- 30) Ash, Elliott and Hansen, Stephen, Text Algorithms in Economics (September 1, 2023). *Annual Review of Economics*, Vol. 15, pp. 659-688, 2023
- 31) Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *J Econ Surv*, 37, 76–111. <https://doi.org/10.1111/joes.12429>
- 32) Chakraborty, Chiranjit and Joseph, Andreas, Machine Learning at Central Banks (September 1, 2017). *Bank of England Working Paper* No. 674



# Литература (3)

- 33) Ahmed, Nesreen & Atiya, Amir & Gayar, Neamat & El-Shishiny, Hisham. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*. 29. 594-621. 10.1080/07474938.2010.481556.
- 34) Kondratyev, Alexei, (April 11, 2018) Learning Curve Dynamics with Artificial Neural Networks
- 35) Petelin, G., Cenikj, G. and Eftimov, T., 2022. Towards understanding the importance of time-series features in automated algorithm performance prediction. *Expert Systems with Applications*, 119023.
- 36) Moshiri, S. and Cameron, N. (2000), Neural network versus econometric models in forecasting inflation. *J. Forecast.*, 19: 201-217. [https://doi.org/10.1002/\(SICI\)1099-131X\(200004\)19:3<201::AID-FOR753>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-131X(200004)19:3<201::AID-FOR753>3.0.CO;2-4)
- 37) Almosova, A., & Andresen, N. (2023). Nonlinear inflation forecasting with recurrent neural networks. *Journal of Forecasting*, 42(2), 240–259. <https://doi.org/10.1002/for.2901>
- 38) Oren Barkan, Jonathan Benchimol, Itamar Caspi, Eliya Cohen, Allon Hammer, Noam Koenigstein, Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks, *International Journal of Forecasting*, Volume 39, Issue 3, 2023, Pages 1145-1162, ISSN 0169-2070, <https://doi.org/10.1016/j.ijforecast.2022.04.009>.
- 39) Marcelo C. Medeiros & Gabriel F. R. Vasconcelos & Álvaro Veiga & Eduardo Zilberman, 2021. "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods," Journal of Business & Economic Statistics, Taylor & Francis Journals, vol. 39(1), pages 98-119, January.
- 40) Chakraborty, Chiranjit and Joseph, Andreas, Machine Learning at Central Banks (September 1, 2017). Bank of England Working Paper No. 674
- 41) Joseph A. Shapley Regressions: A Framework for Statistical Inference on Machine Learning Models // Bank of England Working Paper. – 2019. – N 784.
- 42) Gasparian, M. S. ., Kiseleva, I. A. ., Titov, V. A. ., Sysoev, N. A. ., & Chernysheva, E. N. . (2021). Socioeconomic Development: Search For Optimal Models For Forecasting Inflationary Processes . *International Journal of Criminology and Sociology*, 10, 472–478.
- 43) Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings International Joint Conference on Artificial Intelligence, IJCAI*
- 44) Kondratyev, A., 2018. Learning curve dynamics with artificial neural networks. Available at SSRN 3041232.
- 45) Anna Almosova & Niek Andresen, 2023. "Nonlinear inflation forecasting with recurrent neural networks," Journal of Forecasting, John Wiley & Sons, Ltd., vol. 42(2), pages 240-259, March.