



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

*Образовательная программа «Экономика»*

Золотухина Евгения Олеговна, Рыбин Сергей Васильевич, Халифаева Лейла  
Илхамовна

**НАУКАСТИНГ ДИНАМИКИ УРОВНЯ БЕЗРАБОТИЦЫ В РОССИИ,  
ТУРЦИИ И КИТАЕ**

Курсовая работа студентов 3-го курса

Научный руководитель

Станкевич И.П.

Москва, 2024 год

## СОДЕРЖАНИЕ

1. Введение:.....	3
2. Обзор литературы: .....	3
3. Обзор моделей: .....	5
4. Данные.....	5
5. Обучение и оценка качества моделей. ....	7
5.1. AR(1).....	8
5.2. ARIMA.....	8
5.3. ARIMAX.....	9
5.4. VAR.....	12
5.5. DFM.....	15
5.6. MIDAS .....	177
6. Выводы .....	199
7. Литература .....	22

## **1. Введение:**

Уровень безработицы является одним из ключевых показателей, используемых для оценки экономического благополучия страны и социальных последствий.

Измерение уровня безработицы имеет важное значение, поскольку безработица оказывает серьезное негативное воздействие на людей. Безработица приводит не только к потере дохода и снижению уровня жизни, но и к снижению удовлетворенности жизнью и психическим расстройствам[1].

Измерение безработицы также имеет важное значение с макроэкономической точки зрения, поскольку оно позволяет получить информацию о масштабах недостаточного использования рабочей силы в экономике. Уровень безработицы является ключевым индикатором общего экономического развития и способности экономики создавать рабочие места. Также стоит отметить, что связь между инфляцией и безработицей делает изучение индикаторов рынка труда актуальным. Это было эмпирически подтверждено, например, для России [2]. В дополнение к связи с инфляцией, безработица является одним из ведущих индикаторов экономических рецессий, подтверждение этому также можно найти в литературе [3][4][5].

В нашем исследовании мы сосредотачиваем внимание на трех странах: России, Китае и Турции. Эти государства были выбраны не случайно. Все три страны активно развиваются, и их экономические перспективы привлекают внимание аналитиков и инвесторов со всего мира. Более того в 2024 году они входят в число наиболее быстро развивающихся стран[6], что делает их еще более привлекательными для изучения и сравнительного анализа.

В данной работе мы прогнозируем уровень безработицы в России, Китае и Турции, используя разнообразные методы машинного обучения и социально-экономические показатели. Наукастинг является одним из инструментов, позволяющих достаточно точно оценить динамику в различных сферах экономики и предсказать их состояние в ближайшем будущем. Более того, анализ результатов наукастинга позволяет оценить качество публикуемой статистики, ведь наличие расхождений в прогнозах может говорить о существовании некоторых проблем.

## **2. Обзор литературы:**

В условиях растущей неопределенности проблема безработицы становится все более актуальной. Этот показатель является одним из самых важных индикаторов социально-экономического развития страны, который может оказывать как негативное, так и позитивное воздействие на мир в целом. Именно поэтому важно уметь точно предсказывать уровень безработицы, чтобы в дальнейшем регулировать спрос и предложение на

рынке труда посредством развития информационных технологий, совершенствования системы образования [7].

Несмотря на то, что предсказание будущих трендов и явлений в экономике страны сложнее и требует большего времени, возможно исследовать сам рынок труда посредством ряда показателей и индикаторов, свидетельствующих о его развитии [8]. Таким образом, можно отметить, что для более точного прогнозирования уровня безработицы часто используют модели, которые позволяют дополнительно учесть информацию и тенденции из независимых переменных [9].

На динамику уровня безработицы оказывают влияние ряд макро- и микроэкономических показателей. Одним из самых очевидных является уровень ВВП, воздействие именно этого показателя оценивается в работах многих исследователей, например, в статье Мацкевич А.А. и Гваевой И.В.(2017)[10], в которой говорится о положительной зависимости между показателями. Также безработица часто рассматривается во взаимосвязи с инфляцией, что и демонстрирует кривая Филлипса [11].

При исследовании уровня безработицы в разных странах важно учитывать региональные особенности рынка труда, что отражается в подборе показателей деловой активности населения. К таким индикаторам можно отнести оборот розничной торговли и индекс промышленного производства. Релевантность и значимость этих показателей описывается в работе Щербакова В., Харламовой М. и Гартвич Р., (2023)[12]. Также в некоторых статьях отдельно выделяют взаимосвязь индекса производственной активности PMI и уровня безработицы [13]. Более того, ученые также отмечают важность своевременной оценки состояния рынка труда для принятия мер, направленных на быструю подстройку к шокам в экономике.

Во многих развитых и развивающихся странах экспорт нефти является коммерчески значимым и вносит значительный вклад в ВВП. По этой причине даже небольшие колебания мировых цен на нефть приводят к существенным изменениям не только в экономике в целом, но и на рынке труда. Более того, нефть более волатильная по сравнению с другими активами, а если учесть существующую политическую ситуацию и связанную с ней неопределенность в мире, то роль цен на нефть и нефтепродукты еще больше возрастает. Ряд исследований подтвердил существование данной взаимосвязи на примере экономик Нигерии [14], Пакистана [15], США [16] и др.

Стоит отметить, что множество исследователей оценивали влияние изменения реального обменного курса на уровень безработицы. В работе Nyahokwe и Ncwadi (2013) [17], посвященной изучению данной взаимосвязи на примере стран Южной Африки, делается вывод о том, что именно реальный обменный курс оказался самым значимым для объяснения изменчивости

уровня безработицы, поскольку этот показатель отражает реакцию на шоки процентных ставок, спроса и предложения на мировых рынках.

### **3. Обзор моделей:**

Существует множество статистических моделей, которые применяют как инструмент наукастинга. Во-первых, в случае исследования данных с одинаковой частотностью разумно оценивать авторегрессионные модели, которые прогнозируют данные на основе временных рядов. Например, в работе Куровского Г.С. [18] в качестве базовых методов прогнозирования безработицы используются AR(1) и ARIMA. Эффективность применяемых методов оценивается с помощью RMSE и MAE, а результаты демонстрируют их релевантность при прогнозировании безработицы. В исследовании M.Fajar [19] используется модель ARIMA с экзогенной переменной (ARIMAX), которая основывается не только на исторических данных, но и на информации об экзогенных факторах. Данный метод в результате демонстрирует лучшее качество по сравнению с базовыми и очень точно объясняет существующие закономерности. Еще один инструмент, который используют Sourav Kundu and Rajshekhar Singhania [20] для прогнозирования безработицы, это векторная авторегрессия (VAR). Модель оценивает взаимосвязь между изучаемыми временными рядами, что обеспечивает высокую точность прогноза.

Во-вторых, во множестве исследований, посвященных прогнозированию макроэкономических показателей, применяют динамическую факторную модель (DFM). В работе А.В.Зубарева и К.С.Рыбак [21] есть предположение о том, что влияние ненаблюдаемых переменных на исследуемый показатель можно оценить с помощью информации, содержащейся в наблюдаемых рядах, например, через выделение признаков из экономически содержательных подгрупп данных. В данном исследовании авторы пришли к выводу о том, что DFM модель обладает меньшими ошибками прогноза, по сравнению с AR(1) и VAR. Таким образом, несмотря на относительно несложное построение, метод работает качественно.

В-третьих, при работе с разнородными показателями обычно используют модели с данными смешанной частоты (MIDAS). Например, в работе Schumacher C. [22] рассматриваются различные MIDAS-методы для наукастинга роста ВВП в Еврозоне. Более того, автор сравнивает MIDAS-модели с моделью связующих уравнений (bridge equations), которая является еще одним эффективным методом прогнозирования. В заключении работы отмечается превосходство MIDAS над другими подходами.

### **4. Данные**

Как описывалось ранее, в данной работе исследуется уровень безработицы в России, Турции и Китае. В качестве исследуемой зависимой переменной выступает изменение ежемесячного уровня безработицы в стране месяц к месяцу, измеряемый в %. Период исследования: с января 2015 года по

декабрь 2021 года. Выбран именно этот промежуток времени, поскольку он позволяет учесть больше всего социально-экономических показателей по анализируемым трем странам.

В качестве независимых переменных используется ряд социально-экономических показателей для каждой страны. Данные публикуются ежедневно:

- **ВВП** - изменение показателя по сравнению с предыдущим месяцем (м/м) отражает изменение стоимости товаров и услуг, произведенных экономикой страны по сравнению с соответствующим месяцем предыдущего года, % (gdp).
- **Инфляция** - индекс, который сравнивает изменение цен на товары и услуги месяц относительно месяца, % (infl).
- **Индекс промышленного производства** - показатель динамики объема промышленного производства, как отношение текущего объема производства в денежном выражении к объему в предыдущем месяце % (iip).
- **Оборот розничной торговли** - показатель изменения агрегированной меры розничных продаж товаров и услуг месяц к месяцу, % (retail).
- **Индекс производственной активности** - индикатор, который отражает деловую активность в производственном секторе и основывается на анкетировании 300 компаний отрасли, % (pmi).
- **Цены на нефть BRENT** - основной ценовой ориентир для мировых закупок нефти, USD/Bbl, % (oil).
- **Курс доллара к нац валюте** - спотовый валютный курс доллара США к национальной валюте, USD/нац.валюта, % (usd).

Помимо базовых макроэкономических показателей исследуется также и связь различных фондовых индексов с уровнем безработицы. С одной стороны, используются котировки мировых индексов, которые оказывают влияние на экономики всех стран. Данные публикуются ежедневно:

- **S&P 500** - изменение д/д котировок фондового индекса, в корзину которого включено 503 акции 500 избранных торгуемых на фондовых биржах США публичных компаний, имеющих наибольшую капитализацию, % (SP500).
- **S&P Global 100** - изменение д/д котировок фондового индекса, который измеряет эффективность 100 многонациональных компаний, %(SP100).
- **Nasdaq** - изменение д/д котировок индекса фондового рынка, состоящего из долевых ценных бумаг, выпущенных 100 крупнейшими нефинансовыми компаниями, котирующимися на фондовой бирже Nasdaq, % (NSDQ).

- **CRSP US Total Market Index** - изменение д/д котировок индекса, который включает 4000 компаний с различной капитализацией, % (CRSP).

С другой стороны, оценивается влияние индексов на экономики отдельных, исследуемых нами, стран. Данные публикуются ежедневно:

- **MSCI Russia/Turkey/China** - изменение д/д фондового индекса, который рассчитывается международной аналитической компанией MSCI Inc., % (MSCI\_RUS/MSCI\_TR/MSCI\_CH).
- **Индекс МосБиржи (IMOEX)** — изменение д/д ценового, взвешенный по рыночной капитализации композитного фондового индекса, включающего 50 наиболее ликвидных акций крупнейших компаний, % (IMOEX).
- **BIST 100** - изменение д/д главного индекса Стамбульской биржи, в корзине которого 100 наиболее дорогих компаний, достаточно точно отражающих общую ситуацию, % (BIST).
- **Hang Seng (HSI)** - изменение д/д индекса Гонконгской фондовой биржи, который строится по взвешенной рыночной капитализации 64 крупнейших компаний, представленных на площадке, % (HSI).

В своих исследованиях Юревич М.А. и Ахмадеев Д.Р. [23] и Щербakov В.С, Харламова М.С1, Гартвич .Е. [24] описывают прогнозирование уровня безработицы на основе поисковых запросов. Подобно этим работам были выделены ключевые поисковые запросы, которые могут описывать ситуацию на рынке труда в стране, а именно: работа, вакансии, центр занятости, увольнение, безработица, биржа труда, карьера, поиск работы, резюме, пособие по безработице. После чего была получена ежемесячная статистика по поисковым запросам на заданные тематики для каждой страны с сайта Google Trends [25], который оценивает популярность запроса от 0 до 100 баллов за исследуемых период.

## 5. Обучение и оценка качества моделей.

В начале была проведена проверка стационарности временных рядов уровня безработицы для всех 3 стран. Поскольку наши фактические значения будут использоваться в качестве основы для предсказаний, очень важно уделить внимание данному свойству, ведь, если ряд не стационарен, это может привести к ошибочным выводам и результатам. Для проверки стационарности использовался расширенный тест Дики-Фуллера, нулевая гипотеза утверждает, что ряд является стационарным. В результате проведения теста, для всех стран временной ряд уровня безработицы оказался стационарным, то есть нулевая гипотеза не была отвергнута. Это позволяет нам построить модель ARIMA. Все прогнозы были получены при помощи метода скользящего окна, когда на каждом шаге одна точка добавляется, а одна (самая старая) – удаляется. В исходной выборке было 84 точки, в качестве

оптимального размера окна был выбран период равный 9 месяцам. Также были построены прогнозы на 3 месяца вперед, то есть на будущие периоды. В результате были спрогнозированы уровни безработицы с октября 2015 года по март 2022 года.

В качестве моделей для наукастинга были выбраны ARIMAX (p, d, q), где p - количество членов авторегрессии (количество прошлых наблюдений), d - количество разностей, необходимых для преобразования ряда в стационарный, q - количество запаздывающих ошибок прогноза, VAR, DFM и MIDAS. Кроме того, прогнозы строились для трех разных групп независимых переменных: первая группа - макроэкономические показатели, вторая - поисковые запросы Google Trends и третья - включает все независимые переменные. Для оценки качества полученных прогнозов использовались MAE и RMSE:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|; RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}.$$

При построении прогнозов с применением модели ARIMAX (модели ARIMA с экзогенными переменными) было использовано 3 различных подхода: первый подход - построение модели ARIMAX на основе прогнозов показателей по модели AR(1), второй - на основе прогнозов показателей по модели ARIMA, третий - по фактическим показателям. Последовательно рассмотрим данные модели и качество их прогнозов.

#### 5.1. AR(1).

Для обучения модели использовались фактические значения уровня безработицы. Результаты оценки качества прогнозов представлены в Таблице 1. Наименьшее значение MAE у Турции, что свидетельствует о наибольшей точности модели AR(1) в среднем, стоит отметить, что точность для России тоже неплохая, в отличие от Китая. Наименьшее значение RMSE также у Турции, это говорит о том, что в среднем предсказанные значения отклоняются от фактических меньше по сравнению с Россией и Китаем. Модель AR(1) для Китая демонстрирует значительно большие ошибки в прогнозировании по сравнению с другими, особенно высоко значение RMSE, что говорит о серьезных отклонениях в прогнозах.

Таблица 1. Качество прогнозов модели AR(1).

Страна	MAE	RMSE
Россия	3.43	5.051
Турция	2.671	3.689
Китай	7.286	40.261

#### 5.2. ARIMA.



Чтобы упростить процесс поиска оптимальных параметров, мы использовали ARIMA с автоматическим подбором параметров для каждой страны, параметры подбираются, чтобы минимизировать информационный критерий Акаике (AIC). Затем, используя лучшие параметры, приступали к построению ARIMA и прогнозированию. Таким образом, для модели России и Китая лучшие параметры  $p, d, q$  - это  $(0, 0, 0)$ , а для Турции -  $(2, 1, 2)$ . Результаты оценки качества моделей ARIMA представлены в Таблице 2. Наименьшее значение MAE у Китая, что свидетельствует о наибольшей точности модели ARIMA в среднем, но также стоит отметить, что точность для России и Турции также хорошая. Тем временем можно увидеть, что наименьшее значение RMSE у Турции, это говорит о том, что в среднем предсказанные значения отклоняются от фактических меньше по сравнению с Россией и Китаем. В целом из всех трех моделей ARIMA худшей является модель, построенная для России, но при этом сама по себе она все же имеет достаточно хорошее качество.

Таблица 2. Качество прогнозов моделей ARIMA.

Страна	MAE	RMSE
Россия	3.224	4.729
Турция	2.934	4.016
Китай	2.512	4.634

Сравнивая модели AR(1) и ARIMA, можно с уверенностью сказать, что для данных по России и Китаю модель ARIMA предпочтительнее, а вот для Турции, наоборот, лучшие результаты показывает AR(1).

### 5.3. ARIMAX.

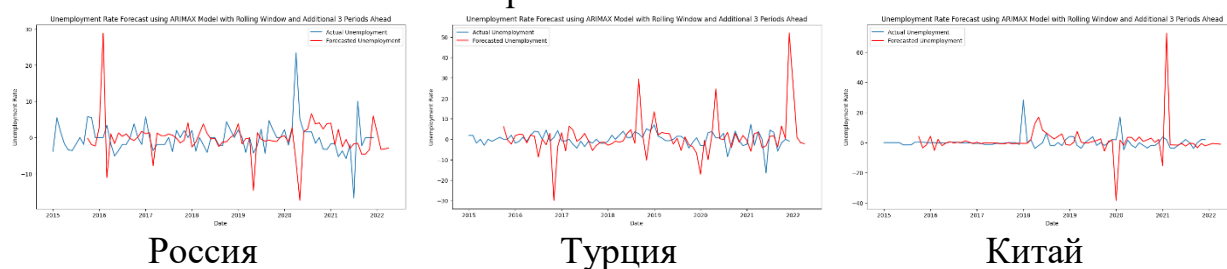
Заметим, что для каждого набора данных, а их у нас 3, как было сказано ранее, для всех моделей были сначала подобраны лучшие параметры при помощи функции ARIMA с автоматическим подбором параметров, а затем обучались модели ARIMAX с лучшими параметрами. Результаты оценки качества для всех моделей с макроэкономическими показателями представлены в Таблице 3. Очевидно, что предсказания по фактическим значениям плохие для всех стран. Модель ARIMAX, построенная для России, показывает наибольшую точность, когда используется модель ARIMA в качестве основы, такой вывод мы делаем, исходя из того, что значения MAE и RMSE в данном случае наименьшие. Для Китая модель ARIMA также является лучшей основой, стоит отметить, что при этом модель AR(1) показывает крайне плохие результаты предсказаний. Наконец, для Турции модель ARIMAX лучше всего работает при использовании модели AR(1) в качестве

основы. Прогнозы лучших моделей с макроэкономическими показателями для каждой страны можно увидеть на Рисунке 1.

Таблица 3. Качество прогнозов моделей ARIMAX для макроэкономических показателей.

Страна	Основа прогнозов	MAE	RMSE
Россия	Факт	14.463	30.082
Россия	AR	7.247	10.727
Россия	ARIMA	4.474	7.128
Турция	Факт	15.733	33.813
Турция	AR	5.632	9.794
Турция	ARIMA	13.337	30.622
Китай	Факт	19.936	55.909
Китай	AR	51.815	156.116
Китай	ARIMA	5.003	11.193

Рисунок 1. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для трех стран с использованием модели ARIMAX по макроэкономическим показателям

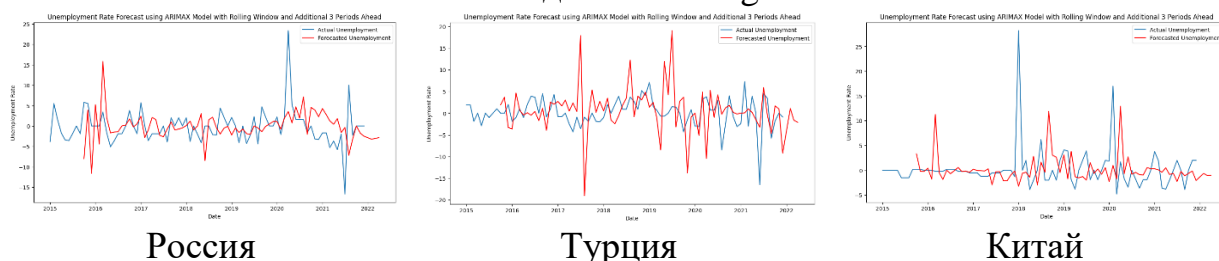


Далее рассмотрим результаты оценки качества моделей для поисковых запросов Google Trends (см. Таблицу 4). Здесь мы снова видим, что предсказания по фактическим значениям плохие для всех стран. В целом, можно сделать вывод, что лучшие модели для всех стран аналогичны тем, что были и для макроэкономических показателей. Прогнозы лучших моделей с поисковыми запросами Google Trends для каждой страны можно увидеть на Рисунке 2.

Таблица 4. Качество прогнозов моделей ARIMAX для поисковых запросов Google Trends.

Страна	Основа прогнозов	MAE	RMSE
Россия	Факт	16.511	29.103
Россия	AR	6.674	9.649
Россия	ARIMA	3.782	5.518
Турция	Факт	10.988	21.66
Турция	AR	4.343	6.068
Турция	ARIMA	5.919	9.973
Китай	Факт	7.025	16.508
Китай	AR	31.493	115.508
Китай	ARIMA	2.98	5.359

Рисунок 2. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для трех стран с использованием модели ARIMAX по данным Google Trends



Наконец, рассмотрим результаты оценки качества моделей для всех независимых переменных, то есть и для макроэкономических показателей, и для поисковых запросов Google Trends (см. Таблицу 5). Из всех ранее обученных нами моделей, только модели ARIMAX по всем независимым переменным дают умеренные значения MAE и RMSE на фактических данных. Можно сказать, что для всех моделей, кроме модели для Китая на основе AR(1), характерны умеренные значения критериев качества. Также стоит отметить, что лучшими моделями для России и Китая остаются ARIMAX на базе ARIMA, а для Турции - на базе AR(1). Прогнозы лучших моделей для каждой страны можно увидеть Рисунке 3.

Таблица 5. Качество прогнозов моделей ARIMAX для всех независимых переменных.

Страна	Основа прогнозов	MAE	RMSE
Россия	Факт	7.23	12.683
Россия	AR	5.191	9.254
Россия	ARIMA	3.32	4.676
Турция	Факт	5.949	7.986
Турция	AR	3.29	4.486
Турция	ARIMA	4.305	6.095
Китай	Факт	3.936	6.259
Китай	AR	13.264	36.019
Китай	ARIMA	2.949	5.072

Рисунок 3. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для трех стран с использованием модели ARIMAX по всем переменным



Рассмотрев различные комбинации независимых переменных и основ для обучения моделей, можно заключить, что лучшей среди всех моделей ARIMAX является модель по всем независимым переменным для Китая (см. Рисунок 3), поскольку она обладает лучшей точностью на основе значений MAE и RMSE. Отметим также, что при помощи функции ARIMA с автоматическим подбором параметров лучшая модель для Китая выглядела как ARIMAX (0, 1, 0).

#### 5.4. VAR.

VAR-модель применяется при построении прогнозов на основе множественных переменных, которые находятся во взаимодействии между собой. Для начала необходимо определять порядок модели (p), который

отвечает за число лагов в модели. Оптимальный параметр  $p$  подбирается на основе критериев AIC и BIC, поэтому мы для каждого прогноза оценивали возможное число лагов и выбирали то, при котором критерии наименьшие. Далее оценивали коэффициенты и строили прогнозы.

Здесь мы также обучали модели на основе 3 разных наборов независимых переменных. Отметим, что здесь в качестве основы прогнозов везде использовались фактические значения уровня безработицы по всем странам. Начнем с рассмотрения модели VAR, которая включает в себя макроэкономические показатели. Результаты оценки качества моделей можно посмотреть в Таблице 6. Заметим, что модель VAR показывает хорошие результаты при прогнозировании уровня безработицы для всех трех стран. Лучше всего данная модель спрогнозировала значения для Турции, именно для данной страны наблюдаются наименьшие значения MAE и RMSE, а это говорит о наибольшей точности модели. Для России критерии качества чуть выше, но они принимают умеренные значения. Значения MAE и RMSE для Китая также показывают приемлемые ошибки прогнозов, однако они немного выше, чем для России и Турции, это может быть связано с особенностями фактических данных для этой страны, поскольку в них изначально было достаточно много нулевых значений.

Таблица 6. Качество прогнозов моделей VAR для макроэкономических показателей.

Страна	MAE	RMSE
Россия	3.966	6.14
Турция	3.094	4.437
Китай	4.277	9.417

Перейдем к анализу моделей VAR, включающих поисковые запросы в Google Trends. Результаты оценки качества моделей можно посмотреть в Таблице 7. В данном случае модель достаточно точно спрогнозировала значения для Турции и Китая. Значения MAE и RMSE для модели России выше (особенно для RMSE), что указывает на то, что эта модель VAR относительно двух других делает плохие прогнозы. Возможно для снижения ошибок прогнозов стоит добавить дополнительные переменные в модель, что мы и сделаем далее.

Таблица 7. Качество прогнозов моделей VAR для поисковых запросов в Google Trends.

Страна	MAE	RMSE
Россия	5.84	10.966

Турция	3.637	5.034
Китай	3.1	5.836

Теперь рассмотрим модель VAR, включающую в себя все независимые переменные. Результаты оценки качества моделей можно посмотреть в Таблице 8. Здесь мы снова можем заметить, что модель VAR показывает хорошие результаты при прогнозировании уровня безработицы для всех трех стран, такие прогнозы можно считать достаточно точными и надежными. Однако, снова по сравнению с другими моделями выделяется значение RMSE модели VAR для России, хотя сам по себе данный критерий стал ниже по сравнению с моделью VAR по поисковым запросам Google Trends, мы все равно приходим к выводу о том, что в данной модели есть некоторые значительные отклонения.

Таблица 8. Качество прогнозов моделей VAR для всех независимых переменных.

Страна	MAE	RMSE
Россия	4.511	7.082
Турция	3.484	4.867
Китай	3.088	5.856

Подводя итог по модели VAR, можно сказать, что рассмотрение различных комбинаций независимых переменных помогло нам выявить лучшие модели: VAR для Турции и России по макроэкономическим показателям и VAR для Китая по всем переменным(см. Рисунок 4), их значения MAE примерно равны и являются наименьшими среди всех, что говорит о высокой точности данных моделей, значение RMSE для модели Турции немного меньше, чем у Китая, но в целом такие значения RMSE означают, что в среднем предсказанные значения отклоняются от фактических не слишком сильно.

Рисунок 4. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для трех стран с использованием модели VAR



### 5.5. DFM

Еще один метод, использованный нами для наукастинга безработицы – это динамическая факторная модель, которая основывается на предположении о том, что наблюдаемые переменные зависят от некоторого числа скрытых факторов. Существует некоторое множество общих факторов, которые описывают группу временных рядов. Таким образом, все наблюдаемые переменные можно линейно выразить через них с учетом случайной ошибки. Прогнозирование осуществляется через предсказание скрытых факторов и дальнейшее восстановление прогнозируемых наблюдаемых переменных. Данная процедура позволяет уменьшить размерность данных, и она строилась нами на основе фактических переменных. В результате нами были оценены модели по трем группам переменных по каждой стране: макроэкономические показатели, поисковые запросы, все признаки.

Рассмотрим DFM, оцененную по макроэкономическим показателям. Результаты оценки качества моделей представлены в Таблице 9. Модель демонстрирует плохие результаты прогнозов на всех выборках, высокие значения RMSE свидетельствуют о том, что модели DFM для макроэкономических показателей сильно отклоняются от фактических. Хуже всего обучилась модель по данным для Китая, это может быть связано с качеством исходных данных, поскольку Китай – страна с высокой экономической волатильностью, это может означать, что макроэкономические показатели могут сильно колебаться, а это затрудняет прогнозирование и приводит к высоким ошибкам. Возможно включение дополнительных переменных позволит улучшить качество прогнозов.

Таблица 9. Качество прогнозов моделей DFM для макроэкономических показателей.

Страна	MAE	RMSE
Россия	10.184	28.309
Турция	7.144	22.727
Китай	24.685	163.367

Рассмотрим DFM, оцененную по поисковым запросам Google Trends. Результаты оценки качества моделей представлены в Таблице 10. Заметим, что для данного набора показателей модель для Китая не обучилась вовсе. Мы предполагаем, что это связано с большим количеством нулевых значений в данных по уровню безработицы, а также аномальные значения, которые могут создавать проблемы для обучения DFM и искажать извлекаемые факторы. Модель DFM показывает хорошую точность прогнозов на данных по

поисковым запросам для Турции. Для России значение MAE и RMSE выше, но в целом модель DFM неплохо справляется с прогнозированием, хотя и присутствуют отклонения, на которые указывает значение RMSE.

Таблица 10. Качество прогнозов моделей DFM для поисковых запросов Google Trends.

Страна	MAE	RMSE
Россия	5.277	7.705
Турция	3.919	5.446

Перейдем теперь к рассмотрению модели DFM, оцененной по все признакам. Результаты оценки качества моделей представлены в Таблице 11. Снова заметим, что для Китая не удалось построить модель DFM. При этом, совмещение двух наборов признаков привело к многократному ухудшению качества прогнозов как для России, так и для Турции. Очень высокие значения MAE и RMSE характерны для обеих моделей, кроме того, значения RMSE намного выше MAE, что свидетельствует о наличии сильных ошибок в отдельных прогнозах. Мы предполагаем, что такое низкое качество прогнозов связано с выбросами в данных, волатильностью макроэкономических показателей, а также ограничениями самой модели DFM: данная модель может быть недостаточно сложной для захвата всех факторов, влияющих на уровень безработицы в России и Турции.

Таблица 11. Качество прогнозов моделей DFM для всех признаков.

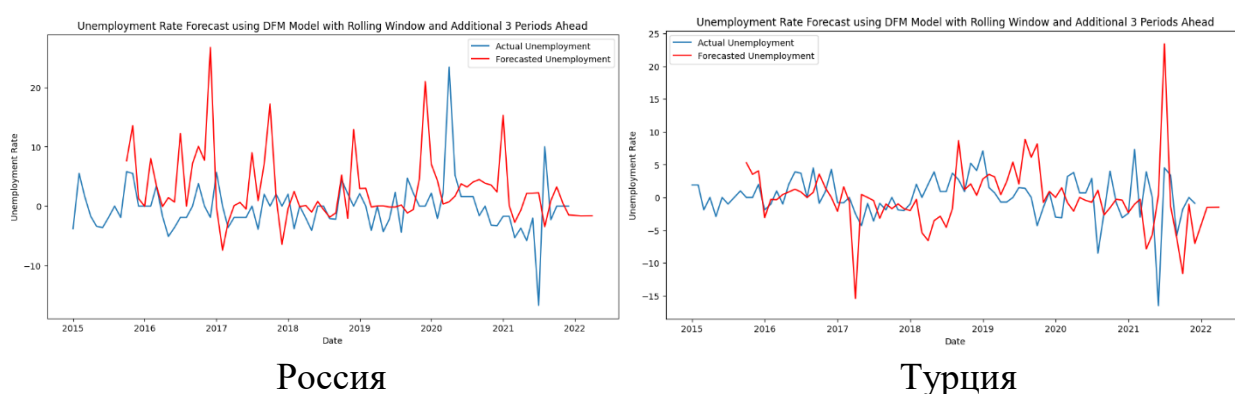
Страна	MAE	RMSE
Россия	79.211	631.873
Турция	29.597	165.225

Таким образом, модели DFM, использующие данные Google Trends, показывают лучшие результаты по сравнению с моделями с другими наборами признаков. В целом можно сказать, что огромное значение на результаты данной модели оказывает качество исходных данных. Лучшей моделью среди всех является DFM по поисковым запросам Google Trends для Турции, она же является лучшей моделью для Турции, соответственно, данная модель обладает самой высокой точностью прогнозов. Для России также лучшей моделью является DMF по поисковым запросам Google Trends. Такой



результат может быть объяснен тем, что в Турции и России скорее всего наблюдается высокая корреляция между изменениями в социально-экономическом статусе людей и их активностью в интернете (см. Рисунок 5). Для Китая не удалось построить 2 из 3 моделей, поэтому мы не будем говорить в данном случае о лучшей модели.

Рисунок 5. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для России и Турции с использованием модели DFM



## 5.6. MIDAS

MIDAS – метод наукастинга, который позволяет использовать данные с разной частотой, что позволяет добавить в анализ различные фондовые индексы, которые оказывают значительное влияние как на экономику в целом, так и на сам уровень безработицы. Мы обучали модель без ограничений, она позволяет учесть динамику показателей, поскольку оценивает не только текущие значения, но и лаги переменных. Стоит отметить, что нами было сделано допущение о том, что MIDAS без ограничений можно представить в виде обычной линейной регрессии, но оцененной по лагам тоже. Для этого сначала привели более частотные данные по фондовым индексам к ежемесячным, как сама целевая переменная, путем нахождения среднего. После этого была оценена модель, построены прогнозы и посчитаны основные метрики. Все собранные данные по индексам были объединены в 3 группы: мировые индексы, страновые индексы, а также датасет со всеми показателями. Далее для оценки моделей к каждой группе данных были добавлены значения макроэкономических показателей и обучены модели.

Для начала нам хотелось оценить, как сильно влияет изменение котировок мировых фондовых индексов на динамику уровня безработицы. Результаты оценки качества моделей можно посмотреть в Таблице 12. Модель для российской экономики имеет наибольшие MAE и RMSE по сравнению с двумя другими. Модель для Турции, наоборот, демонстрирует наименьшие значения ошибок, что говорит о ее высоком качестве. Также значения RMSE немного выше, чем значения MAE для всех стран, что говорит о наличии отклонений в прогнозах.

Таблица 12. Качество прогнозов моделей MIDAS для мировых фондовых индексов.

Страна	MAE	RMSE
Россия	5.076	6.936
Турция	3.494	4.725
Китай	2.763	5.030

Далее модели строились на основе страновых индексов, а затем были так же оценены. Результаты оценки качества моделей можно посмотреть в Таблице 13. Если ориентироваться на MAE, то лучшее качество наблюдается у модели для Китая, а худшее – для России. Однако, значения RMSE лучше для турецкой экономики. Такое различие может быть связано с недостаточно хорошим способом агрегации данных с разной частотой. А также можно предположить, что у китайской модели существуют отклонения в отдельных прогнозах, что и завышает RMSE.

Таблица 13. Качество прогнозов моделей MIDAS для страновых фондовых индексов.

Страна	MAE	RMSE
Россия	5.077	6.936
Турция	3.491	4.721
Китай	2.763	5.030

Наконец, страновые и мировые показатели были объединены для оценки моделей, построенных на всех индексах. Результаты оценки качества моделей можно посмотреть в Таблице 14. Модель, построенная на российских данных, снова имеет худшее качество из всех исследуемых. Это говорит об относительных ошибках и разбросах прогнозов по сравнению с другими странами. У модели для Китая наименьшее значение MAE, что говорит о точности прогнозов. Несмотря на это есть некоторые отклонения, поскольку RMSE немного выше, чем у Турции.

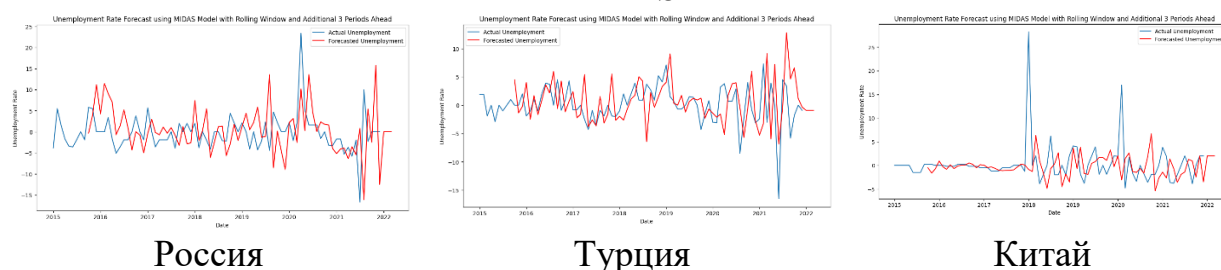
Таблица 14. Качество прогнозов моделей MIDAS для всех фондовых индексов.

Страна	MAE	RMSE
Россия	5.074	6.931
Турция	3.490	4.720

Китай	2.763	5.029
-------	-------	-------

Анализируя все результаты, полученные при оценке различных конфигураций модели MIDAS, стоит отметить, что во всех анализируемых случаях значения ошибок отличаются друг от друга не так сильно. Однако, все же можно выделить лучшие модели для каждой страны. Самые маленькие значения ошибок для России наблюдаются у модели, оцененной по всем собранным котировкам фондовых индексов, а если быть точнее по их динамике. Такую же тенденцию можно наблюдать и для Турции с Китаем, лучшие для них модели также оценены по всем индексам (см. Рисунок 6). Такие результаты демонстрируют важность комплексного подхода при оценке динамики уровня безработицы. Для каждой экономики важно рассматривать как индивидуальные индексы, так и общемировые, чтобы получать наиболее точные результаты.

Рисунок 6. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для трех стран с использованием модели MIDAS



## 6. Выводы

В данной работе мы провели сравнение точности различных моделей для наукастинга уровня безработицы России, Турции и Китая: ARIMAX (на основе фактических значений и прогнозов AR(1), ARIMA), DFM, VAR и MIDAS по 3-м различным наборам признаков. Прогнозы были построены при помощи метода скользящего окна. Точность прогнозов оценивалась при помощи таких метрик качества как MAE и RMSE по последним 75 точкам. Никаких противоречий между значениями MAE и RMSE при оценке качества моделей не было обнаружено.

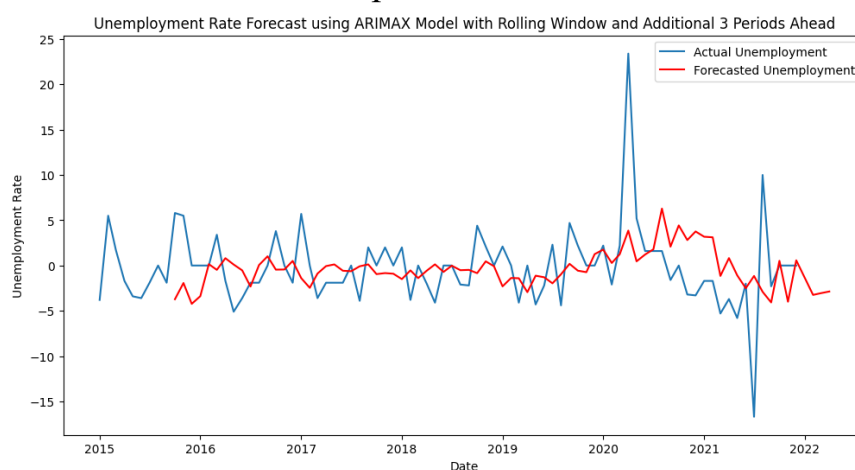
Результаты данной работы показывают важность использования разных методов наукастинга и наборов независимых переменных для получения максимально точных прогнозов, также стоит оценивать первоначальное качество набора данных для построения и обучения моделей. Результаты оценки качества прогнозов показали, что модели, основанные на поисковых запросах Google Trends (здесь имеются в виду как модели по всем признакам, то есть включающие поисковые запросы, так и только по поисковым запросам), показывают более высокую точность по сравнению с моделями,

построенными по макроэкономическим показателям. Из этого следует, что данные Google Trends являются ценным инструментом для прогнозирования уровня безработицы, поскольку они отражают социально-экономическую активность населения и позволяют быстро реагировать на изменения в экономике, основываясь на поведении и «настроении» населения. Единственные модели, хорошо работающие на макроэкономических показателях, - модели векторной авторегрессии (VAR), это подтверждает, что многовариантный подход, учитывающий изменения во времени и различные взаимосвязи между переменными, является эффективным инструментом для прогнозирования уровня безработицы. Так, например, инфляция, уровень безработицы и ВВП могут влиять друг на друга, а модель VAR позволяет учесть это, моделируя каждую переменную как функцию ее собственных лагов и лагов других переменных в системе. Также, основываясь на результатах MIDAS модели, мы можем утверждать, что фондовые индексы также являются важными переменными при прогнозировании уровня безработицы и способствуют улучшению прогнозной точности, так как они отражают динамику финансовых рынков и их влияние на экономику.

На основе метрик точности были выбраны лучшие модели для каждой страны:

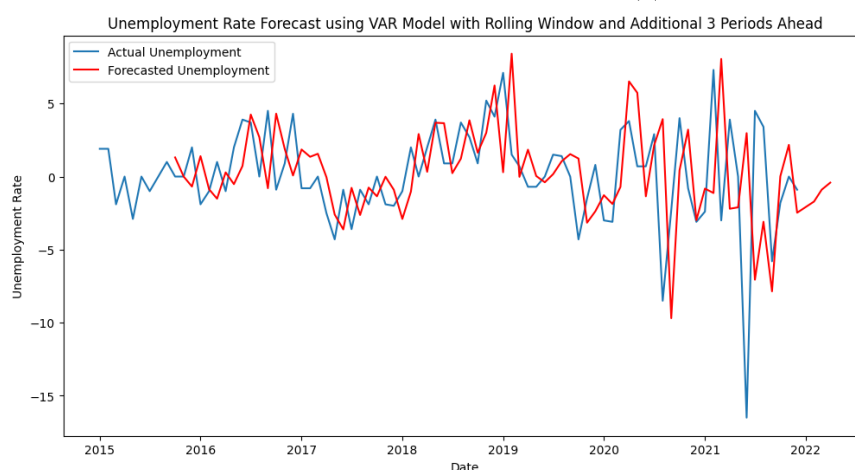
- Для России – модель ARIMAX на основе ARIMA (0, 1, 0) по всем независимым переменным. Модель ARIMAX обладает достаточной объясняющей силой и позволяет учитывать как макроэкономические показатели, так и поисковые запросы, такой широкий охват данных позволяет наиболее точно оценить уровень безработицы для России, на который в равной степени оказывают влияние вышеперечисленные признаки. Результаты прогнозов можно посмотреть на Рисунке 7.

Рисунок 7. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для России с использованием модели ARIMAX на основе прогнозных значений ARIMA (0, 1, 0) по всем независимым переменным



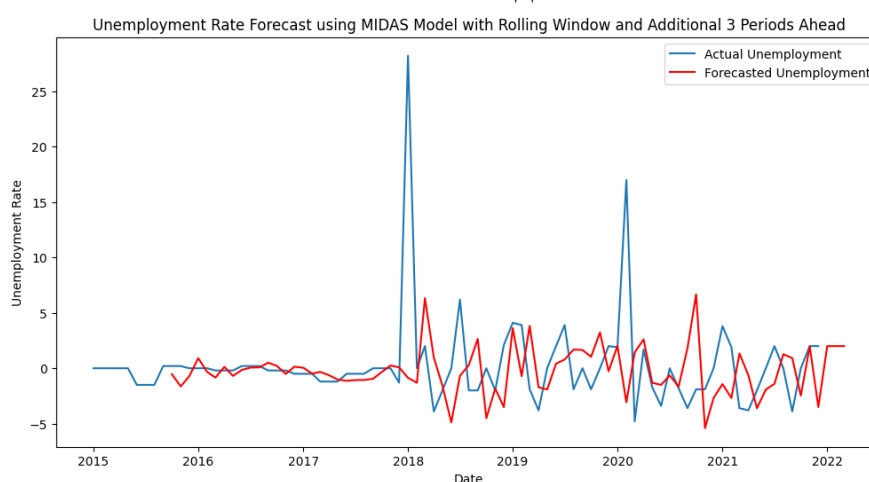
- Для Турции – VAR для макроэкономических показателей. Выше мы уже говорили об особенности модели VAR, связанной с учетом множества взаимосвязей, в том числе между макроэкономическими показателями. Кроме того, хорошее качество модели может обуславливаться тем, что VAR модели могут учитывать влияние краткосрочных шоков и нестабильности, что характерно для экономики Турции. Результаты прогнозов можно посмотреть на Рисунке 8.

Рисунок 8. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для Турции по макроэкономическим показателям с использованием модели VAR



- Для Китая – MIDAS для всех фондовых индексов. Такой результат можно объяснить тем, что китайская экономика интегрирована в глобальную экономику, ранее мы убедились, что учет всех фондовых индексов влияет на качество прогнозов уровня безработицы из-за сильного влияния на экономические процессы. Результаты прогнозов можно посмотреть на Рисунке 9.

Рисунок 9. Динамика фактического и лучших прогнозных значений прироста уровня безработицы за месяц для Китая по всем фондовым индексам с использованием модели MIDAS



## 7. Литература

1. Krueger, A. B., & Mueller, A. I. (2012). Time Use, Emotional Well-Being, and Unemployment: Evidence from Longitudinal Data. *American Economic Review*, 102(3), 594–599. doi:10.1257/aer.102.3.594
2. Орлов Д., Постников Е. (2020). Кривая филлипса: инфляция и *pairu* в российских регионах. Серия докладов об экономических исследованиях, 14-20.
3. Choi, H. (2009). Predicting Initial Claims for Unemployment Benefits.
4. Fleischman, Charles A., and John M. Roberts. "From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model." (2011).
5. Kiley, Michael T. "Financial conditions and economic activity: insights from machine learning." (2020).
6. World Population Review[сайт]. URL: <https://worldpopulationreview.com/country-rankings/emerging-countries>
7. Fathul Muin, Muhamad. (2020). Analysis of Determinants of Unemployment Rate in Indonesia. *Jurnal Perspektif Ekonomi Darussalam*. 6. 145-162. 10.24815/jped.v6i2.16804.
8. Gostkowski, Michał & Rokicki, Tomasz. (2021). Forecasting the Unemployment Rate: Application of Selected Prediction Methods. *European research studies journal*. XXIV. 985-1000. 10.35808/ersj/2396.
9. Ачкасов Ю. К., Шатило Е. Ю. Рынок труда: положение, прогноз, развитие. *Научный вестник ИЭП им. Гайдара.ру (электронный журнал)*. 2016. № 1. С. 53-55.
10. Мацкевич, А. А., И. В. Гваева. "Регрессионный анализ влияния ввп на уровень безработицы в республике Беларусь." *управление информационными ресурсами*. 2017.
11. Ramli, S. F., Fidaus, M., Uzair, H., Khairi, M. & Zharif, A (2018). Prediction of the Unemployment Rate in Malaysia. *International Journal of Modern Trends in Social Sciences*, 1(4), 38-44.
12. Щербаков Василий Сергеевич, Харламова Мария Сергеевна, and Гартвич Роман Евгеньевич. "Анализ безработицы в Сибири на начальном этапе распространения COVID-19" *Вестник Московского университета. Серия 6. Экономика*, no. 6, 2023, pp. 170-191. doi:10.55959/MSU0130-0105-6-58-6-10
13. Habanabakize, Thomas, and Daniel Francois Meyer. "An analysis of the relationships between the purchasing managers' index (pmi), economic growth and employment in the manufacturing sector in South Africa." *International Journal of Economics and Finance Studies* 9.2 (2017): 66-82.

14. Gunu, Umar & Kilishi, Abdulhakeem. (2010). Oil Price Shocks and the Nigeria Economy: A Variance Autoregressive (VAR) Model. *International Journal of Business and Management*. 5. 10.5539/ijbm.v5n8p39.
15. Ahmad, Fawad. "The effect of oil prices on unemployment: Evidence from Pakistan." *Business and Economics research journal* 4.1 (2013): 43.
16. Kocaarslan, Baris, Mehmet Ali Soytas, and Ugur Soytas. "The asymmetric impact of oil prices, interest rates and oil price uncertainty on unemployment in the US." *Energy Economics* 86 (2020): 104625.
17. Bakhshi, Zahra, and Mehrzad Ebrahimi. "The effect of real exchange rate on unemployment." *Marketing and Branding Research* 3 (2016): 4-13.
18. Куровский Глеб Станиславович. "Использование текстовой информации для прогнозирования в макроэкономике" *Вестник Московского университета. Серия 6. Экономика*, no. 6, 2019, pp. 39-58.
19. Fajar, Muhammad & Prasetyo, Octavia. (2020). Forecasting Unemployment Rate in the Time of COVID-19 Pandemic Using Google Trends Data (Case of Indonesia). 6. 29-33.
20. Kundu, Sourav & Singhania, Rajshekhar. (2020). Forecasting the United States Unemployment Rate by Using Recurrent Neural Networks with Google Trends Data. *International Journal of Trade, Economics and Finance*. 11. 135-140. 10.18178/ijtef.2020.11.6.679.
21. Зубарев Андрей Витальевич, and Рыбак Константин Сергеевич. "Наукастинг ввп: динамическая факторная модель и официальные прогнозы" *экономическое развитие России*, vol. 28, no. 12, 2021, pp. 34-40.
22. Schumacher, C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting*, 32(2), 257–270. doi:10.1016/j.ijforecast.2015.07
23. Юревич Максим Андреевич, and Ахмадеев Денис Рашидович. "Возможности прогнозирования уровня безработицы на основе анализа статистики запросов (в поисковых системах)" *Terra Economicus*, vol. 19, no. 3, 2021, pp. 53-64.
24. Щербаков, В. С., М. С. Харламова, and Р. Е. Гартвич. "Методы и модели наукастинга экономических показателей с помощью поисковых запросов." *Развитие экономики регионов: пространственная трансформация, глобальные вызовы и перспективы экономического роста*. 2022.
25. Google Trends[сайт]. URL: <https://trends.google.ru/trends/trendingsearches/daily?geo=RU&hl=ru>