Модель была построена на основе 35 временных рядов, описывающих состояние российской экономики на горизонте 2008–2024 гг.

|  |  |
| --- | --- |
| **Показатель** | **Описание** |
| Wagenom | Среднемесячная номинальная начисленная заработная плата |
| IMOEX | Индекс МосБиржи |
| USD | Курс доллара США |
| EUR | Курс Евро |
| CPI | Индекс потребительских цен |
| IMOEX10 | Индекс МосБиржи (10 ликвидных акций) |
| GOV1Y | Доходность к погашению государственных облигаций сроком 1 год |
| GOV10Y | Доходность к погашению государственных облигаций сроком 10 лет |
| PRODUCERINDEX | Индекс цен производителей |
| TRADE | Розничная торговля в годовом выражении |
| IMPORT | Импорт (по методологии платежного баланса) |
| OILPROD | Добыча нефти |
| М2 | Денежный агрегат М2 |
| M0 | Денежный агрегат М0 |
| UNEMP | Уровень безработицы |
| CONSTRUCTION | Объём строительства в годовом выражении |
| CBRBALANCE | Сумма активов (пассивов) центрального банка в текущих ценах, без исключения сезонного фактора. |
| INFL | Уровень инфляции в годовом выражении |
| EXPORT | Экспорт (по методологии платежного баланса) |
| DEPOSIT | Депозитная ставка по вкладам |
| GOSDEBT | Внутренний государственный долг |
| INDEXRTS | Индекс РТС |
| OILINDEX | Индекс нефти и газа (Мосбиржа отраслевая группа) |
| ELECTROINDEX | Индекс электроэнергетики (Мосбиржа отраслевая группа) |
| FINANCEINDEX | Индекс компаний финансового сектора (Мосбиржа отраслевая группа) |
| POTREBINDEX | Индекс компаний потребительского сектора (Мосбиржа отраслевая группа) |
| METALINDEX | Индекс металлов и добычи (Мосбиржа отраслевая группа) |
| TELEKOMINDEX | Индекс телекоммуникационных компаний (Мосбиржа отраслевая группа) |
| PENSIONINDEX | Индексы активов пенсионных накоплений |
| TRANSPORTINDEX | Индекс транспорта (Мосбиржа отраслевая группа) |
| CHEMINDEX | Индекс химии и нефтехимии (Мосбиржа отраслевая группа) |
| M1 | Денежный агрегат М1 |
| GOV5Y | Доходность к погашению государственных облигаций сроком 5 лет |
| GOV15Y | Доходность к погашению государственных облигаций сроком 15 лет |

Итого, в рамках текущей работы охвачены были следующие группы показателей:

1. Финансовые и фондовые индикаторы
2. Валютные и внешнеторговые показатели
3. Денежные агрегаты и банковские показатели
4. Доходности облигаций
5. Ценовые индексы
6. Реальный сектор экономики

**Кластеризация**

Для анализа временных рядов был использован метод кластеризации TimeSeriesKMeans из библиотеки tslearn. Основная цель заключалась в группировке временных рядов с похожими характеристиками, что позволяет выявить закономерности и структурировать данные для дальнейшего анализа.

В качестве алгоритма был выбран TimeSeriesKMeans с метрикой DTW (Dynamic Time Warping). Эта метрика лучше всего подходит для временных рядов, так как учитывает временные сдвиги и позволяет сравнивать ряды с похожими формами, даже если их пики или паттерны происходят в разное время. Это особенно важно для экономических данных, где сезонность или тренды могут быть смещены.

Перед кластеризацией данные были нормализованы с помощью масштабирования среднего и дисперсии. Нормализация необходима, чтобы временные ряды с разными масштабами сравнивались корректно, а результаты кластеризации зависели от формы данных, а не их абсолютных значений.

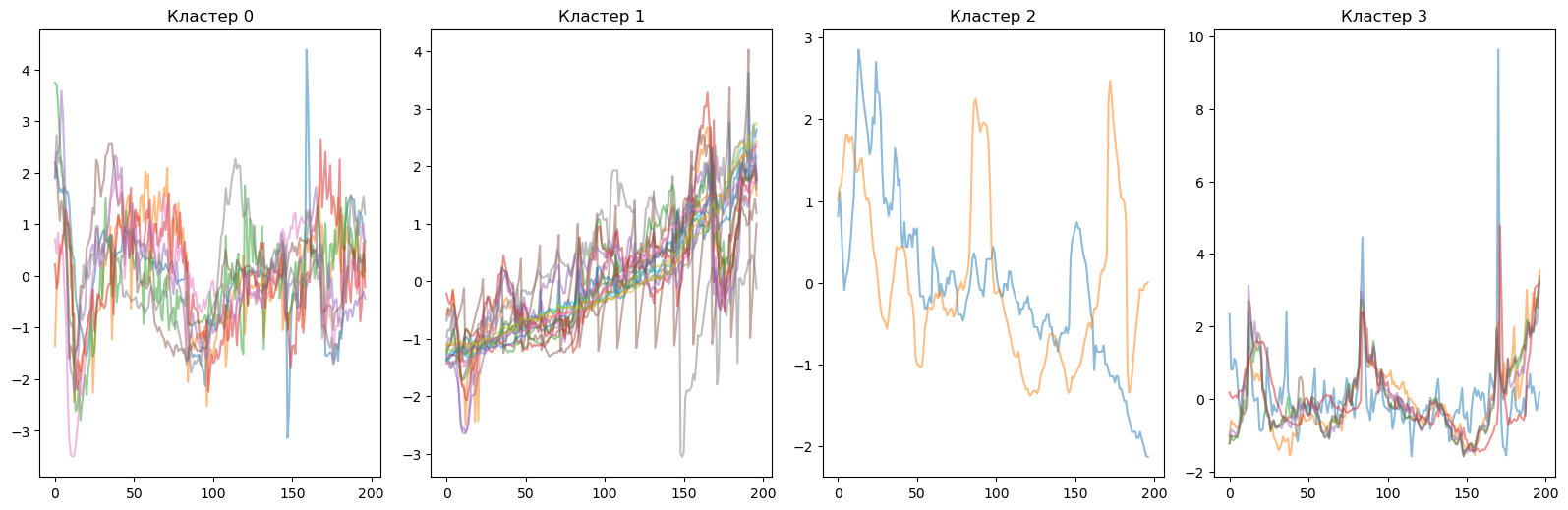
Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Оптимальное количество кластеров было выбрано с использованием метода локтя. Этот метод позволяет определить точку, на которой дальнейшее увеличение количества кластеров перестает значительно уменьшать сумму квадратов расстояний внутри кластеров. В результате оптимальным выбором оказалось 4 кластера.

Результаты кластеризации показали четкое разделение временных рядов на 4 группы. Центроиды кластеров отражают средний паттерн временных рядов внутри каждой группы, что позволило выделить общие тренды и закономерности.

Для наглядного представления результатов были построены графики центроидов кластеров и индивидуальные графики для каждого временного ряда внутри кластеров. Это позволило увидеть, как именно ряды распределились по группам и какие характеристики у них общие.



**1. Кластер 0**

График показывает ряды с высокой изменчивостью на начальных участках и резкими колебаниями по всей длине временной шкалы. Такой паттерн может указывать на неустойчивые или волатильные временные ряды, которые могут быть подвержены внезапным скачкам.

**2. Кластер 1**

В этой группе наблюдается ярко выраженный восходящий тренд с нарастающей динамикой к концу временного периода. Это указывает на ряды с устойчивым ростом, несмотря на небольшие колебания. Такие временные ряды часто характерны для трендовых индикаторов.

**3. Кластер 2**

Этот кластер содержит ряды с выраженными волнообразными колебаниями и устойчивым спадом на финальных отрезках временной шкалы. Паттерн указывает на ряды с сезонной или циклической динамикой, что может быть связано с периодическими факторами или тенденциями.

**4. Кластер 3**

График показывает умеренные колебания с чётко выраженными пиками на определённых участках временной шкалы. Такие пики могут быть вызваны аномальными событиями или временными всплесками активности в рядах.

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

График центроидов наглядно показывает усреднённые паттерны для каждого кластера.

* **Кластер 0** характеризуется значительными флуктуациями.
* **Кластер 1** демонстрирует устойчивый рост.
* **Кластер 2** выделяется волнообразной динамикой.
* **Кластер 3** показывает умеренные колебания с аномальными пиками.

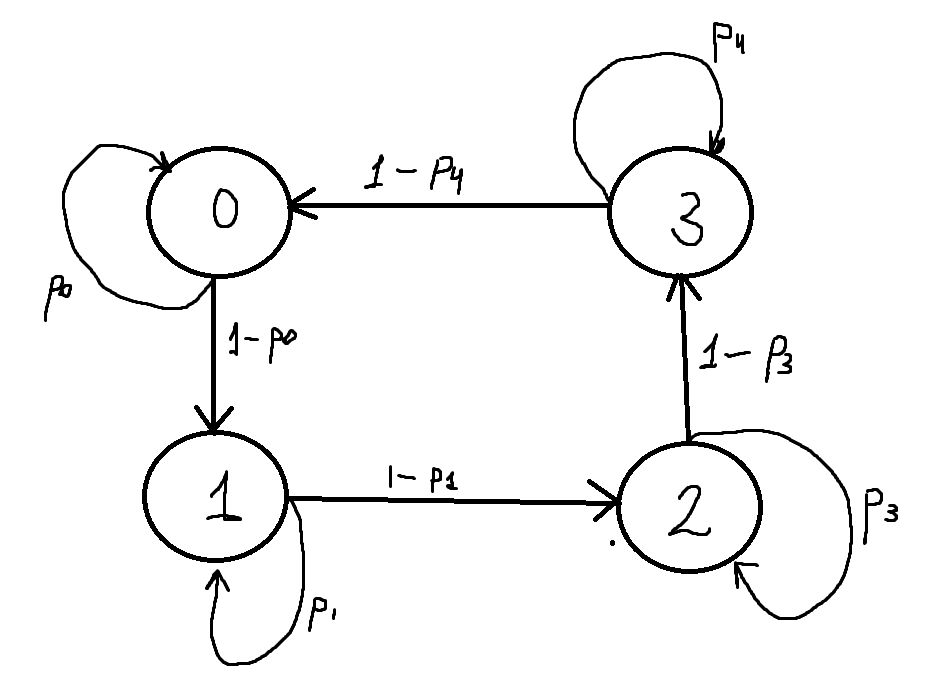
Центроиды помогают обобщить ключевые характеристики каждой группы, позволяя быстро понять структуру и динамику временных рядов в каждом кластере.

**Моделирование**

С помощью PCA рассчитываем компоненту для каждого кластера для каждого месяца. Получаем датафрейм, в котором в каждый момент времени есть 4 компоненты. На нем обучаем скрытую марковскую модель из библиотеки hmm.   
На вход идет наш датафрейм, значения компонент будут использоваться для корректировки скрытого слоя(матрицы эмиссий), матрица переходов, которую мы задаем вручную:

где ,,, – это вероятности системы(в нашем случае экономики) остаться в текущем состоянии

Таким образом мы вводим условие, что из одного состояния можно попасть только в следующее по порядку, то есть пик -> спад -> рецессия -> рост



Также мы подаем на вход начальные вероятности нахождения в каком то состоянии. Для воспроизводимости графика, устанавливаем, что система стартует с 1-го состояния.

Сама модель работает на основе двух алгоритмов: Баума-Велша и Витерби.

Обучение происходит с помощью алгоритма Баума-Велша и состоит из двух повторяющихся шагов:

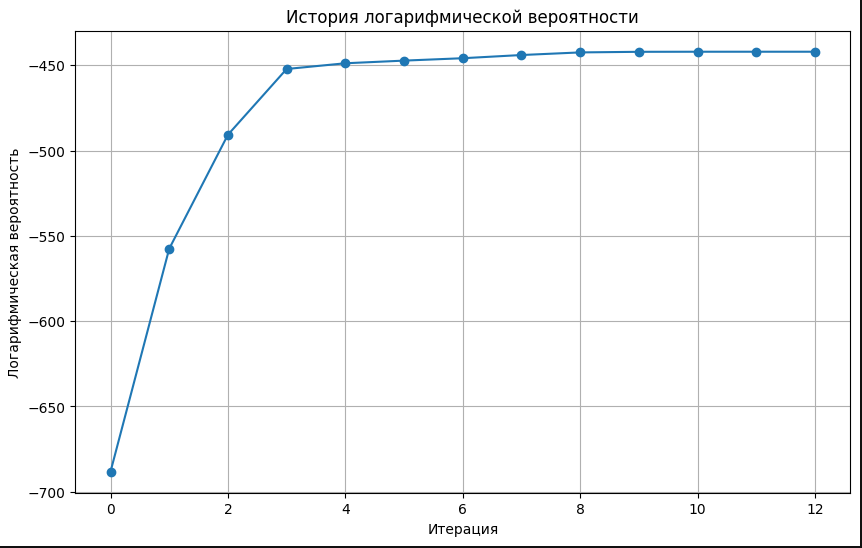
1. На основе начальных вероятностей, значений компонент и начальной инициализации матрицы переходов, рассчитываются вероятности для каждого момента времени нахождения в каком-либо состоянии
2. Корректируется матрица переходов и эмиссий. То есть если переход из пика в спад происходит чаще, чем начальное значение, то мы меняем соответствующее значение в матрице переходов. Соответственно, если при большем значение 1-го кластера(например) чаще мы делаем переход в 1-ое состояние, мы меняем значение в матрице эмиссий

Этот алгоритм повторяется до сходимости tol или по количеству, указываемому, как параметр n\_iter.

После того, как модель обучилась и матрицы скорректированы, мы «предсказываем» - размечаем данные соответствующими состояниями с помощью алгоритма Витерби:

1. Для каждого временного шага и каждого состояния вычисляем максимальную вероятность пути, который заканчивается в этом состоянии
2. Выбираем состояние с максимальной вероятностью
3. Начиная с последнего состояния, восстанавливаем последовательность состояний, используя сохраненную информацию о предыдущих состояниях

Оба алгоритма уже реализованы «под капотом» библиотеки hmm, поэтому остается лишь подобрать необходимые параметры.

Так как алгоритм Баума Велша, по сути, максимизирует логарифмическую вероятность, то ее можно вывести на каждой итерации, чтобы подобрать достаточное количество шагов для сходимости модели  


В нашем случае, модель сошлась за 12 итераций.

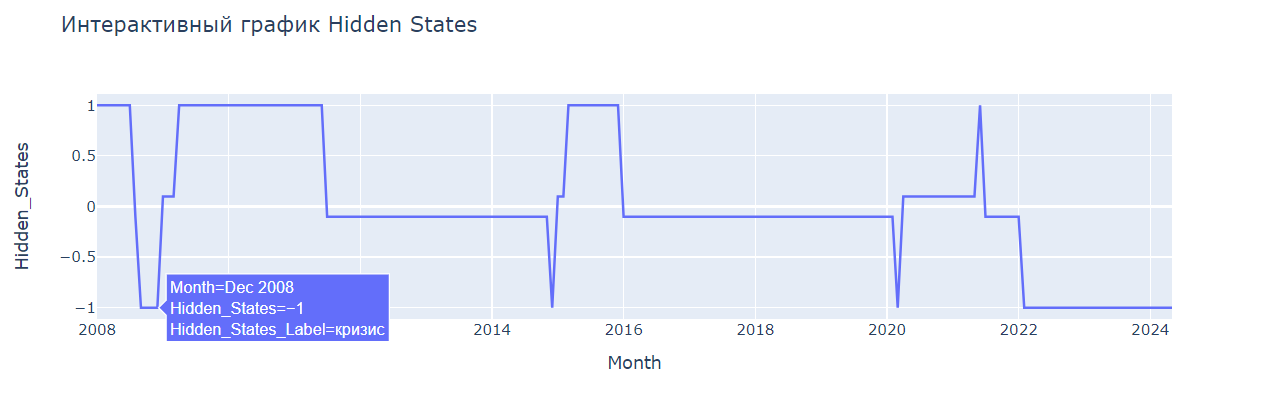
Последним этапом, мы визуализируем полученные предсказания положений бизнес цикла по временным меткам с помощью библиотеки plotly.express  


График интерактивный, можно посмотреть в каждый момент месяца, в каком положении бизнес-цикла находилась Россия.

Далее мы берем все данные за июнь 2024 года, добавляем к обучающему датасету и делаем прогноз с помощью модели на этот месяц на основе откалиброванных матриц перехода и предыдущего состояния экономики. Модель предсказывает продолжение кризиса в этом периоде, что согласуется с реальной экономической ситуацией в этом периоде – продолжающееся давление санкций, напряженная геополитическая обстановка, растущая инфляция, нехватка работников в промышленных отраслях.

Также мы добавили загрузку начальных параметров модели для того, чтобы можно воспроизвести обучение модели, и откалиброванных параметров, чтобы можно было сделать предсказание, без обучения (пример есть в основном скрипте).

**Интерпретация и анализ результатов**

Полученные результаты соответствуют реальным процессам, происходившим в нашей стране за последние 15 лет. На графике четко видны кризисы 2008 года (ипотечный кризис 2007 года, до России дошел через год), кризис 2015 года (валютный кризис, вызванный падением цен на нефть), кризис 2020 года (пандемия covid-19) и кризис 2022 года (начало СВО). Поэтому делаем вывод, что модель получилась адекватной и пригодной для анализа экономического цикла России.