**Введение**

Прогнозирование временных рядов играет значимую роль в современной экономике и позволяет эффективно принимать управленческие решения в условиях неопределённости. Особую актуальность такие исследования приобретают при анализе внешнеэкономической деятельности стран, где показатели экспорта служат важным индикатором экономического развития и интеграции страны в глобальные рынки. В настоящей работе объектом исследования является временной ряд общего объёма экспорта Российской Федерации, выраженный в миллиардах долларов США за период с 1994 по 2021 годы.

Целью данного исследования является выявление наилучшей модели прогнозирования объёмов экспорта из России.

Для достижения поставленной цели решаются следующие задачи:

- проведение предварительного анализа данных;

- выявление и учёт сезонности;

- тестирование временного ряда на стационарность;

- выбор и оценка нескольких моделей прогнозирования, таких как наивная модель, ETS, SARIMA, случайный лес и метод Тета, сравнение их точности;

- построение прогнозов на два года вперёд

Особенностью данной работы является использование широкого спектра методов анализа временных рядов, включая классические статистические подходы и алгоритмы машинного обучения, а также оценку эффективности комбинаций прогнозов. Кроме того, автором проведена диагностика ряда на наличие структурных сдвигов и аномалий, что повышает надёжность прогнозных моделей.

Структура работы включает анализ и визуализацию исходных данных, разделение временного ряда на обучающую и тестовую выборки, построение и оценку моделей, выбор и проверку лучшей модели, а также интерпретацию результатов. Ожидается, что результаты исследования будут полезны для оценки экономических последствий и выработки рекомендаций в сфере международной торговли и внешнеэкономической политики.

**Обработка данных перед моделированием**

В данном исследовании анализируется временной ряд, представляющий собой ежемесячные данные об объёме экспорта из Российской Федерации, выраженные в миллиардах долларов США. Данные охватывают период с января 1994 года по декабрь 2021 года и включают 336 наблюдений.

Исходные данные были представлены в виде таблицы с тремя столбцами: год, месяц и экспорт (в млрд. долларов). Для анализа и прогнозирования данные были преобразованы в формат временного ряда с периодичностью 12 наблюдений в год (ежемесячно).

На предварительном этапе обработки данных были выявлены и устранены выбросы и пропущенные значения. Это позволило стабилизировать данные и снизить влияние экстремальных значений на результаты прогнозирования.

Изображение выглядит как текст, График, линия, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Визуализация исходного временного ряда показывает выраженный восходящий тренд, особенно заметный в период с 2000 по 2008 год, с последующими значительными колебаниями, связанными с глобальным финансовым кризисом и последующими экономическими событиями. Также в данных явно присутствует сезонность с регулярными годовыми колебаниями.

Затем было проведено обычное дифференцирование временного ряда и сезонное дифференцирование.

Изображение выглядит как линия, Шрифт, типография

Контент, сгенерированный ИИ, может содержать ошибки.

Изображение выглядит как текст, линия, График, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

Для дальнейшего исследования было проведено изучение автокорреляционной (ACF) и частичной автокорреляционной (PACF) функций исходного ряда, а также ряда после взятия обычной и сезонной разностей.

|  |  |  |
| --- | --- | --- |
| **Ряд** | **ACF / PACF** | **Комментарий** |
| Исходный временной ряд |  | ACF показывает долгосрочную зависимость, подтверждает нестационарность. |
| Исходный временной ряд |  | Есть значимые пики на первых лагах, что также подтверждает необходимость дифференцирования ряда для достижения стационарности. |
| Временной ряд после обычного дифференцирования |  | Большинство автокорреляций лежат в пределах доверительного интервала, это говорит об успешном устранении долгосрочного тренда. Но остаются значимые автокорреляции на некоторых сезонных лагах, сезонность сохраняется и её необходимо дополнительно учитывать при моделировании. |
| Временной ряд после обычного дифференцирования |  | PACF указывает на сильную связь на первом лаге (p=1). |
| Временной ряд после сезонного дифференцирования |  | ACF показывает пики на сезонных лагах, важен сезонный компонент. |
| Временной ряд после сезонного дифференцирования |  | Значительное снижение автокорреляций. На первых лагах наблюдаются единичные пики, что также подтверждает успешное устранение сезонной составляющей. |

Анализ ACF и PACF исходного ряда подтверждает наличие тренда и сезонности, что делает необходимым проведение процедуры дифференцирования для обеспечения стационарности ряда. После применения обычного дифференцирования автокорреляционная функция существенно изменила свой вид. Значительно сократилось число значимых корреляций. Тем не менее, остаются значимые автокорреляции на некоторых сезонных лагах, что говорит о том, что сезонность сохраняется и её необходимо дополнительно учитывать при моделировании. На графике PACF для обычной разности наблюдаются единичные значимые значения на первых лагах, указывающие на возможные структуры авторегрессии низких порядков. Это подтверждает отсутствие тренда, но указывает на наличие остаточной сезонной компоненты.

После сезонного дифференцирования значения автокорреляционной функции значительно сократились и находятся ближе к нулю. Это значит, что сезонные колебания были эффективно устранены. Остаётся некоторое количество незначительных автокорреляций, которые могут быть скорректированы выбором соответствующих параметров модели SARIMA.

График частичной автокорреляционной функции после сезонного дифференцирования показывает значительное снижение автокорреляций. На первых лагах наблюдаются единичные пики, что также подтверждает успешное устранение сезонной составляющей.

Проведённый анализ ACF и PACF подтверждает необходимость применения как обычного, так и сезонного дифференцирования для достижения стационарности временного ряда экспорта РФ. Применённые преобразования сделали ряд подходящим для дальнейшего прогнозирования с использованием сезонных моделей типа SARIMA, а также других методов, требующих стационарности исходного ряда. На основании данного шага будем использовать модель с учётом тренда и сезонности, что будет учтено в следующем шаге при разделении ряда на обучающую и тестовую выборки.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Статистика Дики-Фуллера | Количество лэгов | p-значение |  |
| Исходный временной ряд | -2.2265 | 6 | 0.4811 | Нестационарный |
| Временной ряд после обычного дифференцирования | -5.298 | 6 | 0.01 | Стационарный |
| Временной ряд после сезонного дифференцирования | -4.3258 | 6 | 0.01 | Стационарный |

Проверка ряда на стационарность с помощью теста Дики-Фуллера (ADF-тест) показала, что исходный ряд является нестационарным (p-value = 0.4811), что подтверждается наличием явного тренда. Для приведения ряда к стационарности были применены обычное и сезонное дифференцирование. Тесты после дифференцирования подтвердили достижение стационарности как после обычного дифференцирования (p-value < 0.01), так и после сезонного дифференцирования (p-value < 0.01).

Разложение ряда на компоненты с помощью метода STL (Seasonal-Trend decomposition using Loess) также помогло отдельно выделить трендовую, сезонную и остаточную компоненты.

Изображение выглядит как линия, График, чек

Контент, сгенерированный ИИ, может содержать ошибки.

Тренд демонстрирует долгосрочные изменения объема экспорта, отражая общую экономическую ситуацию и влияние внешних факторов. Сезонная компонента явно выражена и повторяется из года в год, что обусловлено регулярными колебаниями в торговой активности.

Нам удалость подготовить временной ряд к дальнейшему моделированию и прогнозированию, подтвердив необходимость учёта тренда и сезонности при выборе и построении прогнозных моделей.

**Моделирование данных**

Для построения прогнозных моделей временной ряд был разделён на две выборки (обучающая и тестовая). Разделение выборок позволяет проверить способность моделей прогнозировать будущие значения, минимизируя переобучение и оптимизируя параметры.

Обучающая выборка (период с января 1994 года по декабрь 2018 года) включает данные, которые будут использоваться для построения моделей. Тестовая выборка (период с января 2019 года по декабрь 2021 года) предназначены для оценки точности прогнозов.

Для прогнозирования временного ряда были использованы наивная модель, модель экспоненциального сглаживания, Сезонная авторегрессионная интегрированная модель скользящего среднего (SARIMA), Метод Тета.

|  |  |  |
| --- | --- | --- |
| **Модель** | **Уравнение модели** | **Описание** |
| Наивная модель | Изображение выглядит как Шрифт, типография, символ, белый  Контент, сгенерированный ИИ, может содержать ошибки. | Прогноз строится на основе последнего наблюдения: все будущие значения принимаются равными последнему доступному значению временного ряда. Данная модель служит базовым ориентиром для оценки качества более сложных методов. |
| Модель экспоненциального сглаживания (ETS) | Изображение выглядит как Шрифт, текст, типография, белый  Контент, сгенерированный ИИ, может содержать ошибки. | ETS учитывает ошибку, тренд и сезонность. Модель оптимально адаптируется к данным, определяя компоненты автоматически. |
| Сезонная авторегрессионная интегрированная модель скользящего среднего (SARIMA) |  | Учёт тренда, сезонности и лаговых эффектов делает эту модель мощным инструментом для прогнозирования временных рядов. |
| Метод Тета | Изображение выглядит как Шрифт, типография, рукописный текст, каллиграфия  Контент, сгенерированный ИИ, может содержать ошибки. | Раскладывает временной ряд на трендовую и сезонную компоненты. Данный метод демонстрирует хорошие результаты в соревнованиях по прогнозированию временных рядов. |

Для оценки точности моделей использовались метрики RMSE и MAE. RMSE (корень из среднеквадратичной ошибки) измеряет среднее отклонение прогнозируемых значений от фактических. MAE (средняя абсолютная ошибка) оценивает средний уровень ошибки без учёта её знака.

|  |  |  |
| --- | --- | --- |
| **Модель** | **RMSE** | **MAE** |
| Naive | 9.784297 | 8.649571 |
| ETS | 8.520003 | 7.203654 |
| SARIMA | 11.377975 | 9.739472 |
| Theta | 7.375107 | 6.249774 |

Модель Theta показала наилучший результат по RMSE и MAE, что подтверждает её хорошую способность адаптироваться к характеристикам временного ряда. SARIMA продемонстрировала высокий RMSE, что может быть связано с сложной структурой ряда и необходимостью дополнительной оптимизации параметров. ETS оказалась конкурентоспособной, но уступила Theta в точности. Наивная модель показала наихудшие результаты.

В итоге для прогнозирования временного ряда экспорта из России была выбрана модель Theta, которая продемонстрировала наилучшую точность среди рассмотренных методов (по показателям RMSE и MAE).

Прогноз был построен на 24 месяца вперёд. Основной принцип работы модели Theta заключается в разложении временного ряда на компоненты и построении линейной экстраполяции тренда. Эта модель хорошо зарекомендовала себя в соревнованиях по прогнозированию временных рядов, так как она комбинирует краткосрочные колебания с долгосрочным трендом.

Прогноз был представлен в виде точечного прогноза (наиболее вероятное значение экспорта на каждый месяц прогнозного периода) и доверительных интервалов (80%-й интервал и 95%-й интервал).

Изображение выглядит как текст, черно-белый, Шрифт, снимок экрана

Контент, сгенерированный ИИ, может содержать ошибки.

Прогнозные значения показывают умеренный рост экспорта, однако диапазоны неопределённости увеличиваются по мере продвижения вперёд во времени. Это ожидаемый результат, так как дальние прогнозы обычно имеют более высокую вариативность. Прогноз учитывает сезонность, что отражается в изменениях точечных прогнозов по месяцам. По мере увеличения горизонта прогнозирования доверительные интервалы расширяются, что свидетельствует о повышенной неопределённости. На основе модели можно предположить, что общий тренд сохраняет положительную динамику, несмотря на возможные краткосрочные колебания. Доверительные интервалы позволяют оценить неопределённость прогноза и потенциальные колебания экспорта.

Изображение выглядит как линия, снимок экрана, текст, диаграмма

Контент, сгенерированный ИИ, может содержать ошибки.График остатков модели Theta подтверждает качество модели. Остатки модели колеблются вокруг нуля. Это свидетельствует об отсутствии явного смещения. ACF остатков демонстрирует, что автокорреляция остатков в пределах допустимых границ, что подтверждает их случайное распределение. Гистограмма остатков показывает, что они близки к нормальному распределению, но возможны небольшие выбросы.

График прогноза на 2 года вперёд отражает диапазон доверительных интервалов: чем шире область, тем выше неопределённость. Линия точечного прогноза демонстрирует плавный рост, соответствующий выявленным историческим трендам.

Изображение выглядит как График, линия, снимок экрана, диаграмма

Контент, сгенерированный ИИ, может содержать ошибки.

Прогнозирование экспорта из России с использованием модели Theta позволило получить достоверные оценки будущих значений, учитывая сезонность и долгосрочные тренды. На основе результатов можно сделать вывод о потенциальном росте экспорта в обозримом будущем, однако диапазон неопределённости увеличивается, что требует осторожности при интерпретации дальних прогнозов.

В процессе проведения исследования и построения прогнозных моделей временного ряда экспорта из России возникло несколько значительных проблем, требующих методического подхода к их решению.

Анализ исходных данных выявил наличие выраженных колебаний, что затрудняло применение классических моделей прогнозирования. Значительные скачки экспорта могли быть вызваны внешними экономическими шоками, сезонными изменениями и политическими факторами. Проведена первичная очистка данных, которая позволила сгладить выбросы.

Выбор модели прогнозирования требовал тщательной оценки, поскольку различные подходы (SARIMA, ETS, Theta) обладают разной адаптивностью к структуре временного ряда. Необходимо было не только подобрать модель, обеспечивающую минимальные ошибки, но и оценить устойчивость прогноза. Для каждой модели были рассчитаны ключевые метрики точности: RMSE (корень из среднеквадратичной ошибки) и MAE (средняя абсолютная ошибка).

**Заключение**

Исследование временного ряда экспорта из России за период 1994–2021 гг. позволило провести детальный анализ динамики и разработать прогнозные модели для оценки будущих значений. В ходе работы были выполнены ключевые этапы: очистка и подготовка данных, анализ автокорреляционной структуры, применение и оценка прогнозных моделей, выбор оптимальной модели и построение прогноза. Общий тренд экспорта демонстрирует рост, что согласуется с историческими данными. Прогнозные значения указывают на потенциальное увеличение экспорта, но с высокой неопределённостью. Модель Theta показала наилучшие результаты среди протестированных методов, обеспечив наименьшие ошибки и наиболее адекватное поведение остатков.

Проведённый анализ позволяет сделать обоснованные выводы о будущей динамике экспорта, однако прогноз требует учёта макроэкономических факторов и возможных внешнеэкономических шоков. Для дальнейших исследований возможно расширить модель, включив дополнительные предикторы (цены на сырьевые товары, валютные курсы, санкционные факторы) и это позволит повысить точность прогнозирования.