**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 2](#_Toc58017197)

[1 Основные понятия и определения 4](#_Toc58017198)

[2 Классификация моделей временных рядов 9](#_Toc58017199)

[2.1 Регрессионные модели 10](#_Toc58017200)

[2.2 Адаптивные методы прогнозирования временных рядов 14](#_Toc58017201)

[2.3 Авторегрессионные модели прогнозирования 19](#_Toc58017202)

[2.4 Нейросетевые модели 24](#_Toc58017203)

[2.5 LSTM 27](#_Toc58017204)

[2.6 Модели на базе цепей Маркова 30](#_Toc58017205)

[2.7 Модели на базе деревьев классификации и регрессии 32](#_Toc58017206)

[3 Практическая часть 34](#_Toc58017207)

[ЗАКЛЮЧЕНИЕ 48](#_Toc58017208)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ: 49](#_Toc58017209)

[Приложение А. Листинг программы 51](#_Toc58017210)

[Приложение Б. Графики прогнозов экспорта 52](#_Toc58017211)

[Приложение В. Графики прогнозов импорта 59](#_Toc58017212)

[Приложение Г. Таблица оценки качества модели на данных объёма экспорта 67](#_Toc58017213)

# ВВЕДЕНИЕ

**Актуальность.** Показания приборов говорят нам о таких вещах, как начальная температура, концентрация реагента, процент катализатора и т.д. Некоторые из этих показаний можно отслеживать через равные интервалы времени, например, каждые пять минут, а некоторые можно измерять непрерывно. Полученные данные могут быть зафиксированы в различном виде, чаще всего применяются специальные таблицы, в которых записываются процесс и соответствующее для него значение параметров в конкретный момент времени. Такие данные можно не просто хранить, а можно использовать для анализа и исследования на предмет существования зависимостей.

Для любых задач с изменяющимися количественным переменными особый интерес представляет исследование влияния некоторых переменных на остальные.

Искусственные нейронные сети представляют собой современную технологию, которая способна обучаться на основе данных при участии учителя и без его вмешательства. Нейронные сети хорошо применяются не только для задач распознавания зрительных образов, задач классификации, но и для задач прогнозирования. Помимо нейросетевой модели для задачи прогнозирования будут рассмотрены некоторые модели прогнозирования временных рядов, которые наиболее актуальны на сегодняшний день.

**Цели и задачи.** Целью работы является проверка возможности прогнозирования суммарного торгового трафика импорта для некоторых крупных стран нейросетевыми методами. Разработка новой модели и соответствующего ей метода прогнозирования.

Для этого необходимо решить следующие задачи:

1. проанализировать предметную область;
2. осуществить обзор моделей и методов прогнозирования временных рядов, выявить достоинства и недостатки каждого класса моделей;
3. выявить наиболее эффективные классы моделей прогнозирования, определить перспективные подходы, которые могут позволить устранить недостатки авторегрессионного класса моделей;
4. разработать собственную модель для прогнозирования на основе нейронных сетей; выполнить программную реализацию алгоритмов;
5. оценить эффективность предложенной модели прогнозирования при решении задачи прогнозирования различных временных рядов.

**Объект и предмет исследования.** Объектом исследования работы являются временные ряды. Предметом исследования является модель, которая построена нейросетевыми методами для прогнозирования временного ряда.

**Метод исследования.** Для решения поставленных задач в работе использованы метода математического моделирования, анализ временных рядов, регрессионный анализ, методы объектно-ориентированного программирования.

**Результаты исследования.** В ходе исследования были получены результаты:

1. выявлены недостатки моделей прогнозирования;
2. разработана модель и метод прогнозирования временных рядов.

**Структура работы.** Работа включает введение, три главы, заключение, список использованных источников.

В первой главе описаны основные понятия и определения: временные ряды, сезонность, цикличность, тренды.

Во второй главе рассмотрена классификация моделей прогнозирования, описаны статические и структурные модели прогнозирования, рассмотрены их недостатки и достоинства.

В третьей главе описана программная реализация нейросетевой модели, которая была сделана с использованием библиотек tensorflow/keras, языка программирования Python. Также в этой главе проанализирована область применения реализованной нейросетевой модели.

# 1 Основные понятия и определения

Пусть дан временной ряд где – элемент временного ряда. – модель временного ряда, где , – горизонт прогнозирования, – вектор параметров модели.

Одним из главных требований к временному ряду является его стационарность, которая состоит в том, что распределение его значений является инвариантным относительно момента времени, для которого оно построено. Для характеристики стационарности используется то, что для двух выборок, которые были построены в разные моменты времени, закон распределения должен оставаться тем же. Если же элемент рассматриваемых временных рядов многомерен, то следует анализировать каждый компонент элемента временного ряда и проверить, что он имеет равномерное распределение на отрезке , где – математическое ожидание, – её дисперсия. [1]

Далее будут рассматриваться эконометрические временные ряды, для которых характерны явления:

1. тренды;

2. сезонная компонента;

3. циклическая компонента.

Рассмотрим каждое из явлений более подробно.

Под трендом понимается длительная тенденция изменения показателей временного ряда. Наиболее распространённым способом моделирования тенденций временных рядов является внедрение и построение различных аналитических функций, которые характеризуют зависимость уровней ряда от времени. Используются линейные, степенные или экспоненциальные функции. Например, уравнение прямой линии , парабола 2-го порядка , логарифмическая , степенная , показательная , гиперболическая , логистическая , тригонометрическая . Также возможно использование комбинированных функций.

Тип тренда устанавливают на основе данных временного ряда, путём осреднения показателей динамики ряда, на основе статистической проверки гипотезы о постоянстве параметров графика.

Методы оценки, как правило, разделяют на параметрические и непараметрические. Вторые не применяются для прогнозирования, они необходимы в том случае, когда не удаётся найти подходящую функцию. Параметрические же, в свою очередь, рассматривают временной ряд как гладкую функцию от : , ; после чего различными методами оцениваются параметры функции , например, методом наименьших квадратов. Метод аналитического выравнивания [2] позволяет оптимально подобрать функцию, которая будет наилучшим образом описывать зависимость уровней именно этого временного ряда от времени.Параметры функции определяются на основе метода наименьших квадратов, который можно представить следующей формулой:

Этот метод заключается в минимизации квадрата евклидова расстояния между двумя векторами – вектором восстановленных значений зависимой переменной и вектором фактических значений зависимой переменной. Этот метод также называют методом нахождения оптимальных параметров линейной регрессии, таких, что сумма квадратов ошибок (регрессионных остатков) минимальна.

Циклическая компонента отражает циклические изменения уровней временного ряда для периодов свыше 1 года. Также стоит отметить, что циклическая компонента связана с циклами деловой активности, её периодичность составляет от 2 до 10 лет. Эту компоненту сложно идентифицировать, если анализировать данные за непродолжительный, относительно цикла, период времени. В этом случае компоненту невозможно отделить от трендовой.

Под сезонностью понимают периодические колебания, наблюдаемые на временных рядах. В экономике многие явления характеризуются периодически повторяющимися сезонными эффектами. Например, розничные продажи растут с приближением новогодних праздников, а после них спад. Соответственно, временные ряды, отражающие эти сезонные эффекты, содержат периодические колебания.

Перед выделением сезонных колебаний необходимо вычислить период сезонности. В большинстве случаев период известен из контекста задачи (если рассматривать розничные продажи, то период будет равен году). Однако если период не известен заранее, то его можно найти с помощью автокорреляционной функции. Таким образом, сезонная компонента может отражать квартальные, месячные и недельные циклы.

Функции обнаружения сезонности встроены во многие программы, предназначенные для работы со статистическими данными, такие как Statistica.

Также иногда выделяют нерегулярную компоненту, которая отражает нерегулярные колебания уровней временного ряда, которые невозможно предсказать. Эта компонента является следствием однократных, а не систематических событий, влияющих на уровни ряда.

Существует два способа, с помощью которых компоненты временных рядов могут взаимодействовать:

1. аддитивная модель:

;

1. мультипликативная модель:

,

где – уровень ряда динамики, – трендовая компонента, – циклическая компонента, – сезонная компонента, – нерегулярная компонента. [3]

Выбор той или иной модели зависит от характера исходных данных. Например, если каждый год амплитуда циклических и сезонных изменений носит постоянный характер, то чаще всего используют аддитивную модель, при росте амплитуды этих изменений вместе с ростом показателей – мультипликативную. Стоит отметить, что в практике прогнозирования чаще всего используют мультипликативную модель.

Адаптивная модель с мультипликативной сезонностью была предложена П. Р. Уинтерсом [4]. Из-за экспоненциальной схемы модель становится сложнее, зато и точность прогнозов для большинства товаров существенно возрастает.

Говоря о прогнозировании временных рядов, необходимо различить два взаимосвязанных понятия: метод прогнозирования, модель прогнозирования.

Метод прогнозирования – последовательность действий, которые нужно совершить для получения модели прогнозирования временного ряда. Модель прогнозирования – функциональное представление, которое адекватно описывает временной ряд и является основой для получения будущих значений процесса.

Метод прогнозирования содержит последовательность действий, в результате выполнения которой определяется модель прогнозирования конкретного временного ряда. Метод прогнозирования также включает действия по оценке качества прогнозных значений.

Общий итеративный подход к построению модели прогнозирования состоит из следующий шагов.

Шаг 1. На первом шаге на основании предыдущего собственного или стороннего опыта выбирается общий класс моделей для прогнозирования временного ряда на заданный горизонт.

Шаг 2. Определенный общий класс моделей обширен. Для непосредственной подгонки к исходному временному ряду, развиваются грубые методы идентификации подклассов моделей. Такие методы идентификации используют качественные оценки временного ряда.

Шаг 3. После определения подкласса модели, необходимо оценить ее параметры, если модель содержит параметры, или структуру, если модель относится к категории структурных моделей. На данном этапе обычно используется итеративные способы, когда производится оценка участка (или всего) временного ряда при различных значениях изменяемых величин. Как правило, данный шаг является наиболее трудоемким в связи с тем, что часто в расчет принимаются все доступные исторические значения временного ряда.

Шаг 4. Далее производится диагностическая проверка полученной модели прогнозирования. Чаще всего выбирается участок или несколько участков временного ряда, достаточных по длине для проверочного прогнозирования и последующей оценки точности прогноза. Выбранные для диагностики модели прогнозирования участки временного ряда называются контрольными участками (периодами).

Шаг 5. В случае если точность диагностического прогнозирования оказалась приемлемой для задач, в которых используются прогнозные значения, то модель готова к использованию. В случае если точность прогнозирования оказалось недостаточной для последующего использования прогнозных значений, то возможно итеративное повторение всех описанных выше шагов, начиная с первого.

# 2 Классификация моделей временных рядов

Обычно модели временных рядов разделяют на статистические и структурные.

В статистических моделях зависимость будущего значения от прошлого задается в виде некоторого уравнения. К ним относятся:

1. регрессионные модели (линейная регрессия, нелинейная регрессия);
2. авторегрессионные модели (ARIMAX, GARCH, ARDLM);
3. модель экспоненциального сглаживания;
4. модель по выборке максимального подобия;

В структурных моделях зависимость будущего значения от прошлого задается в виде некоторой структуры и правил перехода по ней. К ним относятся:

1. нейросетевые модели;
2. модели на базе цепей Маркова;
3. модели на базе классификационно-регрессионных деревьев.

Классификация моделей временных рядов представлена на рисунке 1.

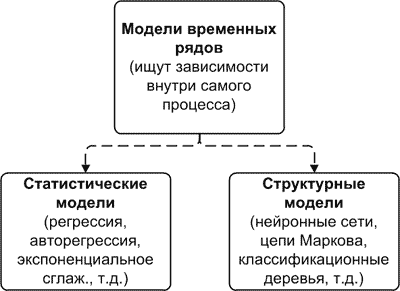


Рисунок 1 – Классификация моделей временных рядов

Далее рассмотрим более подробно каждый из разделов моделей.

# 2.1 Регрессионные модели

Регрессия – зависимость математического ожидания (например, среднего значения) случайной величины от одной или нескольких других случайных величин (свободных переменных):

,

где – зависимая переменная, – объясняющая переменная.

Регрессионным анализом называется поиск такой функции , которая наилучшим образом приближает данные.

Критерием качества приближения (целевой функцией) обычно является среднеквадратичная ошибка:

*,*

где – фактический ожидаемый результат, – прогноз модели.

Регрессия может быть представлена в виде суммы неслучайной и случайной составляющих:

,

где – функция регрессионной зависимости, – аддитивная случайная величина с нулевым матожиданием.

Относительно характера распределения этой величины делаются предположения, называемые гипотезой порождения данных. Для подтверждения или опровержения этой гипотезы выполняются статистические тесты, называемые анализом остатков. При этом предполагается, что независимая переменная не содержит ошибок. Регрессионный анализ используется для прогноза, анализа временных рядов, тестирования гипотез и выявления скрытых взаимосвязей в данных.

Линейная регрессия – одна из важнейших и широко используемых техник регрессии. Эта самый простой метод регрессии. Одним из его достоинств является лёгкость интерпретации результатов. Как было замечено ранее, в основе модели регрессии лежат исследования зависимостей одной переменной от другой. Иногда две переменные связаны точным уравнением прямой линии. Уравнение прямой может быть полезно во многих ситуациях для обобщения наблюдаемой зависимости одной переменной от другой. Такое уравнение можно получить методом наименьших квадратов, когда существуют экспериментальные данные.

Пусть линия регрессии переменной . Тогда можно записать линейную модель

,

так что для данного соответствующее значение состоит из величины плюс добавка в виде , при учёте которой любой индивидуальный получает возможность не попасть на линию регрессии.

Уравнение линейной модели, которое было записано выше, представляет из себя модель, в которую «мы верим». Начинать необходимо с предположения, что эта модель установлена, но она требует проверки на последующих стадиях. Первоначальное допущение верности установленной модели называют постулированием модели или допущение о её правильности. Величины , – параметры модели.

Когда говорят о том, что модель линейна или нелинейна, то имеют в виду линейность или нелинейность по параметрам. Величина наивысшей степени предикатора в модели называется порядком модели. Под предикаторами понимают прогностический параметр, средство прогнозирования. [5] Например, модель, которая имеет вид:

представляет из себя регрессионную модель второго порядка (по ) и линейная (по ). Обозначение вида чаще всего используется в полиномиальных моделях. [6]

Под многомерной регрессией понимают линейную регрессию в -мерном пространстве, где объекты и признаки являются -мерными векторами. Пусть существует множество объектов , множество ответов , набор вещественнозначных признаков , . Введём матричные обозначения: матрица информации , целевой вектор , вектор параметров и диагональная матрица весов:

*, , ,, .*

Алгоритм можно записать формулой:

.

Функционал квадрата ошибки в матричном виде:

Для точности оценки регрессии как линейной, так и многомерной модели можно воспользоваться методом наименьших квадратов. Основной проблемой многомерной линейной регрессии является вырожденность, или, в более общем случае, мультиколлинеарность матрицы , которую приходится обращать. Подобные проблемы возникают, когда среди признаков есть почти линейно зависимые. Мультиколлинеарность матрицы определяется её числом обусловленности:

,

где — собственные значения матрицы . Чем больше число обусловленности, тем ближе матрица к вырожденной и тем неустойчивее обратная к ней матрица. Плохая обусловленность матрицы: . Матрицу принято считать плохо обусловленной, если её число обусловленности превышает . [7]

Нелинейная регрессия – это способ нахождения нелинейной модели взаимосвязи между зависимой переменной и набором независимых переменных. В отличие от традиционной линейной регрессии, которая ограничена оценкой линейных моделей, нелинейная регрессия может оценивать модели с произвольными взаимосвязями между независимыми и зависимыми переменными. Это достигается при помощи итерационных алгоритмов оценки.

Нелинейная модель , .

Функционал среднеквадратичного отклонения:

Если функция – непрерывно дифференцируема, то можно применить метод стохастического градиента [8]. Для получения наилучших результатов с меньшими затратами на время применяют более трудоёмкий с точки зрения вычислений метод. Этот метод Ньютона-Рафсона можно описать следующими этапами:

1. Начальное приближение

.

2. Итерационный процесс

*,*

где – градиент функционала в точке , матрица вторых производных называется гессианом функционала в точке , – величина шага (чаще полагают, что ).

Линеаризация в окрестности текущего :

Метод Ньютона-Гаусса также хорошо помогает в сведении к линейной регрессии, как и метод Ньютона-Рафсона. Однако стоит отметить, что скорость сходимости зависит от выбора , от выбора самой модели, в частности её характеристика по приближению выборки.

Матричные обозначения для метода Ньютона-Гаусса:

– -матрица первых производных;

– вектор значений .

Формула t-й итерации метода Ньютона-Гаусса:

*.*

*–* это решение задачи многомерной линейной регрессии:

*.*

# 2.2 Адаптивные методы прогнозирования временных рядов

Адаптивные методы прогнозирования временных рядов представляют из себя методы, цель которых заключается в построении самокорректирующихся (самонастраивающихся) экономико-математических моделей, которые способны отражать изменяющиеся во времени условия, учитывать информационную ценность различных членов временной последовательности и давать достаточно точные оценки будущих членов данного ряда. Такие модели предназначаются, прежде всего, для краткосрочного прогнозирования.

Последовательность процесса адаптации в основном выглядит следующим образом. Пусть модель находится в некотором исходном состоянии (т.е. определены текущие значения ее параметров) и по ней делается прогноз. Выжидаем, пока истечет одна единица времени (шаг моделирования), и анализируем, насколько далек результат, полученный по модели, от фактического значения ряда. Ошибка прогнозирования через обратную связь поступает на вход системы и используется моделью в соответствии с ее логикой для перехода из одного состояния в другое с целью большего согласования своего поведения с динамикой ряда. На изменения ряда модель должна отвечать "компенсирующими" изменениями. Затем делается прогноз на следующий момент времени, и весь процесс повторяется.

Предполагаем, что задан временной ряд:  , где  - значение временного ряда в момент времени .  - прогноз значения временного ряда в момент времени , сделанное в момент времени .

Простейшие адаптивные модели:

1. экспоненциальное сглаживание, Модель Брауна;

2. модели линейного роста, которые включают в себя Модель Хольта, модель линейного роста Брауна, модель прогнозирования Дж. Бокса и Г. Дженкинса;

3. сезонные модели: модель Хольта-Уинтерса, модель Тейла-Вейджа.

При экспоненциальном сглаживании, в частности при Модели Брауна, предполагается, что ряд генерируется моделью, которая записана следующей формулой:

*.*

В этой формуле – есть варьирующий по времени средний уровень ряда, выступает в качестве белого шума. Прогноз временного ряда получается по формуле:

,

где – значение экспоненциальной средней в момент времени . можно найти, используя:

где , представляет из себя параметр сглаживания.

Главное достоинство такой прогнозной модели состоит в том, что она способна последовательно адаптироваться к новому уровню процесса без значительного реагирования на случайные отклонения.

Недостаток: экспоненциальная средняя дает систематическую ошибку, когда временной ряд имеет тенденцию линейного роста

Прогноз для моделей линейного роста может быть получен по уравнению:

,

где ,  – текущие оценки коэффициентов адаптивного полинома первого порядка.

В разных моделях эти коэффициенты вычисляются различными способами. Рассмотрим три модели и методы вычисления оценок коэффициентов адаптивного полинома первого порядка.

В [модели Хольта](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C_%D0%A5%D0%BE%D0%BB%D1%8C%D1%82%D0%B0) используются формулы:

,

,

где  выступают в качестве параметров адаптации, для которых должно выполняться: .

Если рассматривать модель линейного роста Брауна, стоит отметить, что эта модель является частным случаем модели Хольта и для вычисления оценок применяют:

;

,

где  – ошибка прогноза,  - коэффициент дисконтирования, характеризующий обесценивание данных наблюдения за единицу времени.

Для модели прогнозирования Дж.Бокса и Г.Дженкинса справедливы:

.

.

Данная модель не дает преимуществ перед моделью Хольта, так как коэффициент  часто оказывается близким к нулю.

Рассмотрим сезонные модели. А именно Модель Хольта-Уинтерса, которая учитывает мультипликативный тренд и сезонность, и Модель Тейла-Вейджа, которая, в свою очередь, будет брать во внимание аддитивный тренд и сезонность.

Рассматривая Модель Хольта-Уинтерса, нельзя не упомянуть о модель Хольта, ведь именно на основе этой модели Уинтерс создал свою прогностическую модель, которая способна учитывать экспоненциальный тренд и аддитивную сезонность.

В модели Хольта также необходимо решить задачу прогнозирования временного ряда. При существовании линейного тренда на данных временного ряда использование модели Брауна не имеет смысла. Для учёта влияния линейного тренда используют модель Хольта. Однако эта модель не учитывает сезонность.

Для описания математических формул для Модели Хольта-Уинтерса зададим временной ряд где . Для решения задачи прогнозирования временного ряда будет верно:

,

,

,

,

где – период сезонности, , – сезонный профиль, – параметр тренда, – параметр прогноза, который «очищен» от влияния тренда и сезонности.

Параметры , , должны находиться в диапазоне от до . Чаще всего их предлагается находить экспериментальным путём. Уинтерс также отмечал, что один такой набор весов можно использовать для широкого класса продуктов. [4]

Теперь рассмотрим усложнённую модель Хольта, которая учитывает сезонность и аддитивный тренд. Для неё справедливы формулы:

,

,

,

,

где – период сезонности, , – сезонный профиль, – параметр тренда, – параметр прогноза, который «очищен» от влияния тренда и сезонности. Используя метод минимизации квадратичной ошибки, выбор параметров , , происходит экспериментально, диапазон сохраняется, как и для модели Хольта-Уинтерса.

Рассмотрим стохастический процесс Тейла и Вейджа. При изучении экспоненциальной средней они предложили применить двухпараметрический предикатор Хольта для прогнозирования некоторого вероятностного процесса, который ярко характеризуется трендом. Процесс Тейла-Вейджа аналитически записывается так:

*;*

где – значение уровня исследуемого временного ряда в момент времени , – прирост уровня от момента к моменту , , – временный последовательности с нулевым математическим ожиданием, постоянными дисперсиями и отсутствуем ковариации. Временной ряд не является стационарным и не имеет строго определённой автоковариационной функции. Но Нерлов и Вейдж показали, что из вышеупомянутых уравнений следует стационарность вторых разностей процесса .

Схема составления прогноза выглядит следующим образом:

;

*,*

где . Если ошибку прогноза, сделанного в момент времени на шаг вперёд, обозначить через , то уравнения адаптации примут вид:

;

.

Ошибка прогноза является суммой трёх компонент: ошибки оценки уровня процесса в момент , ошибки оценки прироста уровня в момент и комбинации случайных компонент и в момент времени . Очевидно также, что определение оптимальных и определению оптимальных и .

Используя соотношение:

,

Нерлов и Вейдж показали, что проблема прогнозирования эквивалентна задаче прогнозирования второй разности и что при ограничениях, наложенных на параметры адаптации, ошибка прогноза является линейной комбинацией текущего и прошлых значений стационарного ряда :

*,*

где – сходящий ряд весов.

В результате минимизации дисперсии ошибки прогноза на один шаг вперёд были получены следующие результаты:

;;; ;

; *.* [9]

#### **2.3 Авторегрессионные модели прогнозирования**

#### Это широчайший и один из двух наиболее широко применимых классов моделей:

* ARIMAX (autoregression integrated moving average extended);
* GARCH (generalized autoregressive conditional heteroskedasticity);
* ARDLM (autoregression distributed lag model).

Авторегрессионная (AR) модель – модель временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда. Авторегрессионный процесс порядка *p* (AR(*p*)-процесс) определяется следующим образом:

,

где  {\displaystyle a\_{1},\ldots ,a\_{p}}– параметры модели (коэффициенты авторегрессии), {\displaystyle c}   – постоянная (часто для упрощения предполагается равной нулю), а {\displaystyle \varepsilon \_{t}} – белый шум.

Простейшим примером является авторегрессионный процесс первого порядка AR(1)-процесс:

.

Для данного процесса коэффициент авторегрессии совпадает с коэффициентом автокорреляции первого порядка.

Другой простой процесс – процесс Юла  – AR(2)-процесс:

.

Под ARIMAX-моделью (autoregression integrated moving average extended) понимают некую математическую модель для анализа временных рядов, которая объединяет в себе интегрированную авторегрессию, скользящее среднее и возможность учёта дополнительных внешних факторов. Также стоит отметить, что модели ARIMAX часто используют при решении задач, в которых требуется построить прогноз на основе имеющихся данных и вычислить соответствующие последующие значения ряда на основе предыдущих. Важным примечанием является тот факт, что для того чтобы получить прогнозируемые данные необходимо предварительное обучение узла.

Обучение узла может производиться вручную или автоматически. В некоторых компонентах используются самообучающиеся алгоритмы. Например, кластеризация, самоорганизующиеся карты, квантование. Такие узлы требуют первоначального обучения. Ручное обучение или переобучение позволяет контролировать параметры переобучения Модели с возможностью просмотра полученных результатов. Автоматический способ, в свою очередь, гораздо быстрее и хорошо подходит при незначительных изменениях в исходных данных.

Особенности настройки с помощью ARIMAX:

1. Настройка входных столбцов – заключается в правильном задании назначений столбцов входного набора данных. Необходимо определить, какой из столбцов является прогнозируемым, какой выступает в качестве входного, какой отвечает за те данные, которые не участвуют в построении модели.

2. Нормализация данных – этап, который обычно пропускают для прогнозируемых данных.

3. При настройке ARIMAX структура определение структуры модели происходит автоматически. Параметры в процессе вычисления подбираются таким образом, чтобы минимизировать значение AIC.

Порядок *p* авторегрессионной части AR определяет число предыдущих значений ряда, учитываемых при построении модели.

Порядок *d* разностей ряда, порядок интегрирования, задает порядок при необходимости привести исходный ряд к [стационарному](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%B0%D1%80%D0%BD%D0%BE%D1%81%D1%82%D1%8C).

Порядок *q* части скользящего среднего, части MA, определяет размер скользящего окна для сглаживания исходного ряда. Для все трёх порядков верно, что устанавливается значение целого типа, при чём это значение больше 0.

4. Есть возможности включения расчёта сезонности. Это позволит задать параметры для сезонной составляющей.

Горизонтом прогнозирования можно назвать временной интервал, в пределах которого прогноз выполняется с заданной точностью. Кроме этого, под горизонтом прогнозирования понимают параметр модели прогнозирования на основе метода скользящего окна. Выделяют три уровня горизонта прогнозирования:

* долгосрочный;
* среднесрочный;
* краткосрочный.

Особенности прогнозирования временных рядов по средствам ARIMAX:

1. Горизонт прогноза задает количество значений, которые будут спрогнозированы и добавлены в выходной набор в конце исходного временного ряда.

2. Важно рассчитать ошибку аппроксимации. Будет добавлен в выходной набор столбец со средними отклонениями прогнозируемых значений от фактических.

3. Доверительный интервал прогноза указывается в процентах от 0 до 100, имеет вещественный тип, обычно равен 95.

Процесс авторегрессии порядка , который обозначается AR(p), описывается формулой:

*,*

где – вещественная константа, – коэффициенты, – ошибка прогнозирования. Для определения , используют метод наименьших квадратов или метод максимального правдоподобия.

Также используется другой тип модели и часто применяется совместно с авторегрессией. Эта модель скользящего среднего порядка и описывается формулой:

*,*

где – порядок скользящего среднего, – ошибка прогнозирования.

Целесообразно объединить в одной модели авторегрессию и скользящее среднее. Общая модель обозначается ARMA(p,q). Если в качестве входных данных используются не сами значения временного ряда, а их разность *q*-того порядка, то модель носит название авторегрессии проинтегрированного скользящего среднего. Такую модель называют ARIMA(p,d,q). Развитием модели ARIMA(p,d,q) является модель ARIMAX(p,d,q), которая записывается уравнением:

*,*

где – коэффициенты внешних факторов . [10]

GARCH-модель (generalized autoregressive conditional heteroskedasticity) – модель, которая чаще всего используется для прогнозирования ситуации на финансовых рынках в условиях нестабильности. Для таких ситуаций характерно такое понятие, как дисперсия на различных интервалах наблюдения, которую также называют гетероскедастичностью. Линейную регрессионную модель в таких случаях не применяют.

Говоря о модели GARCH, нельзя не вспомнить о ARCH-модели, в которой используется условная, зависимая от времени дисперсия, выражаемая через квадрат значений показателей прошлых периодов. Для ARCH-модели верно:

*,*

где – коэффициент задержки.

Позднее была предложена GARCH-модель. Она является обобщенной авторегрессионной моделью гетероскедастичности. GARCH-модель предполагает, что на текущую изменчивость дисперсии влияют предыдущие изменения показателей и предыдущие оценки дисперсии. Расчёт дисперсии для GARCH-модели:

,

где – количество предшествующих оценок, которые влияют на текущее значение, – весовые коэффициенты,  отражающие степень влияния предыдущих оценок на текущее значения. Модель обозначают GARCH. [11]

ALRDM-модель (autoregression distributed lag model) – авторегрессионная модель с распределённым лагом. Часто при моделировании процессов на изучаемую переменную влияют и текущие значения процесса, и его лаги, то есть значения временного ряда, предшествующие изучаемому моменту времени. ALRDM-модель описывается уравнением:

*,*

где – коэффициенты, – величина лага. Модель обозначают ALRDM, чаще всего применяют при экономических процессах. [12]

# 2.4 Нейросетевые модели

Нейронная сеть – это математическая модель, а также её программное или аппаратное воплощение, которая построена по принципу организации и функционирования биологических нейронных сетей, которые, в свою очередь определяются как сети нервных клеток живого организма. [13]

Другими словами, под понятием нейронных сетей подразумевают компьютерные системы, которые собраны из сотен, тысяч или миллионов нейроподобных элементов, способных обучаться и действовать по принципу похожему на то, как работает мозг человека. Обычная искусственная нейронная сеть состоит из сотен, тысяч или миллионов искусственных нейронов, названных блоками, которые выстроены в слои, где каждый блок соединён с соседним как в собственном слое, так и в ближайшем. Некоторые из блоков выступают в качестве блоков ввода и призваны получать из внешнего мира информацию. Эти блоки соединяются со скрытыми блоками, которые обрабатывают полученные данные и занимают большую часть искусственной нейронной сети. Также существует блок вывода, который, в свою очередь, занимается извлечением полученной и обработанной информации. Соединение между блоками характеризуется числом (весом), которое может быть как положительным, так и отрицательным. Чем больше вес связи, тем сильнее один блок влияет на другой. [14]

Нейросети часто стали применяться в практических целях: для решения задачи прогнозирования, задачи распознавания образов, задачи управления и других. Стоит также отметить, что нейросети не программируются в привычном смысле этого слова, – они обучаются.

В общем случае выделяют три фундаментальных класса нейросетевых архитектур:

1) однослойные сети прямого распространения;

2) многослойные сети прямого распространения;

3) рекуррентные сети.

Дадим общее описание для каждой архитектуры.

В многослойной нейронной сети нейроны располагаются по слоям. В самом простом случае в такой сети существует входной слой узлов источника, информация от которого передаётся на выходной слой нейронов. Такая сеть носит название сеть прямого распространения (ацикличная сеть). Так на рисунке 2 показана структура ацикличной сети для случая четырёх узлов во входном слое и четырех вычислительных элементов. Эта нейронная сеть называется однослойной, так как вычисления производятся только в одном слое.

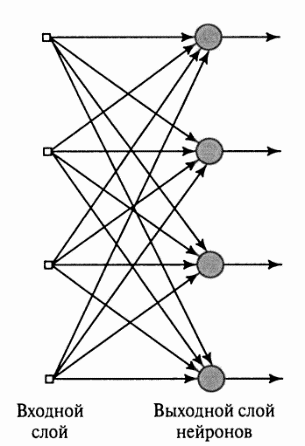


Рисунок 2 – Ацикличная сеть с 1 слоем нейронов

Следующий класс нейронных сетей прямого распространения характеризуется существованием одного или нескольких скрытых слоёв, узлы которого называют скрытыми нейронами. Добавляя один или несколько скрытых слоёв, из входных данных можно выделить статистики высокого порядка. Такая сеть позволяет выделять глобальные свойства данных с помощью локальных соединений за счёт наличия дополнительных синаптических связей и повышения уровня взаимодействия нейронов.

Слой сети, в котором все узлы связаны со всеми узлами следующего слоя, называется полносвязным. Так на рисунке 3 показана нейронная сеть с полносвязными слоями. Если некоторые из синаптических связей отсутствуют, такая связь называется неполносвязной.

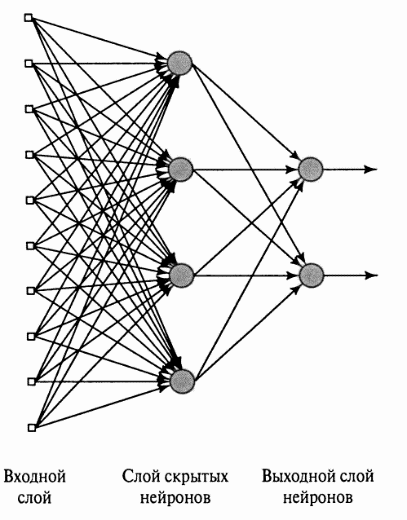


Рисунок 3 – Полносвязная сеть прямого распространения с 1 скрытым и 1 выходным слоем

Рекуррентная нейронная сеть (RNN) отличается от сети прямого распространения наличием, по крайней мере, одной обратной связи. Так рекуррентная сеть может состоять из единственного слоя нейронов, каждый из которых направляет свой выходной сигнал на входы остальных нейронов слоя.

Такой тип нейронной сети хорошо подходит для временных рядов. RNN обрабатывают временной ряд шаг за шагом, поддерживая внутреннее состояние от временного шага к временному шагу.

Архитектура RNN представлена на рисунке 4. Как можно заметить, в приведённой структуре отсутствуют обратные связи нейронов с самими собой. RNN на рисунке 4 также не имеет скрытых нейронов. На рисунке 5 показан другой класс RNN, который включает скрытые нейроны. Обратные связи для данного класса исходят из скрытых и из выходных нейронов.

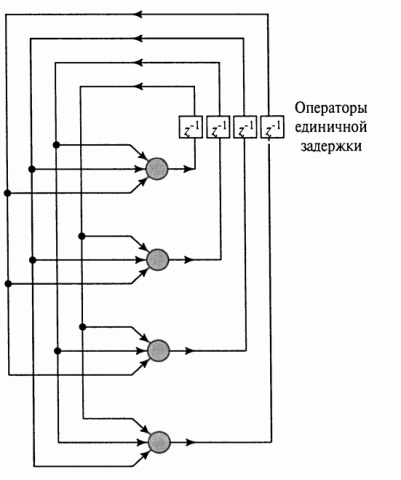


Рисунок 4 – RNN без скрытых нейронов и обратных свзяей нейронов с самими собой

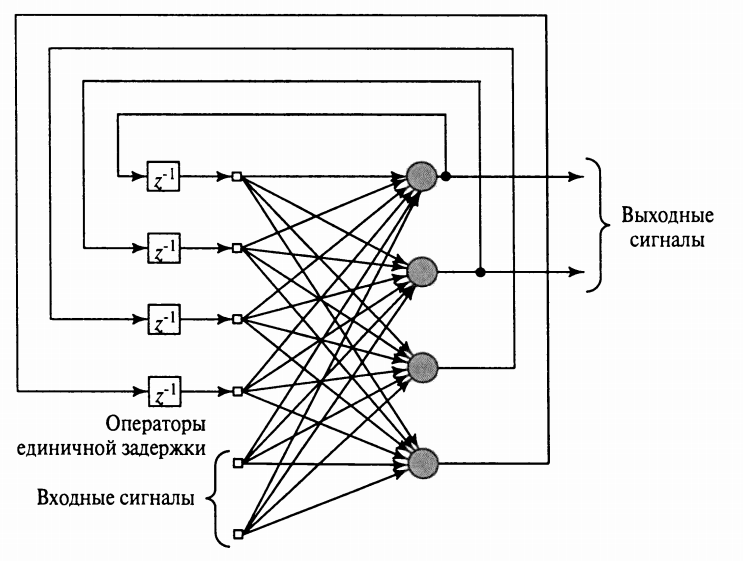


Рисунок 5 – RNN со скрытыми нейронами

Обычно обратная связь подразумевает использование элементов еденичной задержки, это приводит к нелинейному динамическому поведению при существовании в сети нелиненых нейронов. [15]

Рекуррентные НС при вычислении очередного значения используют не только входной вектор, но и вектор своего состояния вычисленный в предыдущую итерацию функционирования. Это обогащает вычислительную модель и позволяет решать более сложные задачи.

# 2.5 LSTM

LSTM RNN (Long Short-Term Memory Recurrent Neural Network) – рекуррентная нейронная сеть с долгой краткосрочной памятью. LSTM была изобретена в 1977 году З. Хохрайтером и Ю. Шмидхубером. Как и все другие модификации нейронных сетей, LSTM является универсальной моделью, которая способна решать большой круг различных проблем.

В отличие от традиционных RNN, LSTM специально разработаны для устранения проблемы долгосрочной зависимости. Основная особенность LSTM – запоминание информации в течение длительных периодов времени, отсюда следует, что они практически не требуют обучения. Это достигается за счет того, что повторяющийся модуль модели имеет комбинацию четырех слоев, взаимодействующих друг с другом.

Другими словами, LSTM отлично походит для задач, когда необходимо с течением времени накапливать знания об успешности прошлых предсказаний. LSTM применяется, когда существует сильная зависимость текущих значений временного ряда от предыдущих, но величина лага, имеющий наибольшее влияние на показатель, динамична, то есть коэффициенты, стоящие перед лаговыми переменными постоянного изменяются.

Далее опишем принцип работы сети LSTM.

Важным и ключевым понятием является состояние ячейки, которая напоминает конвейерную ленту. Оно проходит сквозь всю цепочку, подвергается при этом незначительным линейным преобразованиям.

Гейты, выступая в качестве структуры, предназначены для изменения количества информации в состояниях этих ячеек. Гейты состоят из сигмовидного слоя нейронной сети, на выходе которого выдаются «0» и «1», и операции поточечного умножения. [16]

LSTM содержит три таких гейта для контроля состояния ячеек.

Во-первых, слой утраты. Здесь решается вопрос о том, какую информацию необходимо удалить из состояния ячейки, сигмовидный слой принимает решение. Слой получает на вход , и выдает «0» или «1» для каждого номера состоянии ячейки. Формула для «слоя гейта утраты»:

.

Во-вторых, слой сохранения. Здесь решается вопрос о том, какую информацию необходимо сохранить в состоянии ячейки. Процесс разбивается на две части. Сначала сигмовидный слой решает, какие значения требуется обновить. Затем слой tanh создаёт вектор новых значений-кандидатов, которые добавляются в состояние. На следующем шаге происходит объединение двух значений для обновления состояния.

.

*,*

где – вектор новых значений-кандидатов.

В-третьих, формирование нового состояния. Теперь необходимо обновить предыдущее состояние ячейки для получения нового состояния. Для этого:

,

где – старое состояние, второе слагаемое выступает в качестве новых значений кандидатов, которые масштабируются в зависимости от того, как мы решили обновить каждое значение состояния.

Необходимо решить, какой результат ожидаем нами на входе. Результат будет являться отфильтрованным состоянием ячейки:

,

,

где – выходной сигнал сигмовидного гейта, необходим для того, чтобы разместить все значения в интервале . [17]

На рисунке 6 изображена схема LSTM, рисунок взят из источника [18].

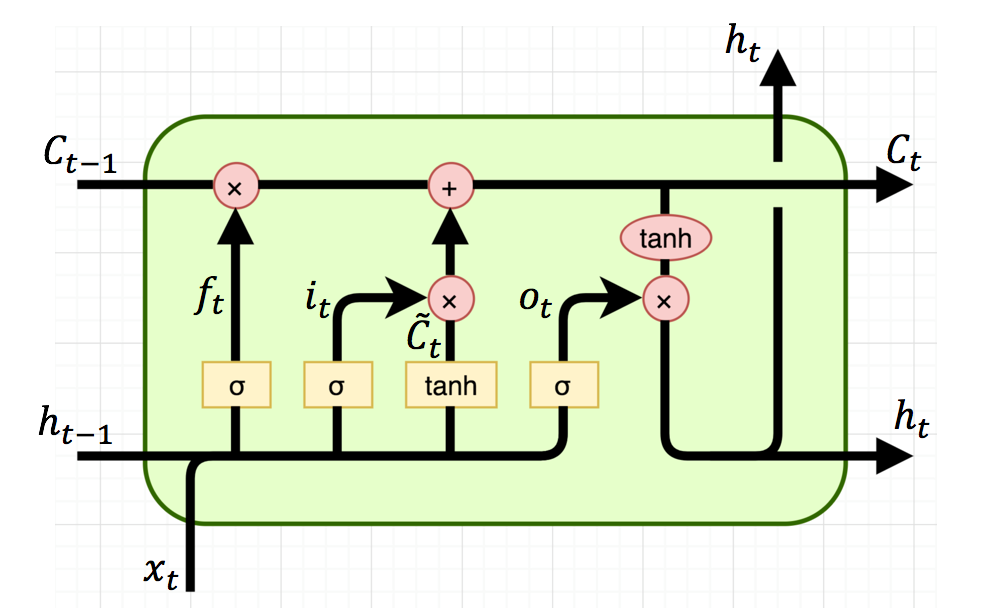


Рисунок 6 – Схема LSTM

Модуль LSTM имеет состояние ячейки и три шлюза, что дает им возможность выборочно изучать, отключать или сохранять информацию от каждого из модулей.

# 2.6 Модели на базе цепей Маркова

Рассмотрим систему, которая состоит из семейства случайных переменных и определяется следующей формулой:

.

Значение , которое характеризует случайную переменную в дискретный момент времени , называет состояние системы в этот момент. Пространство состояний – пространство всех возможных значений, которые могут принимать эти случайные переменные. Если структура стохастического процесса такова, что условное распределение вероятности зависит исключительно от предыдущего значения и не зависит от всех более ранних значений, то такой процесс называется цепью Маркова, или Марковской цепью.

Формально можно записать так:

*.*

Это соотношение называется свойством Маркова. Формально это свойство описывают следующим образом.

Последовательность случайных переменных принимает форму цепи Маркова, если вероятность нахождения системы в состоянии в некоторый момент времени зависит исключительно от вероятности нахождения системы в момент времени в состоянии .

Можно сделать вывод о том, что цепи Маркова имеет смысл рассматривать как порождающие модели, которые состоят из множества попарно связанных друг с другом состояний. Каждый раз, когда система переходит в конкретное состояние, именно с ним ассоциируется выход системы.

Другими словами, модели прогнозирования на основе цепей Маркова предполагают, что будущее состояние процесса зависит только от его текущего состояния. Процессы, которые моделируются цепями Маркова, относятся к процессам с короткой памятью.

Также затронем понятие вероятности перехода, так как в цепи Маркова переход из одного состояния в другое является вероятностным. Однако при этом выход системы является детерминированным. Пусть вероятность перехода:

*,*

где являются условными вероятностями, подчиняющимися:

*,* для всех .

Предполагается, что вероятности перехода фиксированы и не изменяются во времени. Из этого следует, что последнее условие выполняется при в любой момент времени. Такую цепь называют гомогенной по времени.

Если система имеет конечное множество возможных состояний , вероятности перехода формируют матрицу размерности :

.

Элементы матрицы удовлетворяют условиям , последнее из которых гарантирует равенство суммы всех элементов каждой из строк единице. Такая матрица называется стохастической. Любая стохастическая матрица может служить матрицей вероятностей перехода.

# 2.7 Модели на базе деревьев классификации и регрессии

Рассмотрим деревья классификации и регрессии – CART (Classification and Regression Tree). CART представляет из себя алгоритм бинарного дерева решений, который предназначен для решения задач классификации и регрессии.

Стоит также отметить, что деревья CART всё чаще применяют для прогнозирования временных рядов. Модели CART используют при моделировании процессов, на которые влияют внешние и категориальные факторы. При влиянии внешних непрерывных факторов используют деревья регрессии, при категориальных – деревья классификации. Также имеет место быть смешение этих деревьев при влиянии обоих факторов.

По модели CART прогнозируемое значение временного ряда зависит от предыдущих значений и некоторых независимых переменных. Рассмотрим бинарное классификационно-регрессионное дерево, которое описано в источнике [19].

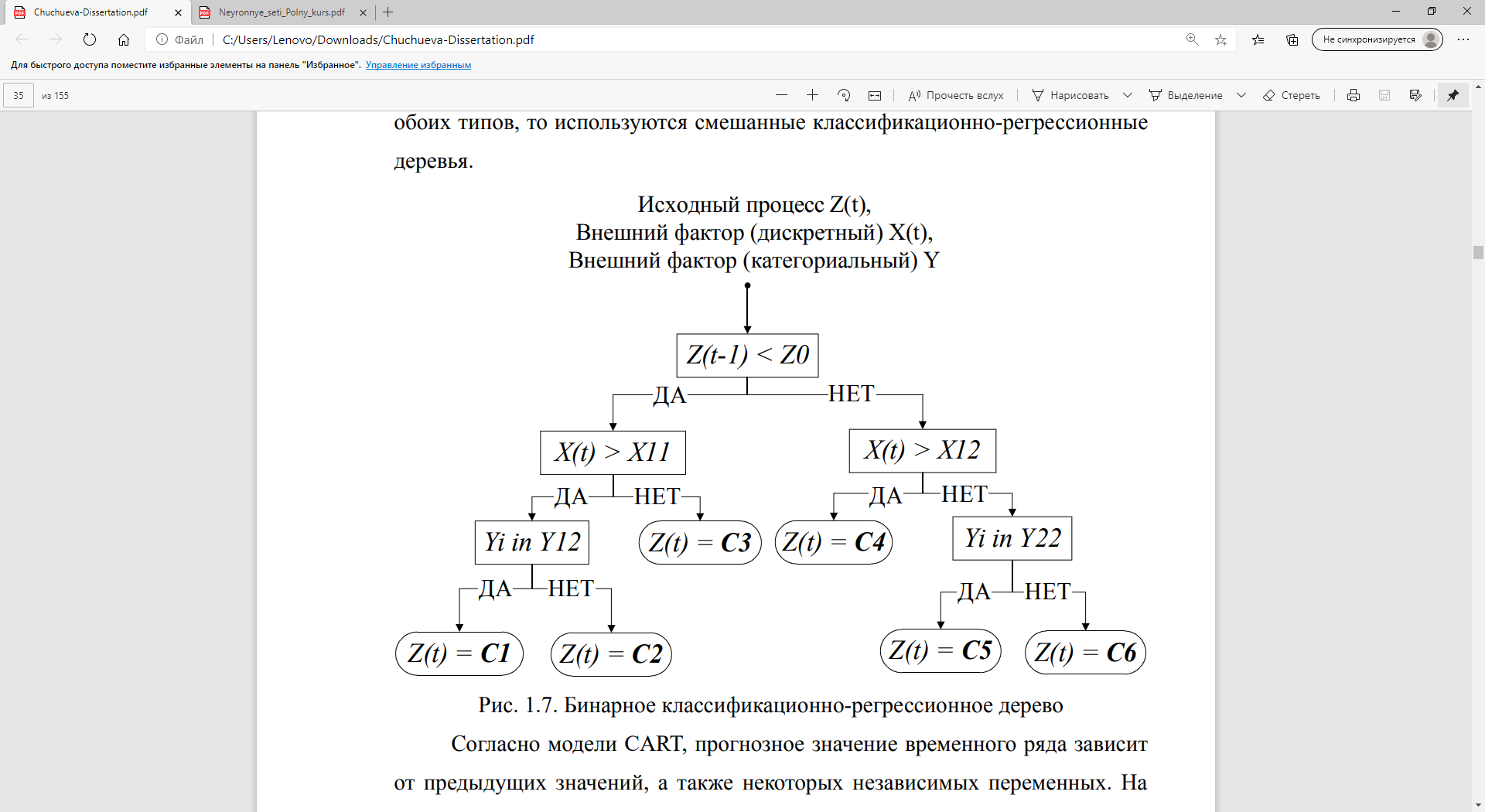


Рисунок 7 – CART

Сначала предыдущее значение процесса сравнивается с константой . При выполнении неравенства проверяется неравенство , при невыполнении проверяется неравенство . Проверки продолжаются до того момента, пока не будет обнаружен лист дерева, в котором происходит определение будущего значения процесса . При поиске этого листа используются непрерывные и категориальные переменные, для которых выполняется проверка присутствия значения в одном из заранее определенных подмножеств. Значения пороговых констант и подмножеств выполняется на этапе обучения дерева.

Так можно сделать вывод о том, что CART – модель, в основе которой лежит зависимость будущей величины процесса посредствам построения структуры дерева. [20]

# 3 Практическая часть

Из источника [21] были предварительно скачаны данные, – таблица MBSComtrade.csv. В файле находятся данные, характеризующие суммарный торговый трафик по месяцам, указанный в долларах, для различных стран с 1957 года по 2020 год. Таблица состоит из 25 столбцов (параметров). Лишь часть из них потребуется для решения поставленной задачи. Эти параметры можно условно поделить на три категории:

* параметры, отвечающие за дату: год (year), тип периода (period\_type), номер месяца (period\_number), дата (period, period\_in\_date);
* параметры, характеризующие страну: название (country\_english\_name), код (country\_code);
* параметры, отвечающие за товарооборот: поле, в котором прописывается вид товарооборота - импорт/экспорт (trade\_flow\_desc), объём товарооборота (value).

Объём товарооборота (value) представлен в этой таблице суммарным объёмом экспорта/импорта по всем торговым группам в долларах для различных стран.

Целью работы является проверка возможности прогнозирования суммарного торгового трафика импорта для некоторых крупных стран нейросетевыми методами.

Решать задачу будем построением и обучением нейросетевой модели.

**3.1 Описание программных и математических инструментов**

Программа была написана на высокоуровневом языке программирования Python.

Python поддерживает структурное, обобщённое, объектно-ориентированное, функциональное и аспектно-ориентированное программирование. Динамическая типизация, автоматическое управление памятью, полная интроспекция, механизм обработки исключений – основные архитектурные черты. На сегодняшний момент Python является активно развивающимся языком программирования. Стандартная библиотека Python включает набор полезных функций.

Pandas представляет из себя программную библиотеку на языке Python для обработки и анализа данных.  Работа Pandas с данными строится поверх библиотеки NumPy, являющейся инструментом более низкого уровня. Предоставляет специальные структуры данных и операции для манипулирования числовыми таблицами и временными рядами.

NumPy является фундаментальным пакетом для научных вычислений на языке программирования Python. Эта библиотека обеспечивает поддержку больших многомерных массивов и матриц, вместе с большой библиотекой высокоуровневых математических функций для операций с этими массивами.

TensorFlow представляет из себя мощную библиотеку для создания нейронных сетей. TensorFlow предназначен для проектирования, создания, изучения моделей глубокого обучения, и численных вычислений.

Scikit-learn основан на [NumPy](https://datastart.ru/blog/read/4-priema-python-numpy-kotorye-dolzhen-znat-kazhdyy-novichok) и SciPy. Библиотека поддерживает:

* предварительную обработку данных;
* уменьшение размерности;
* выбор модели;
* регрессии;
* классификации;
* кластерный анализ. [22]

Scikit-learn также используют для подготовки данных к алгоритмам машинного обучения: стандартизация или нормализация данных и многое другое.

Для реализации программы на языке Python была использована отдельная, виртуальная среда Anaconda, выступающая в качестве платформы для обработки данных.

Anaconda – дистрибутив языков программирования Python и R, включающий набор популярных свободных библиотек, которые в свою очередь объединены проблематиками науки о данных и машинного обучения. Основная цель – развёртывание единого согласованного комплекта наиболее востребованных тематических моделей с разрешением возникающих зависимостей и конфликтов. На 2020 год содержит более 1500 модулей.

Воспользовавшись источником [23], был скачан дистрибутив Anaconda3-2020.11-Windows-x86\_64.exe для работы с нейронными сетями. При запуске скачанного дистрибутива происходит установка Anaconda и Python (версия 3.8.0).

После успешной установки Anaconda3 и Python нужной версии иногда возникает необходимость добавить в переменные среды системы пути:

* Anaconda3\Library\bin;
* Anaconda3;
* Anaconda3\Scripts.

Далее нужно установить некоторые библиотеки. Сделать это можно с использованием командной строки, где указывается через пробел название программы (conda), команда (install), наименование соответствующей библиотеки. Пример установки библиотеки показан на рисунке 8. Также можно установить все библиотеки сразу, указав их после команды install через пробел.

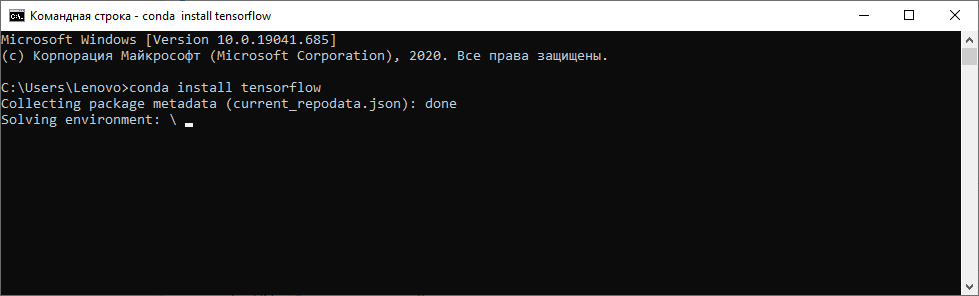


Рисунок 8 – Установка библиотеки tensorflow с помощью командной строки

В программе используются следующие библиотеки и модули: matplotlib (3.2.1), tensorflow (2.2.0), pandas (1.0.3), scikit-learn (0.22.1), numpy (1.18.1).

Jupyter Notebook выступает в качестве инструмента для интерактивной разработки и представления проектов в области Data Science. Для работы программы необходимо открыть «Jupyter Notebook» из папки «Anaconda3», либо воспользоваться меню «Пуск» и запустить ярлык Jupyter, который изображен на рисунке 9.

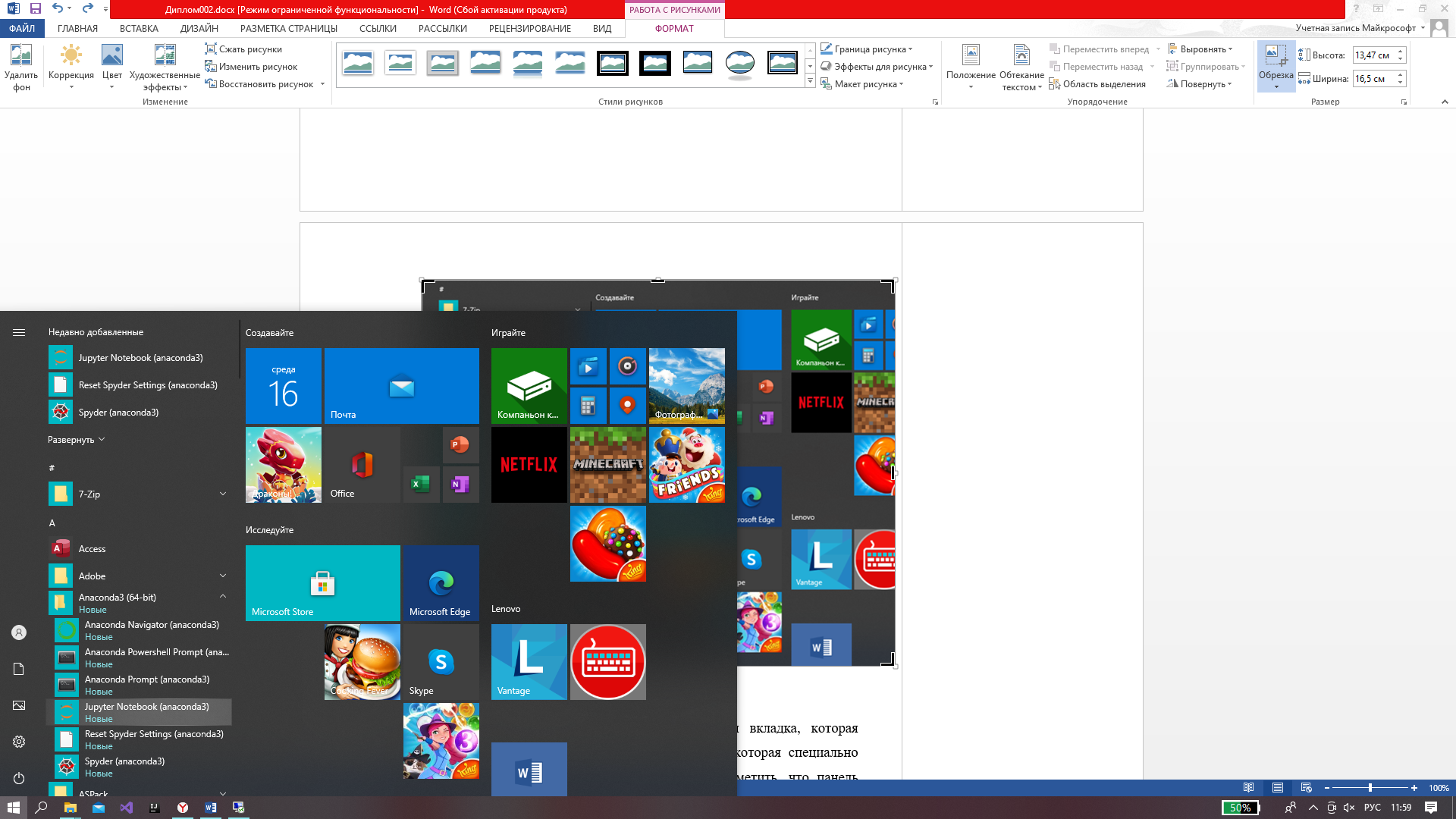


Рисунок 9 – Папка «Anaconda3» в меню «Пуск»

В веб-браузере по умолчанию откроется новая вкладка (URL: http://localhost:8888/tree), которая представляет из себя панель инструментов ноутбука, которая специально разработана для управления ноутбуками. Но стоит отметить, что панель управления предоставит доступ к файлам и подпапкам, которые содержатся в каталоге запуска Jupyter. В файловой системе можно выбрать и открыть нужный файл или создать новый. Интерфейс панели представлен на рисунке 10.

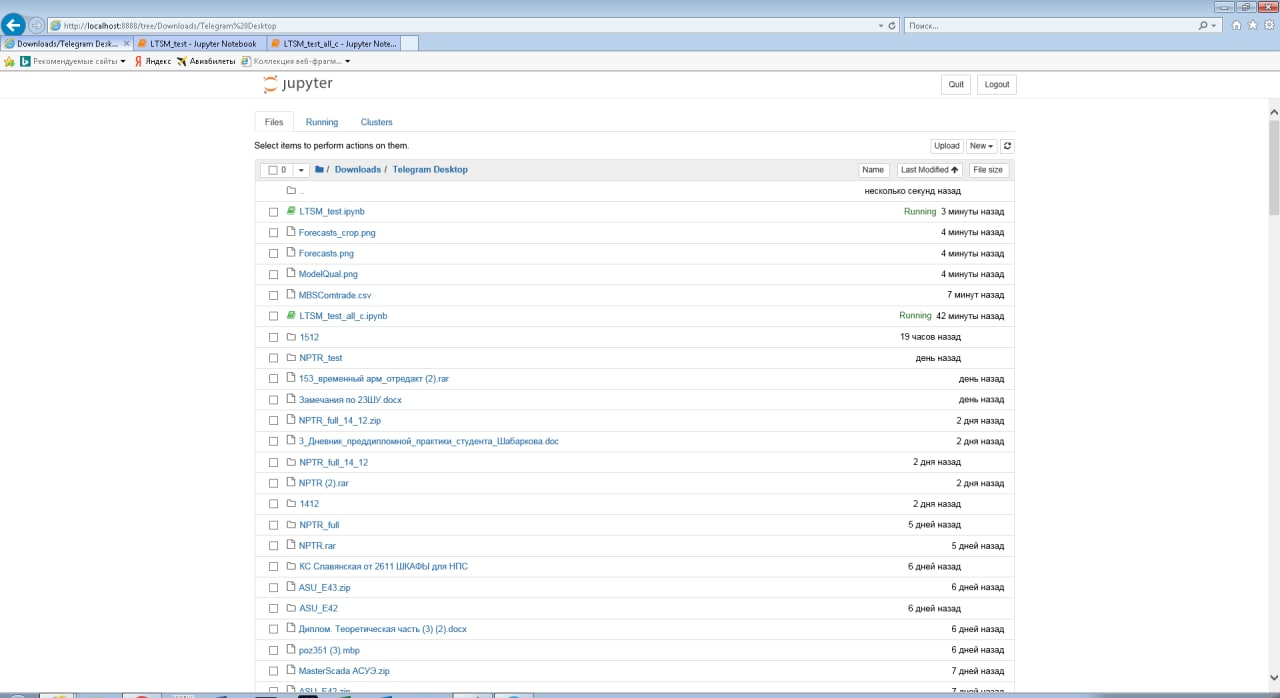


Рисунок 10 – Интерфейс панели инструментов ноутбука Jupyter

Для начала работы программы необходимо открыть файл с расширением «.ipynb», тем самым будет открыта интерактивная среда. Для выполнения определенного блока программы необходима комбинация клавиш «Shift» и «Enter». При выполнении блока в квадратных скобках соответствующего блока будет стоят символ «\*» как на рисунке 11.

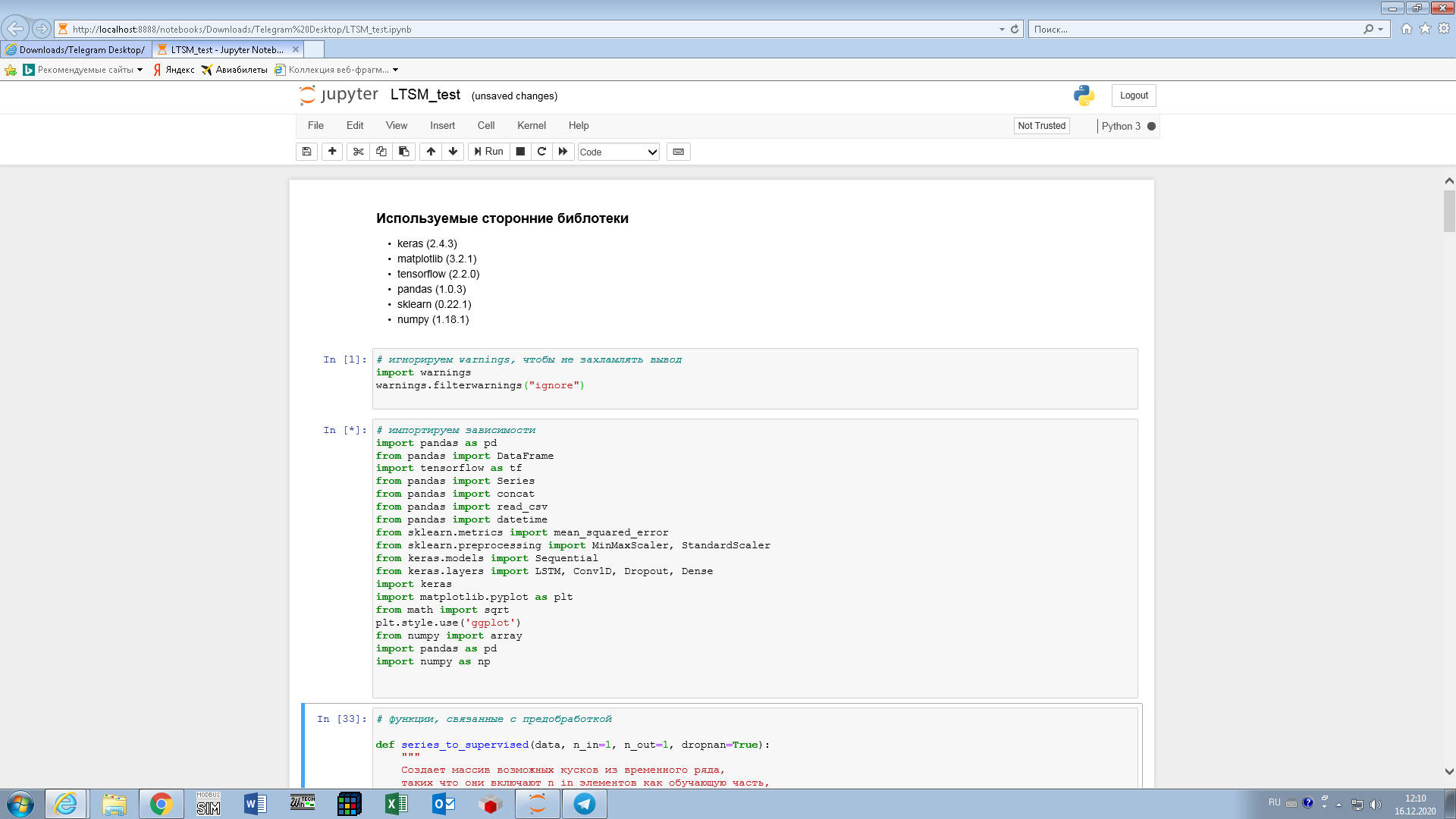


Рисунок 11 – Запуск блока для импортирования зависимостей

Для корректной работы программы необходимо, чтобы документ с исходными данными, в нашем случае это таблица MBSComtrade.csv, лежал в папке с исполняемым файлом, как показано на рисунке 10. В ходе работы было создано два исполняемых файла:

1. LTSM\_test.ipynb;
2. LTSM\_test\_all\_c.ipynb.

В первом файле модель применяли, заменяя test\_c на соответствующее название страны при выполнении начального блока программы, для двух крупных стран: США, Российская Федерация, во втором – алгоритм циклично применяется для оставшихся стран с качественными данными. Результаты работы представлены в главе 3.5.

Перед началом использования модели на данных их необходимо было нормализовать, так как они имеют чёткий тренд. Кратко опишем функции, связанные с предобработкой:

1. функция «series\_to\_supervised» создаёт массив возможных кусков из временного ряда, создана для устранения «краевых» эффектов;
2. функция «difference» предназначена для подсчёта разницы между элементами временного ряда;
3. функция «prepare\_data» преобразует данные с помощью нахождения разности и шкалирования, и формирует тренировочную и обучающую выборку.

Далее была построена нейронная сеть с помощью функции «fit\_lstm», которая создаёт и обучает нейросеть «model» с учётом выбранным параметров. Описание архитектуры сети указано в главе 3.3, про обучение – в главе 3.4.

Затем происходило предсказывание предстоящих значений после заданной точки с помощью функции «forecast\_lstm» и предсказывание с помощью переданной модели для всех временных участков с помощью функции «make\_forecasts».

Также был произведена численная и визуальная оценка качества модели с помощью трёх метрик.

В результате работы LTSM\_test.ipynb в папке исполняемого файла сохраняются два рисунка: «Forecasts.png» – график предсказаний из каждой точки тестовой выборки, включает весь «train sample»; и часть этого графика «Forecasts\_crop.png» для лучшей читаемости. Пример для данных импорта из Российской Федерации показан на рисунке 12.

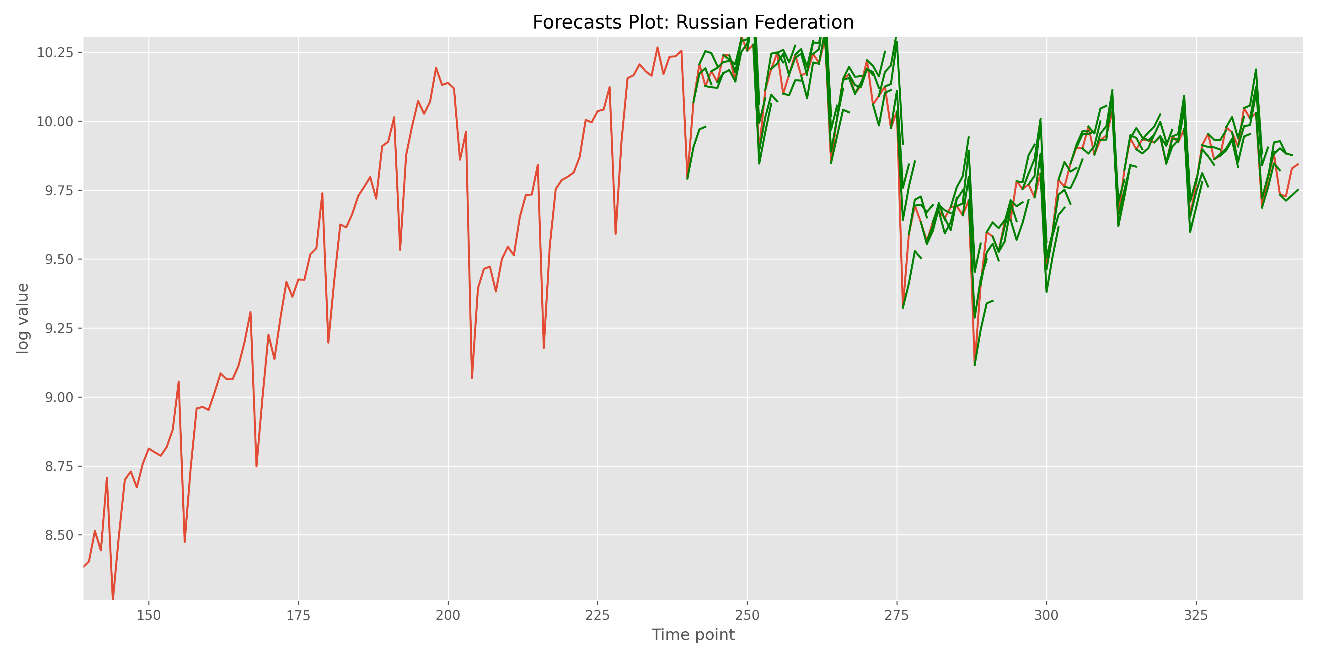
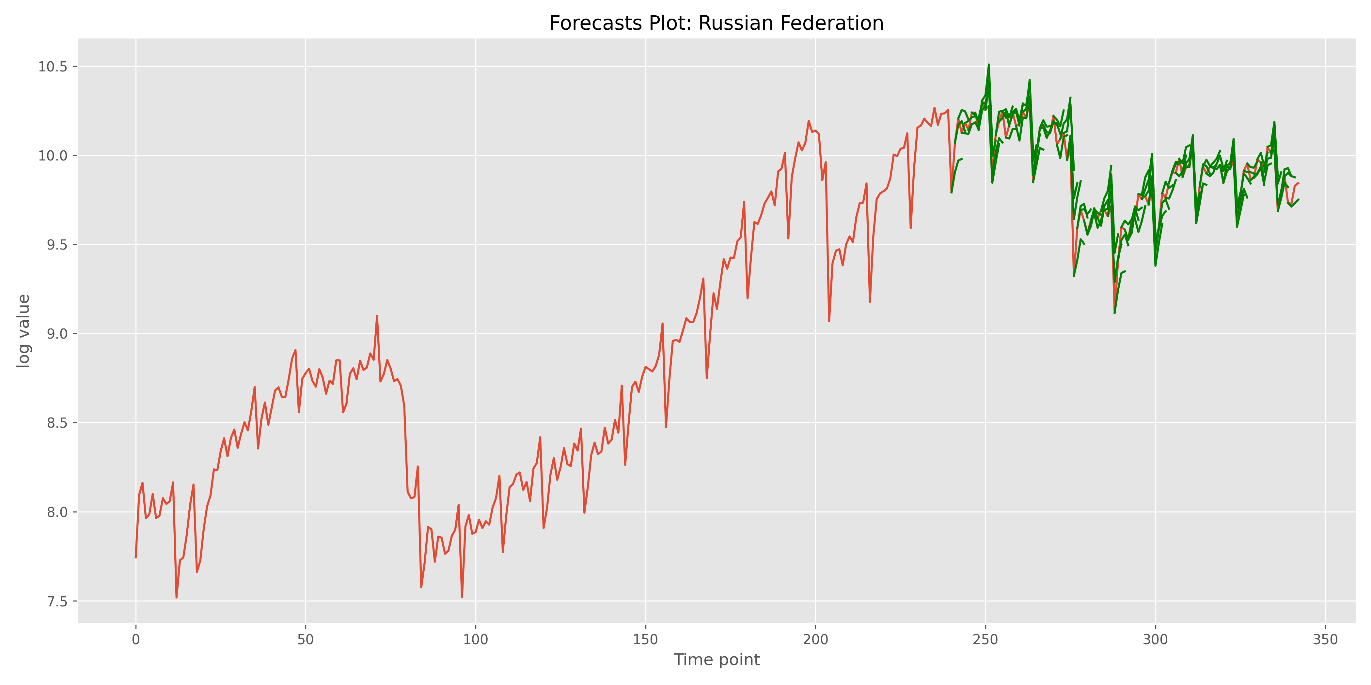


Рисунок 12 – «Forecasts.png» и «Forecasts\_crop.png» соответственно



Также в результате работы программы были созданы два текстовых файла

1. «EXPORT\_RESULTS.txt» для оценки качества модели на данных объёма экспорта разных стран;
2. «IMPORT\_RESULTS.txt» для оценки качества модели на данных объёма импорта разных стран.

Они включают в себя три оценки модели: значение функции потерь (loss), точности (accuracy) и средней абсолютной ошибки (MAE).

**3.2 Структура данных**

В практической части дипломной работы была реализована нейронная сеть с архитектурой LSTM-RNN.

Так как в файле MBSComtrade.csv много параметров, большую часть из которых не имеет смысла применять в нашей задаче, было произведено извлечение нужных параметров из файла с помощью методов pandas.

Данные товарного оборота представляют из себя классический пример временного ряда, с выраженной цикличностью и, как правило, растущим трендом. Для корректного построения модели данные были приведены к виду стационарного временного ряда.

Для этого на первом этапе производилось логарифмирование значений временного ряда для устранения экспоненциально растущей амплитуды. Далее каждое значение в временной точке было заменено на разницу между -ой и -ой точкой (нулевая точка теряется), что позволяет избавится от тренда.

Полученные данные разбивались на небольшие вектора с длиной равной

,

где - это длина временного интервала, используемая для предсказания (по умолчанию 36 месяцев), – это длина предсказываемого участка (по умолчанию 3 месяца).

Каждый вектор был разбит на два фрагмента, длиной и соответственно. Левый фрагмент был определен как исходные данные, правый фрагмент как целевые (target). Таким образом, мы получили готовый датасет для обучения предсказательной модели.

**3.3 Описание модели**

Нами была предложена предсказательная модель на основе рекуррентных нейронных сетей. Архитектура сети представлена на рисунке 13.

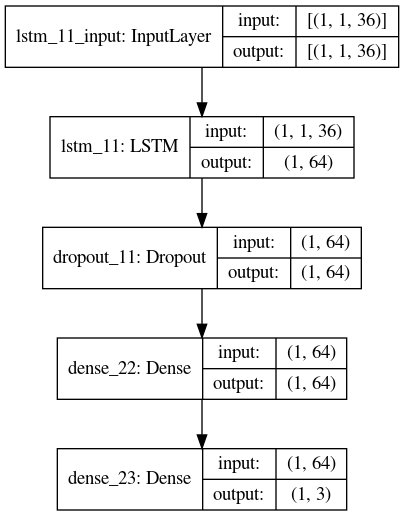


Рисунок 13 – Архитектура нейронной сети

В качестве основного слоя был добавлен блок из 64 ячеек с архитектурой LSTM. Слой Dropout был добавлен для регуляризации (с долей обнуляемых весов ). Необходимость регуляризационного слоя была очевидна, так как без него наблюдался рост значения функции потерь на валидационной части выборки уже с ранних итераций обучения, что явно говорит о наличии переобучения. Dense слой из 64 нейронов был использован для более аккуратного предсказания сложной зависимости.

В качестве функции активации был использован гиперболический тангенс (tanh). Рекуррентная функция активации в LSTM блоке – сигмоида.

**3.4 Обучение модели**

Для обучения сети был выбран метод обратного распространения ошибки. Этот метод используется при обновлении весов в нейронной сети. Данный способ обучения нейронной сети помогает отрегулировать каждый вес пропорционально тому, насколько он способствует общей ошибке. При итеративном уменьшении ошибки каждого веса получится ряд весов, которые дают хорошие прогнозы.

Метод обратного распространения ошибки является специфической реализацией градиентного спуска в пространстве весов многослойных сетей прямого распространения. Основная идея этого метода заключается в вычислении частных производных функции сети по всем элементам настраиваемого вектора весов для данного вектора . [15]

Для поиска весов был применен оптимизационный алгоритм Adaptive moment estimation (Adam), который описан в источнике 21. В качестве метрик оценки модели были использованы значение функции потерь (loss), точности (accuracy) и средней абсолютной ошибки (MAE). MAE – это мера ошибок между парными наблюдениями, выражающими одно и то же явление:

,

где – среднее арифметическое абсолютных ошибок, – предсказанное значение, – истинное значение.

Каждая из метрик рассчитывалась как на тренировочной, так и на валидационной части для отслеживания эволюции модели в ходе итераций обучения. Все данные метрик для каждой итерации сохранялись для дальнейшего анализа.

Были реализованы два критерия остановки процесса обучения. Первый – явное задание пользователем числа итераций обновления весов сети. При использовании такого подхода при задании слишком большого числа итерации наблюдалось незначительное переобучение модели даже при наличии регуляризационного слоя. Поэтому был реализован второй подход, основанный на анализе изменений значения функции потерь на валидационной части выборки (val\_loss). При стабилизации этой величины в диапазоне процесс обучения прерывался с сохранением наилучших по величине val\_loss параметров модели.

Обучаемые параметры в модели – непосредственные веса нейронной сети. Необучаемыми (задаваемыми пользователем) параметрами стали количество LSTM нейронов, число нейронов в слоях Dense, доля обнуляемых весов в регуляризационном слое Dropout, длина используемого для предсказания фрагмента временного ряда.

На вход модель принимает вектор значений с размерностью, равной выбранной длине участка временного ряда, на основе которого строится предсказание (по умолчанию 36 месяцев). На выходе модель возвращает вектор значений с размерностью, равной выбранной длине прогнозируемого временного интервала (по умолчанию 3 месяца). Для оценки качества модели была дополнительно написана функция расчета значения среднеквадратичной ошибки (RMSE) для каждого шага прогноза. RMSE вычисляется по следующей формуле:

,

где – прогноз модели, – фактический ожидаемый результат.

Также была реализована функция расчета точности по данным направления изменения. Предсказанные изменения для каждой временной точки были представлены в виде двух классов («значение растет» и «значение падает»). На основе полученных реальных и предсказанных бинарных векторов считалась метрика точности (accuracy) по классической формуле:

,

где – число точек в тестовой части временной выборки.

Обучение модели происходило силами центрального процессора Intel Core i7 9750H, при наличии 16 Gb ОЗУ. Максимальный объем потребляемой памяти в процессе обучение не превысил 1 Gb. Время обучения модели зависело от характера и размерности данных и находилось в диапазоне 15-162 секунды для одной страны. Так, на выборке данных импорта для Российской Федерации в тестовом запуске время обучения модели составило 52 секунды. Среднее число эпох обучения составило 282 эпохи, в зависимости от страны значения на тестовом запуске находились в диапазоне 18-414 эпох. Максимальный объем потребляемой памяти при использовании модели составил не более 500 Mb, включая интерпретатор Python и данные.

**3.5 Результаты вычислений**

В качестве примера будет рассмотрено применение модели на данных ежемесячного импорта для Российской Федерации. На рисунке 14 приведен график ежемесячного импорта в период с 1992 по 2020 год. Отметим, что временной ряд имеет выраженную сезонность и тренд.

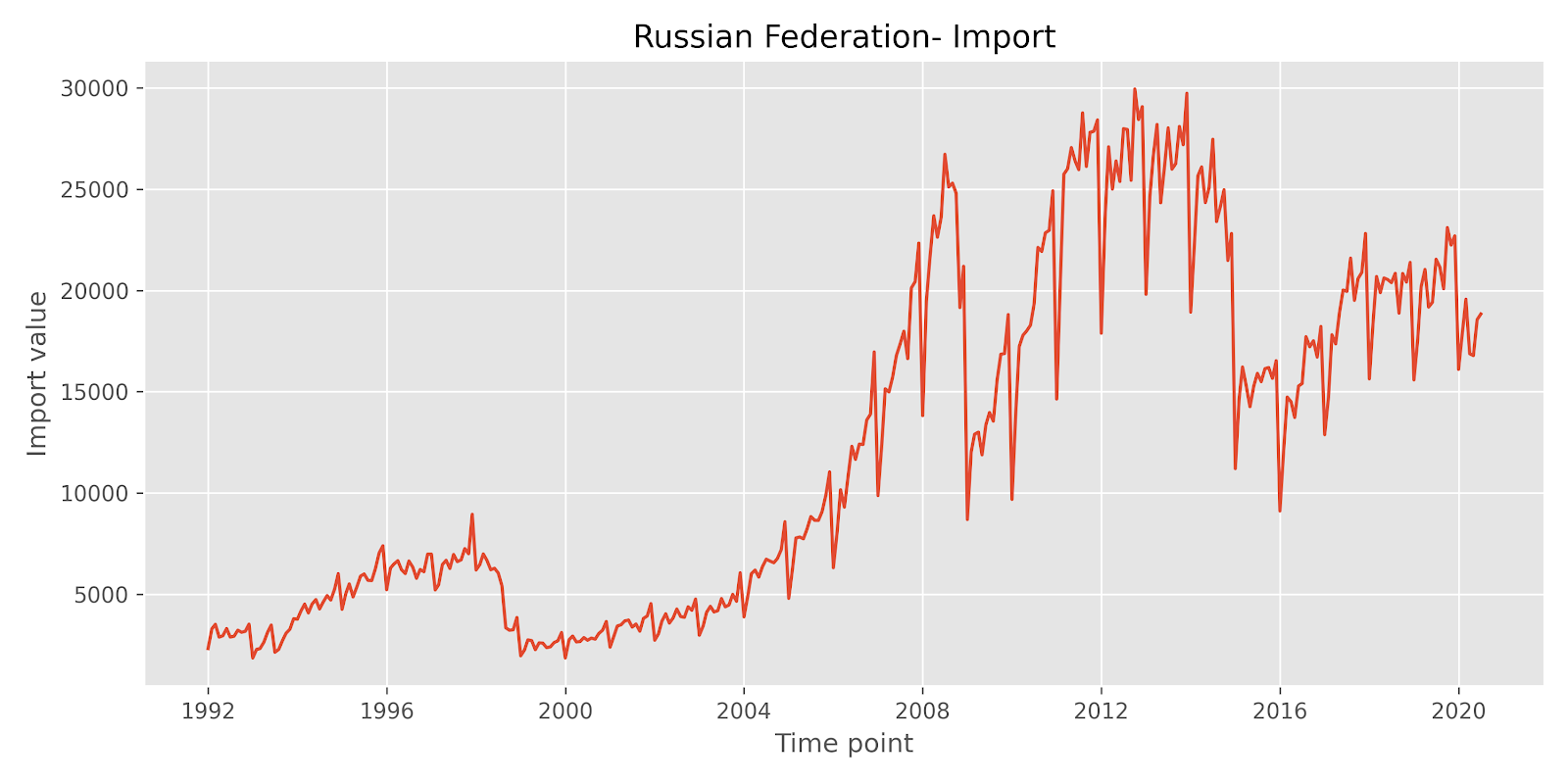


Рисунок 14 – График суммарного объёма импорта в Республику Бенин по всем товарным группам в долларах

На рисунке 15 приведена эволюция модели в процессе обучения. В качестве метрик выбраны значение функции потерь (loss) и средняя абсолютная ошибка (MAE). Видно, что значение val\_loss выходит на плато примерно после 200 итераций обновления весов. Значение MAE при этом так же стабильно снижается.

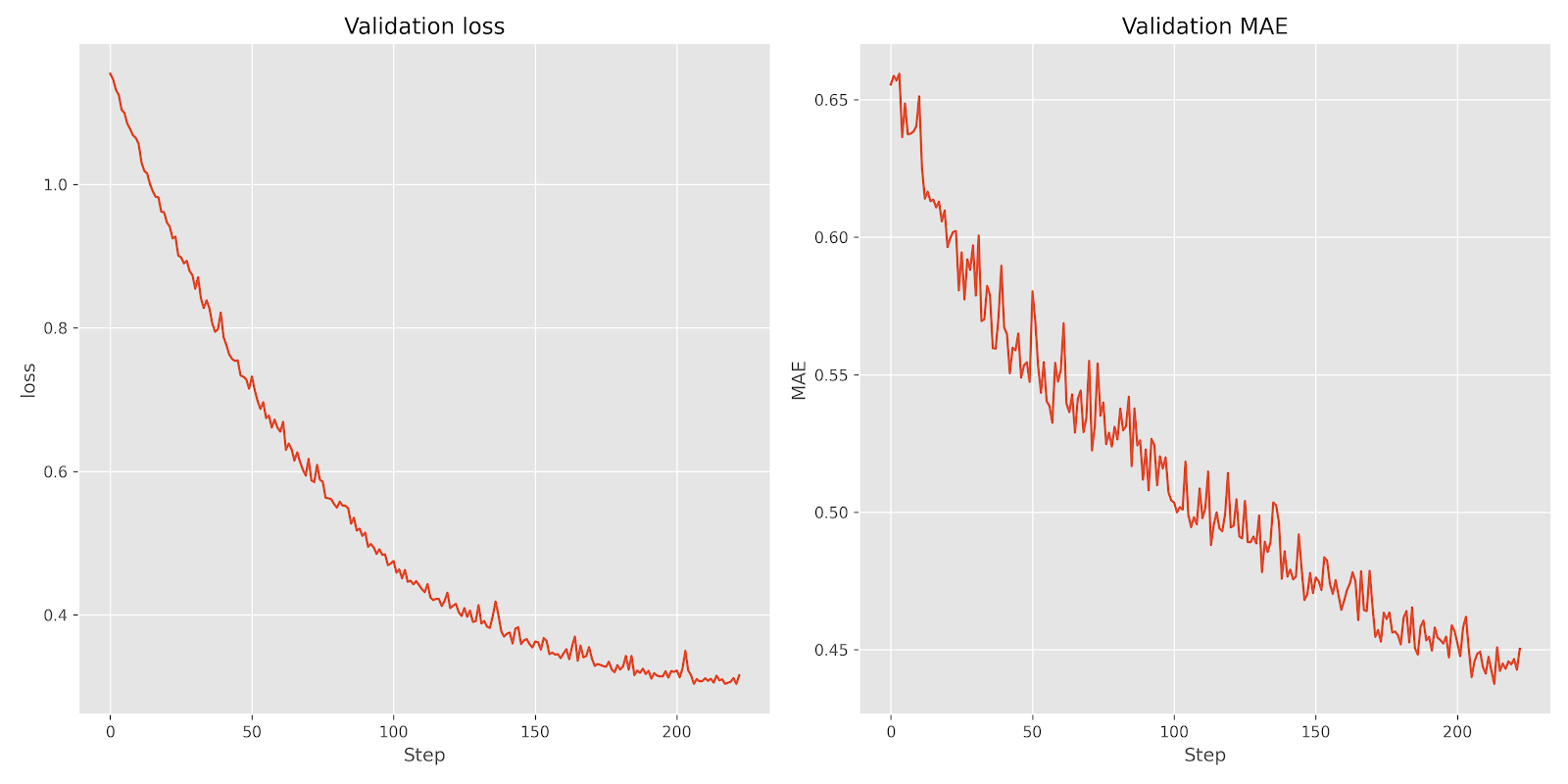


Рисунок 15 – Изменение метрик качества (loss и MAE) в процессе обучения модели на данных импорта для Российской Федерации

На рисунке 16 показан результат работы предсказательной модели на тренировочной части выборки, где прогнозы отрисованы зеленым цветом, красным – реальные значения. Из каждой временной точки тренировочной выборки построен прогноз на три временные точки вперед. Как видно, модель хорошо описывает сезонные колебания, но ожидаемо плохо предсказывает аномальные изломы в линии тренда.

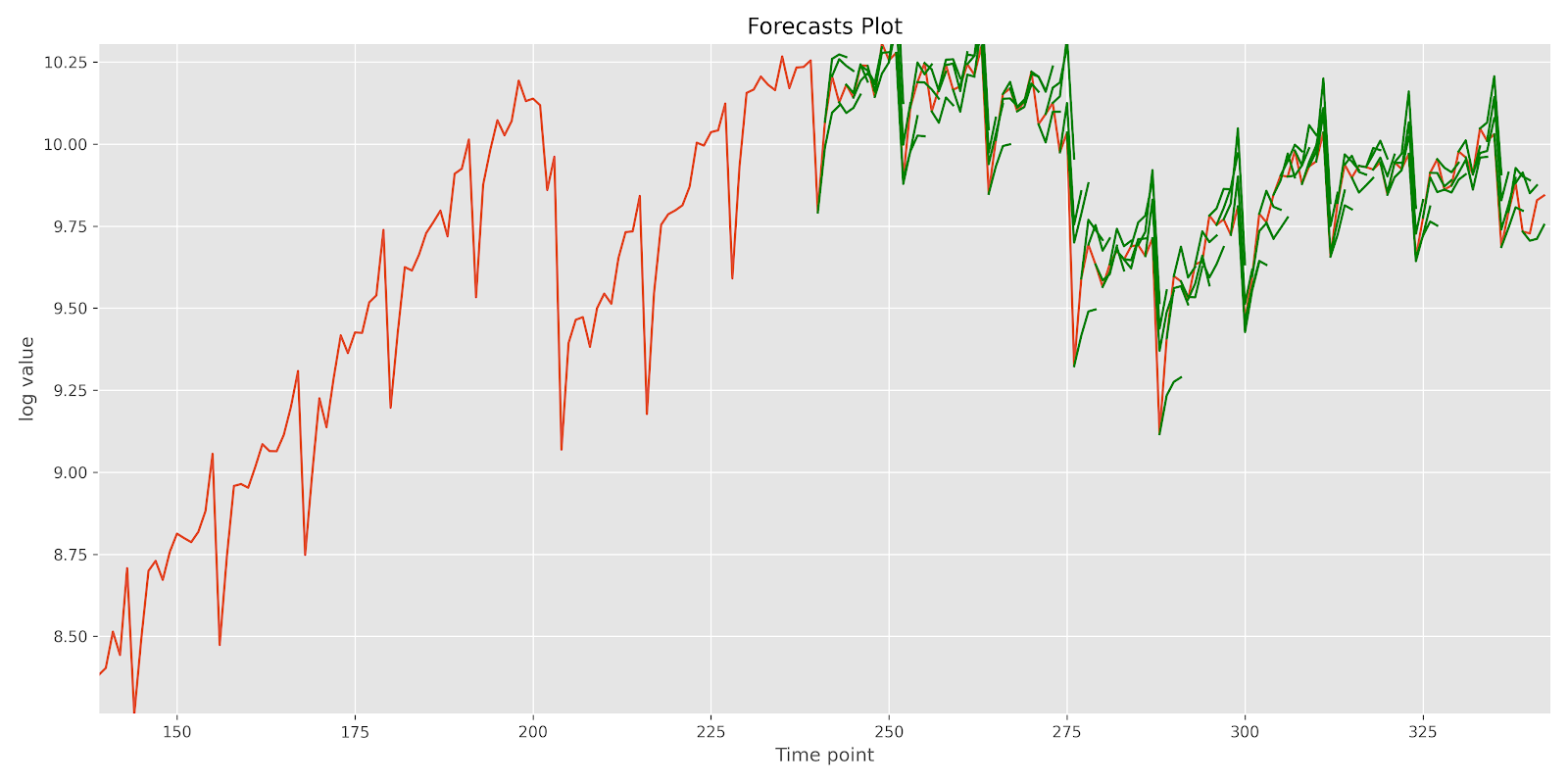


Рисунок 16 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Российской Федерации

Для каждого шага («дальности») прогноза было посчитано среднеквадратичное отклонение по прогнозам из всех точек тестовой выборки:

1 шаг прогноза (t+1) RMSE: 0.089396

2 шаг прогноза (t+2) RMSE: 0.121422

3 шаг прогноза (t+3) RMSE: 0.131092

Ожидаемо, что отклонение от реальных значений растет с каждым шагом, поскольку растет эффект накопленной ошибки. В среднем приемлемое качество предсказания достигается на интервалах не более одного сезонного цикла, и сильно зависит от расположения временной точки в пределах цикла (иными словами, точность прогноза зависит от месяца).

Подобные результаты были получены для 15 стран, для которых имелись качественные данные по товарообороту:

1. Российская Федерация;
2. США;
3. Великорбритания;
4. Германия;
5. Китай, Гонконг;
6. Канада;
7. Япония;
8. Турция;
9. Дания;
10. Мальдивы;
11. Швеция;
12. Южная Корея;
13. Швейцария;
14. Тунис;
15. Малайзия.

Под качеством данных здесь понимается достаточно длинный период и отсутствие пропусков и явных технических аномалий. Для каждой страны модель обучалась независимо. Пример результатов анализа приведен в таблице 1. Здесь представлены оценки качества модели для данных импорта 15 взятых в исследование стран. Аналогичная таблица для экспорта представлена в Приложении Г.

Таблица 1 – Оценка качества модели на данных объёма импорта разных стран

|  |  |  |  |
| --- | --- | --- | --- |
|  | val\_loss | val\_accuracy | val\_mae |
| Российская Федерация | 0.3106 | 0.4286 | 0.4399 |
| Binary accuracy: 0.6868686868686869 |  |  |  |
| США | 0.4077 | 0.6984 | 0.4812 |
| Binary accuracy: 0.569023569023569 |  |  |  |
| Великобритания | 0.3197 | 0.3651 | 0.4157 |
| Binary accuracy: 0.4882154882154882 |  |  |  |
| Германия | 0.5432 | 0.6222 | 0.5841 |
| Binary accuracy: 0.6902356902356902 |  |  |  |
| Китай, Гонконг | 0.4966 | 0.4603 | 0.5391 |
| Binary accuracy: 0.5925925925925926 |  |  |  |
| Канада | 0.3112 | 0.6825 | 0.411 |
| Binary accuracy: 0.6902356902356902 |  |  |  |
| Япония | 0.5868 | 0.7302 | 0.5511 |
| Binary accuracy: 0.7104377104377104 |  |  |  |
| Турция | 0.19 | 0.5556 | 0.3308 |
| Binary accuracy: 0.5892255892255892 |  |  |  |
| Дания | 0.377 | 0.5873 | 0.4878 |
| Binary accuracy: 0.6734006734006734 |  |  |  |
| Мальдивы | 0.1179 | 0.4286 | 0.2897 |
| Binary accuracy: 0.48148148148148145 |  |  |  |
| Швеция | 0.4075 | 0.6667 | 0.5022 |
| Binary accuracy: 0.6296296296296297 |  |  |  |
| Южная Корея | 0.1987 | 0.5079 | 0.3252 |
| Binary accuracy: 0.5387205387205387 |  |  |  |
| Швейцария | 0.3455 | 0.6667 | 0.4511 |
| Binary accuracy: 0.7171717171717171 |  |  |  |
| Тунис | 0.365 | 0.4138 | 0.4575 |
| Binary accuracy: 0.5454545454545454 |  |  |  |
| Малайзия | 0.5475 | 0.5849 | 0.5836 |
| Binary accuracy: 0.5892255892255892 |  |  |  |
| Аргентина | 0.4014 | 0.5556 | 0.476 |
| Binary accuracy: 0.7239057239057239 |  |  |  |

**Область применимости**

Область применимости моделей ограничена временными рядами, которые имеют годовую сезонность. Наличие единого тренда в данных ожидаемо повышает точность, аномальные изломы в тренде ухудшают качество прогноза.

В качестве примера аномального с точки зрения модели поведения временного ряда приведём график импорта для США. Так на рисунке 17 изображен график ежемесячного объёма импорта для США.

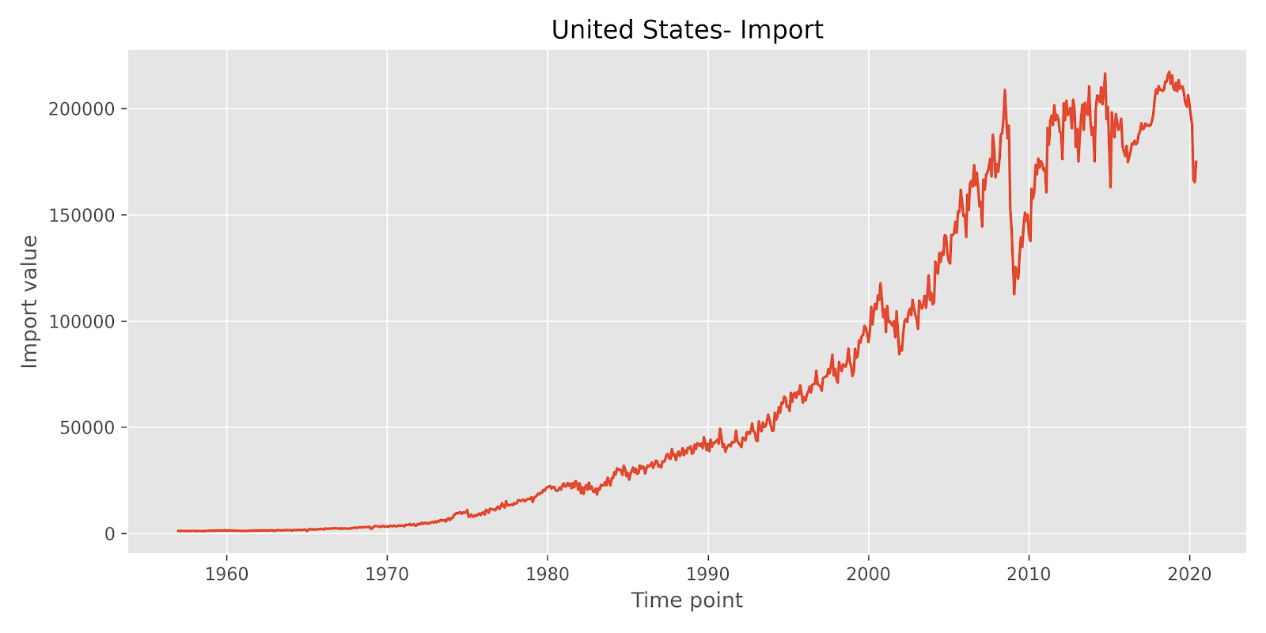


Рис 17 – График ежемесячного объем импорта для США

На участке, соответствующем 2017-2020 годам, заметно изменение поведения временного ряда: пропадает четкая цикличность, отсутствует единый тренд. Поскольку модель обучается на всем временном промежутке, такие отклонение от обычного поведения приводят к худшей предсказательной силе, что можно видеть на рисунке 18.

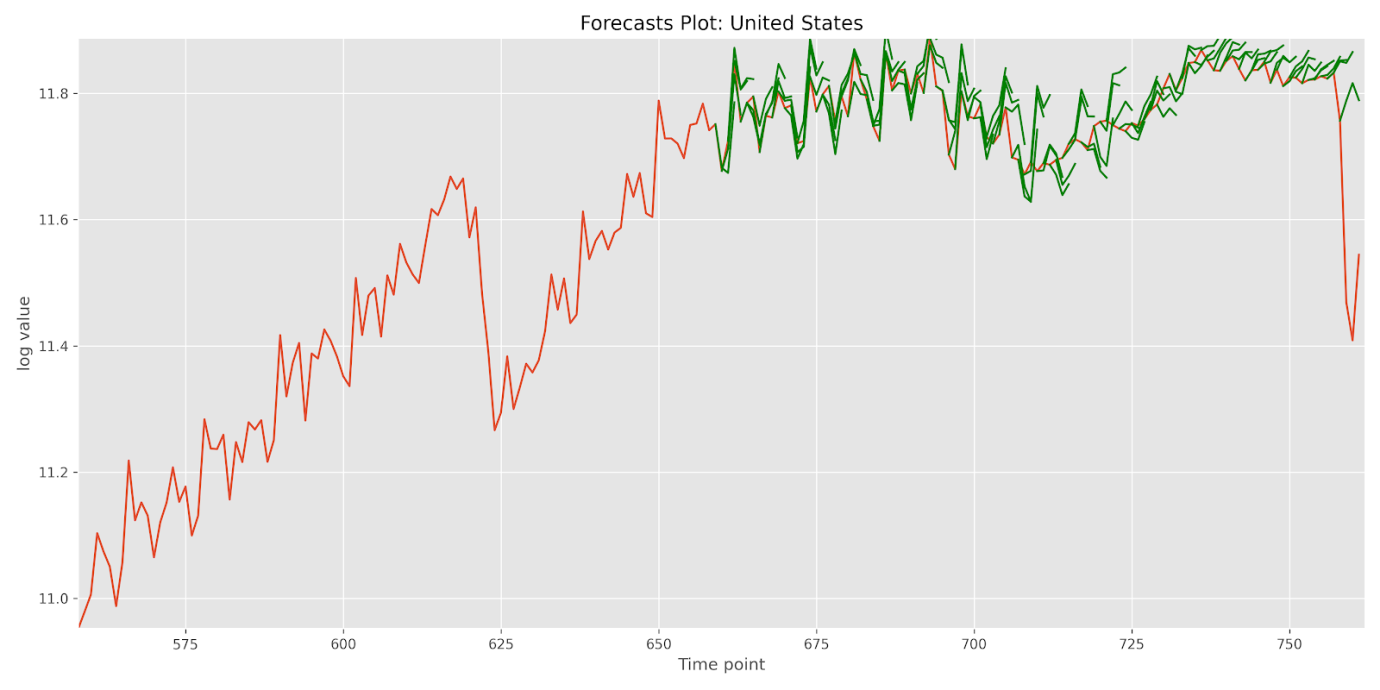


Рисунок 18 – Прогноз ежемесячного объема импорта для США

Также была предпринята попытка прогнозирования таких временных рядов, как отношение EUR/USD, стоимость сырьевых продуктов и акций. Данные были взяты из источника [25]. В результате качество прогноза было на уровне случайного предсказания. Такой слабый результат – следствие специфики изменений подобных временных рядов. Например, в исследовании [26] на примере нигерийского фондового рынка было показано, что поведение продуктов, активно продаваемых и покупаемых на фондовой бирже, приближено к случайному блужданию, отсутствует четкая цикличность и тренд, поэтому простой технический анализ не в состоянии эффективно строить прогнозы только по данным предыдущих временных интервалов. Более формально, такие временные ряды представляют из себя марковский процесс, когда состояние в ближайшем будущем определяется только состоянием в настоящем, и не зависит от прошлых состояний.

# ЗАКЛЮЧЕНИЕ

Задача прогнозирования временных рядов имеет высокую актуальность для многих предметных областей и является неотъемлемой частью повседневной работы многих компаний. На сегодняшний день разработано множество моделей для решения задачи прогнозирования временного ряда, наибольшую применимость среди которых  имеют авторегрессионные и нейросетевые модели.

В результате выполнения работы была спроектирована и реализована нейросетевая модель для прогнозирования торгового трафика. Также были рассмотрены понятия и определения из области нейронных сетей, проанализированы достоинства и недостатки моделей и методов прогнозирования временных рядов.

Для достижения цели были пройдены основные этапы проектирования от изучения предметной области до апробации модели.

В ходе работы был сделан вывод о том, что область применимости созданной модели ограничена временными рядами, которые имеют годовую сезонность.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ:

1 <http://www.ysu.am/files/Paper4.pdf>

2 https://laws.studio/uchebniki-statistika/metod-analiticheskogo-vyiravnivaniya.html

3 <https://bigenc.ru/economics/text/2087489>

4 Winters P.R. Forecasting sales by exponentially weighted moving averages // Management Science. – 1960. – Vol. 6. – № 3.  
5 https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%B5%D0%B4%D0%B8%D0%BA%D1%82%D0%BE%D1%80

6 Draper N., Smith H. Applied regression analysis

7 <http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BD%D0%BE%D0%B3%D0%BE%D0%BC%D0%B5%D1%80%D0%BD%D0%B0%D1%8F_%D0%BB%D0%B8%D0%BD%D0%B5%D0%B9%D0%BD%D0%B0%D1%8F_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F>

8 http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4\_%D1%81%D1%82%D0%BE%D1%85%D0%B0%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B3%D0%BE\_%D0%B3%D1%80%D0%B0%D0%B4%D0%B8%D0%B5%D0%BD%D1%82%D0%B0

9 *Лукашин Ю. П.* Адаптивные методы краткосрочного прогнозирования временных рядов

10 <https://help.loginom.ru/userguide/processors/datamining/arimax.html>

11 https://wiki.loginom.ru/articles/garch-model.html

12 <https://www.mbureau.ru/articles/dissertaciya-model-prognozirovaniya-vremennyh-ryadov-glava-1>

13 <https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C>

14 <https://www.youtube.com/watch?v=L1tsgJ9m1zM>

15 Нейронные сети. Полный курс Автор: Саймон Хайкин

16 <https://neurohive.io/ru/osnovy-data-science/lstm-nejronnaja-set/>

17 https://web.archive.org/web/20161123045043/http://deeplearning.cs.cmu.edu:80/pdfs/Hochreiter97\_lstm.pdf

18 https://polarwinco.com/2019/03/06/gru-vs-lstm/

19 Hannes Y.Y., Webb P. Classification and regression trees: A User Manual for IdentifyingIndicators of Vulnerability to Famine and Chronic Food Insecurity

20 <https://techrocks.ru/2018/07/22/5-key-libraries-and-packets-for-data-analysis-in-python/>

21

22 <http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%91%D0%BE%D0%BA%D1%81%D0%B0-%D0%9A%D0%BE%D0%BA%D1%81%D0%B0>

23 <https://www.anaconda.com/>

24 http://arxiv.org/abs/1412.6980

25 <https://ru.investing.com/>

26 <http://www.thescipub.com/abstract/10.3844/jmssp.2010.342.346>

27 <https://www.tensorflow.org/tutorials/structured_data/time_series?hl=en>

28 <http://www.machinelearning.ru/wiki/images/archive/c/cb/20160412121749%21Voron-ML-forecasting-slides.pdf>

29 <http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%B0%D1%82%D0%B5%D0%B3%D0%BE%D1%80%D0%B8%D1%8F:%D0%9F%D1%80%D0%BE%D0%B3%D0%BD%D0%BE%D0%B7%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%B2%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D1%85_%D1%80%D1%8F%D0%B4%D0%BE%D0%B2>

30 <https://www.tensorflow.org/api_docs/python/tf/keras/layers/Conv1D>

31 <https://www.python.org/>

32 <https://pythonru.com/osnovy/operatory-python>

33 <https://keras.io/>

34 <https://habr.com/ru/company/ods/blog/325432/>

35 <https://python.ivan-shamaev.ru/keras-tutorial-beginner-guide-to-deep-learning-in-python/>

# Приложение А. Листинг программы

# Приложение Б. Графики прогнозов экспорта

В этом приложении представлены графики прогнозов товарооборота экспорта, которые были построены на 3 месяца вперёд для соответствующих стран.

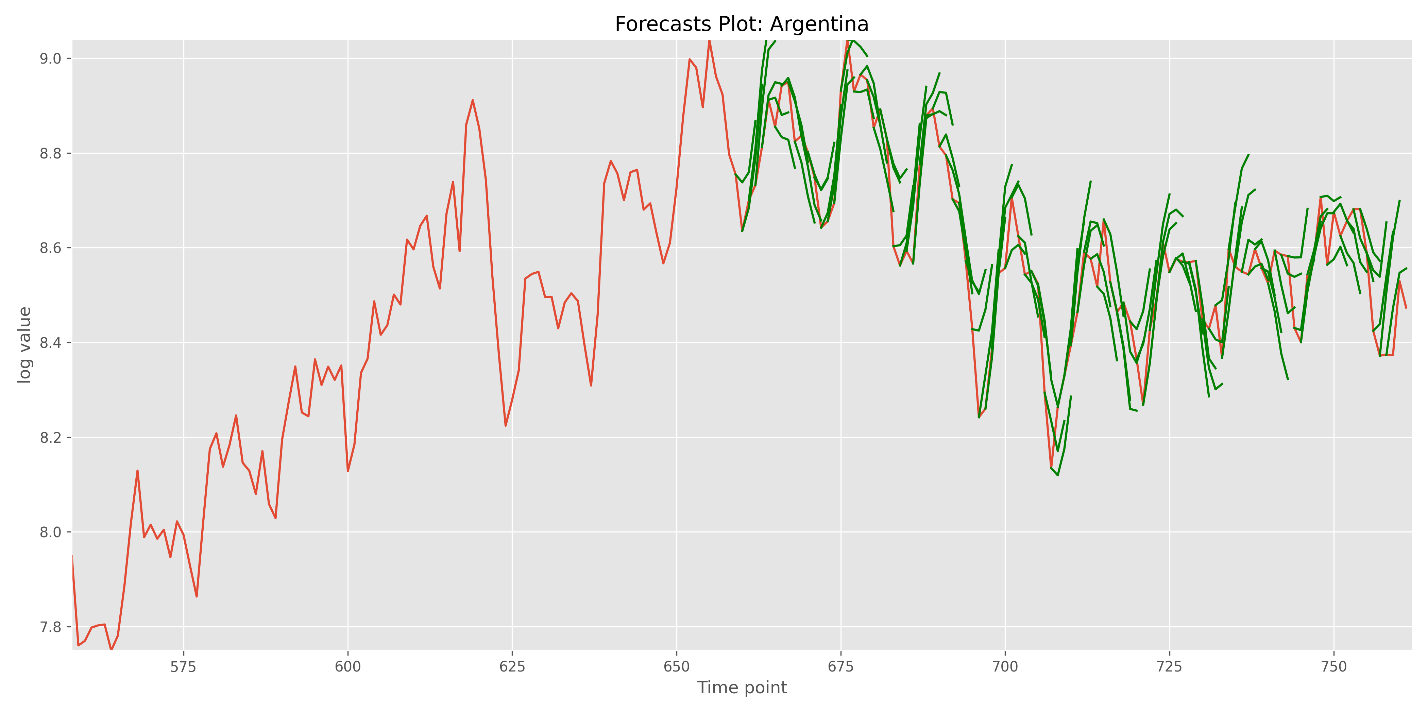


Рисунок 19 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Аргентины

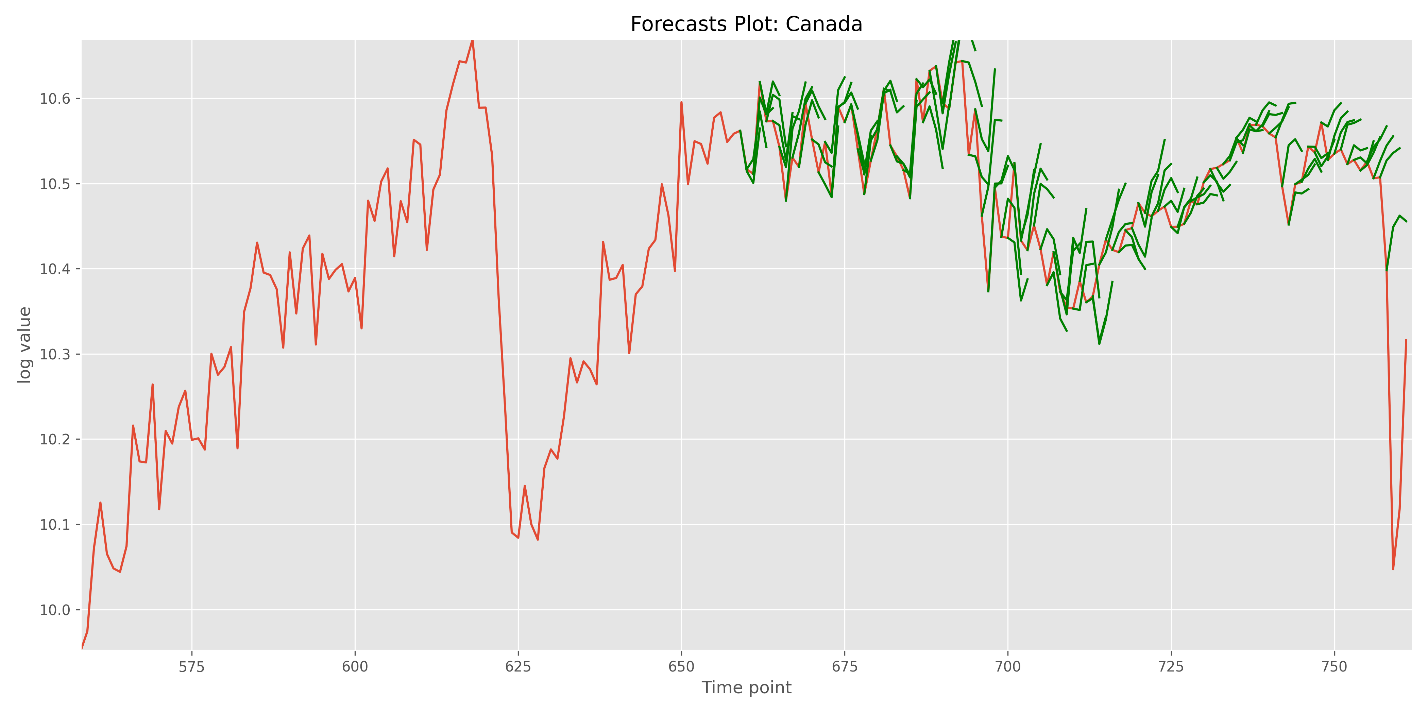


Рисунок 20 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Канады

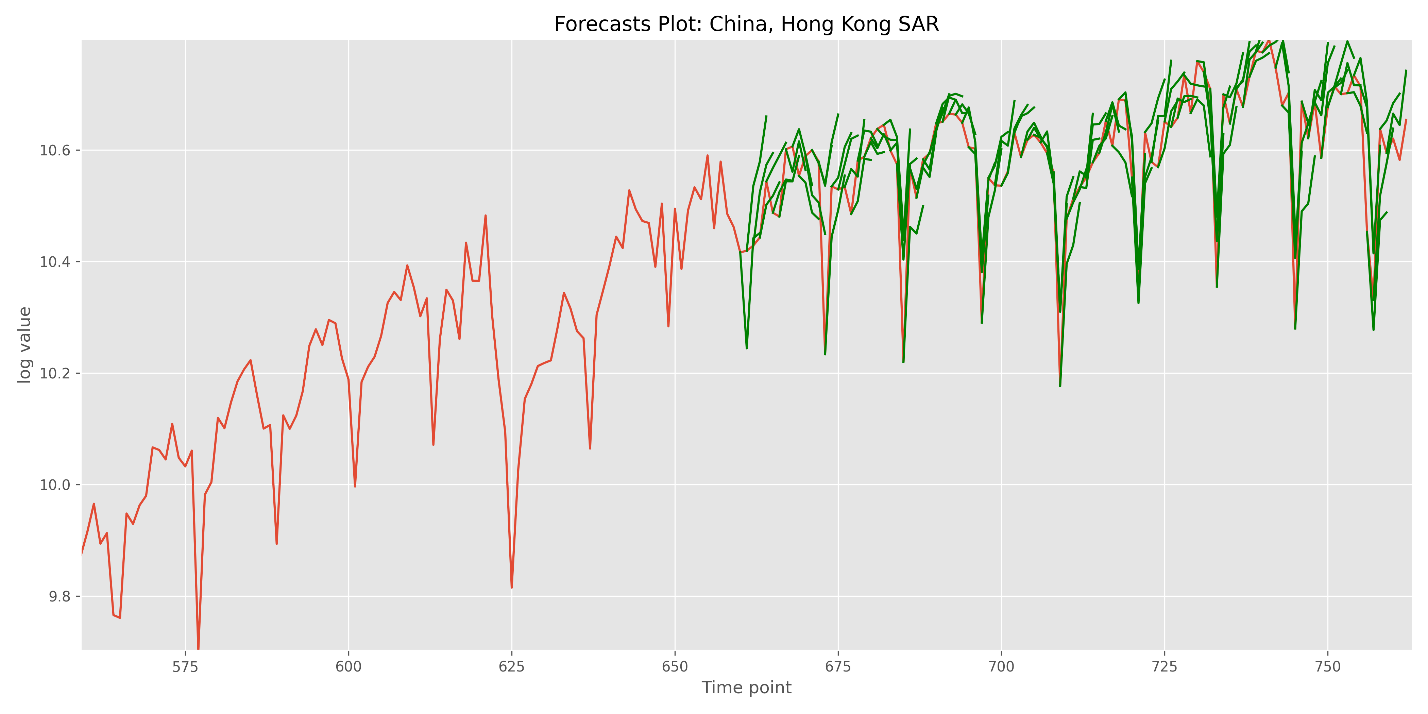


Рисунок 21 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Китая, Гонконга

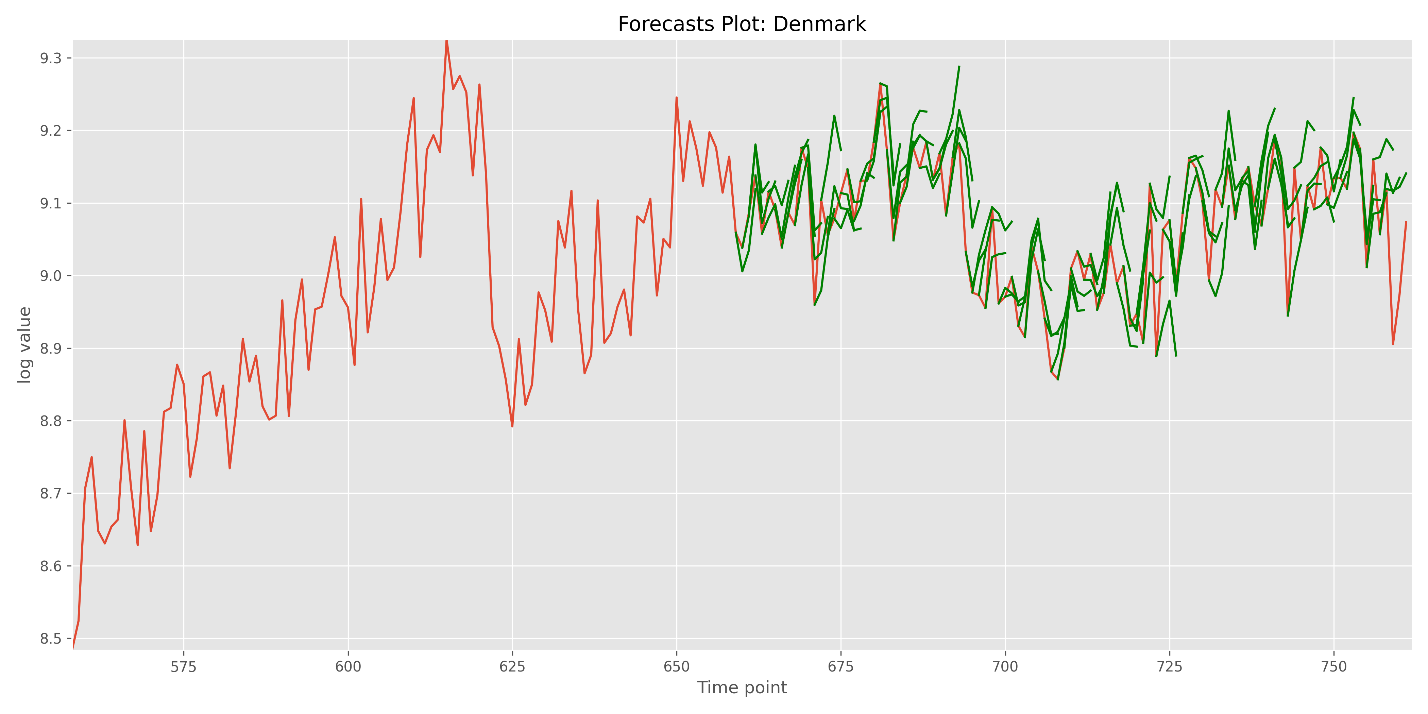


Рисунок 22 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Дании

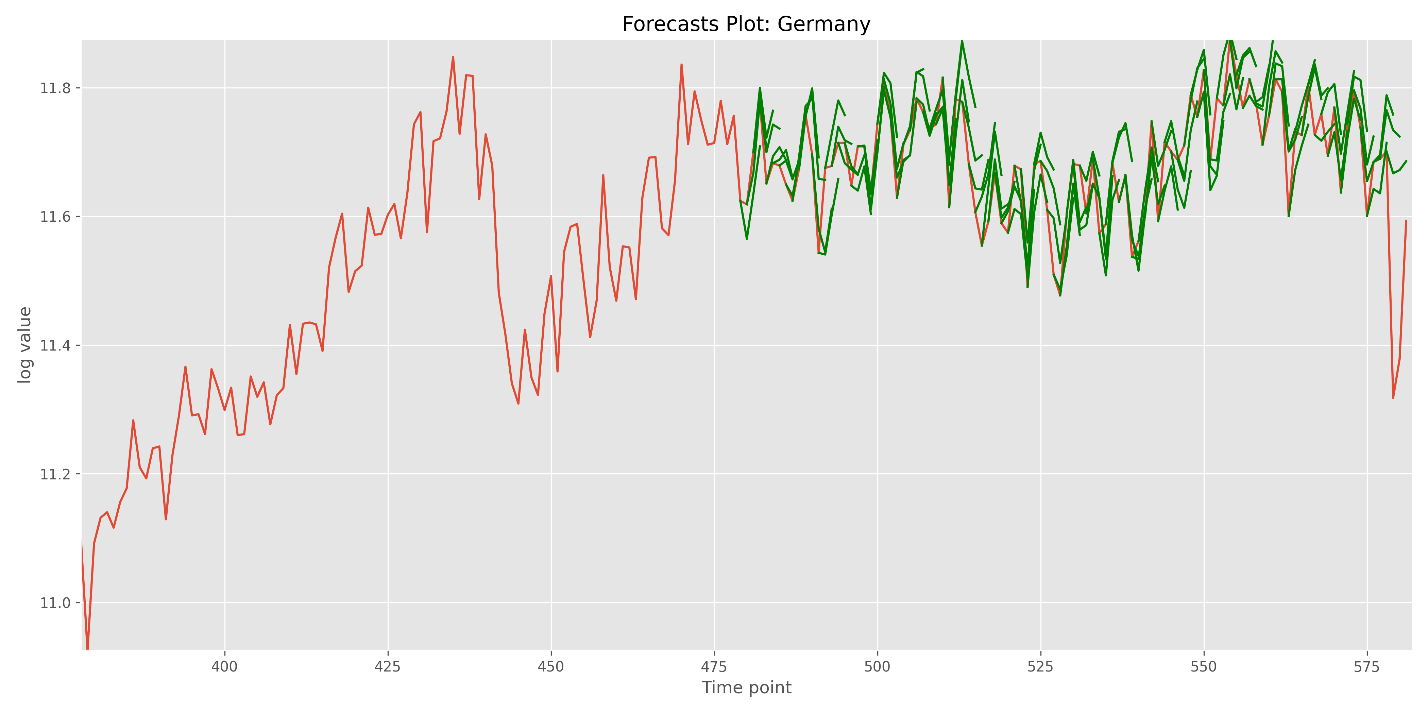


Рисунок 23 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Германии

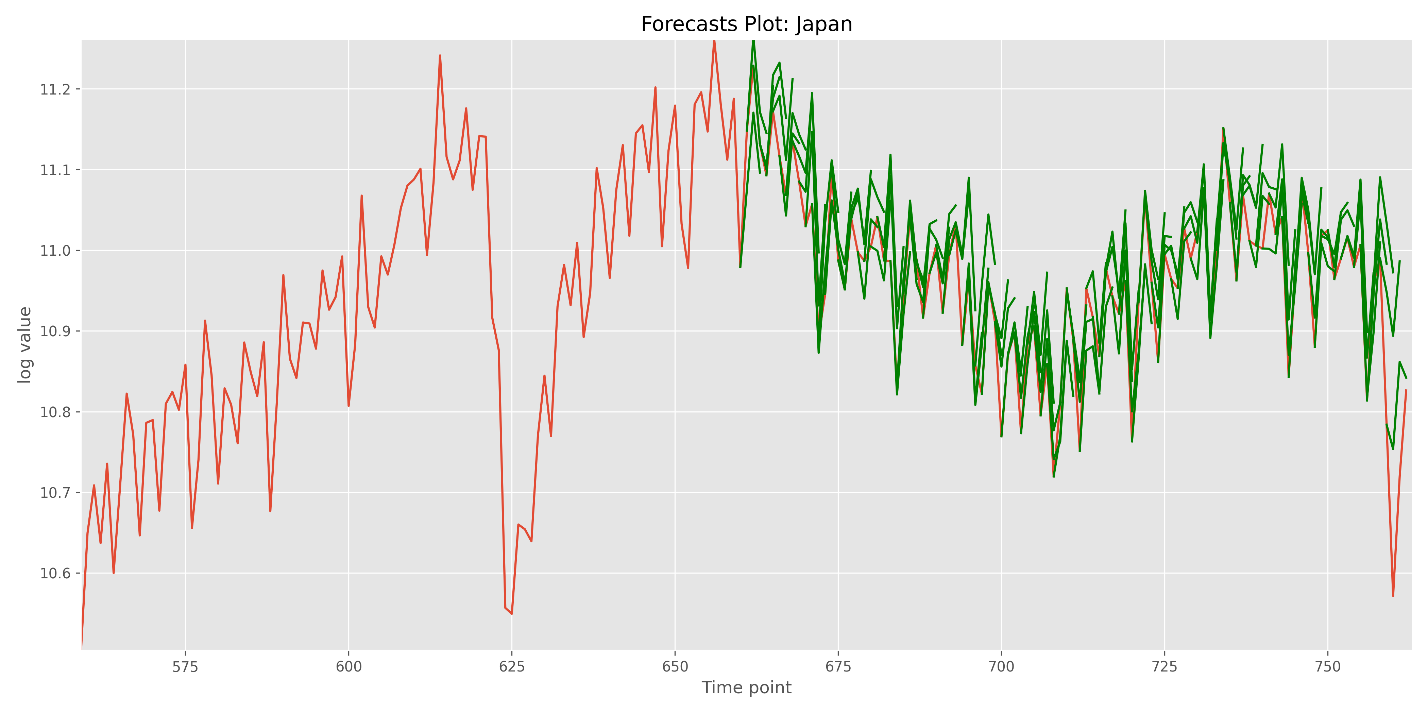


Рисунок 24 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Японии

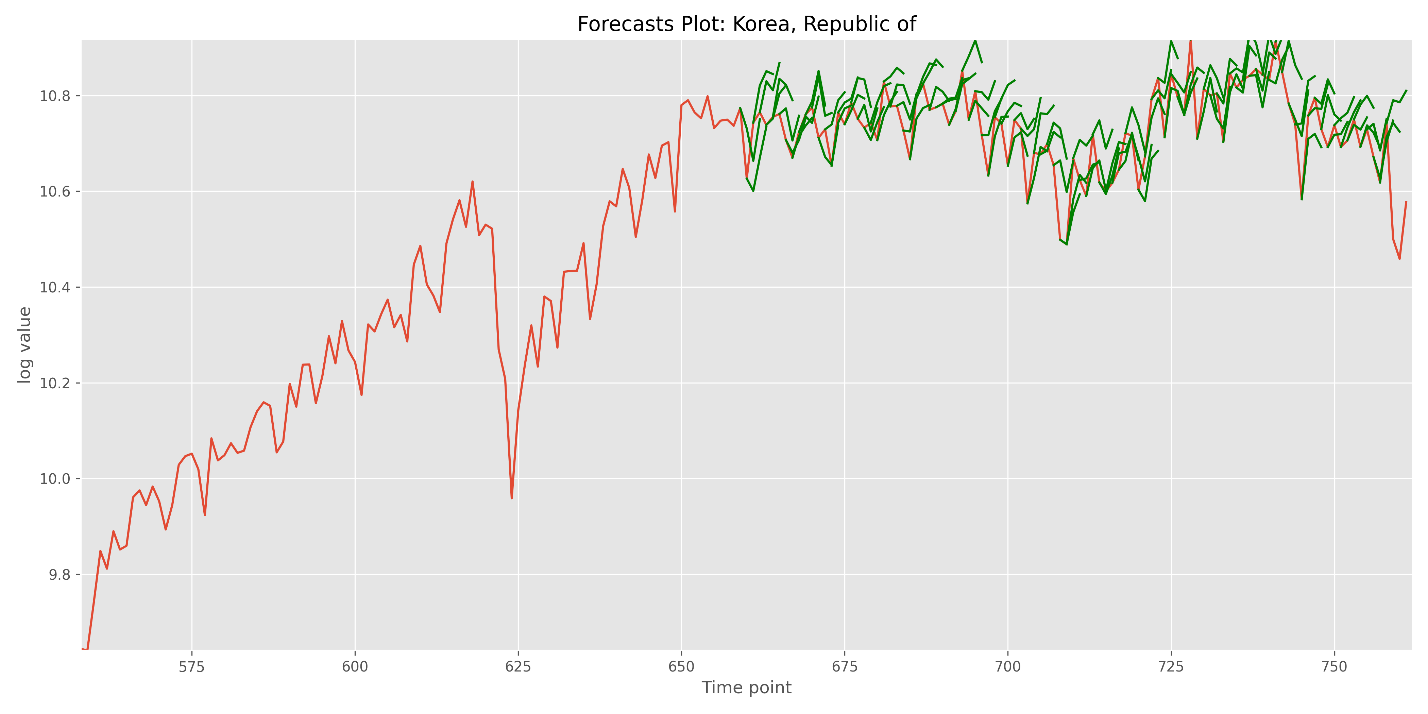
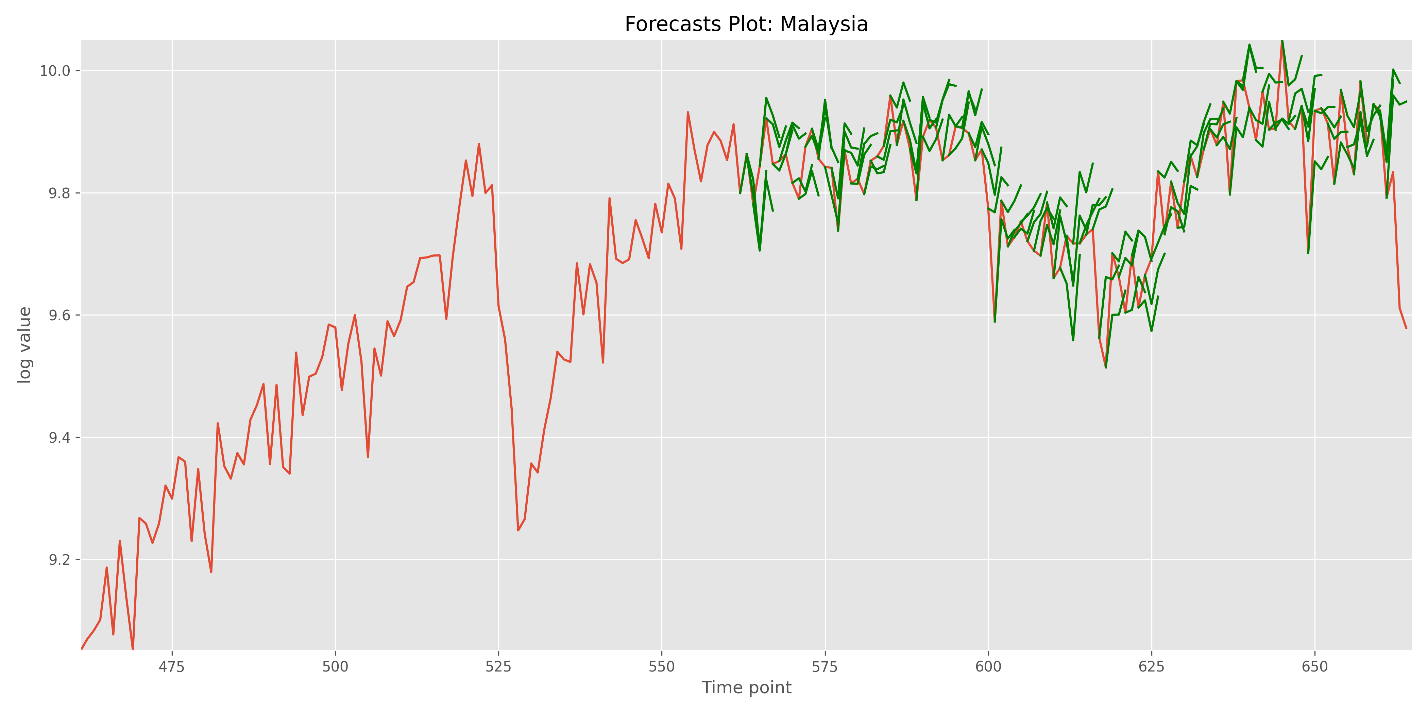


Рисунок 25 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Кореи

 Рисунок 26 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Малайзии

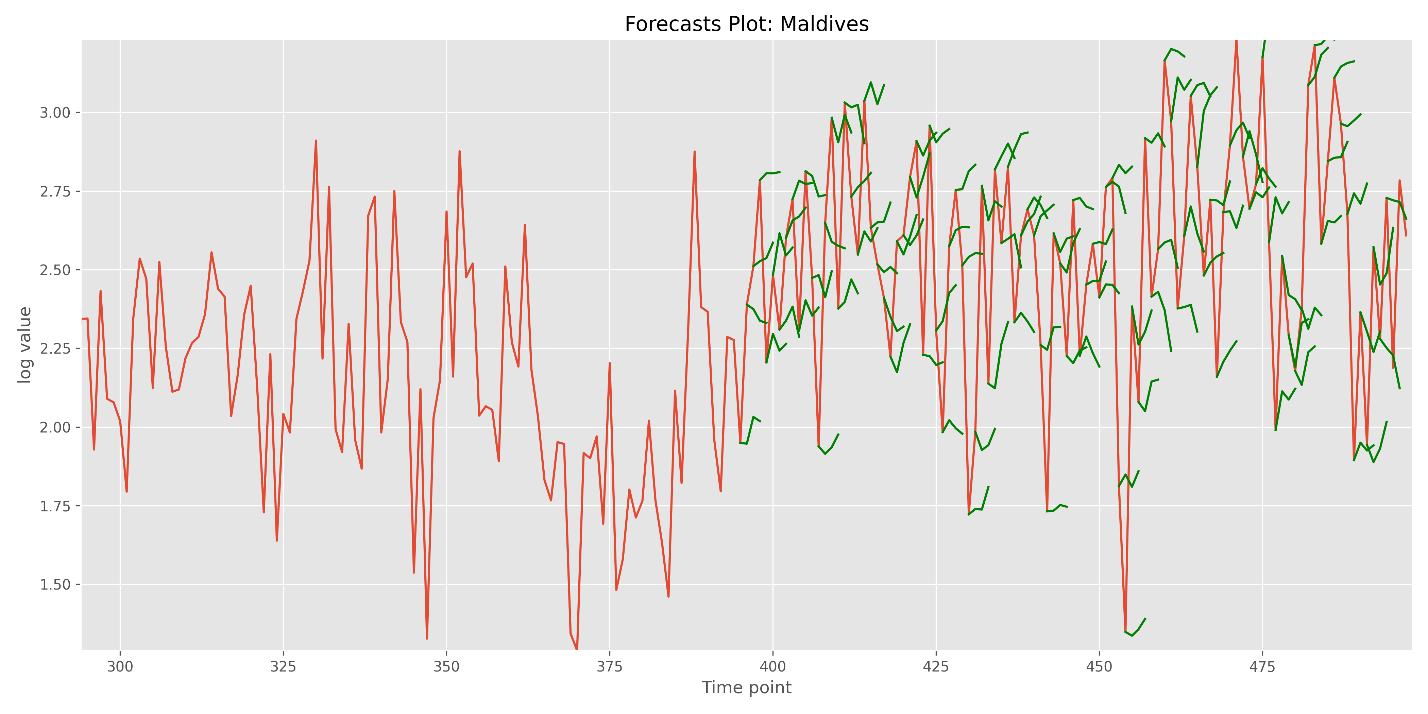


Рисунок 27 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Мальдив

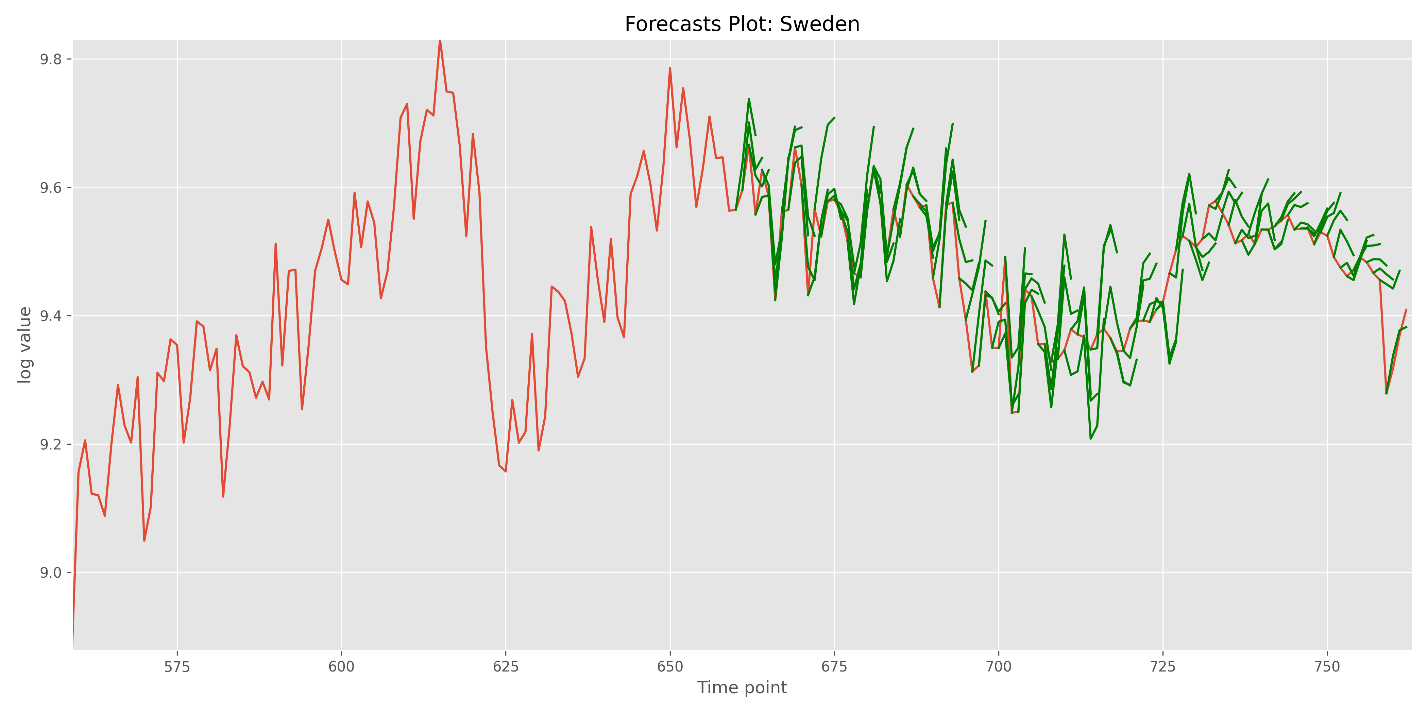


Рисунок 28 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Швеции

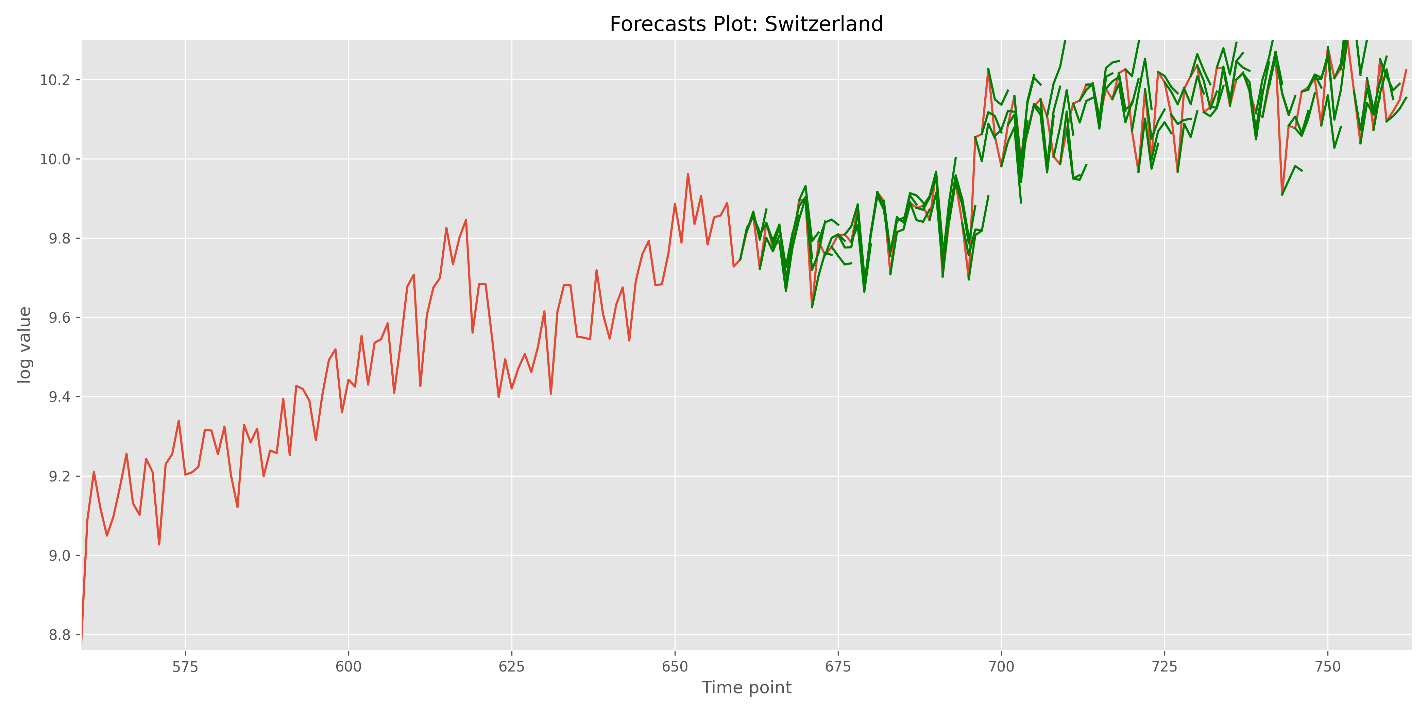


Рисунок 29 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Швейцарии.

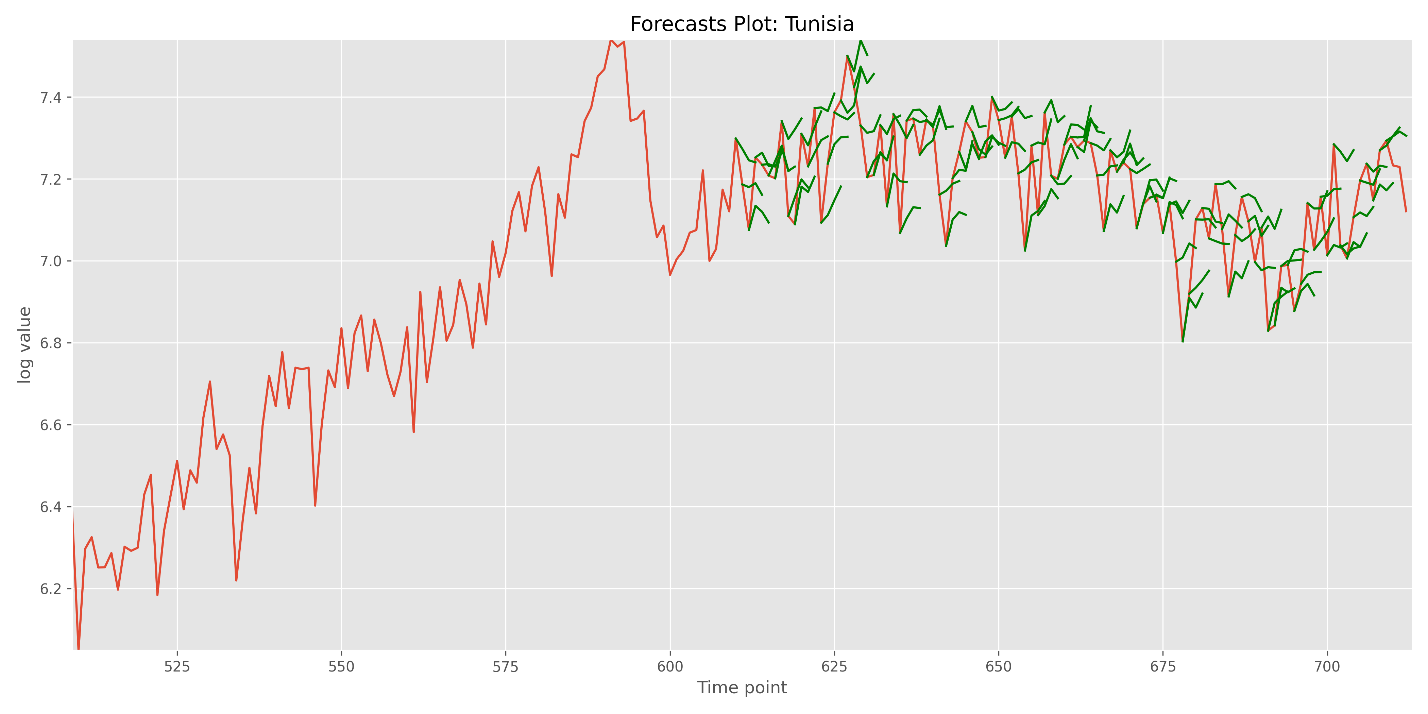


Рисунок 30 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Туниса.

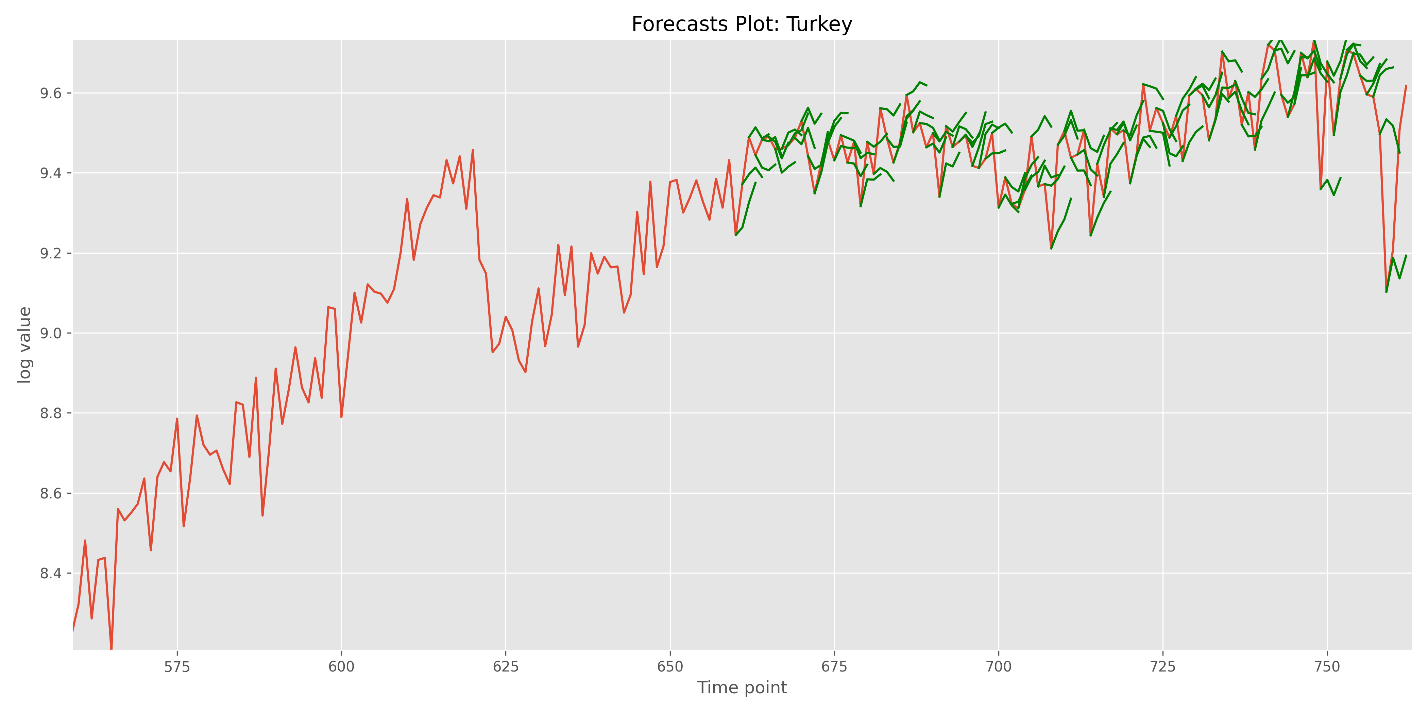


Рисунок 31 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Турции.

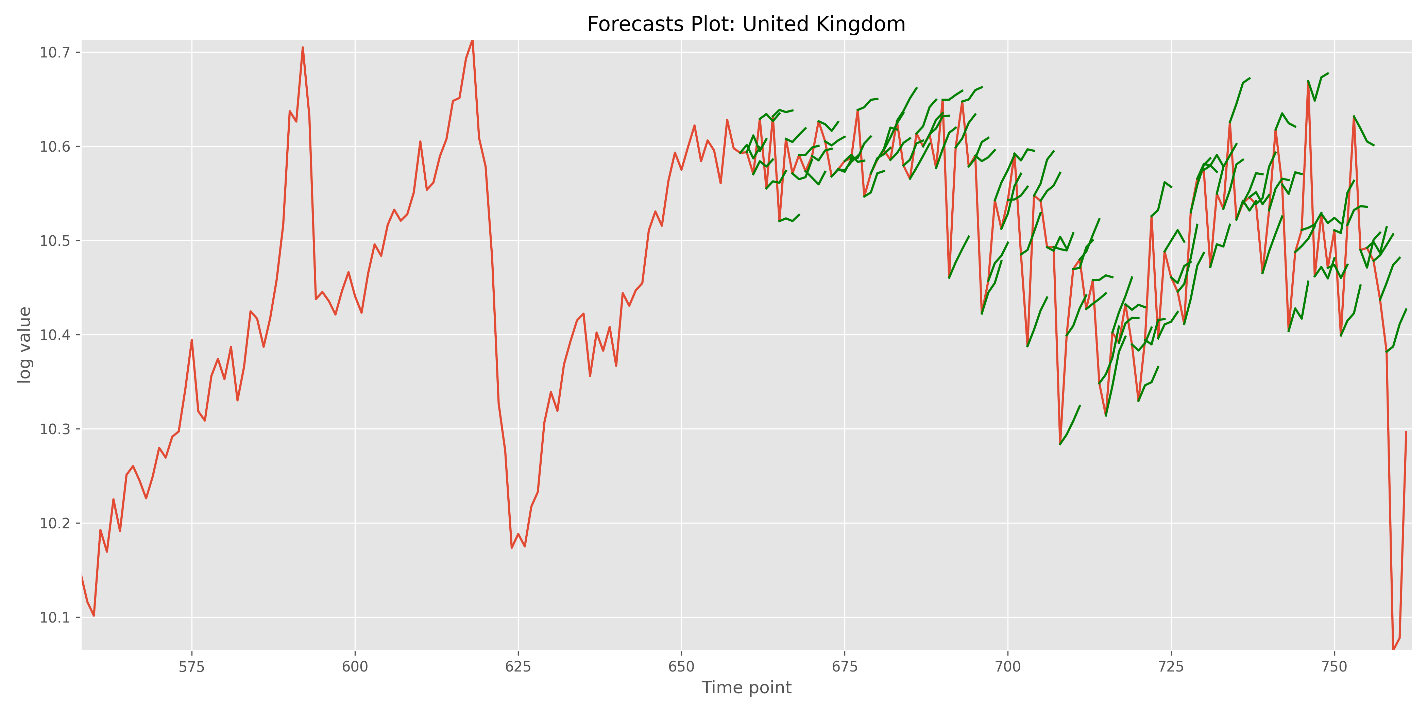


Рисунок 32 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Великобритании.

# Приложение В. Графики прогнозов импорта

В этом приложении представлены графики прогнозов товарооборота импорта, которые были построены на 3 месяца вперёд для соответствующих стран.

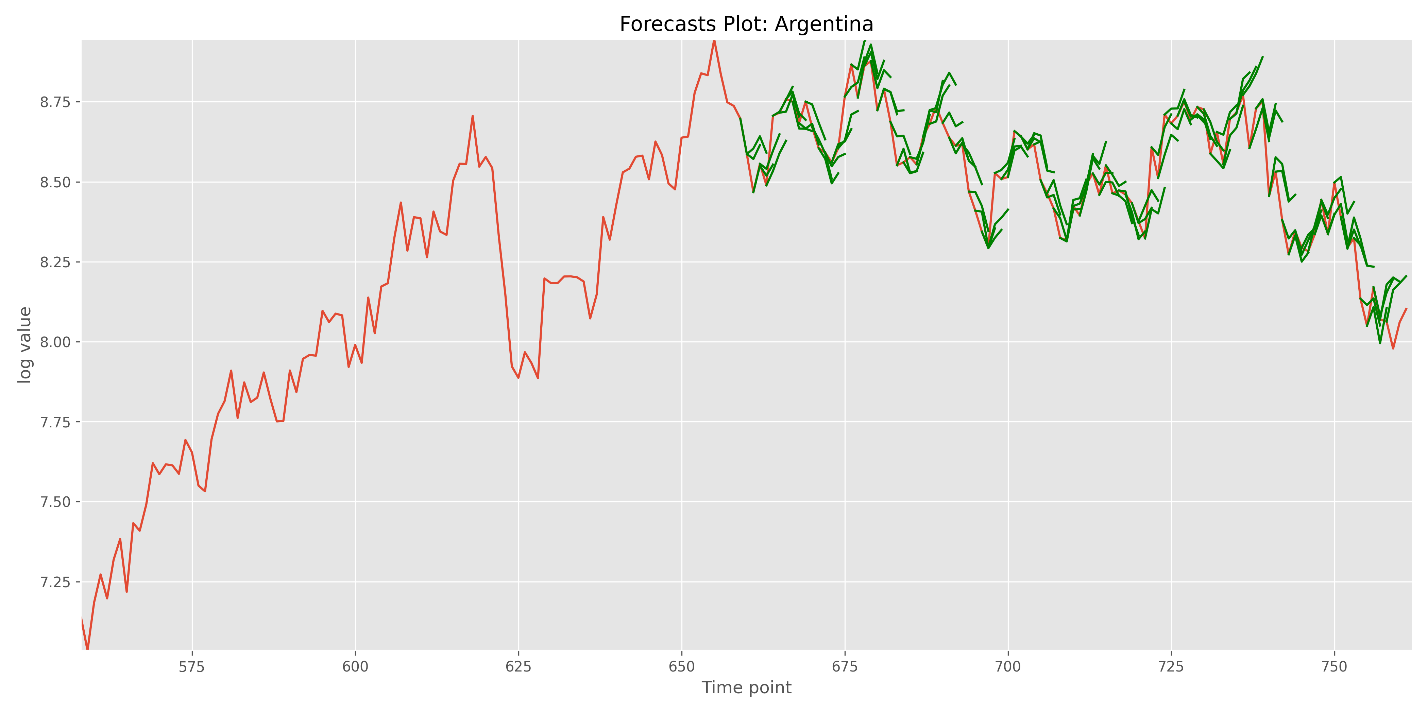


Рисунок 33 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Аргентины.

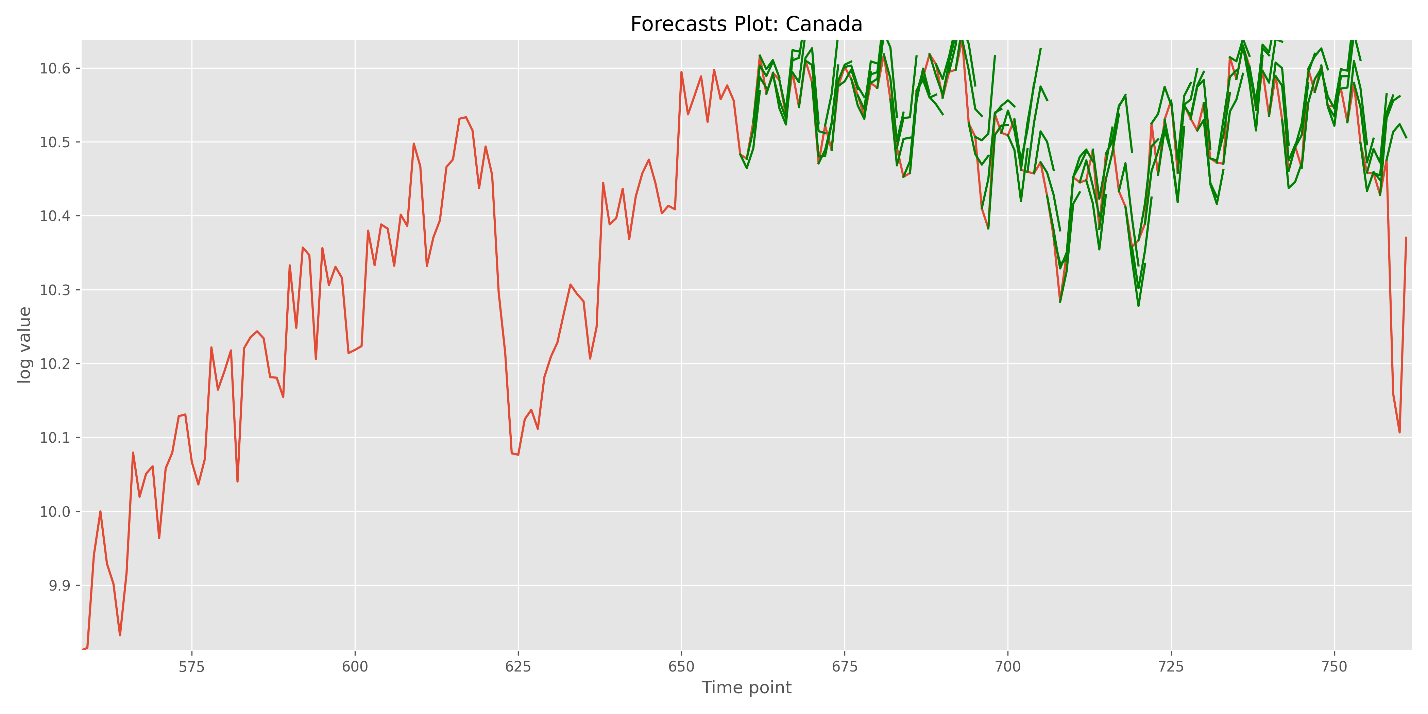


Рисунок 34 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Канады.

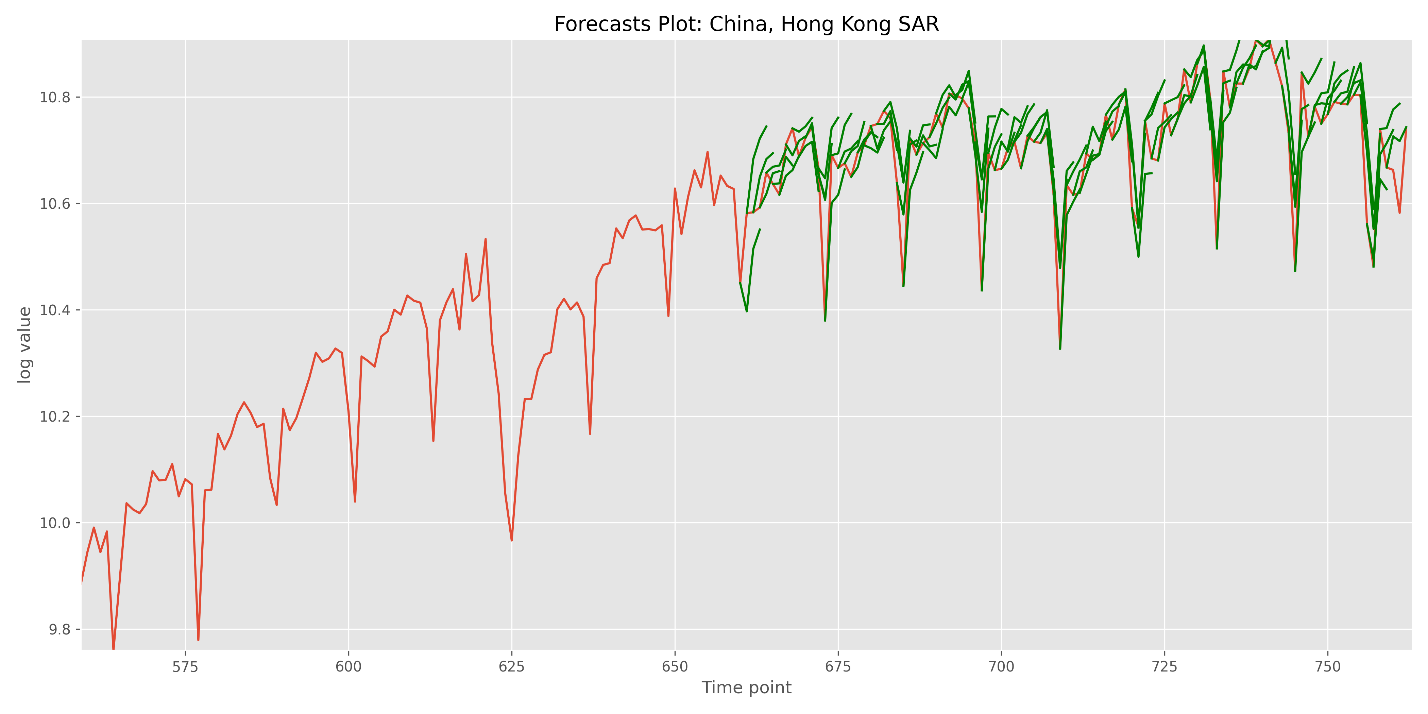


Рисунок 35 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Китая, Гонконга.

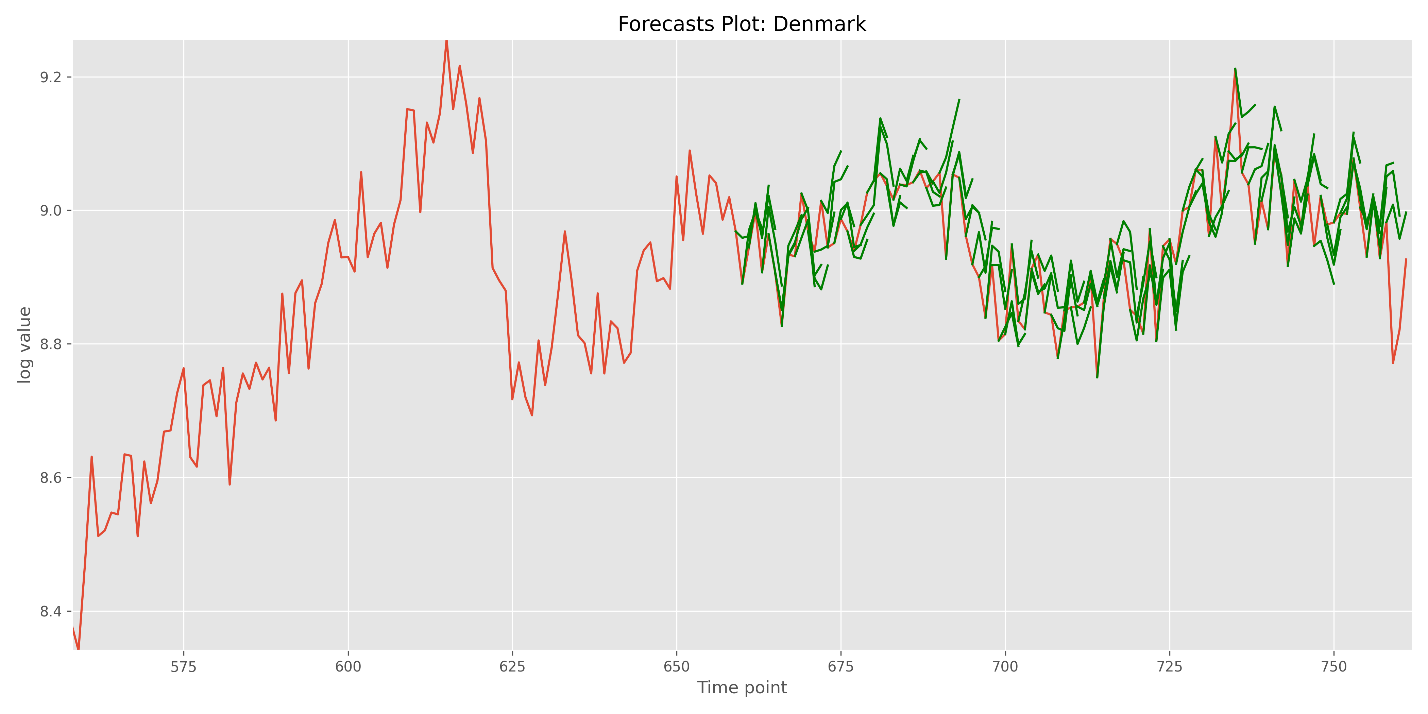


Рисунок 36 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Дании.

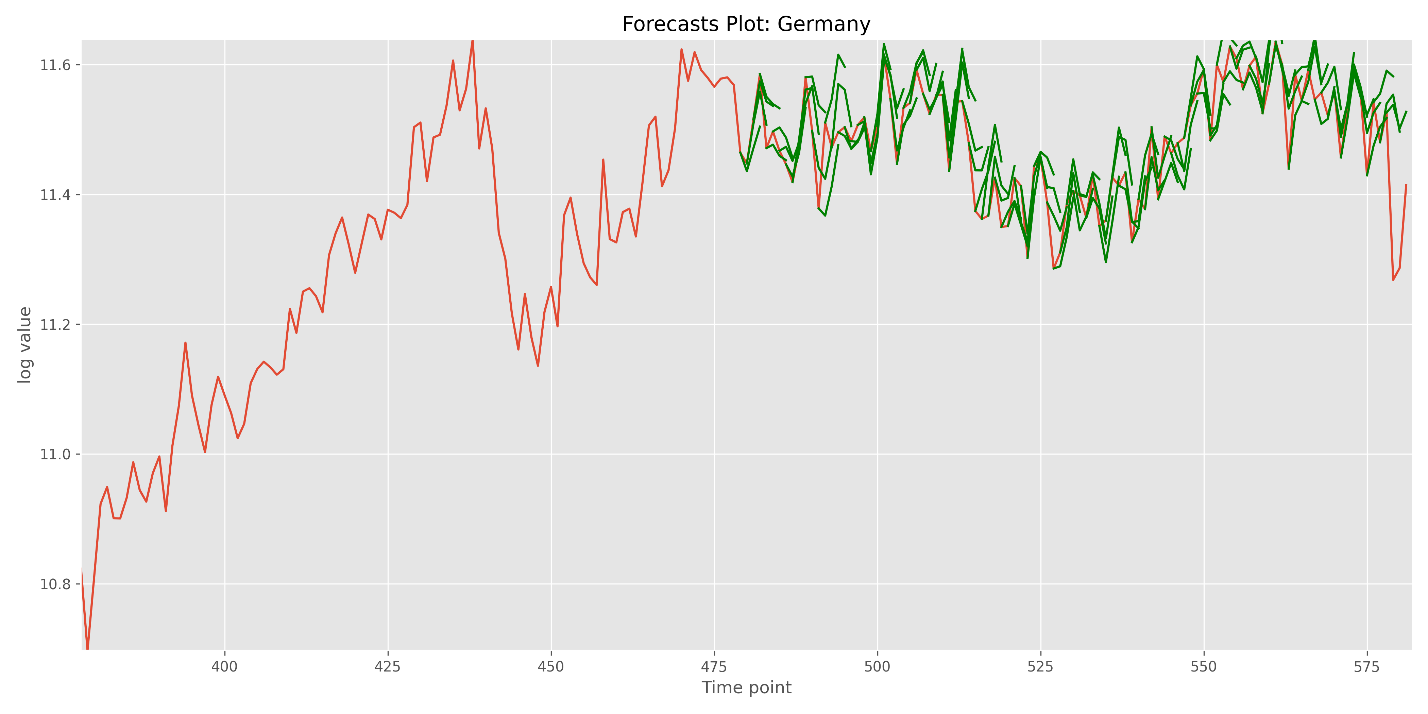


Рисунок 37 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Германии.

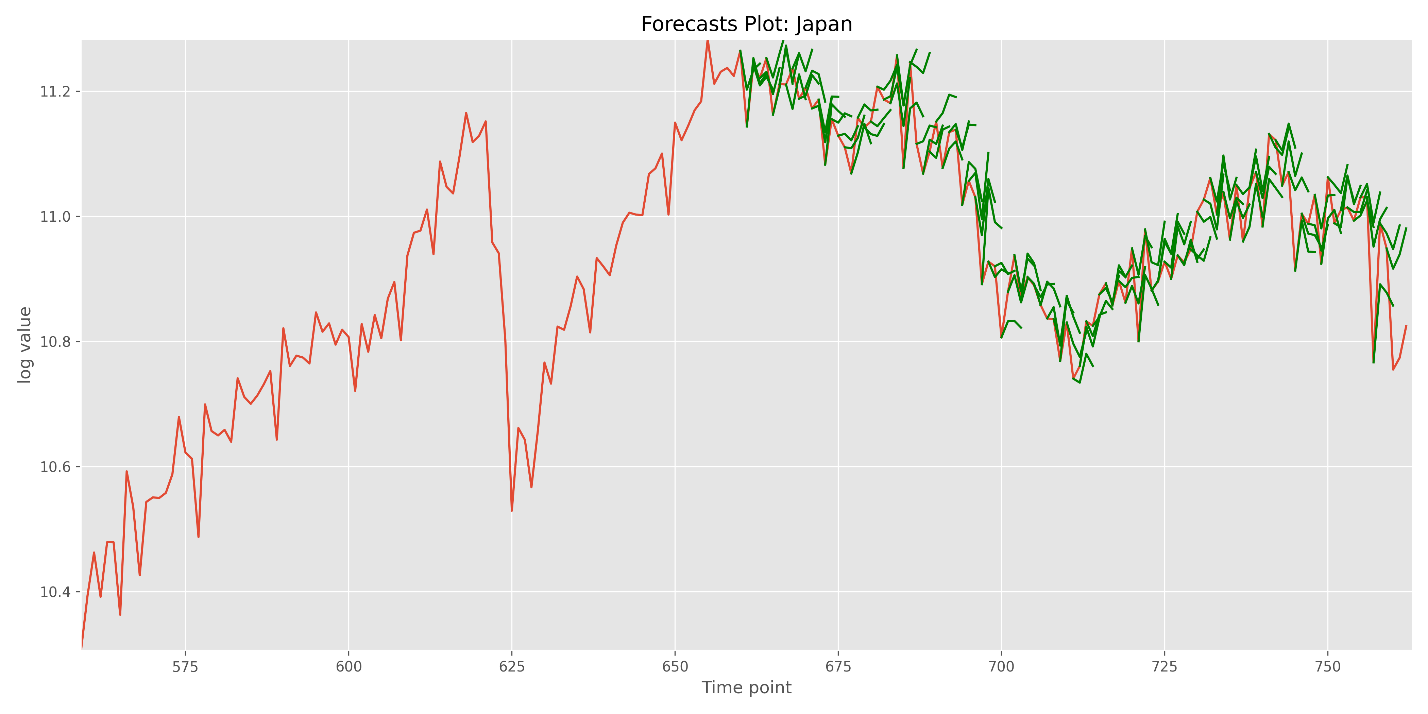


Рисунок 38 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Японии.

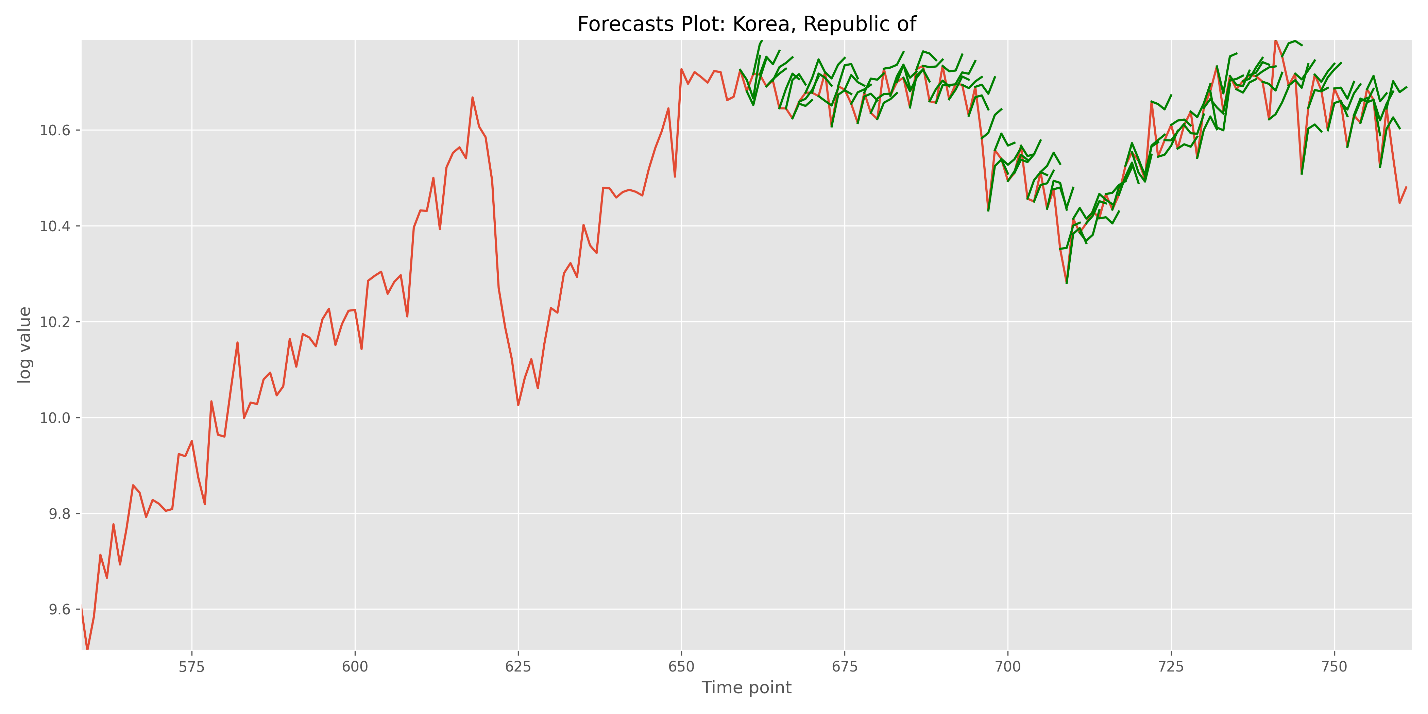
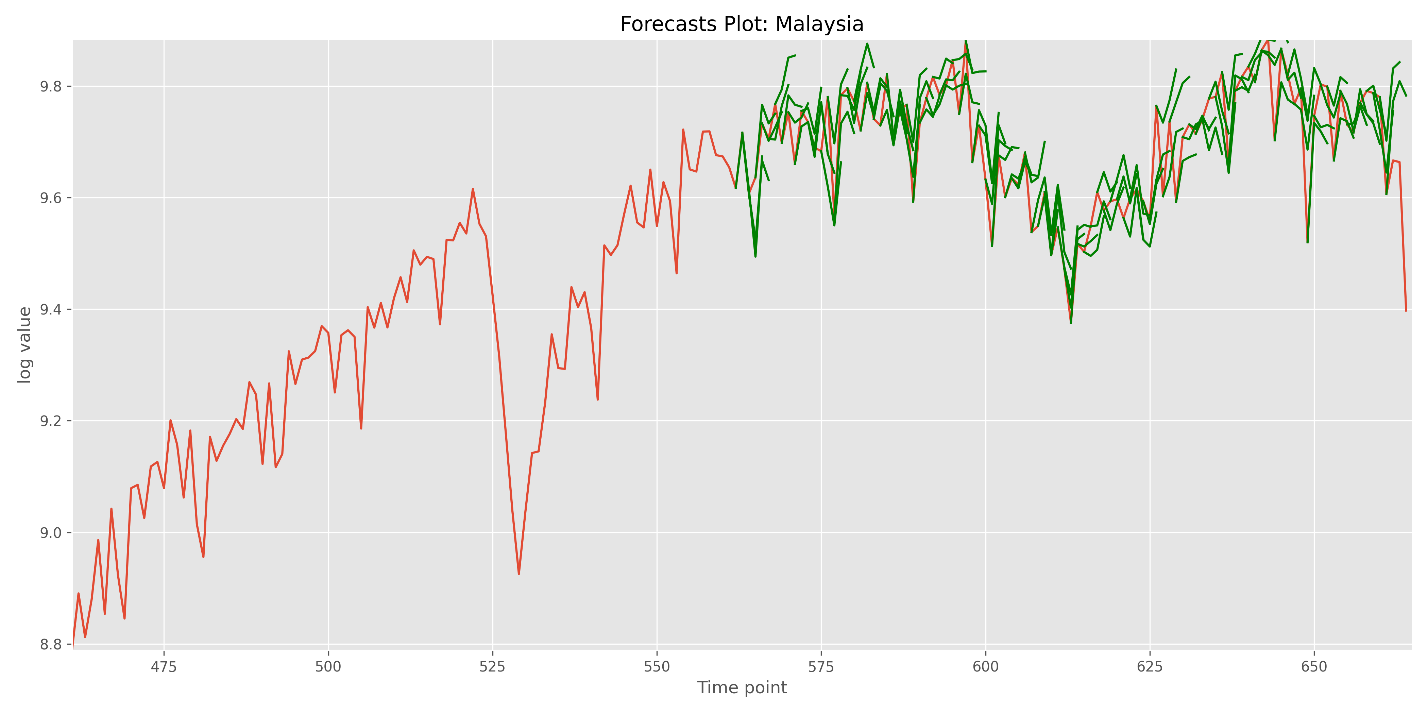


Рисунок 39 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Кореи.

 Рисунок 40 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Малайзии.

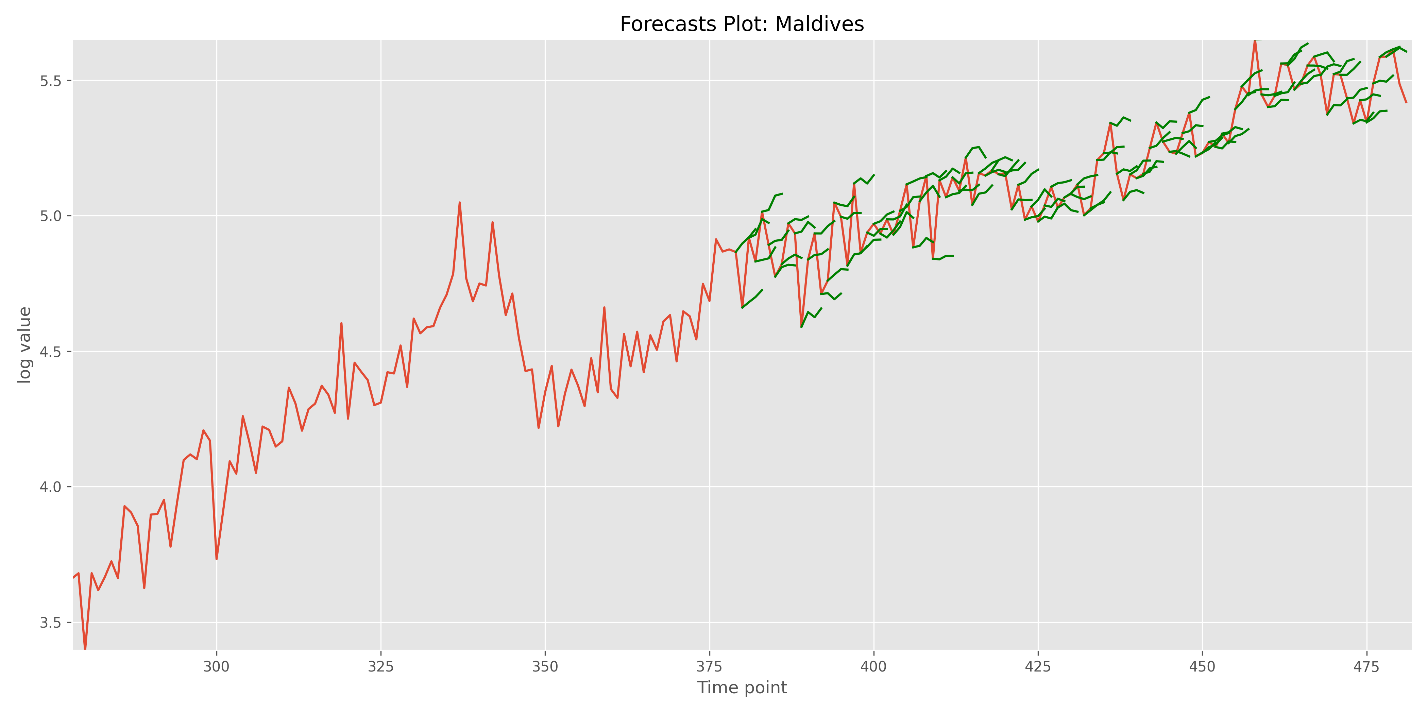


Рисунок 41 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Мальдив.

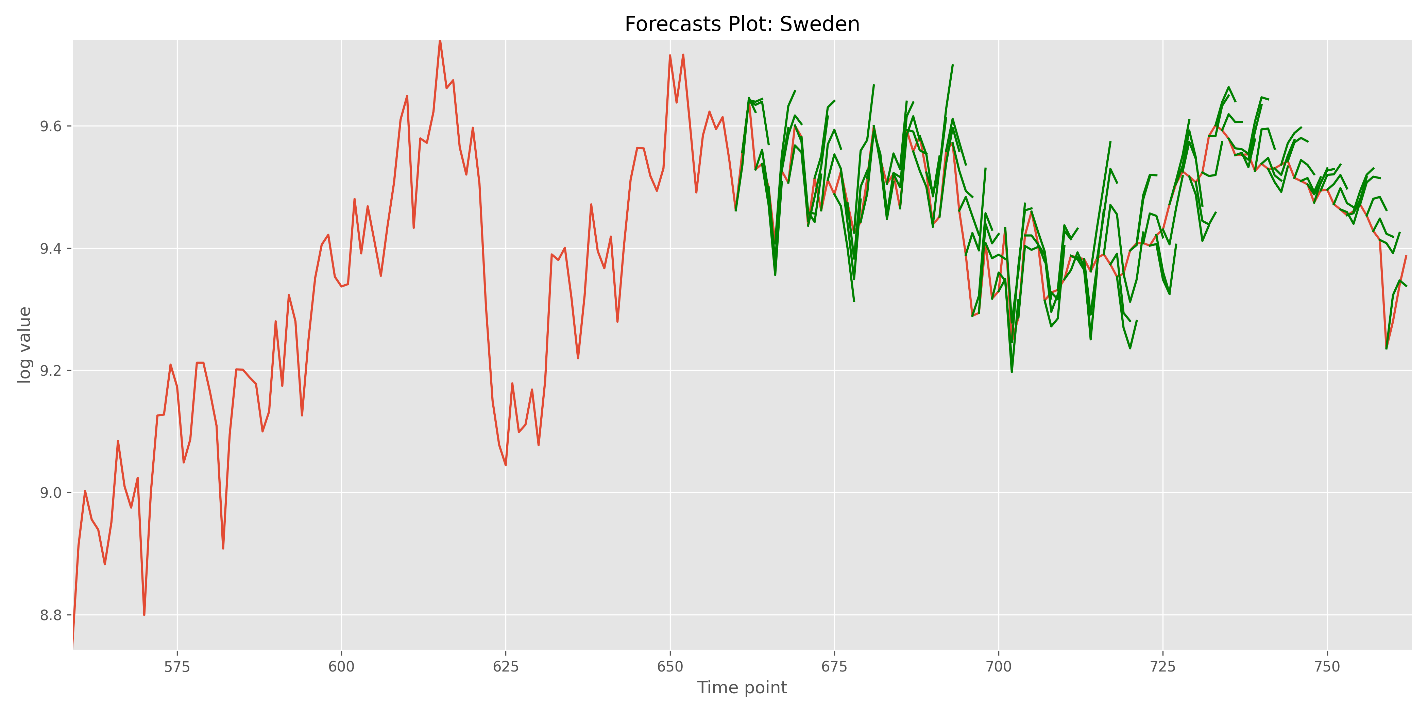


Рисунок 42 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Швеции.

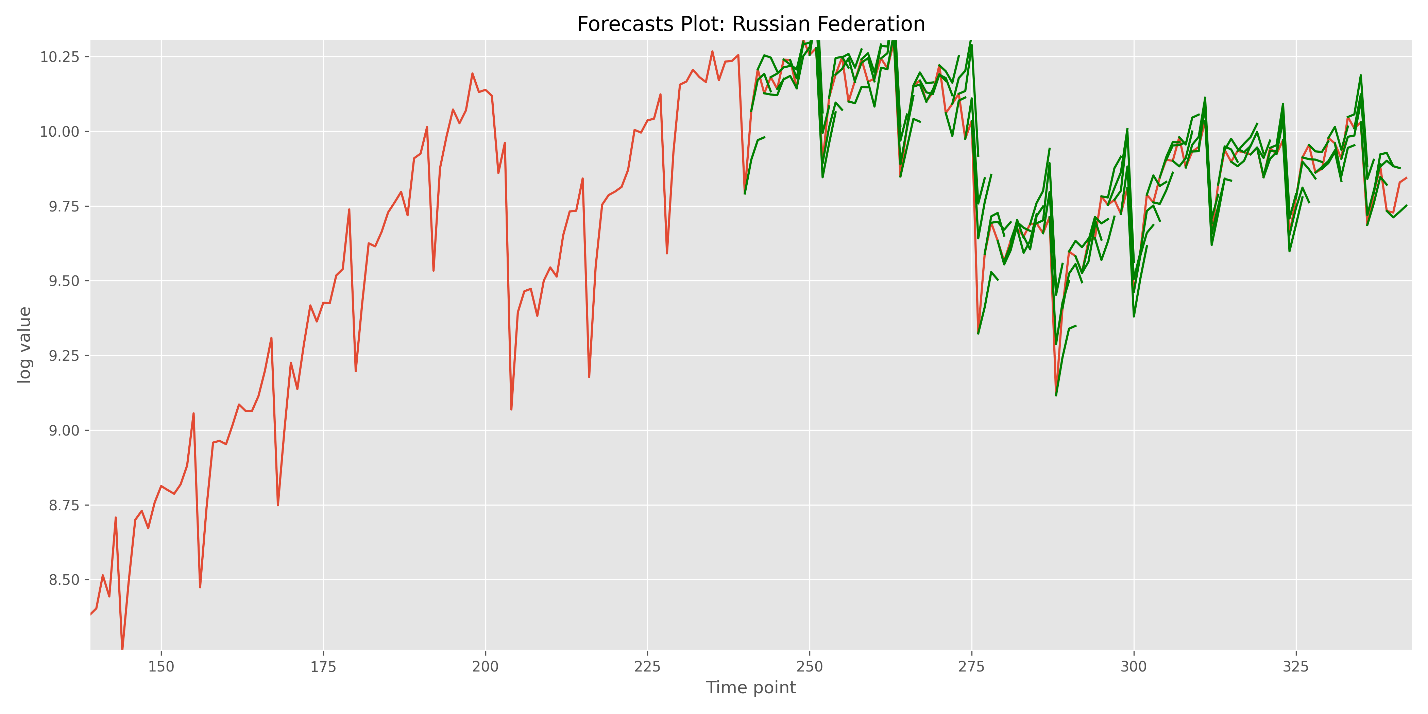


Рисунок 43 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Российской Федерации.

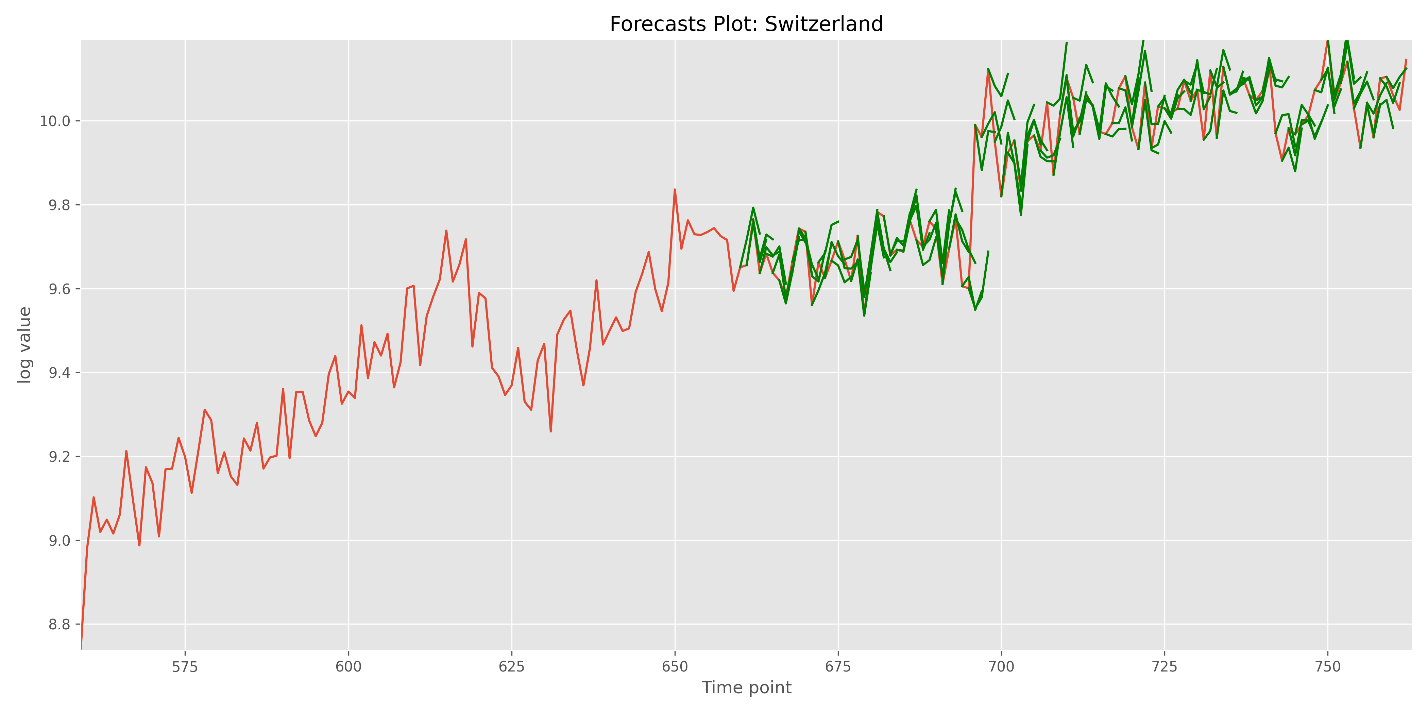


Рисунок 44 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Швейцарии.

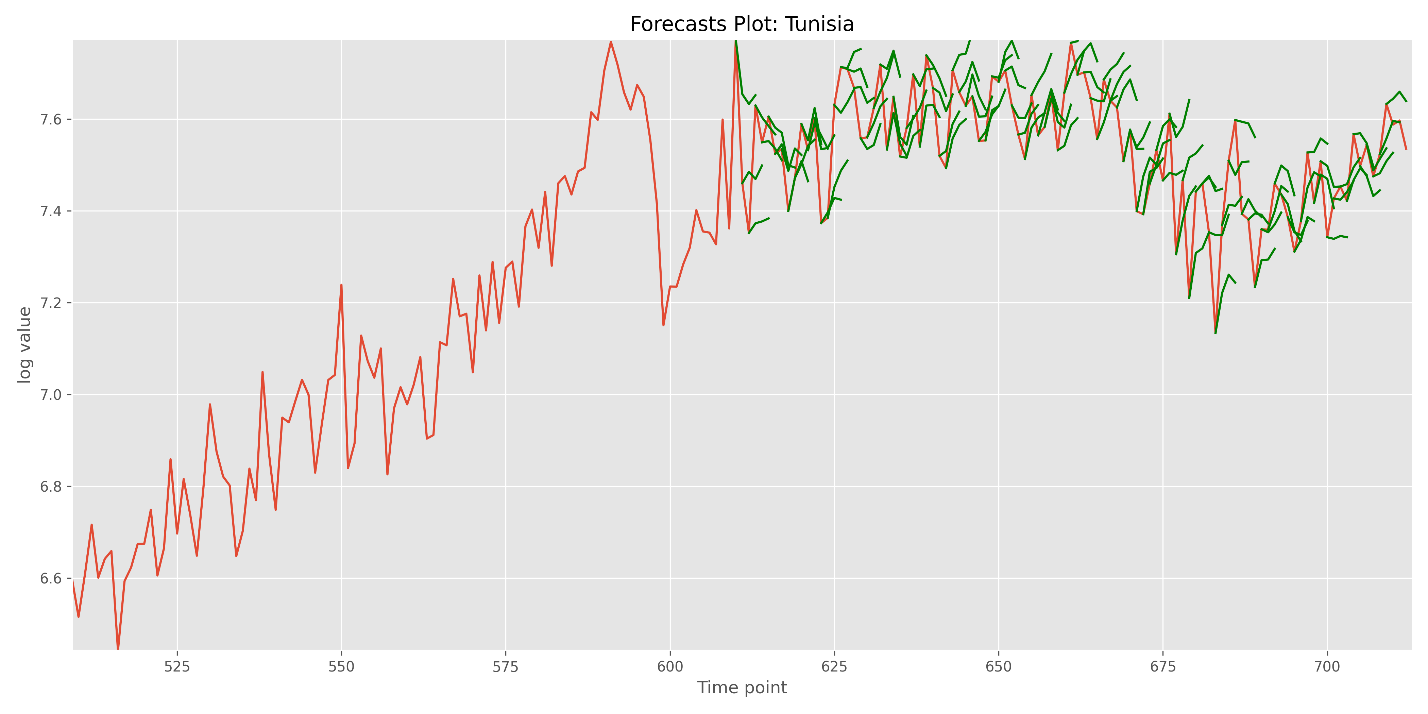


Рисунок 45 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Туниса.

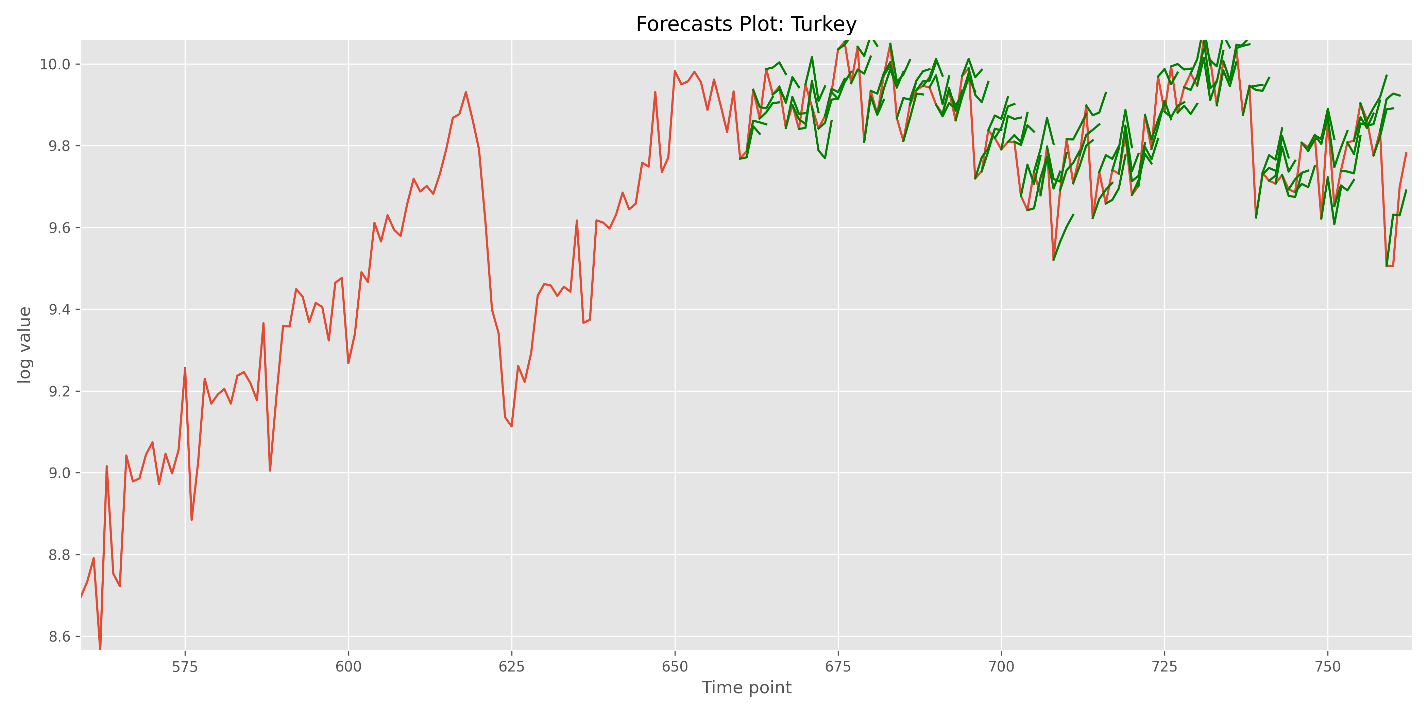


Рисунок 46 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Турции.

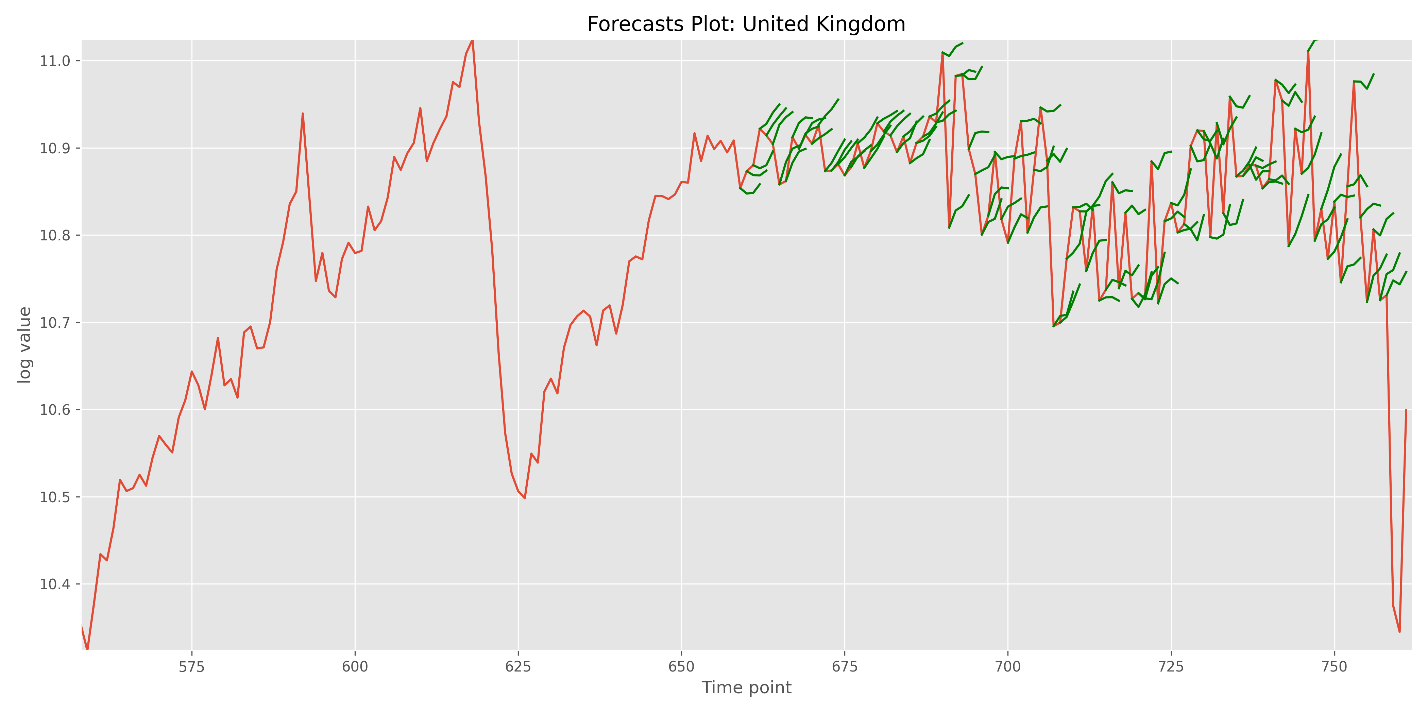


Рисунок 47 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для Великобритании.

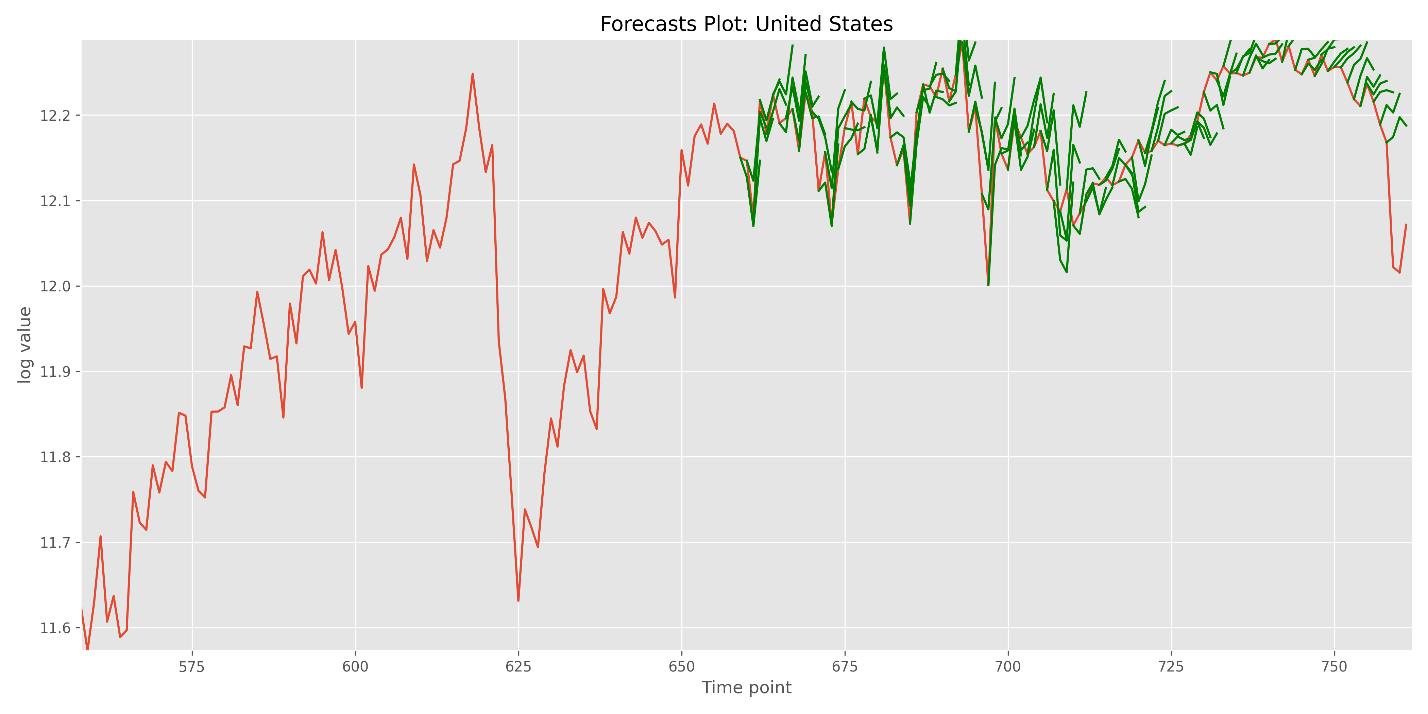


Рисунок 48 – Построенные на 3 месяца вперед прогнозы из каждой временной точки тренировочной выборки для США.

# Приложение Г. Таблица оценки качества модели на данных объёма экспорта

Оценка качества модели на данных объема экспорта разных стран представлена в таблице 2.

Таблица 2 – Оценка качества модели на данных объёма экспорта разных стран

|  |  |  |  |
| --- | --- | --- | --- |
| Country | val\_loss | val\_accuracy | val\_mae |
| Российская Федерация | 0.4719 | 0.2857 | 0.4975 |
| Binary accuracy: 0.5555555555555556 |  |  |  |
| США | 0.2308 | 0.6984 | 0.371 |
| Binary accuracy: 0.5959595959595959 |  |  |  |
| Великобритания | 0.3426 | 0.3016 | 0.4241 |
| Binary accuracy: 0.468013468013468 |  |  |  |
| Германия | 0.5252 | 0.7111 | 0.5816 |
| Binary accuracy: 0.7373737373737373 |  |  |  |
| КитайЮ Гонконг | 0.3401 | 0.6032 | 0.4442 |
| Binary accuracy: 0.6127946127946128 |  |  |  |
| Канада | 0.5554 | 0.5714 | 0.5467 |
| Binary accuracy: 0.5521885521885522 |  |  |  |
| Япония | 0.1774 | 0.7778 | 0.3233 |
| Binary accuracy: 0.8585858585858586 |  |  |  |
| Турция | 0.2865 | 0.3333 | 0.4425 |
| Binary accuracy: 0.5892255892255892 |  |  |  |
| Дания | 0.7538 | 0.5397 | 0.7162 |
| Binary accuracy: 0.6195286195286195 |  |  |  |
| Мальдивы | 0.5778 | 0.3333 | 0.617 |
| Binary accuracy: 0.46464646464646464 |  |  |  |
| Швеция | 0.4067 | 0.6667 | 0.51 |
| Binary accuracy: 0.6127946127946128 |  |  |  |
| Южная Корея | 0.178 | 0.4762 | 0.3147 |
| Binary accuracy: 0.5420875420875421 |  |  |  |
| Швейцария | 0.2523 | 0.6825 | 0.38 |
| Binary accuracy: 0.7138047138047138 |  |  |  |
| Тунис | 0.1787 | 0.2586 | 0.3472 |
| Binary accuracy: 0.5555555555555556 |  |  |  |
| Малайзия | 0.7266 | 0.6226 | 0.6542 |
| Binary accuracy: 0.5892255892255892 |  |  |  |
| Аргентина | 0.4208 | 0.5238 | 0.5184 |
| Binary accuracy: 0.632996632996633 |  |  |  |