

Министерство цифрового развития, связи и
массовых коммуникаций Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего
образования «Сибирский государственный университет телекоммуникаций и
информатики» (СибГУТИ)

Отчёт
по лабораторной работе №5
по дисциплине «**Прикладные задачи теории вероятностей**»

Выполнил:

студент гр. ИС-142

«__» декабря 2023 г.

/Наумов А.А./

Проверил:

профессор кафедры В.С.,

«__» декабря 2023 г.

/Родионов А.С./

Оценка « _____ »

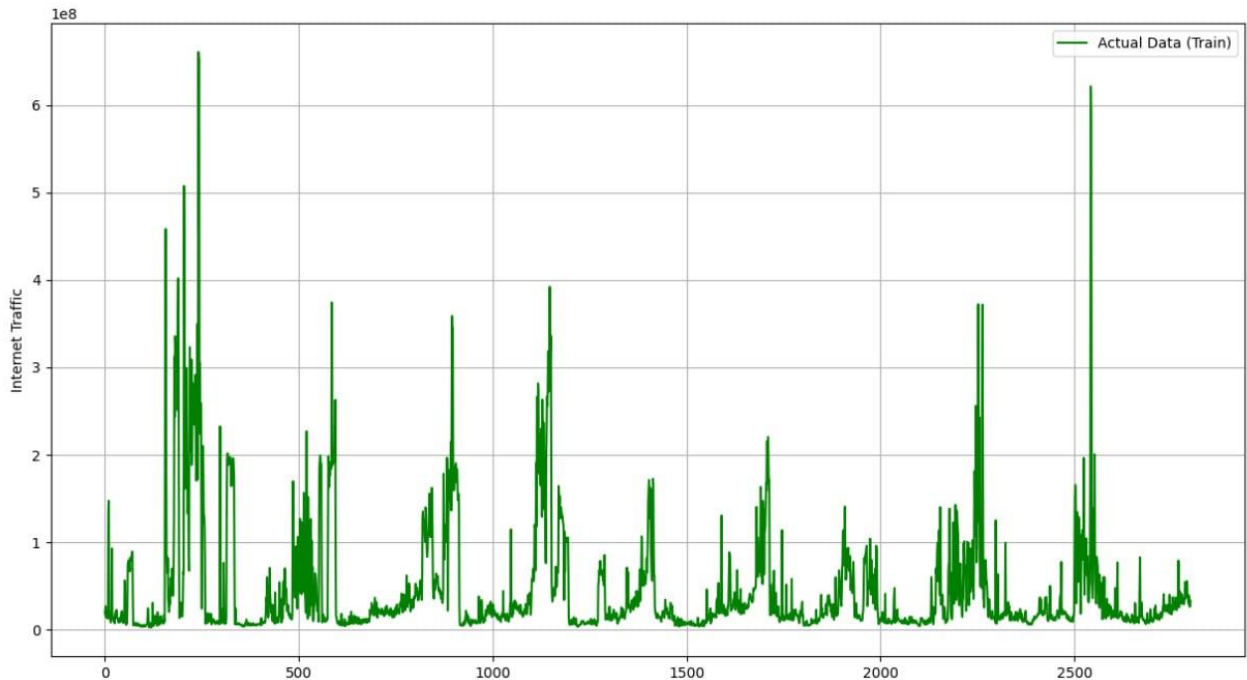
Новосибирск 2023

ВЫПОЛНЕНИЕ РАБОТЫ

Цель работы

Провести анализ временного ряда для входного интернет-трафика и спрогнозировать различными методами следующие 5 точек в этом ряду.

Визуализация исходного трафика



Для построения прогнозов воспользуемся следующими методами прогноза:

Вот краткое описание методов прогнозирования:

1. Простое скользящее среднее (Simple Moving Average, SMA):

Простое скользящее среднее - это один из самых простых методов прогнозирования временных рядов. Он использует среднее значение наблюдений за определенное число периодов, чтобы создать прогноз. Например, если вы хотите прогнозировать продажи товара на следующий месяц, вы можете взять среднее значение продаж за последние несколько месяцев. SMA позволяет сгладить шумы в данных и выделить общий тренд, но он может быть неэффективным при прогнозировании временных рядов с сезонными или циклическими компонентами.

2. Экспоненциальное сглаживание:

Экспоненциальное сглаживание - это метод, который уделяет большее внимание последним наблюдениям и уменьшает влияние более старых наблюдений по экспоненциальной функции. Этот метод подходит для временных рядов с изменяющимися трендами и сезонными компонентами. Он включает в себя параметр сглаживания (обычно обозначаемый как α), который определяет вес, присваиваемый последним наблюдениям. Чем ближе значение α к 1, тем

большой вес уделяется последним наблюдениям.

3. ARIMA (Авторегрессионная интегрированная скользящая средняя):

ARIMA - это более сложный метод прогнозирования, который учитывает авторегрессию (AR), интегрирование (I) и скользящую среднюю (MA). ARIMA-модель предназначена для анализа и прогнозирования временных рядов, которые могут иметь тренды и сезонные компоненты. Она включает в себя параметры p , d и q , где:

- p - порядок авторегрессии, определяющий, сколько предыдущих значений временного ряда учитываются для прогноза.
- d - порядок интегрирования, обозначающий, сколько раз нужно разностно суммировать ряд, чтобы сделать его стационарным.
- q - порядок скользящей средней, определяющий, сколько предыдущих ошибок модели учитываются для прогноза.

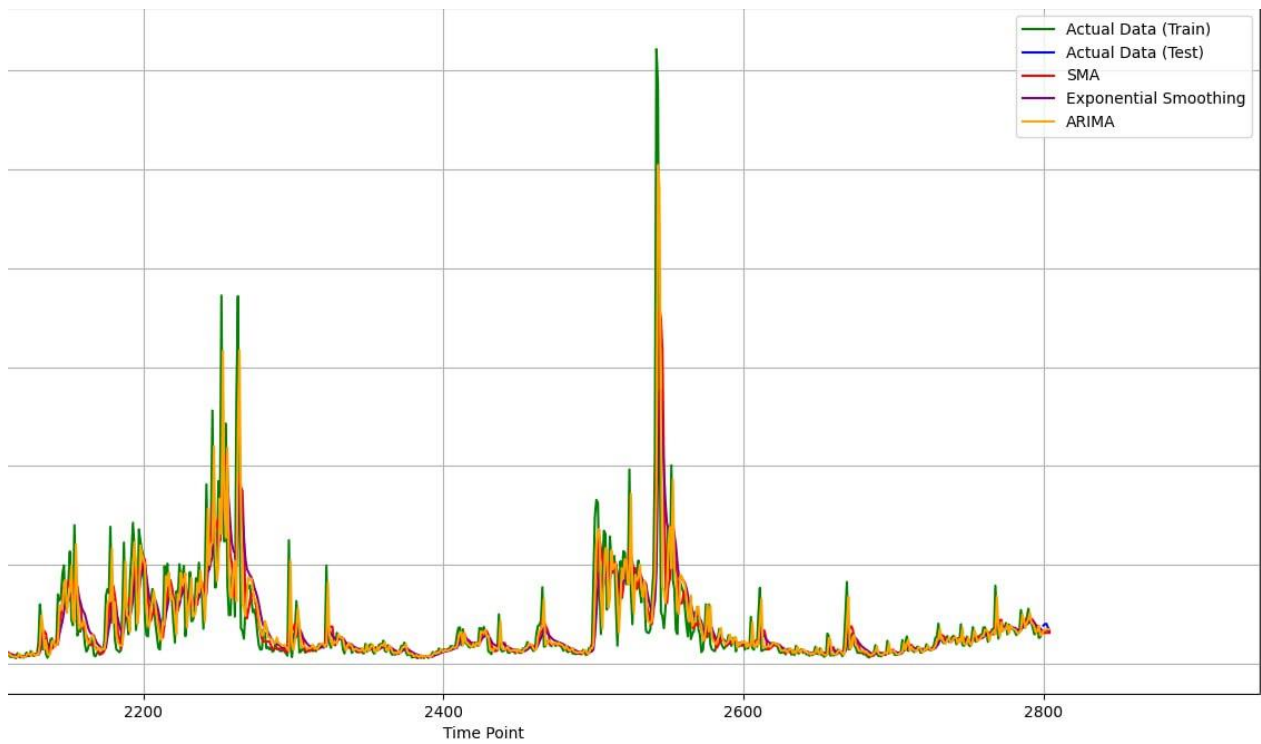
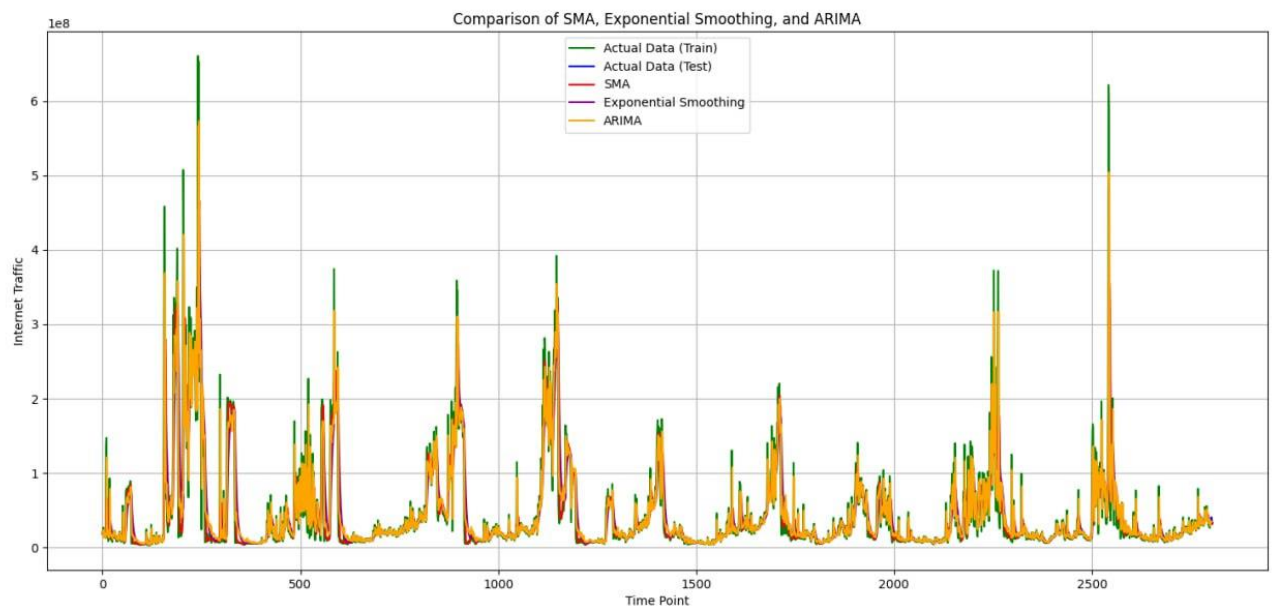
ARIMA-модель может быть эффективной при анализе и прогнозировании временных рядов, но ее настройка и интерпретация могут быть сложными задачами.

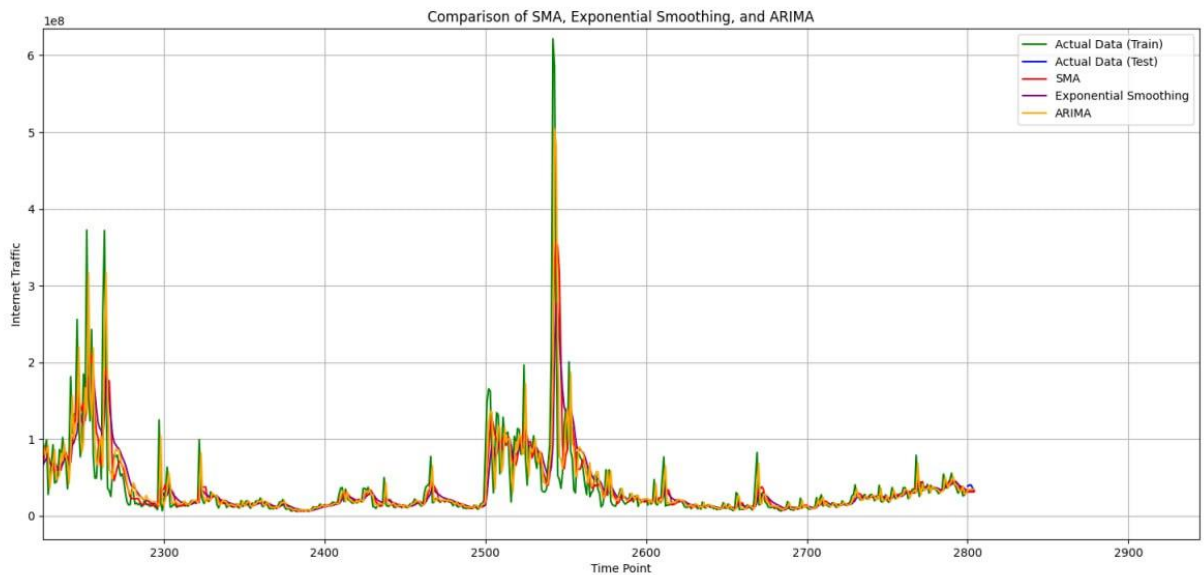
Посмотрим, что будет происходить с прогнозами при разных значениях параметров в предложенных методах.

На представленных графиках будут отображены результаты прогнозирования для первых 2800 точек входящего интернет-трафика, используя методы Простого скользящего среднего (SMA), Экспоненциального сглаживания и ARIMA, а также прогнозы для следующих 5 точек.

- Зеленая линия показывает реальные данные обучающей выборки (2800 точек).
- Синяя линия - фактические значения тестовой выборки (5 точек).
- Красная линия - прогнозы SMA, которые являются сглаженными значениями для первых 2800 точек, дополненными одинаковыми прогнозными значениями.
- Фиолетовая линия - прогнозы, сделанные с помощью метода экспоненциального сглаживания, как для обучающих, так и для прогнозируемых точек.
- Оранжевая линия - прогнозы, полученные с помощью ARIMA, также для обучающих и прогнозируемых точек.

Вариант 1 с параметрами (SMA_window = 5, EXP_smooth(alpha) = 0.2, ARIMA = 2(p),1(d),2(q))

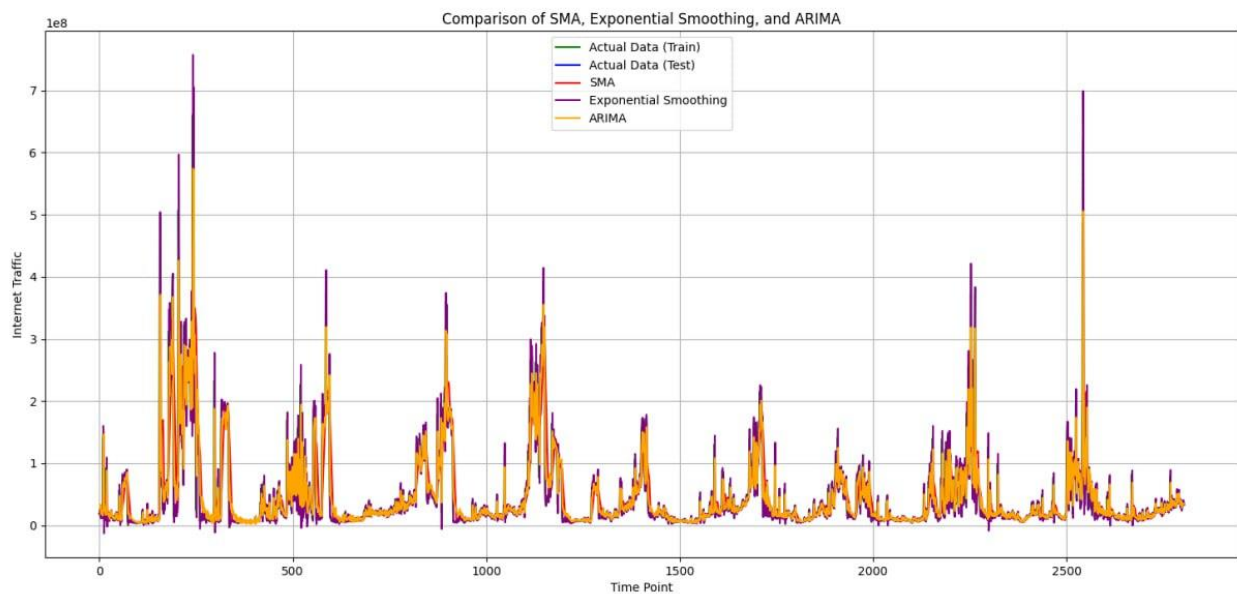


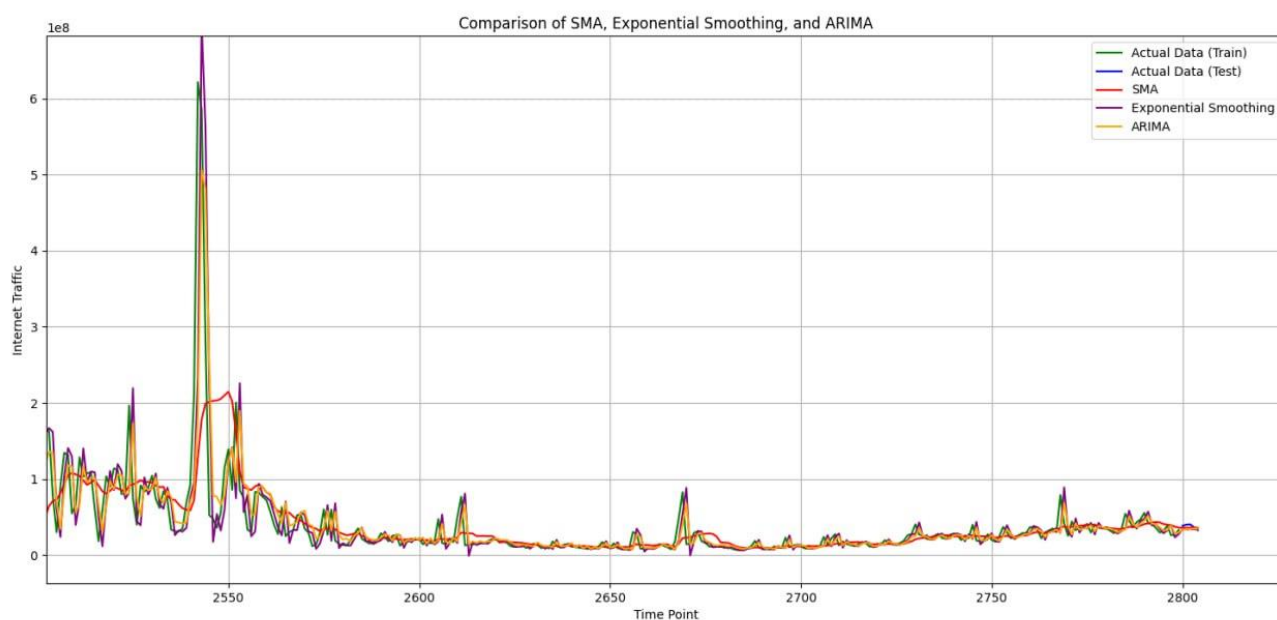
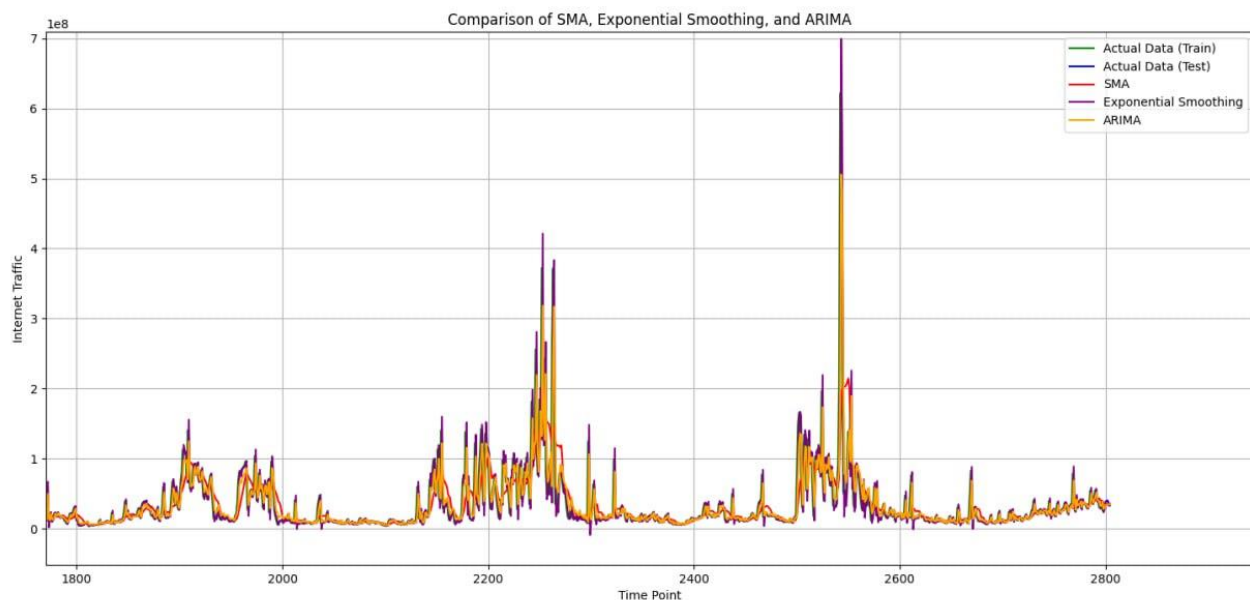


В данный момент графики ведут себя достаточно предсказуемо в области реальных данных, но на этапе прогноза ведут себя как константные функции. ARIMA показывает легкий рост. При обычных значениях особо сильных изменений на графике не наблюдается. Экспоненциальное сглаживание не выделяется, для наглядности нужно увеличить переменную `exp_smooth`.

Вариант 2 (SMA_window = 10, EXP_smooth = 1.2, ARIMA=4,2,4)

Увеличим SMA вдвое, `exp_smooth` на 1 единицу и параметры ARIMA вдвое. Посмотрим что произойдет с прогнозом.





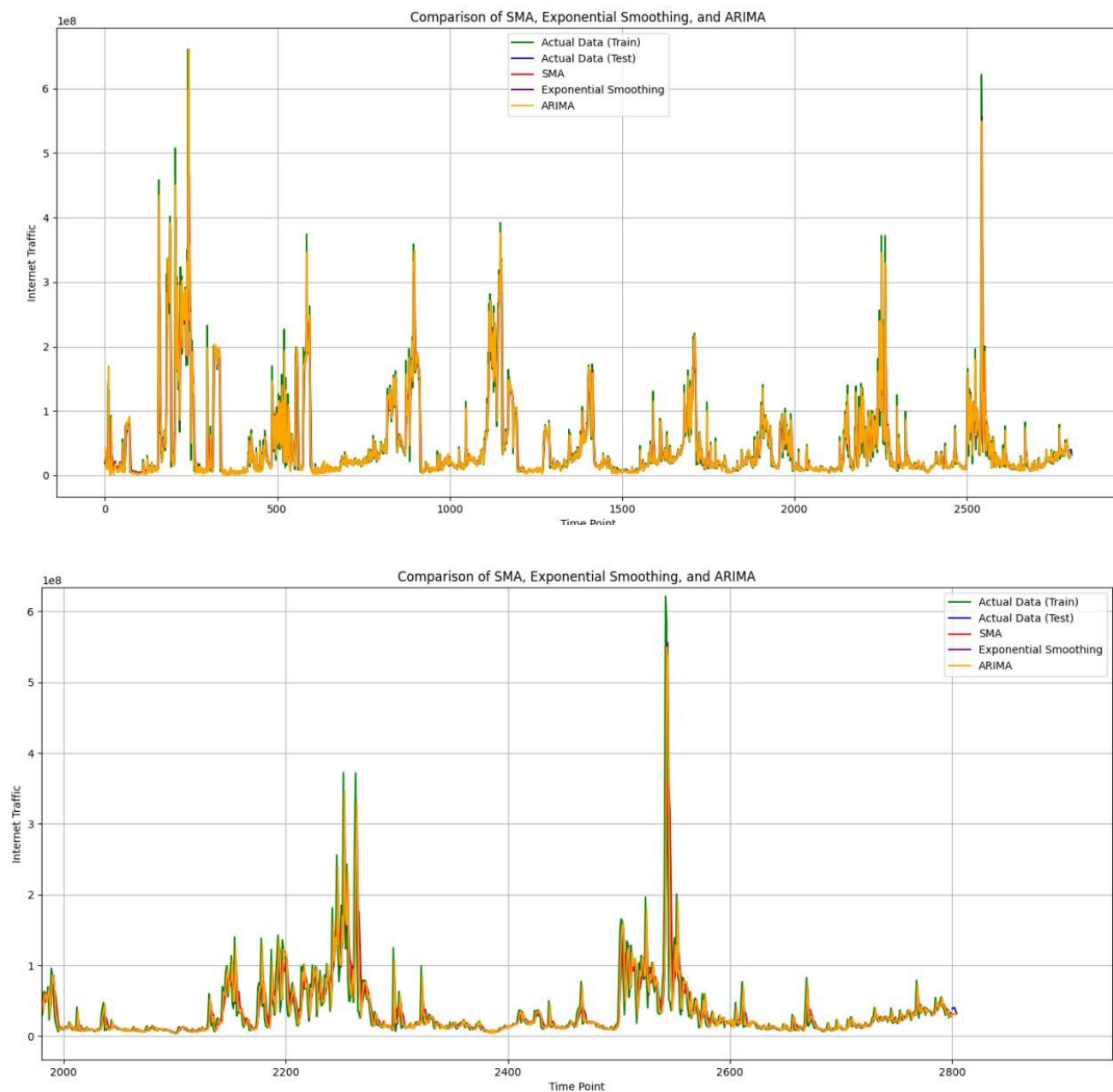
Экспоненциальное сглаживание на пиках выходит выше основного трафика, что говорит о хаотичности, но дает похожий прогноз, но не точный, ведь трафик особо не изменяется на данных значениях.

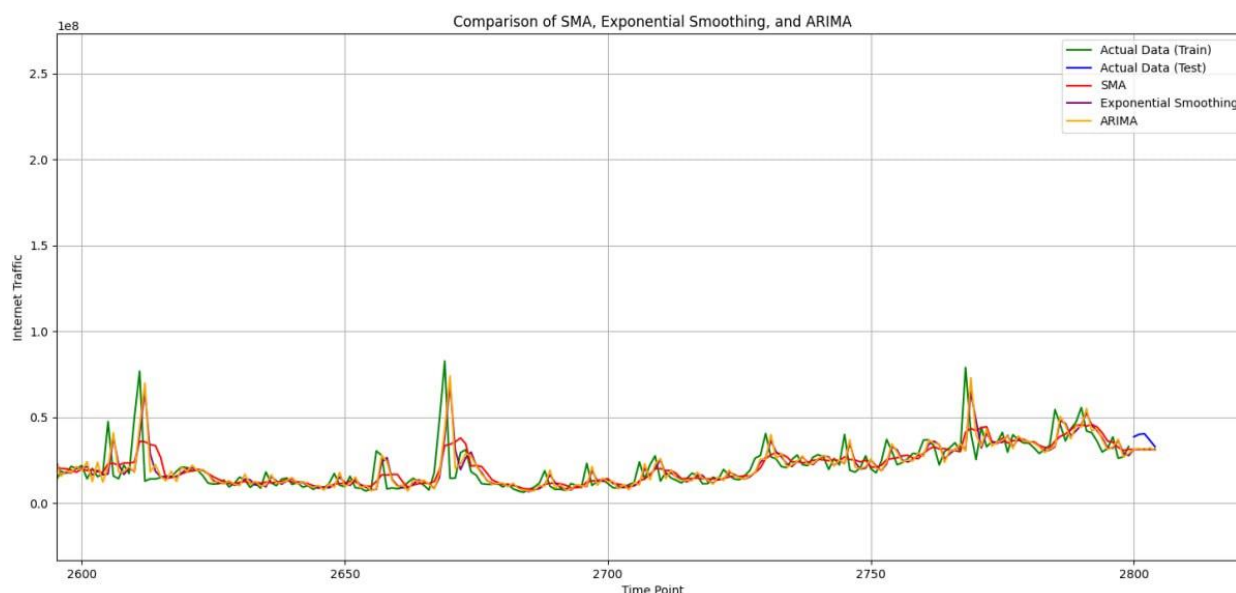
ARIMA достаточно хорошо ведет себя при данных значения, но видно что график начинает запаздывать.

Простое скользящее среднее из-за лага начало гораздо больше запаздывать и показывать похожий результат при прогнозе, но пики графика полностью сглаживаются при этих значениях.

Параметры у ARIMA (4,2,4) достаточно стабильны и показывают неплохой прогноз, но можем еще увеличить параметр на (5,3,5), чтобы посмотреть, станет ли лучше или хуже. У SMA слишком большой параметр, нужно обратно урезать до 5, а экспоненциальное сглаживание, к 1.0, но не обязательно точно единицу.

Вариант 3 с параметрами (SMA_window = 5, EXP_smooth(alpha) = 0.7, ARIMA = 5(p),3(d),5(q))





в данном варианте экспоненциальное сглаживание почти полностью повторяет ARIMA.

Выводы о прогнозах:

Простое скользящее среднее (SMA) представляет собой метод прогнозирования, который усредняет данные на определенном временном интервале, и все прогнозы, сделанные на этом интервале, будут иметь одно и то же значение. Это может быть недостаточно гибким и точным методом, так как он не адаптируется к изменениям в данных. Если увеличить размер окна (интервала), на котором выполняется усреднение, то это может улучшить предсказания для сглаживания долгосрочных трендов, но это также приводит к задержке в реакции на изменения в данных. Напротив, при малом размере окна, SMA может лучше отражать краткосрочные пики данных, но это часто приводит к плохим прогнозам из-за чувствительности к шумам. Таким образом, SMA может давать недостаточно точные прогнозы, особенно при нестабильных данных.

Экспоненциальное сглаживание, в отличие от простого скользящего среднего (SMA), является более гибким методом прогнозирования. Этот метод более чувствителен к резким изменениям в данных, что позволяет ему лучше адаптироваться к краткосрочным изменениям. Он также способен поддерживать нисходящий тренд в данных после определенного отрезка.

Однако экспоненциальное сглаживание имеет свойство быть "немасштабируемым", что означает, что его точность может сильно зависеть от настройки параметра сглаживания (обычно обозначаемого как α). Если параметр сглаживания выбран неправильно, это может привести к недооценке или переоценке будущих значений временного ряда. Таким образом, хотя экспоненциальное сглаживание может быть очень точным, его эффективность

требует тщательной настройки и мониторинга.

ARIMA (Авторегрессионная интегрированная скользящая средняя) - это метод прогнозирования временных рядов, который обладает высокой адаптивностью к последним изменениям в данных и учитывает потенциальные тренды и сезонность. ARIMA может быть особенно эффективным в анализе и прогнозировании временных рядов, которые имеют сложную структуру.

ARIMA включает в себя параметры, такие как порядок авторегрессии (AR), порядок интегрирования (I) и порядок скользящей средней (MA). При настройке этих параметров ARIMA может уловить и автоматически адаптироваться к различным паттернам в данных, включая авторегрессию, тренды и сезонность. Это позволяет ему обнаруживать и предсказывать различные пики и особенности в данных.

Однако ARIMA не всегда будет лучшим методом в сравнении с экспоненциальным сглаживанием или другими методами в зависимости от конкретной структуры данных и требований прогнозирования. Он может обеспечивать более масштабируемые и точные прогнозы, но также может потребовать более сложной настройки и анализа данных для достижения оптимальных результатов.

С учетом проведенного анализа, можно сделать вывод о том, что для прогнозирования временного ряда входящего интернет-трафика наиболее точным методом оказалась модель экспоненциального сглаживания. Тем не менее, следует подчеркнуть, что выбор метода прогнозирования зависит от специфических характеристик временного ряда и целей анализа. В долгосрочной перспективе метод ARIMA предоставляет лучшие результаты благодаря способности учитывать периодичность данных и сложные временные структуры.