

Анализ изменения стоимости биржевых индексов

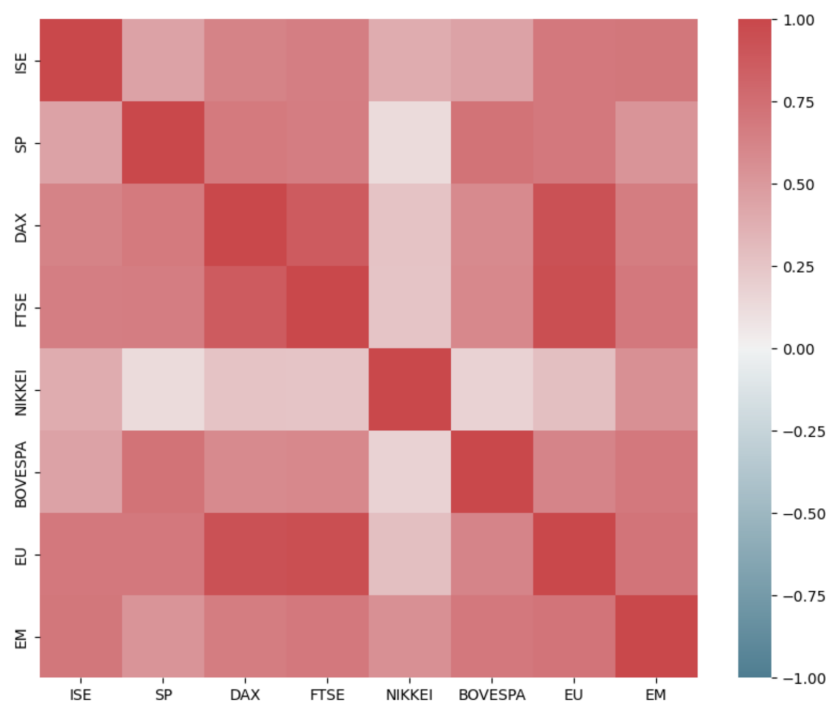
Дудиев Хетаг, Степанов Никита

1 Вступление

В данной работе мы занимались исследованием датасета, содержащего в себе информацию об изменении биржевых индексов. Мы попытались предсказать значения индексов на основе предыдущих данных, используя модели для анализа временных рядов. Было сделано два предсказания с помощью разных моделей: AM в Python библиотеке Prophet и ARIMA в Sktime.

2 EDA

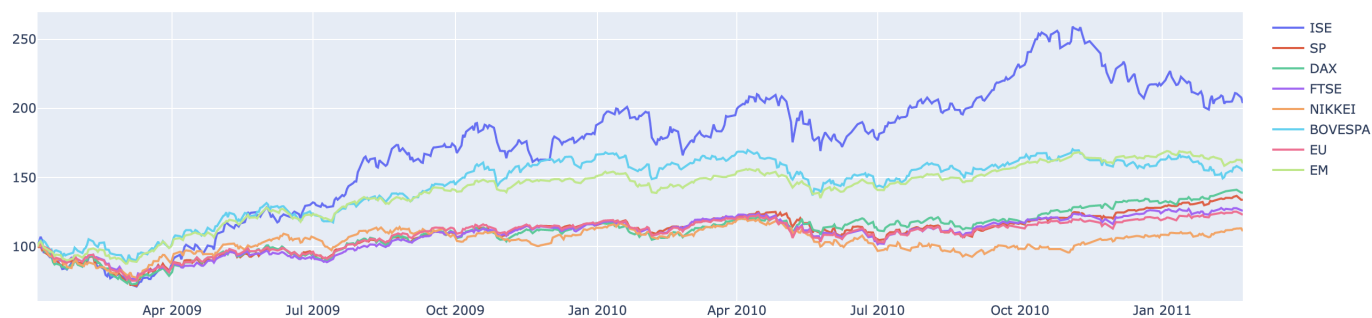
В этой части мы занимались неким предварительным анализом. В первую очередь мы занялись преобразованием исходного датасета, чтобы с ним было удобнее работать. Например, переименовали столбцы, а также оставили только индексы, выраженные в долларах. Далее мы посчитали корреляцию между индексами, чтобы посмотреть, как они соотносятся друг с другом.



Тепловая карта корреляции биржевых индексов

Видно, что сильнее всего коррелируют индексы EU и DAX, EU и FTSE, DAX и FTSE. Это логично и ожидаемо, так как все эти индексы — европейские.

Также мы обратили внимание на суммарное процентное изменение индексов за период в два года, представленный в датасете.



Изменение индексов

3 Prophet

Мы использовали Additive model из библиотеки Prophet. Предсказание переменной y от времени t вычисляется по формуле:

$y = g(t) + s(t)$, где $s(t)$ - сезонная компонента (зависимость y от дня недели, месяца, или от какого-то другого цикла), в нашей модели есть только сезонная компонента, которая вычисляется линейной регрессией зависимой переменной от дней недели (1 или 0, в зависимости от того день недели или нет), $g(t)$ - тренд, возрастающая с насыщением функция (как правило логистическая функция). Мы обучали модель на всех значениях кроме последних 120, которые пытались предсказать. Точность оценивали с помощью Mean Absolute Percentage Error, процентом дней когда мы угадали вырастет акция или нет, а также конечным значением акции. Лучше всего, с точки зрения MAPE, получилось предсказание для DAX, EM и ISE: 7.97, 9.87, 10.10. В целом у нас неплохо получилось предугадать конечное значение ISE (наиболее выросшая акция):

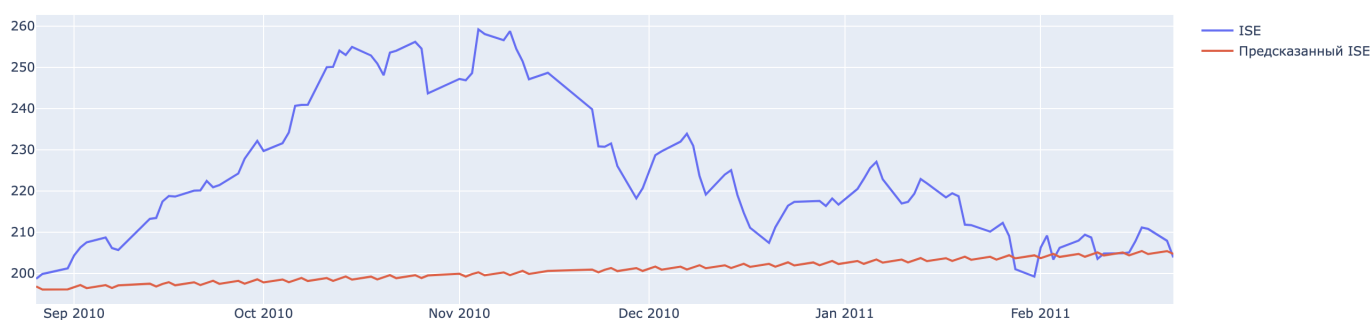


График для ISE и его предсказанных значений

4 Sktime

В данной части мы использовали модель AutoARIMA из библиотеки Sktime для предсказания значения индексов. Модель $ARIMA(p, d, q)$ для нестационарного временного ряда имеет вид:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t, \text{ где } \varepsilon_t \text{ — стационарный временной ряд, } c, a_i, b_j \text{ — параметры}$$

модели, Δ^d — оператор разности временного ряда порядка d (последовательное взятие d раз разностей первого порядка — сначала от временного ряда, затем от полученных разностей первого порядка, затем от второго порядка и т. д.)

В качестве обучающего датасета мы взяли все значения, кроме последних 120, а оставшиеся, как раз, попытались предсказать. Для каждого индекса мы посчитали Mean Absolute Percentage Error. Наименьшая ошибка получилась у индексов ISE, BOVESPA и EM — 8.63, 5.57 и 5.46 соответственно.



5 Заключение

Предсказания получились не такими точными, потому что индексы состоят из множества различных акций, поэтому у индексов нет каких-то определенных циклов изменений, они в большей степени зависят от событий в определенном регионе или в мире, которые модель предсказать не может, лучше отработала ARIMA модель. Prophet отработала хуже, потому что не смогла найти какую-то определенную сезонную зависимость.