Краткая формулировка задачи:
даны были данные о продаже конкретного товара с полями:
- Num – номер элемента выборки
- y – результат продаж товара
- year – год продаж
- week – неделя продаж в году
- shift – сдвиг от настощей недели (без задержки по времени)
-f1 … f60 – срезы по результатам продаж товара в разное время

Итоговое решение, которое набрало больше всего быллов:

1. Подготовка данных
   Все остальные поля оставила.
   Сортировка для кросс валидации по времени.
2. В качестве таргета использовались результаты продаж – y
3. model = XGBRegressor(n_estimators=100, max_depth=14,
   random_state=95)

По сути: просто GradientBoosting с подобранными путем множественной проверки в циклах критериев. (при чем он давал лучше результаты, чем GradientBoostingRegressor из sklearn)

Пробовались следующие подходы:

1. Линейная регрессия, проведенная с квадратами признаков, попарными произведениями, а также кубами в разных вариациях.
2. Градиентный бустинг с разными количеством деревьев, глубинами и лернинг рейтами. Искались лусшие путем переборов в циклах.
3. Рандом Форест с разными количеством деревьев, глубинами. Искались лусшие путем переборов в циклах. Также с различнами критериями в листьях. Например, с Mean Absolute Error вместо MSE (для улучшения SMAPE)
4. Так же была проделана попытка найти закономерность в данных. Для этого было сделано предположение, что f0 … f60 – срезы отвечающие за 1 недели продаж. Для этого для каждого item_id было (по возможности) составлено соответствие между колличеством продаж y и неделей подаж week и каким-то срезом fi. Что выводилось как ответ в итоге это известный ответ, смещенный вдоль срезов на разницу между известной неделей и обрабатываемой неделей. В итоге это решение, совмещенное с РандомФорест почему-то оО не набрало много баллов.

   PS: draft*.ipynb

Код итогово решения: final_version.ipynb

Оценка качества:

Кроссвалидация на отсортированных по времени данных, делалась путем:
Первая четверть – Вторая четверть
Первые две четверти – Третья четверть
Первые три четвери – Четвертая четверть
Очевидно со скорингом SMAPE.

Кросс валидация примерно совпадала с той, что на leaderboard. Но была немного меньше у меня все-таки.