**Описание данных:**

* 'status' - статус продажи
* 'private pool’ и 'PrivatePool' - наличие частного бассейна,
* 'propertyType' - тип недвижимости, поместья, свойства,
* 'street' - адрес дома
* 'baths' - количество ванных комнат
* 'homeFacts' - сведения о строении дома, содержит несколько отдельных данных, имеющих влияние на оценку недвижимости, дома, квартиры
* 'fireplace' - сведения о наличии камина (ов) в доме
* 'city' - город
* 'schools' - сведения о школах в районе
* 'sqft' - площадь дома в футах
* 'zipcode' - почтовый индекс
* 'beds' - количество спален в доме
* 'state' - государство/штаты
* 'stories' - этажность недвижимости
* 'mls-id' и 'MlsId' - идентификатор MLS (Multiple Listing Service - Служба множественного листинга)

*Хотя веб-сайт MLS не является общедоступным, вы можете найти ограниченную информацию о листинге на веб-сайте Realtor.com. Национальная ассоциация риэлторов, которая управляет всеми ассоциациями MLS, владеет и поддерживает этот веб-сайт.*

*После выбора Поиск по идентификатору MLS, Опция поиска и введя уникальный номер MLS для свойства, вы увидите следующую информацию:*

* *Информация об адресе, включая город, штат и почтовый индекс*
* *Название района, включая ссылку на другие дома для продажи в том же районе*
* *Описание недвижимости, в том числе тип дома и его площадь в квадратных футах, количество спален и ванных комнат, а также размер участка в квадратных футах*
* *Листинг риэлторской компании*
* 'target' - целевой признак, который необходимо спрогнозировать - цена недвижимости

[4:07](https://sfexpertsworkspace.slack.com/archives/D0262L20CN7/p1679317639078489)

**Важные примечания о данных:**

Признаки 'homeFacts' и 'schools' представляют собой словари и содержат несколько отдельных сведений. Необходимо десериализовать содержимое этих признаков. Если имеются полезные данные, то необходимо из них создать новые признаки.

[Типы домов, жилищ, и их основные характеристики](https://www.rocketmortgage.com/learn/types-of-houses) (может пригодиться при обработке признаки propertyType)

* Apartment - апартаменты (в них нельзя прописаться)
* Condo - квартира (кондоминиум)
* Co-Op - совместная (кооперативная квартира)
* Single-Family (Detached, Tiny Home) - односемейный (маленький, отдельный)
* Townhome, Townhouse - таунхауз
* Cape Cod - черепица, деревянный сайдинг, центральная дверь с окнами по бокам, 1-2 этаж.
* Colonial home - два этажа с лишним, симметричный, центральная лестница, формальный вид.
* Contemporary - основные характеристики: Чистые, простые линии; нейтральные цвета; натуральные текстуры
* Cottage - уютная, небольшая веранда; небольшие жилые помещения; индивидуальный характер.
* Craftsman - "ремесленник" элементы ручной работы из дерева, выступающие балки, большие колонны.
* Greek Revival - большие белые колонны, украшения в греческом стиле, парадный вход.
* Farmhouse - деревенское, прямоугольное, большое крыльцо, крыши в стиле сарая и особенности.
* French Country - влияние прованса, каменный внешний вид, состаренный вид.
* Mediterranean - белая штукатурка, теплое дерево и камень, черепичные крыши, есть элементы испанских и итальянских вилл.
* Midcentury Modern - сочетание органики и геометрии, гладкий, лаконичный, большие окна.
* Ranch - один этаж, низкая крыша, открытые жилые помещения, задний двор.
* Split-Level - жилые помещения разделены короткими лестничными пролетами.
* Tudor - деревянный каркас, каменная кладка, ассимметричные, крутые двускатные крыши.
* Victorian (Queen Anne Houses) - два-три этажа, богато украшенные крутые двускатные крыши, небольшие башни, яркие фасады.
* Europian Houses style - часто используют кирпич или камень и включают в себя высокие крутые крыши, высокие окна, часто со ставнями, и традиционные декоративные детали, такие как фронтоны и замковые камни.
* log home - деревянный дом

**Некоторые жаргонные сокращения, для исправления дублирования (может пригодиться при обработке признака propertyType):**

* 'manufactured home (house)' = 'mobile' = 'prefab' = 'modular'
* 'mobile' = 'mo2 le' = 'mo2le'
* 'cabin' = 'ca2 n' = 'ca2n'
* midcentury = mid century = mid-century
* single family = single-family home
* **Некоторые термины, используемые в сфере продажи-покупки недвижимости:**
* status - статус, состояние
* estate - недвижимость
* 'foreclosed' ('foreclosure') - 'исключено', 'лишен права выкупа' — это процесс, когда ваша недвижимость переходит к кредитору из-за неуплаты Вами ипотечных выплат. Продажа foreclosure недвижимости отличается от традиционных продаж. Переход недвижимости в собственность залогодержателя если, по какой-либо причине, вы не можете выплачивать ваш mortgage Mortgage - ссуда в банке на недвижимость (Ипотека).
* 'pre foreclosure' ('pre foreclosure auction') - «до выкупа» - «аукцион до выкупа»
* 'under contract showing' ('under contract show', 'under contract', 'active under contract') - 'по договору', - «действующий по контракту»,
* 'under contract backups' - резервное копирование по контракту, 'active backup' - «активное резервное копирование» 'backup contract' - «резервный контракт»
* 'contingency' - непредвиденные обстоятельства
* 'contingency contract' - «контракт на случай непредвиденных обстоятельств», 'active contingency' - 'активное непредвиденное обстоятельство', 'insp inspection contingency' - 'непредвиденные обстоятельства инспектирования'
* 'pending escape clause' - 'ожидающая оговорка об освобождении'
* 'pending backup wanted' - «требуется резервное копирование в ожидании» 'pending take backups' - «ожидание создания резервных копий»
* 'pending continue show financing' - 'в ожидании продолжения шоу финансирования' 'pending continue show' - 'ожидание продолжения шоу'
* 'pending insp finance' ('pending inspection') - "в ожидании финансовой проверки"
* 'due diligence period' - «период должной осмотрительности»
* 'activated' - активирован (active)
* 'active with contract' - 'активен с контрактом' 'active option' = = active with contract = = 'active option contract' = = contingency period = = due diligence period «активный вариант» - «активный опционный контракт», In some states, the active option contract is referred to as a contingency period or due diligence period.
* 'active with offer' - 'активен с предложением'
* 'active auction', 'auction active', 'auction'
* contingency, contingencies - случайность, непредвиденные обстоятельства contingent - условный. В сфере недвижимости, когда дом указан как условный, это означает, что предложение было сделано и принято, но до завершения сделки должны быть выполнены некоторые дополнительные критерии. В сфере недвижимости, когда дом указан как условный, это означает, что предложение было сделано и принято, но до завершения сделки должны быть выполнены некоторые дополнительные критерии.
* cooling-off period - период обдумывание (период охлаждения) conditional offers - условные предложения

****

Мария Жарова

15:23

Добрый день! Пересмотрела ещё раз Вашу работу, то, что подразумевается под реализацией в продакшене, не нашла:( Просто сохранять результаты в файл будет недостаточно, здесь подразумевается создание интерфейса, при помощи которого пользователь бы не видел код совсем, а только вводил бы данные - получал предсказания. Это можно реализовать как раз через docker и flask, или сделать обёртку в телеграм-бота:) Метрики примемлемые, но насколько их можно улучшить - сказать сложно, т.к. для каждых данных они свои, дело может быть уже не в модели, а в "потолке", который достигнут на конкретном датасете. В таком случае можно рассматривать задачу более глобально, дополнять новыми источниками и улучшать показатели. По поводу декомпозиции временных рядов можно попробовать seasonal\_decompose из библиотеки stats

Василий Бухаров

21.04.2023, 14:08

[@zharova.ma](https://app.pachca.com/chats/4734693?thread_id=839707) Подскажите пожалуйста после загрузки ссылки на платформе на github с решением - нужно ещё и индивидуально ментору её скидывать? Почему спрашиваю, потому что там было сказано - отправте ссылку ментору на проверку...

2 комментария

22 апреля 2023 г.

****

Мария Жарова

11:48

Добрый день! Приношу извинения да долгий ответ🙏🏻 Ничего дополнительно делать и сообщать не нужно, просто загружайте ссылку на платформу и всё😊

Василий Бухаров

19.03.2023, 16:03

[@zharova.ma](https://app.pachca.com/chats/4734693?thread_id=505910) Подскажите пожалуйста, а на гитхабе должна быть, наряду с ноутбуком, модель обернутая в докерконтейнер, готовая к работе в продакшене?

1 комментарий

20 марта 2023 г.

****

Мария Жарова

00:28

Добрый день! Желательно да, чтобы можно было поставить полный балл по этому критерию🙃

А Вы случаем не стандартизировали y\_train перед подачей в модель? Просто если стандартизировали то округление и даст такую большую ошибку. На трэйне не должен RFR давать такую большую mape

Я про целевую переменную

Сколько уникальных значений в y\_train и в rf\_prediction\_train?

Я Ваши данные не вижу, но есть ощущение, что проблемма где то здесь

Сделайте датафрейм из реальных значений и предсказаний и посмотрите на него, я думаю это поможет найти то место, где проблема зарылась

**В**

Василий Бухаров

11:07

Если у вас среди фичей есть что то коррелируюшее с целевой переменной больше чем на 0.3 то полюбому на тренировочный выборке предсказания не смогут дать такую ошибку

Точнее даже так - даже если нету вообще никаких корреляций, то по алгоритму rfr все предсказания сойдутся тогда примерно к среднему, и соответственно mape будет в районе 0.5 ну или меньше, в зависимости от дисперсии в данных. Если mape приближается к 1, то проблемма где то в преобразовании данных, и плюс у mape есть такой момент, что в случае наличия в данных большого количества значений целевой переменной близких к 0, то она будет расти непропорцианально mae