

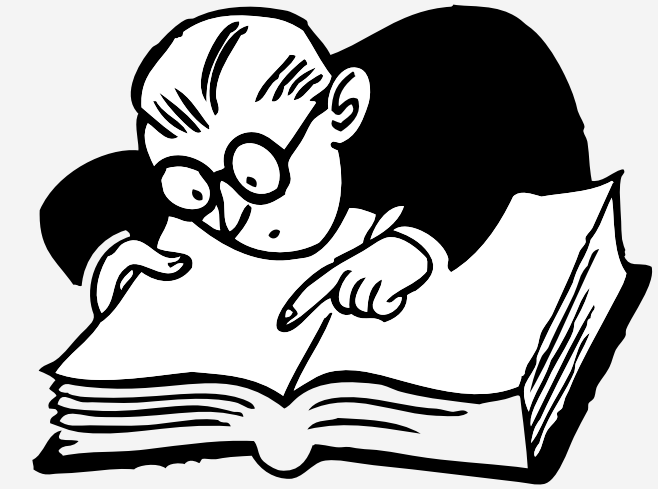
Прогнозування ціни на ноутбук

на основі його
ключових компонентів

Роботу виконав
Студент групи ІП-22
Нижник Дмитро Сергійович



Мета курсової роботи



Основна мета - розробка програми, що здатна передбачати ціну на ноутбуки на основі їх характеристик

Для досягнення основної мети необхідно обрати, оцінити та покращити моделі машинного навчання

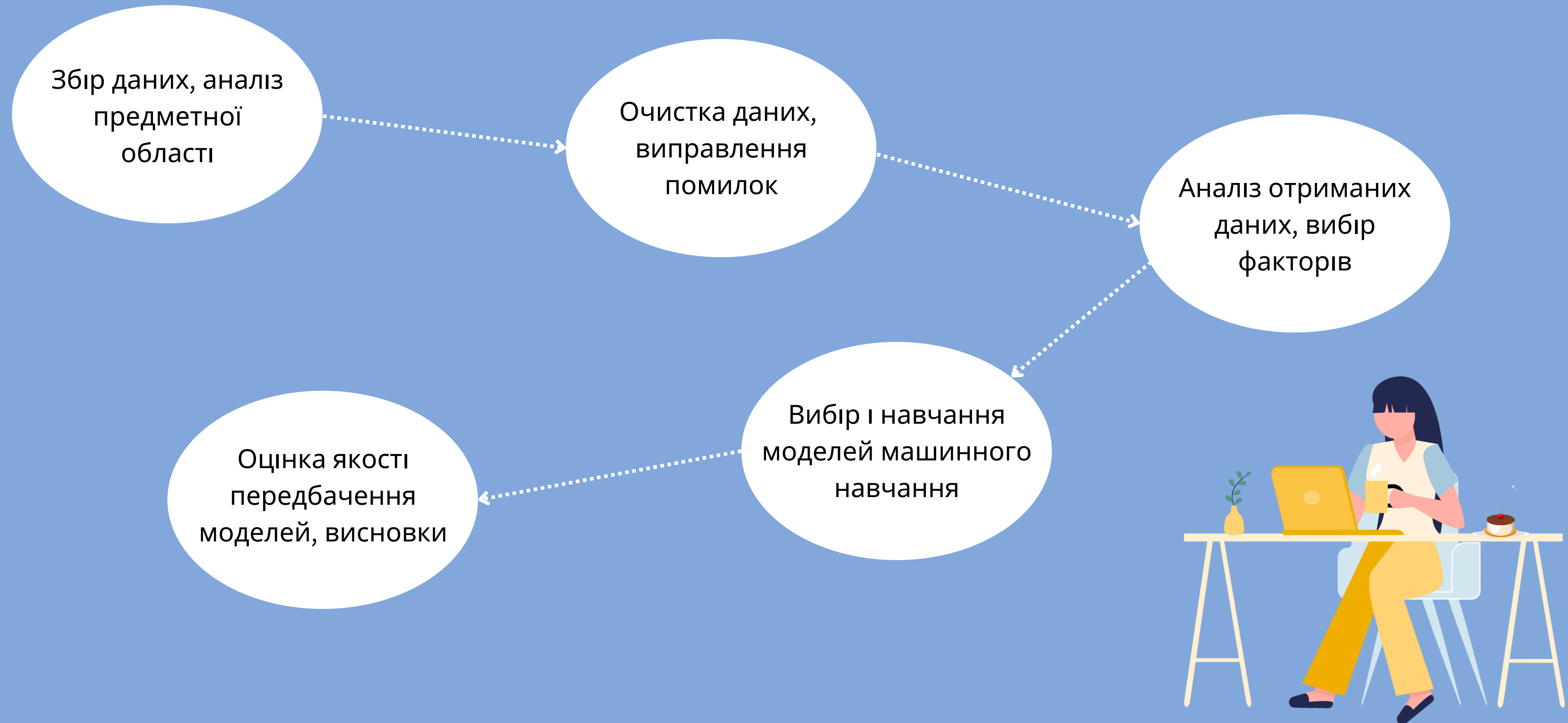
Для ефективного аналізу зібрані дані треба підготувати: очистити, виправити та оцінити

Інструменти виконання роботи - мова програмування Python 3 та її бібліотеки

Результати дослідження можуть бути корисними і продавцям, і покупцям



Послідовність дій



Вибір предикторів



Вибір методів машинного навчання

Linear Regression

Має наступні переваги:

- проста для розуміння
- проста для використання
- результати легко інтерпретуються

Має наступні недоліки:

- чутлива до перенавчання
- чутлива до викидів
- чутлива до мультиколінеарності

Ridge Regression

Має наступні переваги:

- не складна у реалізації
- нижча чутливість до мультиколінеарності
- зменшений ризик перенавчання

Має наступні недоліки:

- результати складніше інтерпретуються
- вимагає підбору α

Random Forest

Має наступні переваги:

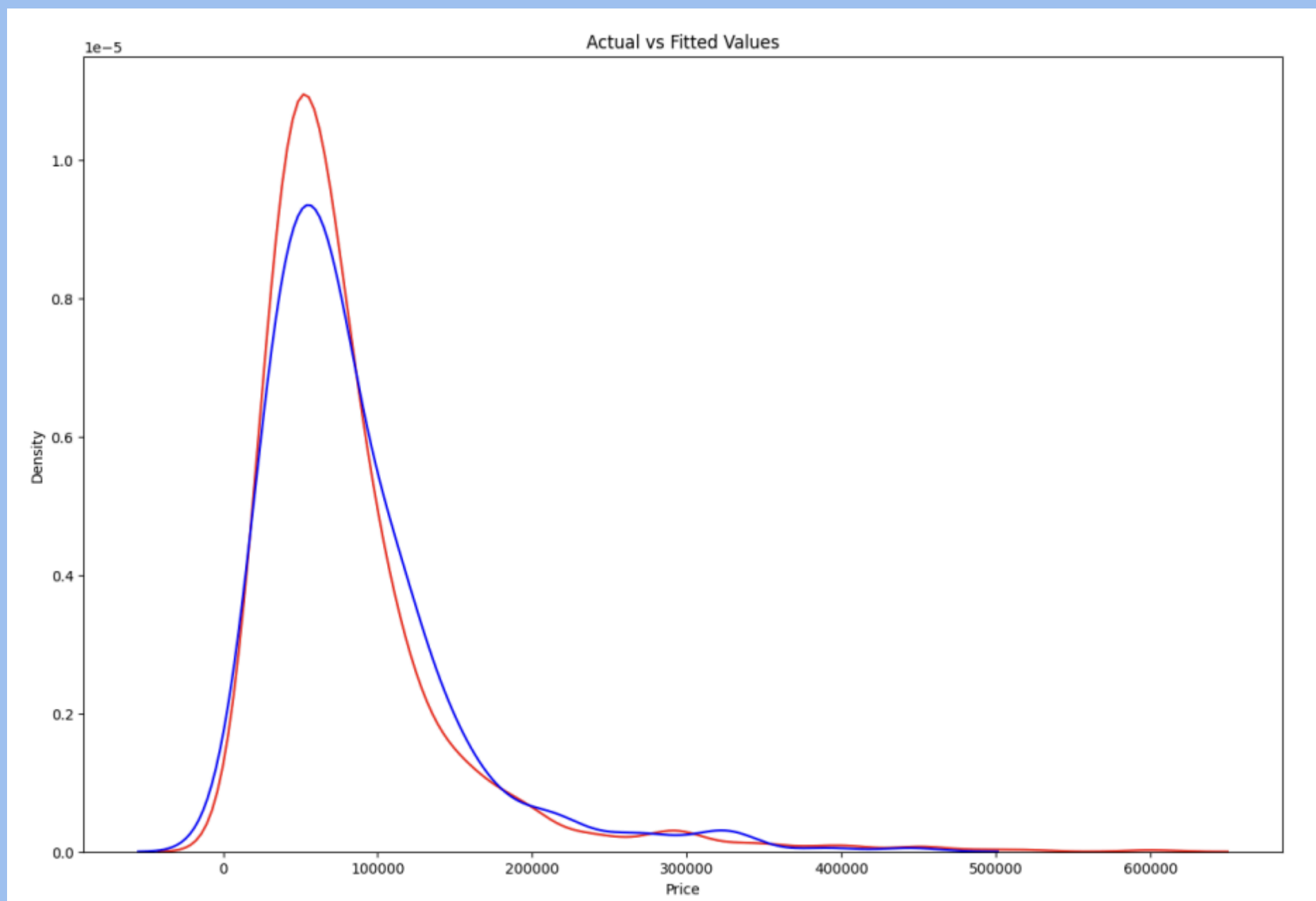
- стійка до перенавчання
- стійка до шуму та викидів
- висока точність передбачень

Має наступні недоліки:

- довго навчається
- підбір гіперпараметрів дуже довгий та затратний
- вимагає великого обсягу пам'яті



Аналіз результатів. Лінійна регресія



Висновки побудови та оцінки моделі:

- найлегша у реалізації
- найменш точна модель
- швидше всього навчається

Значення оцінок:

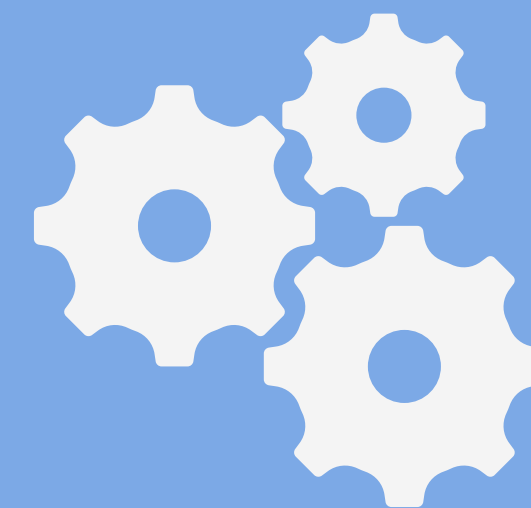
MAPE = 20.5%

RMSE = 25128

$R^2 = 0.863$



Аналіз результатів. Гребенева регресія



Висновки побудови та оцінки моделі:

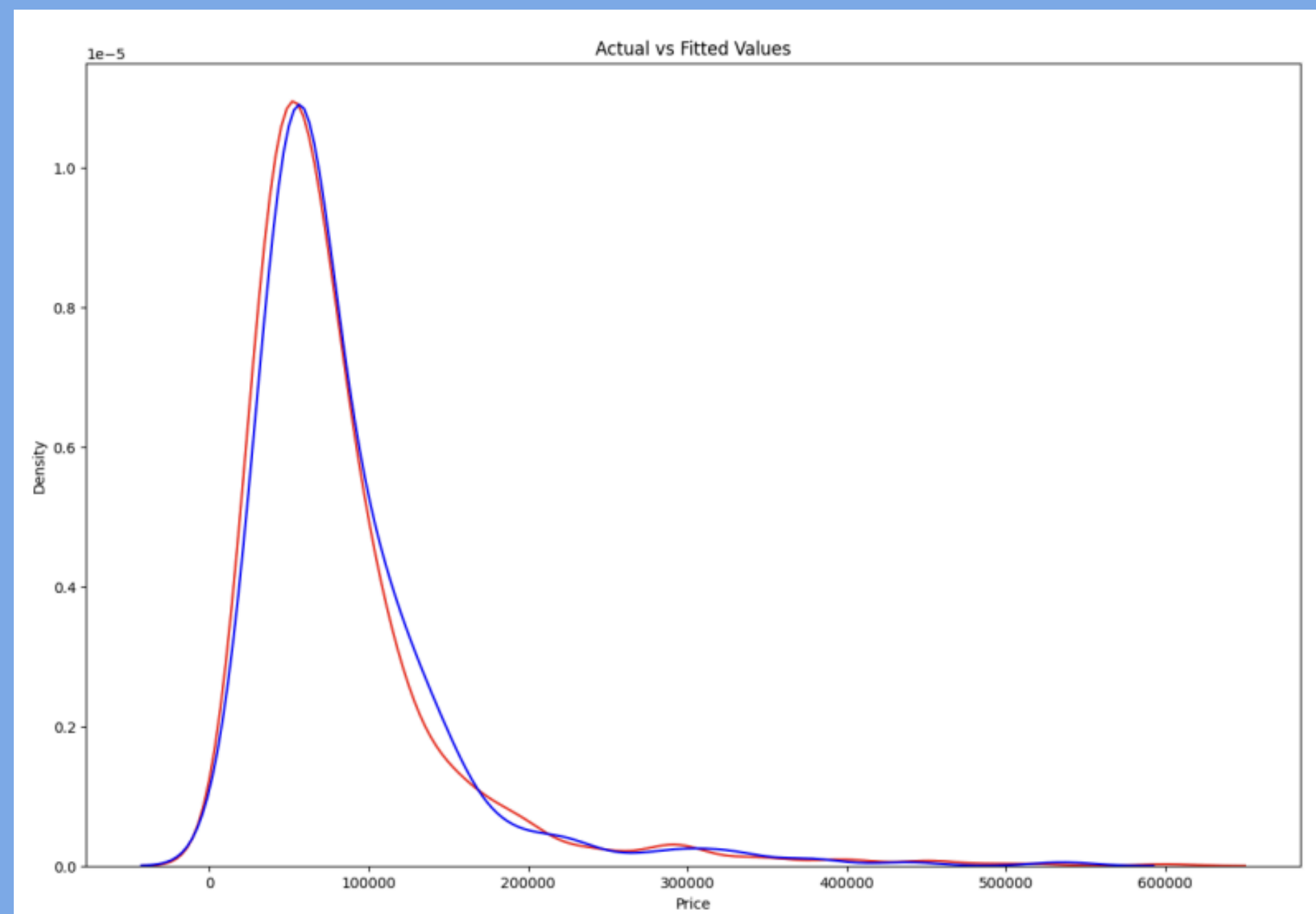
- найбільш точна модель
- достатньо швидко навчається
- підбір гіперпараметру, стандартизація та поліноміальна трансформація даних покращили результати

Значення оцінок:

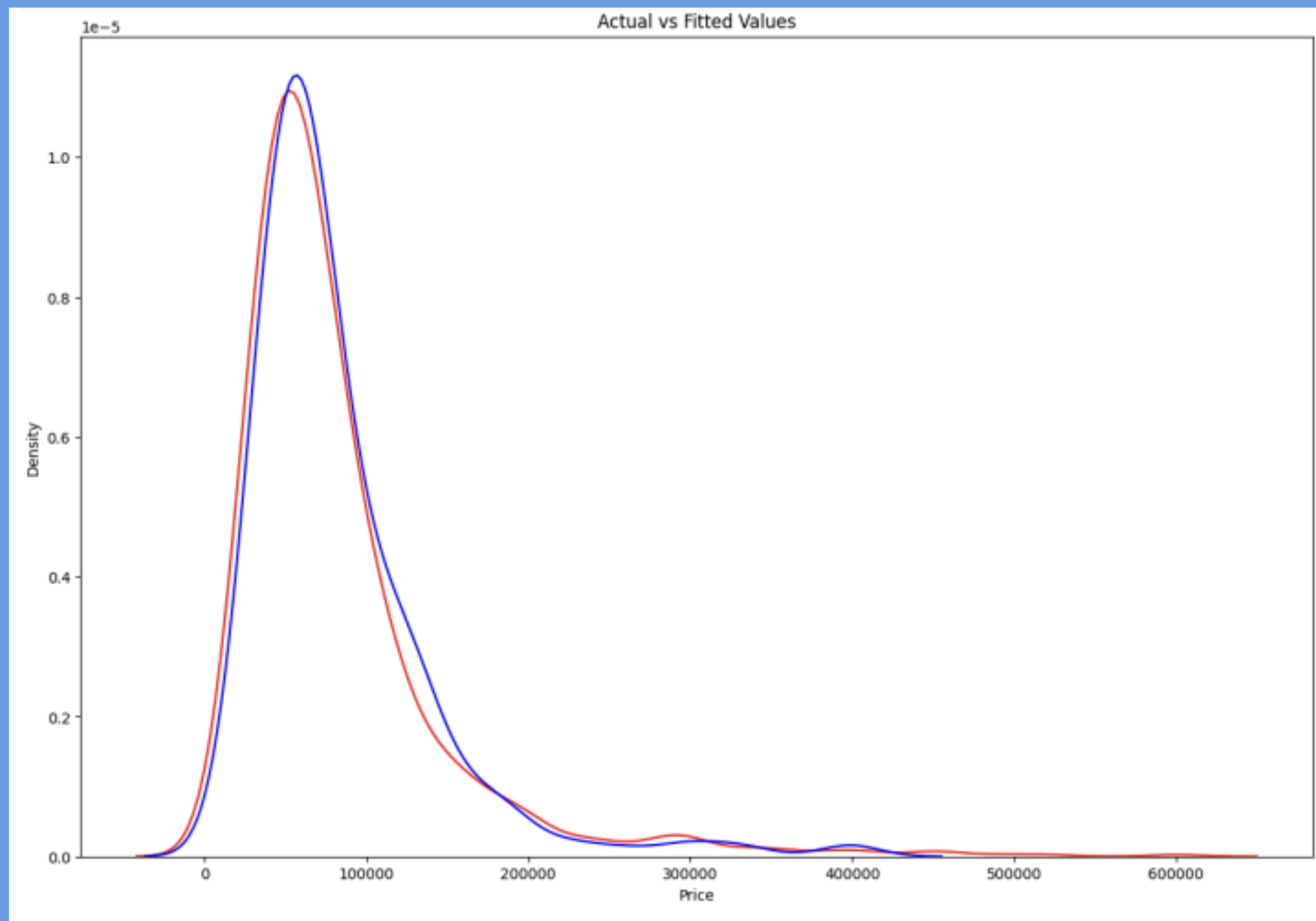
MAPE = 17.1%

RMSE = 23082

$R^2 = 0.884$



Аналіз результатів. Випадковий ліс



Висновки побудови та оцінки моделі:

- достатньо точна модель
- довго навчається, довго підбираються параметри
- підбір гіперпараметрів не покращив результати

Значення оцінок:

MAPE = 17.1%

RMSE = 26105

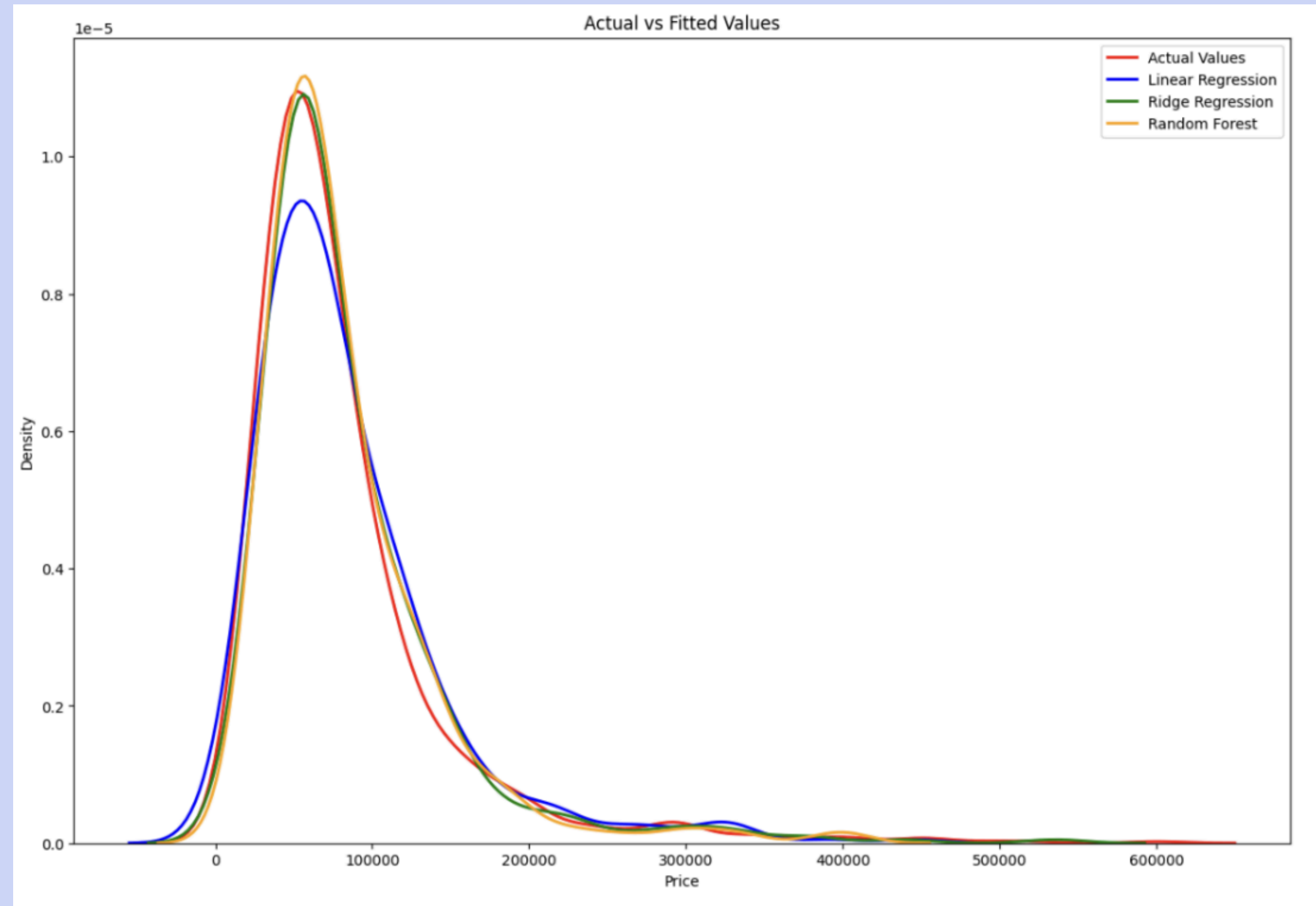
$R^2 = 0.852$

Висновки

Під час виконання курсової роботи було зроблено наступне:

- обрано, очищено, виправлено та проаналізовано дані
- обрано числові та категоріальні фактори для аналізу даних
- побудовано та оцінено 3 моделі машинного навчання

Найбільш якісні результати дала модель Ridge Regression з підбором параметру α , стандартизацією та поліноміальним перетворенням даних





Дякую за увагу!

На цьому все