МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное автономное образовательное учреждение высшего образования

«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

КАФЕДРА № 43

ОТЧЕТ   
ЗАЩИЩЕН С ОЦЕНКОЙ

ПРЕПОДАВАТЕЛЬ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Доцент |  |  |  | В. Ю. Скобцов |
| должность, уч. степень, звание |  | подпись, дата |  | инициалы, фамилия |

|  |
| --- |
| ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ №3 |
| классификация табличных данных на основе нейросетевых моделей |
| по курсу: интеллектуальный анализ данных на основе методов машинного обучения |
|  |
|  |

РАБОТУ ВЫПОЛНИЛ

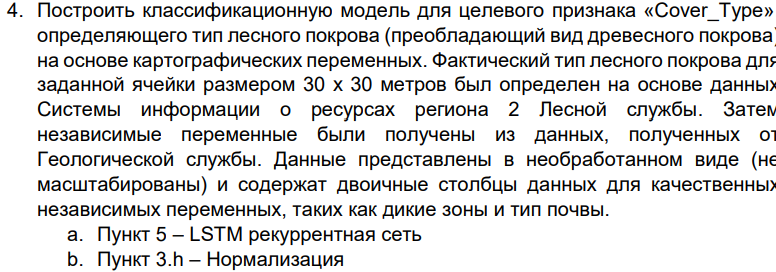
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СТУДЕНТ гр. № | 4134К |  |  |  | Д.В. Самарин,  Р. Р. Усов |
|  |  |  | подпись, дата |  | инициалы, фамилия |

Санкт-Петербург 2024

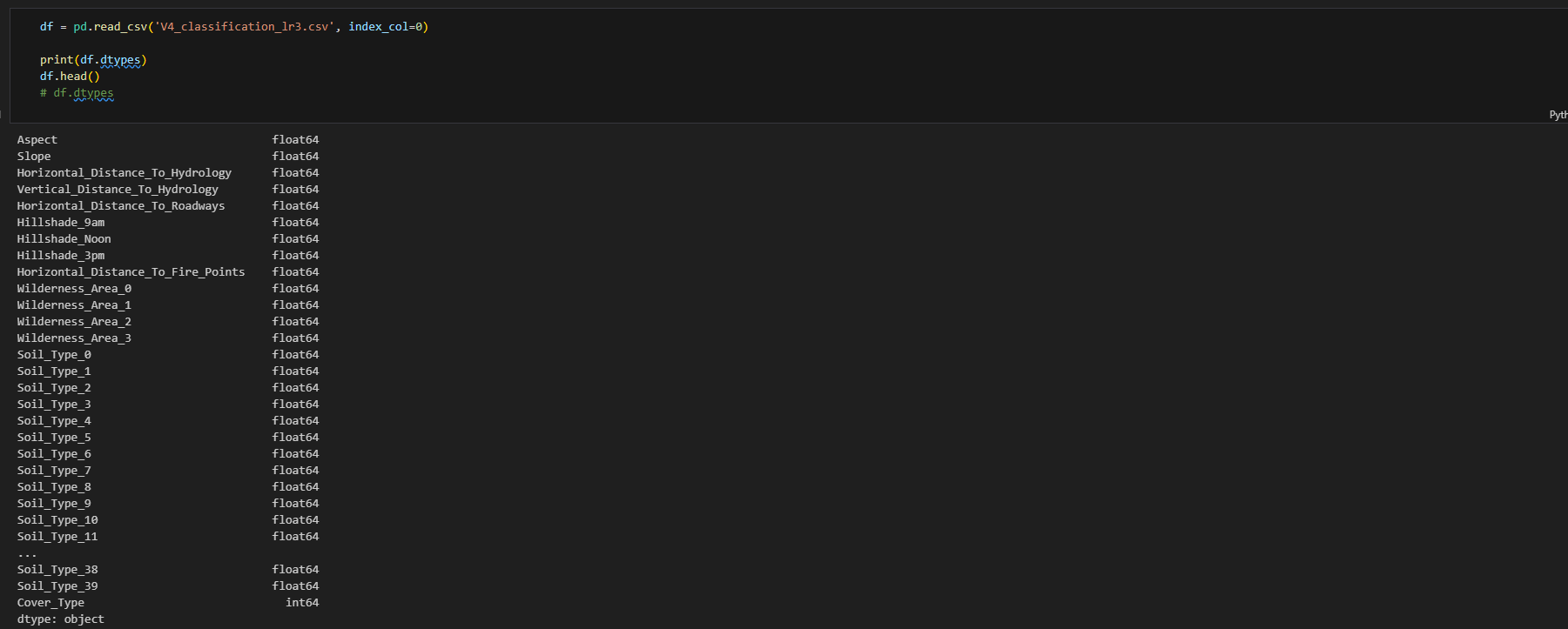
1. **Цель работы**

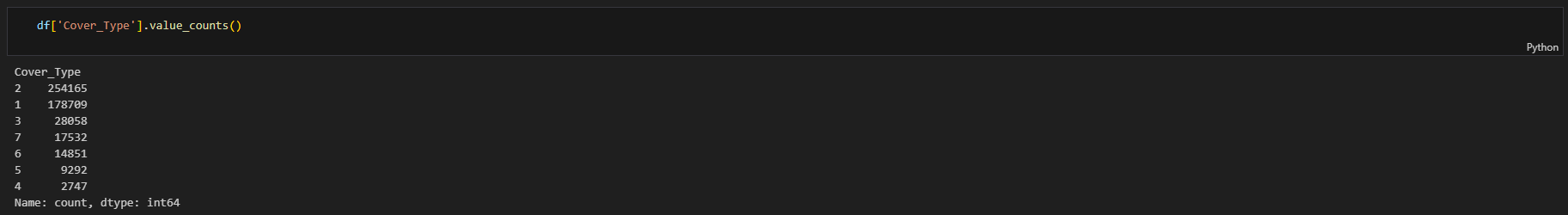
Необходимо выполнить классификационный анализ данных по указанному целевому признаку

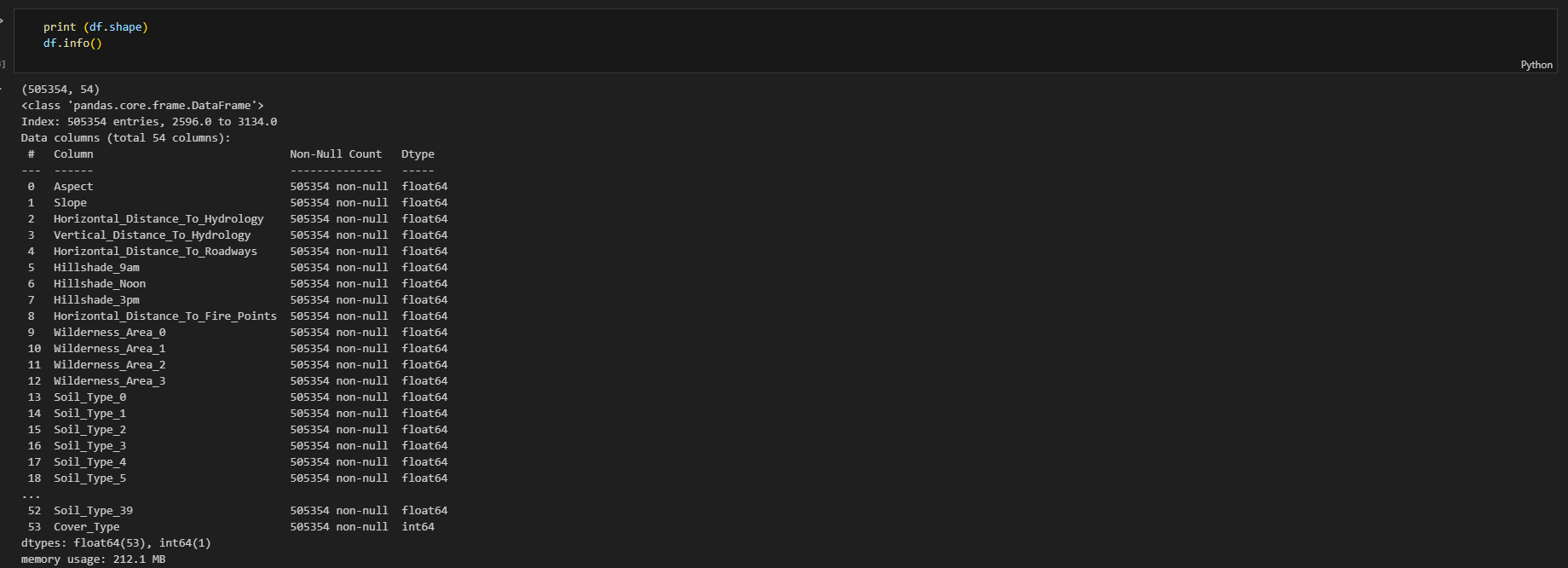
**Вариант:** 4



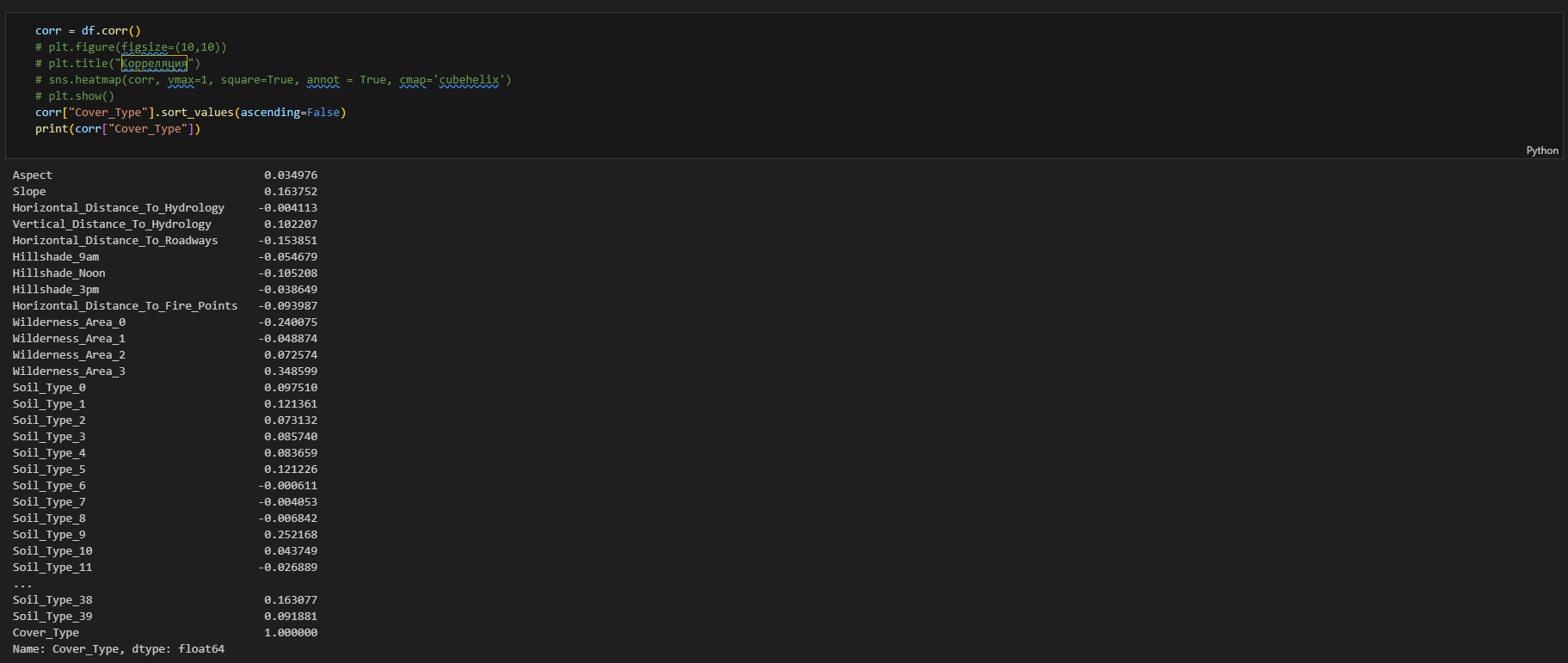
1. **Ход работы**
2. %matplotlib inline
3. import joblib
4. import pandas as pd
5. import numpy as np
6. import seaborn as sns
7. import matplotlib.pyplot as plt
8. import tensorflow as tf
9. from sklearn.model\_selection import train\_test\_split
10. from sklearn.preprocessing import Normalizer
11. from sklearn.linear\_model import LinearRegression
12. from sklearn.metrics import mean\_squared\_error, r2\_score
13. from tensorflow.keras.models import Sequential
14. from tensorflow.keras.layers import LSTM, Dense, Input
15. from pandas.plotting import scatter\_matrix
16. from tensorflow.keras.utils import to\_categorical
17. from sklearn.metrics import cohen\_kappa\_score
18. from tensorflow.keras.optimizers import Adam
19. df = pd.read\_csv('V4\_classification\_lr3.csv', index\_col=0)
20. print(df.dtypes)
21. df.head()
22. # df.dtypes



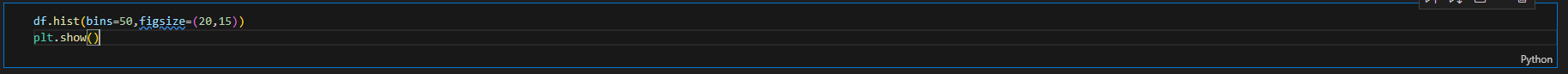


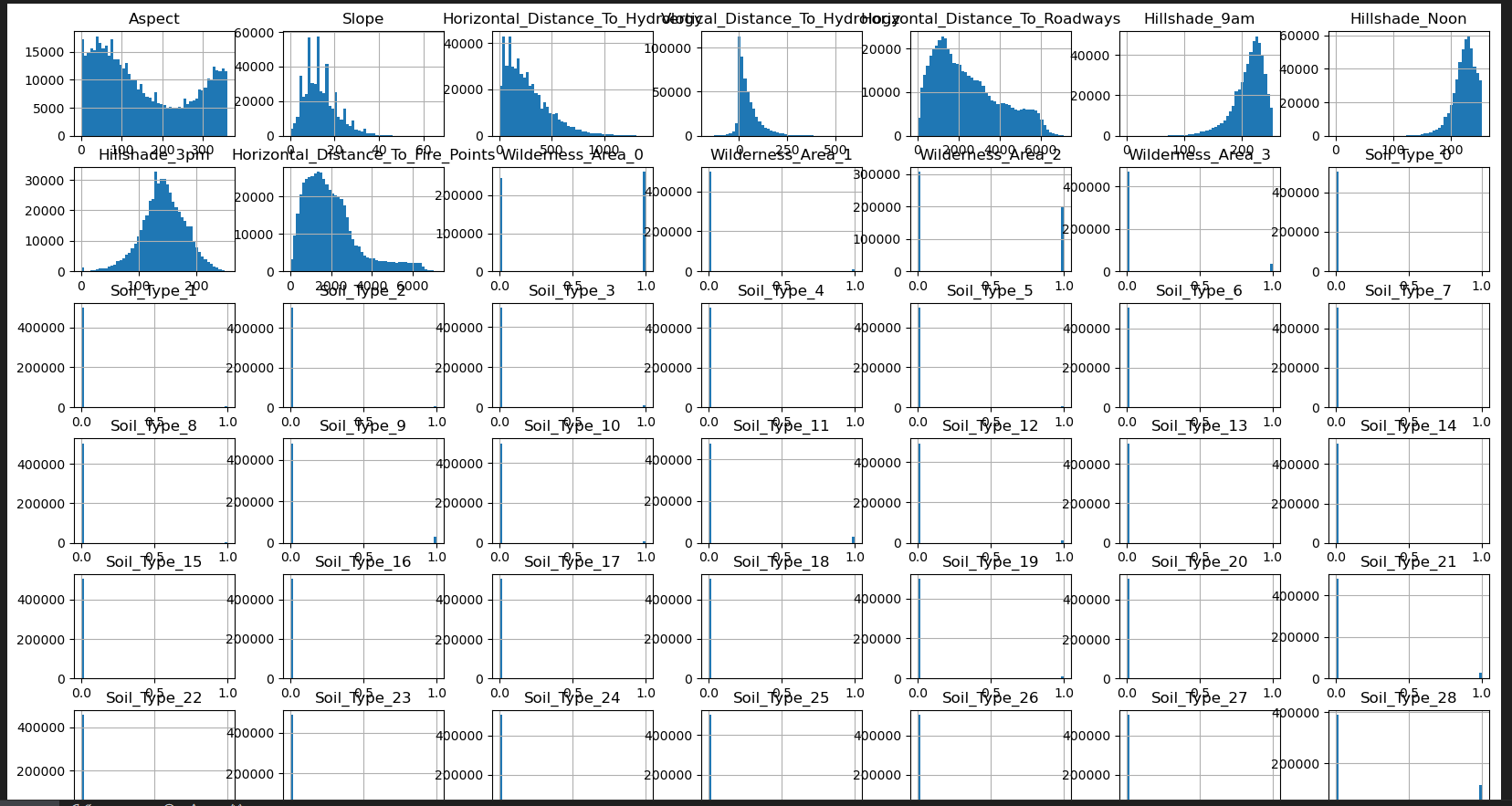


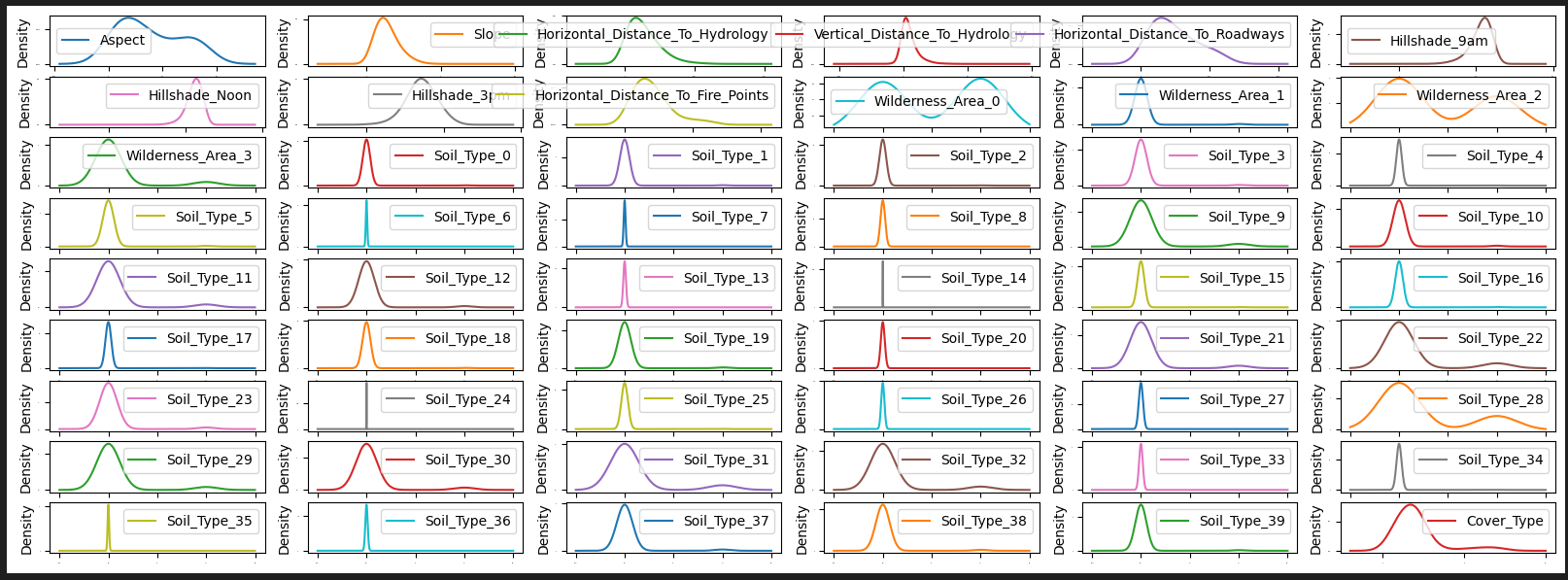
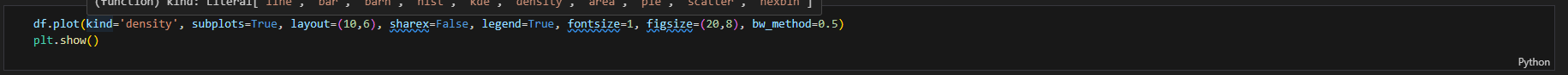
Выполним анализ корреляционной зависимости для целевого признака (Cover\_Type)

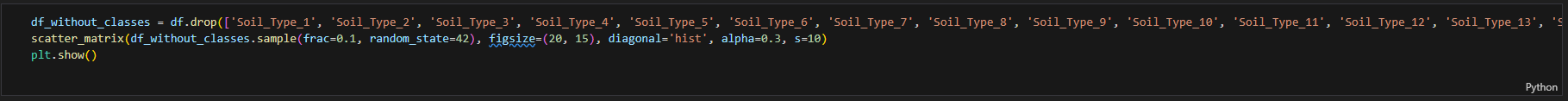


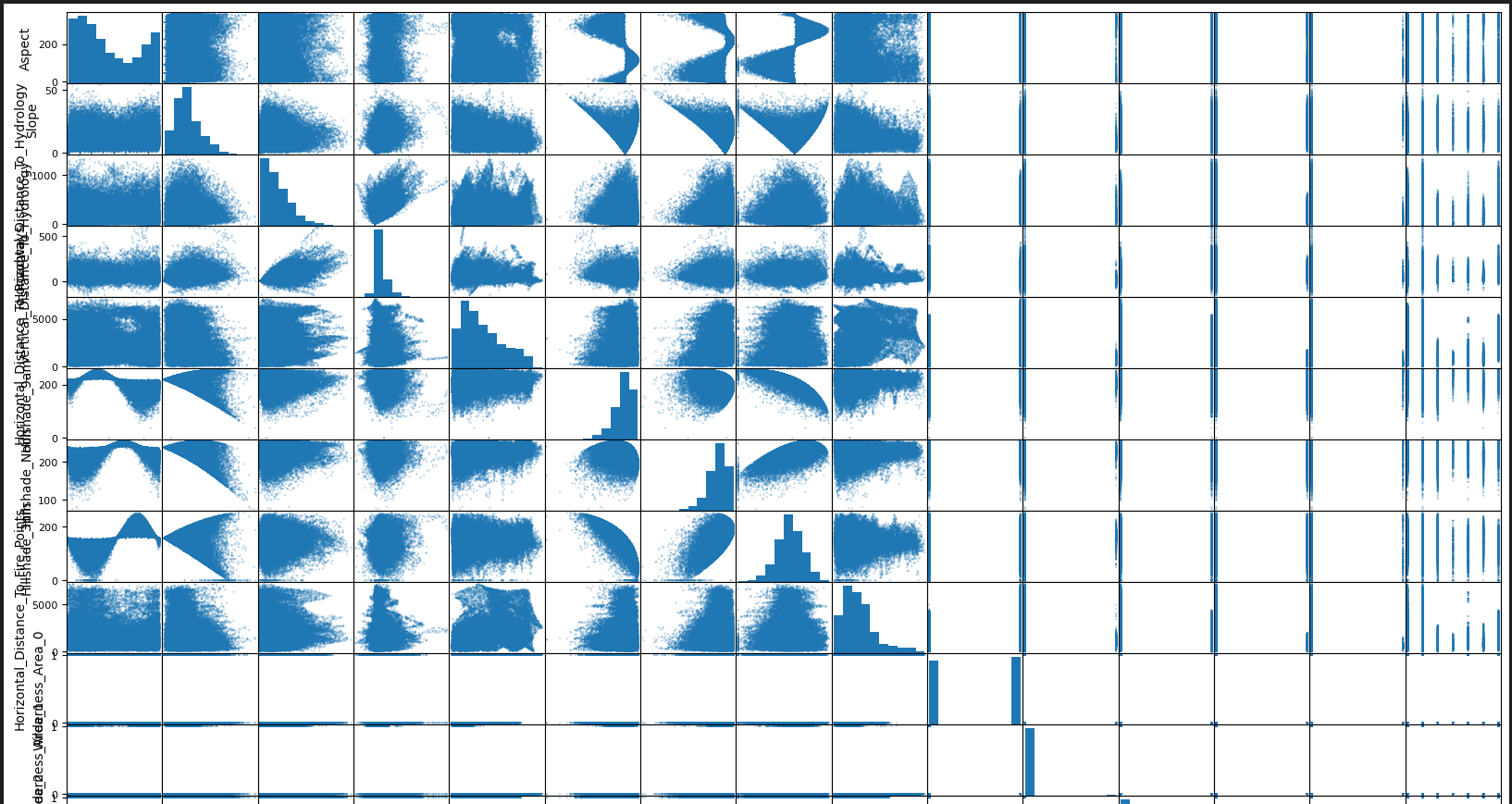
Построим гистограммы распределения и плотнотей и зависимостей признаков





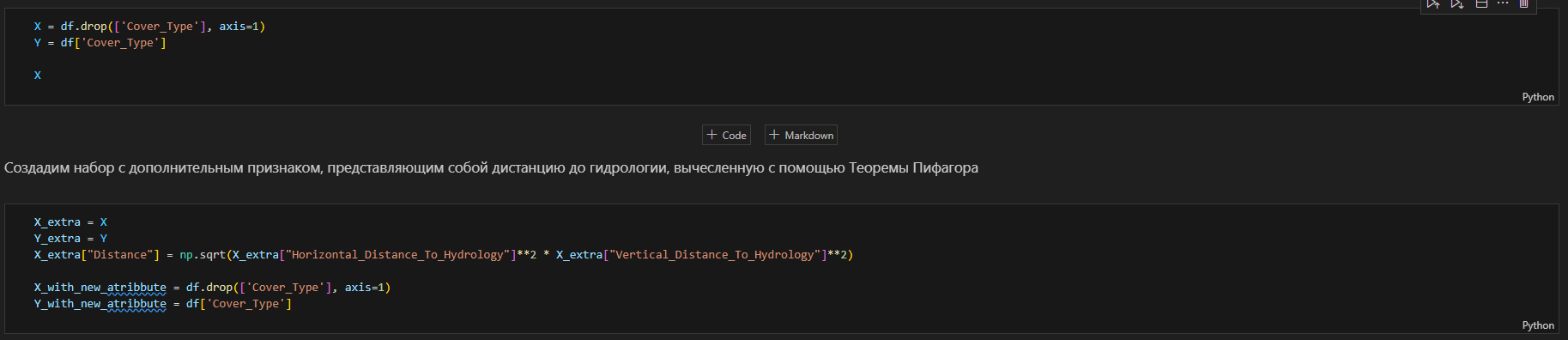




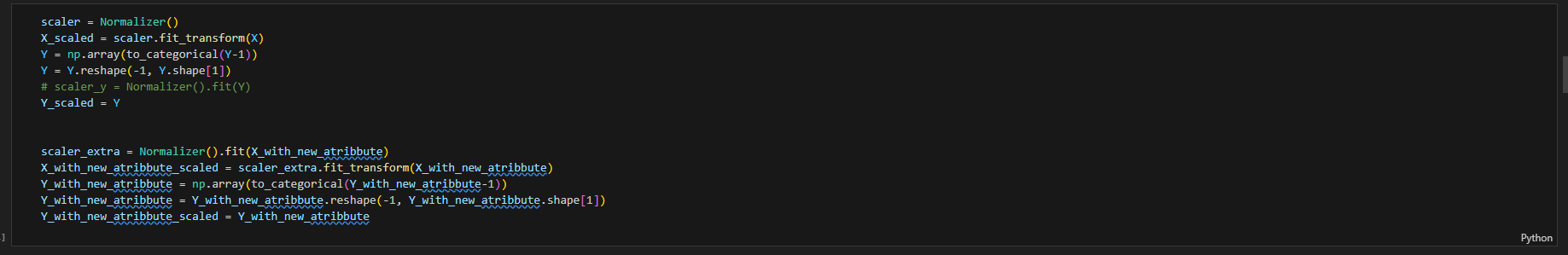


Как можно заметить, самые несущественные признаки - тип почвы в некоторых столбцах. Поскольку этот признак является важным, и представляет собой категориальный признак, мы не можем от него избавиться.

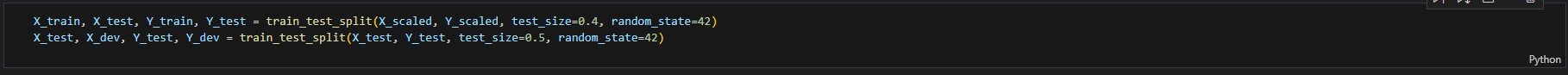
Разделим данные на входные и выходные значения



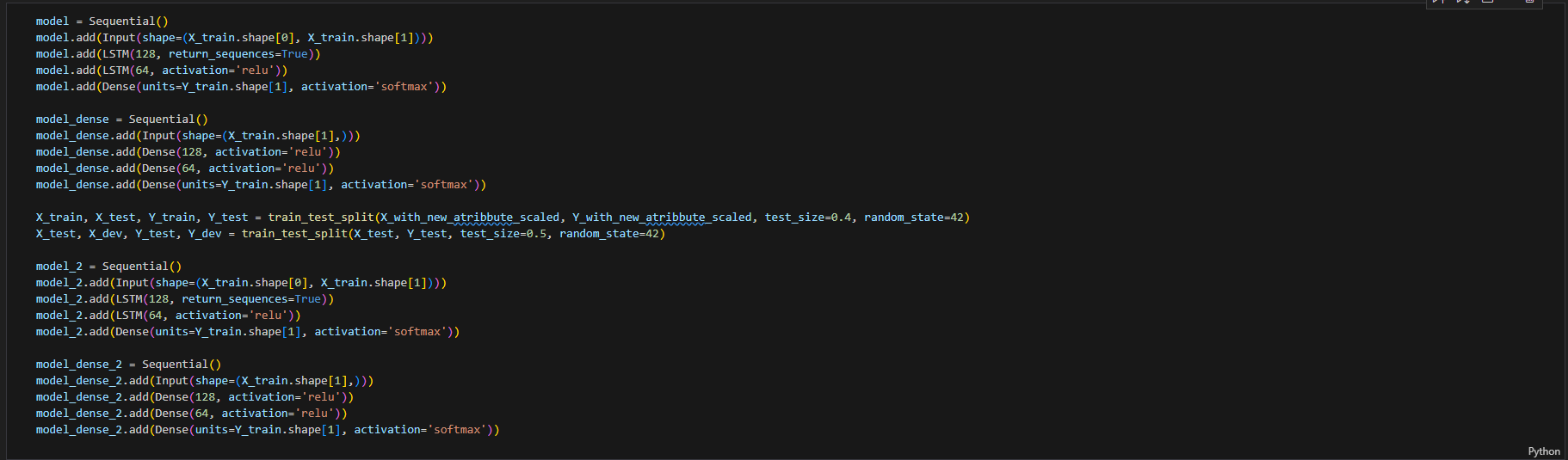
Используем нормализаю для обоих наборов данных.



Разобъём выборку на обучающую, тестовую и валидационную в отношении 60/20/20 %

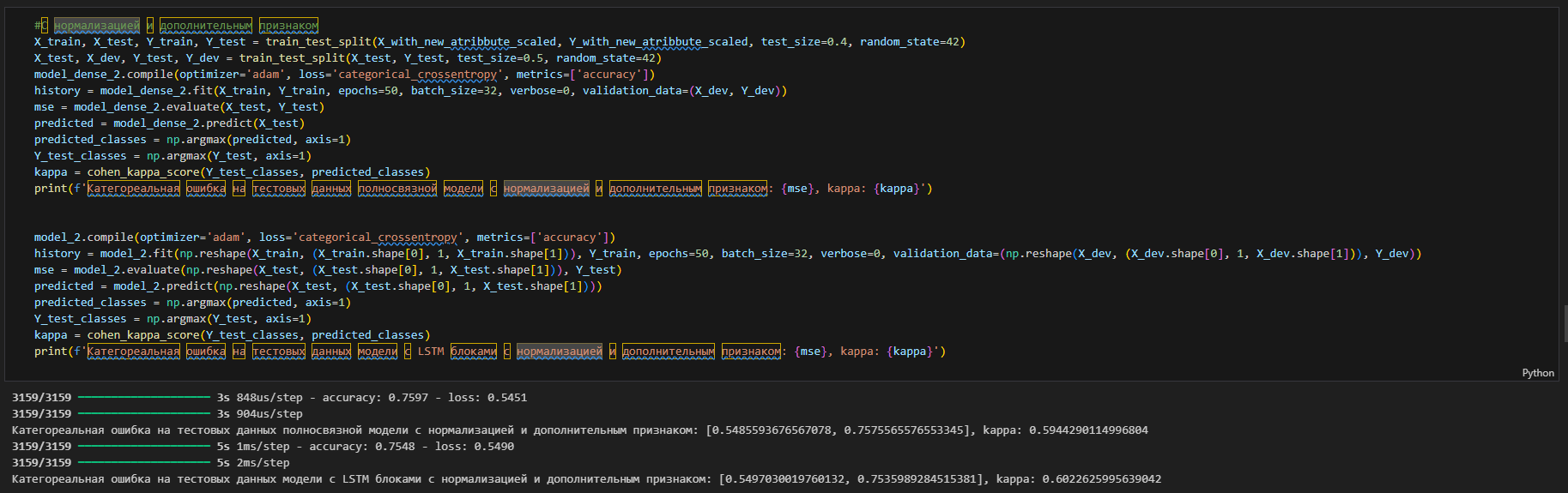


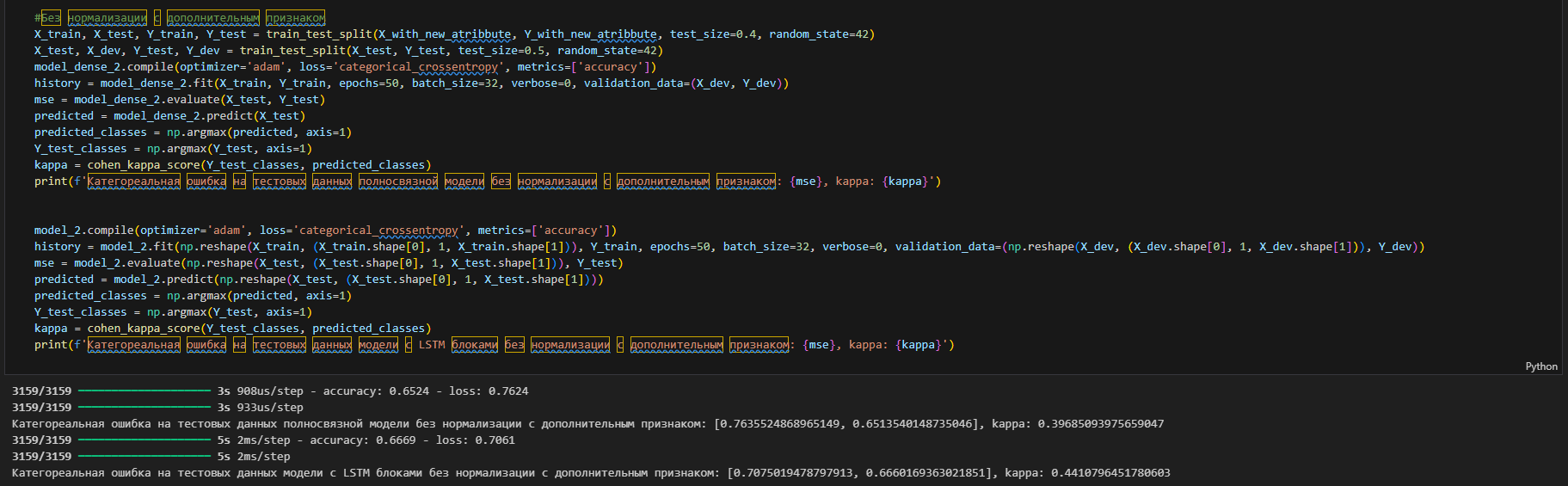
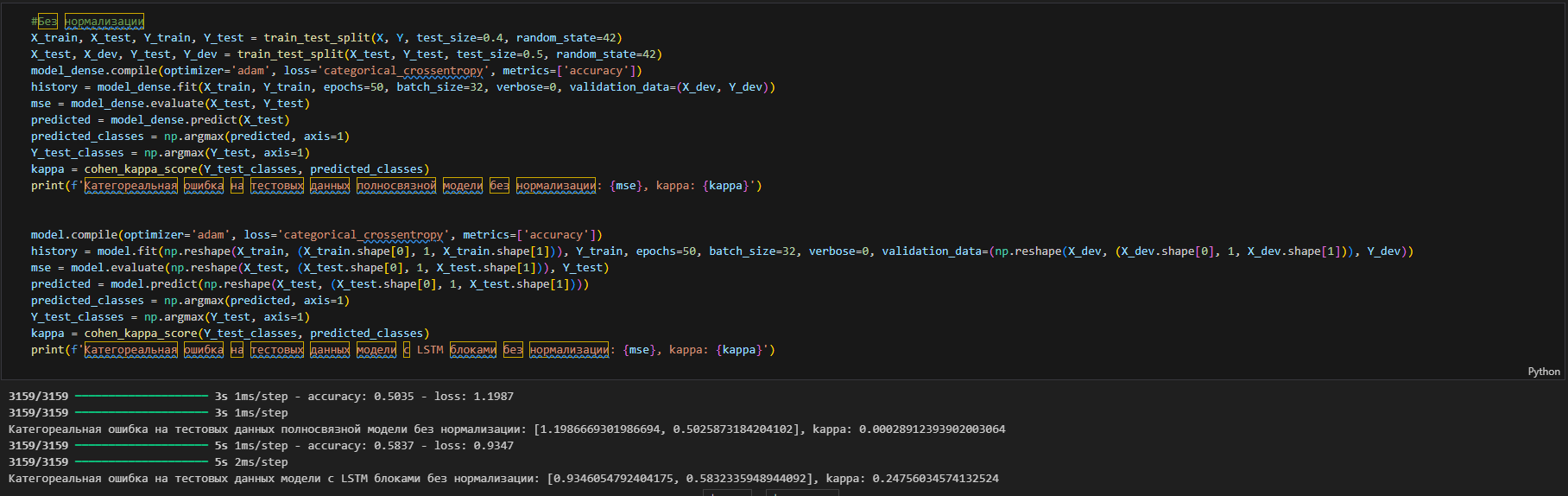
Создадим модель с LSTM блоками и полносвязную модель для последующего обучения



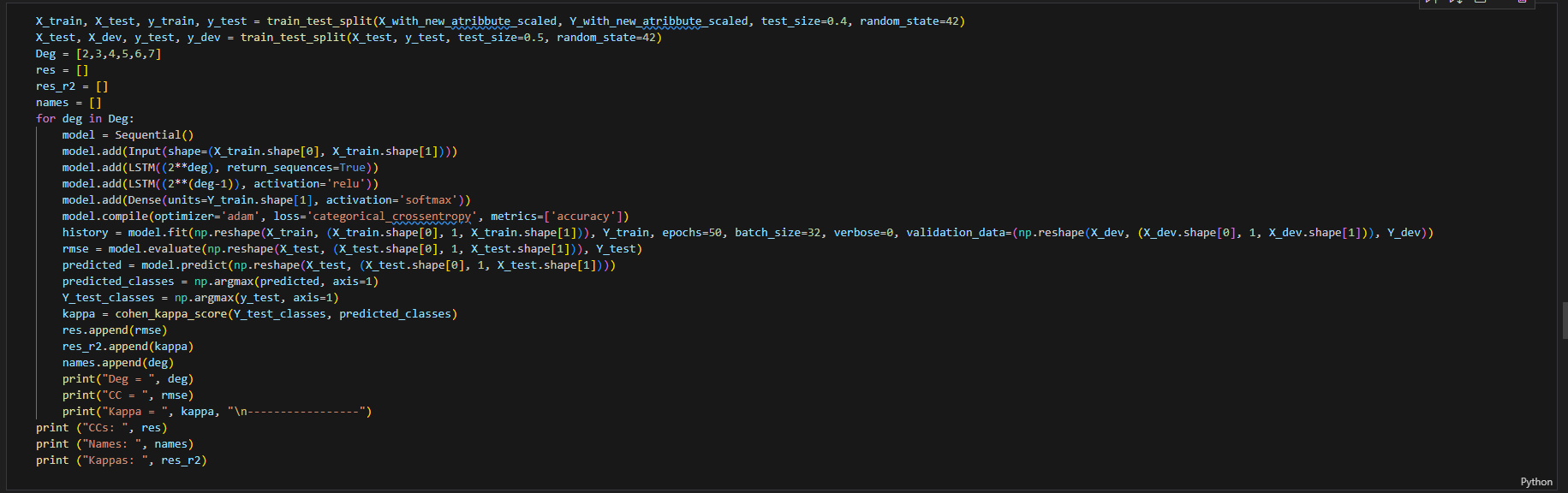
Обучим модели





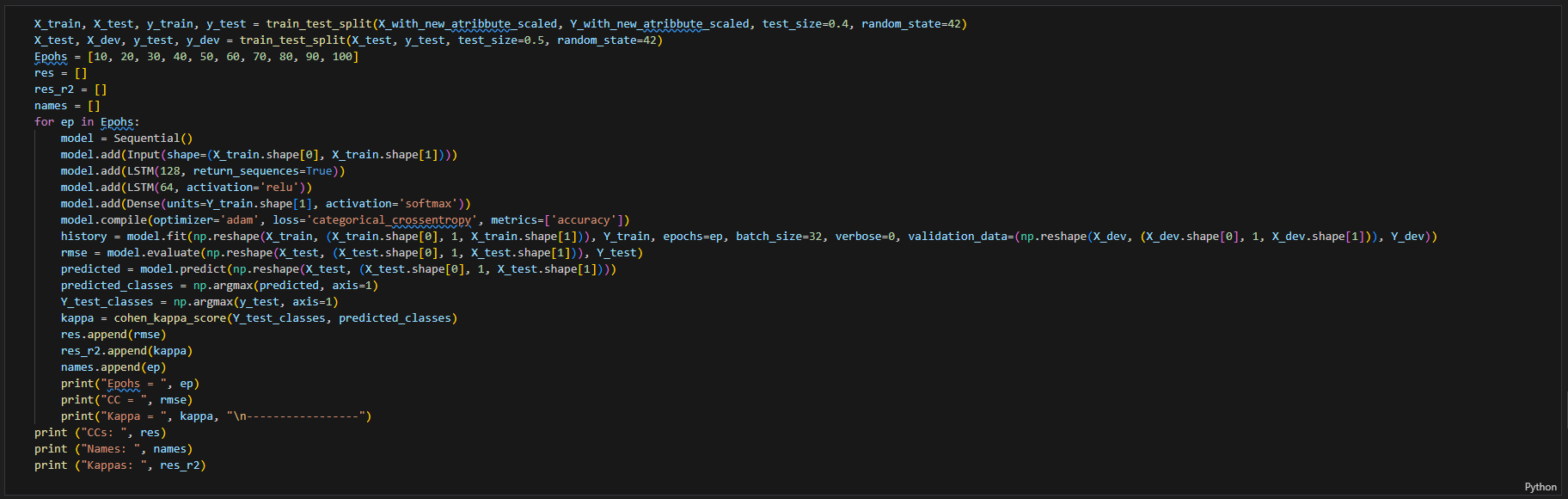


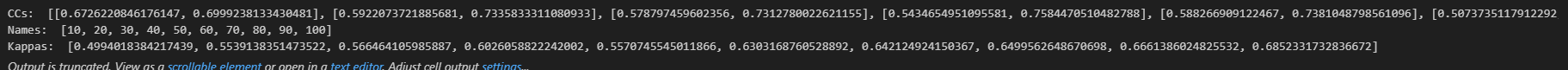
Лучший результат показала модель с LSTM блоками, нормализацией и дополнительными признаками. Выполним Grid поиск гиперпараметров для неё



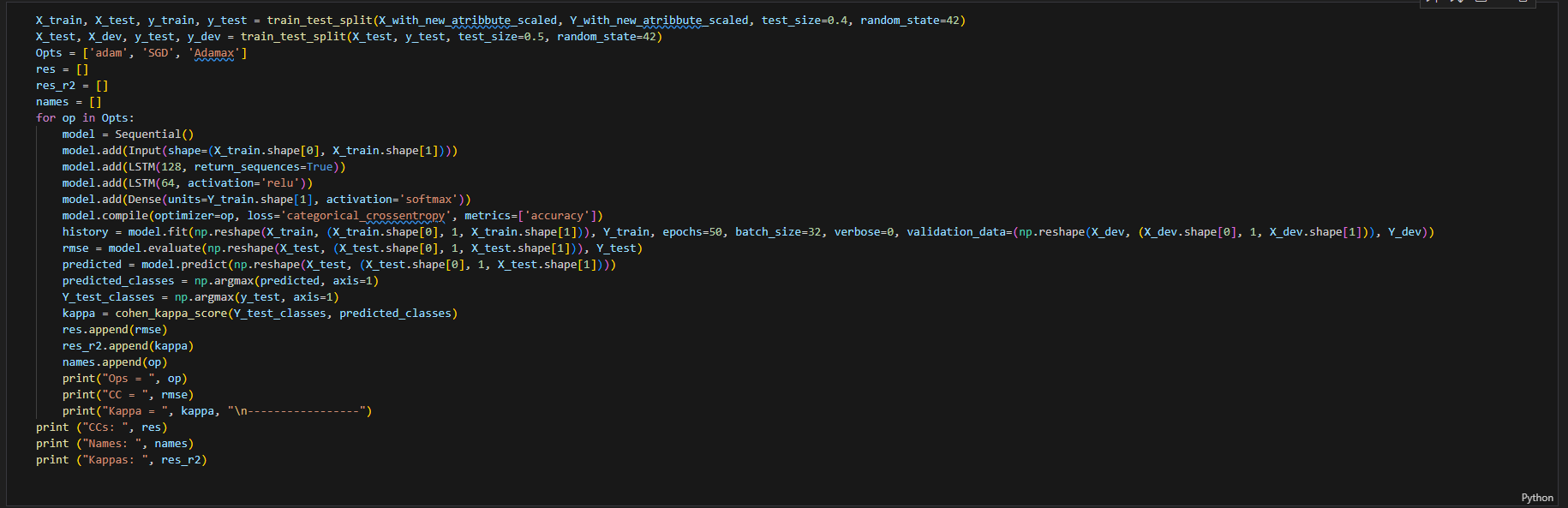


Лучший результат при количестве нейронов: 128 и 64 в первом и втором слоях соответсвенно



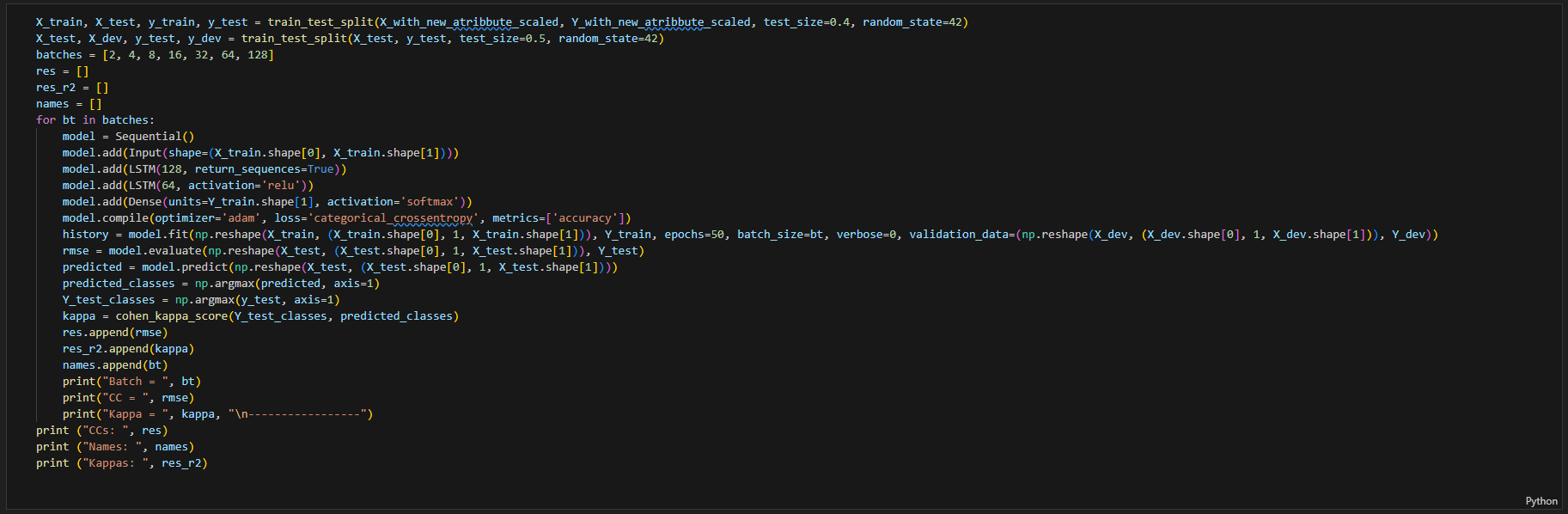


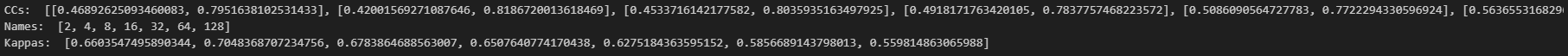
Лучший результат при количестве эпох, равном 100



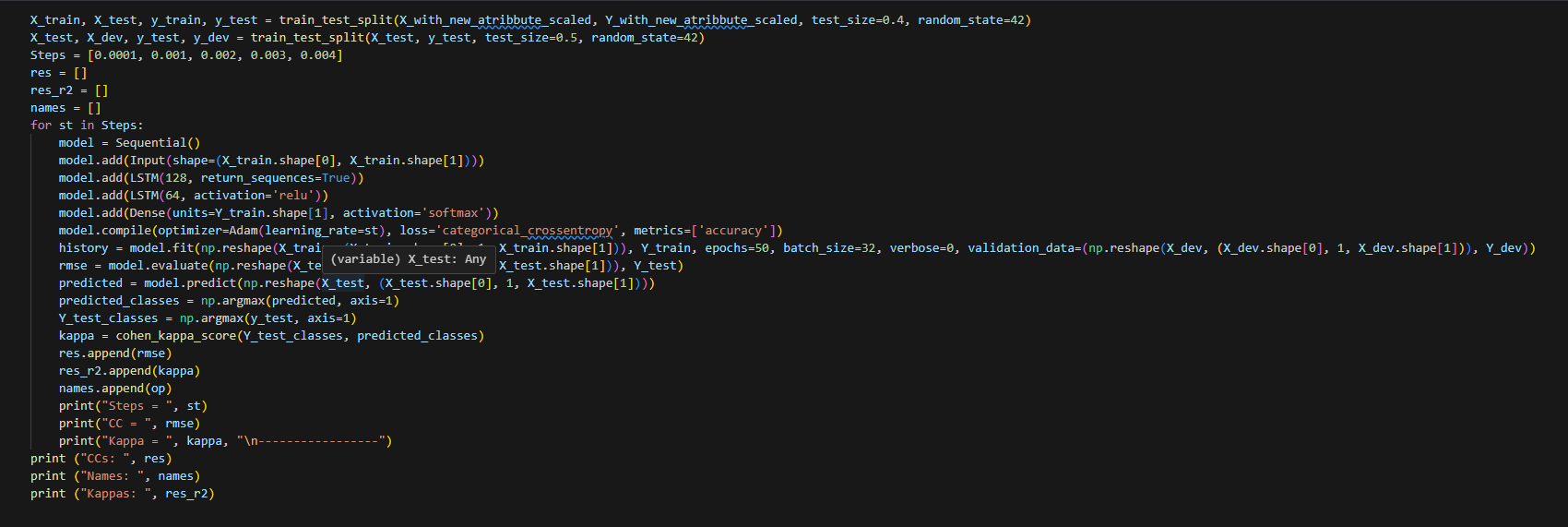


Лучший результат у оптимизатора adam





Лучший результат с размером батча 4





Лучший результат со скоростью обучения, равной 0.004

Итоговая модель, на основе LSTM блоков с нормализованным датасетом с дополнительным признаком и следующими гиперпараметрами:

- 128, 64 нейронов в первом и втором слоях соответсвенно

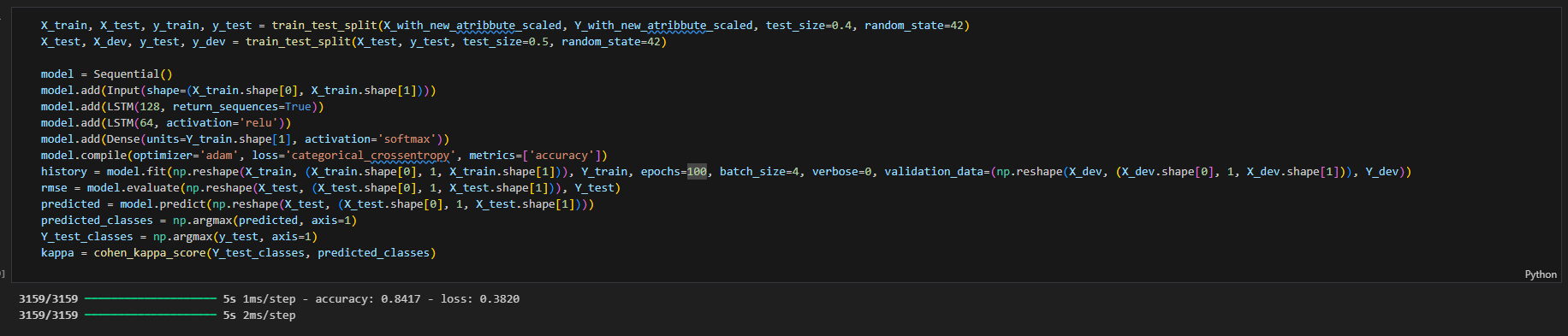
- Оптимизатор Adam

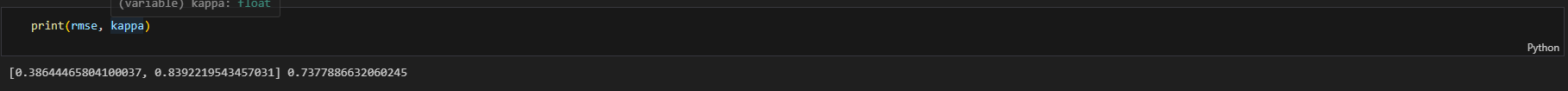
- Размер батча 4

- Количество эпох 100

- Скорость обучения 0.004

Выполним прогнозирования для модели с лучшими параметрами:





Лучшая точность данной модели:

- Kappa: 0.7377

- CC: 0.3864

- Accuracy: 0.8392

1. **Выводы**

В ходе данной лабораторной работы был произведён поиск лучшей модели для датасета с целевым признаком тип покрытия. Задача состояла в классификации данного признака. Лучший результат показала LSTM модель с нормализованным датасетом с дополнительным признаком и следующими гиперпараметрами:

- 128, 64 нейронов в первом и втором слоях соответсвенно

- Оптимизатор Adam

- Размер батча 4

- Количество эпох 100

- Скорость обучения 0.004

Её точность получилась:

- Kappa: 0.7377

- CC: 0.3864

- Accuracy: 0.8392