**МИНИCTEPCTBO НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ**

**РОССИЙСКОЙ ФЕДЕРАЦИИ**

**Федеральное государственное автономное**

**образовательное учреждение высшего образования**

**«Северо-Кавказский федеральный университет»**

**Кафедра инфокоммуникаций**

**«Обучающая, проверочная и тестовая выборки. Переобучение НС»**

**Отчет по лабораторной работе №2**

**по дисциплине «Обработка данных для построения систем искусственного интеллекта»**

|  |
| --- |
| Выполнила студентка группы ПИЖ-б-о-21-1 |
| Халимендик Я.Д. «    »\_\_\_\_\_\_\_\_\_20\_\_г. |
| Подпись студента\_\_\_\_\_\_\_\_\_\_\_ |
| Работа защищена «    »\_\_\_\_\_\_\_\_\_\_\_\_20\_\_г. |
| Проверил канд. технических наук,  доцент, доцент кафедры инфокоммуникаций  Воронкин Р. А. \_\_\_\_\_\_\_\_\_\_\_\_  (подпись) |

Ставрополь, 2024 г.

**Цель работы:** изучить типы выборок (обучающая, проверочная и тестовая), способы разделения данных на выборки. Изучить причины возникновения переобучения сети и способы решения данной проблемы. Применить полученные знания для решения практических задач.

**Ход работы:**

**Задача 1.** Создайте 9 моделей нейронной сети с различными архитектурами и сравните в них значения точности на проверочной выборке (на последней эпохе) и на тестовой выборке. Используйте следующее деление: обучающая выборка - 50000 примеров, проверочная выборка - 10000 примеров, тестовая выборка - 10000 примеров.

Создайте сравнительную таблицу в конце ноутбука, напишите свои выводы по результатам проведенных тестов.

**Решение.**

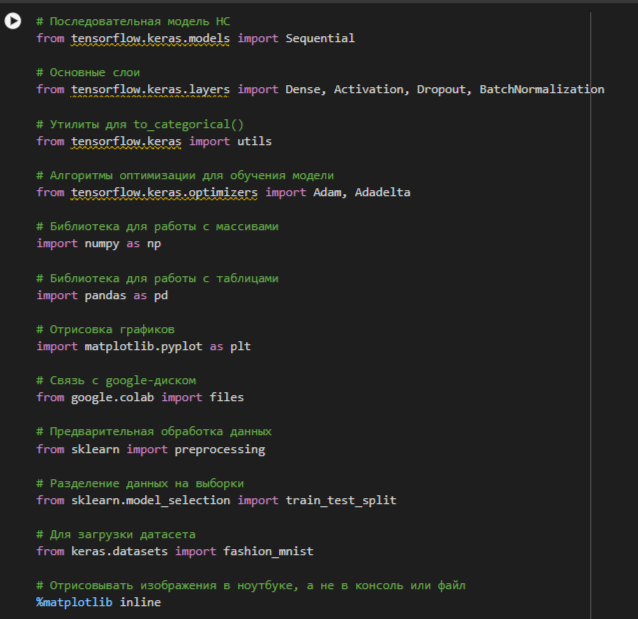


Рисунок 1.1 – Подключение библиотек

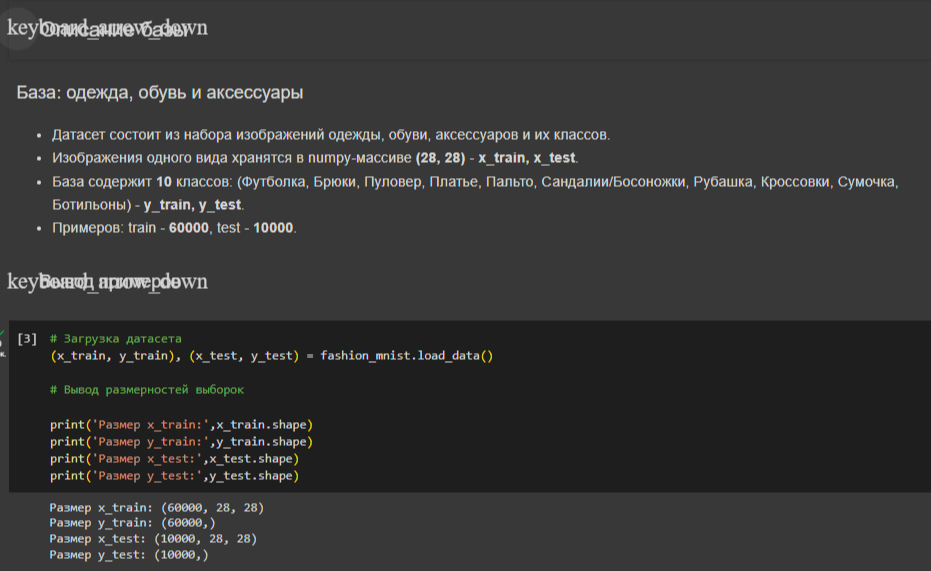


Рисунок 1.2 – Загрузка изображений

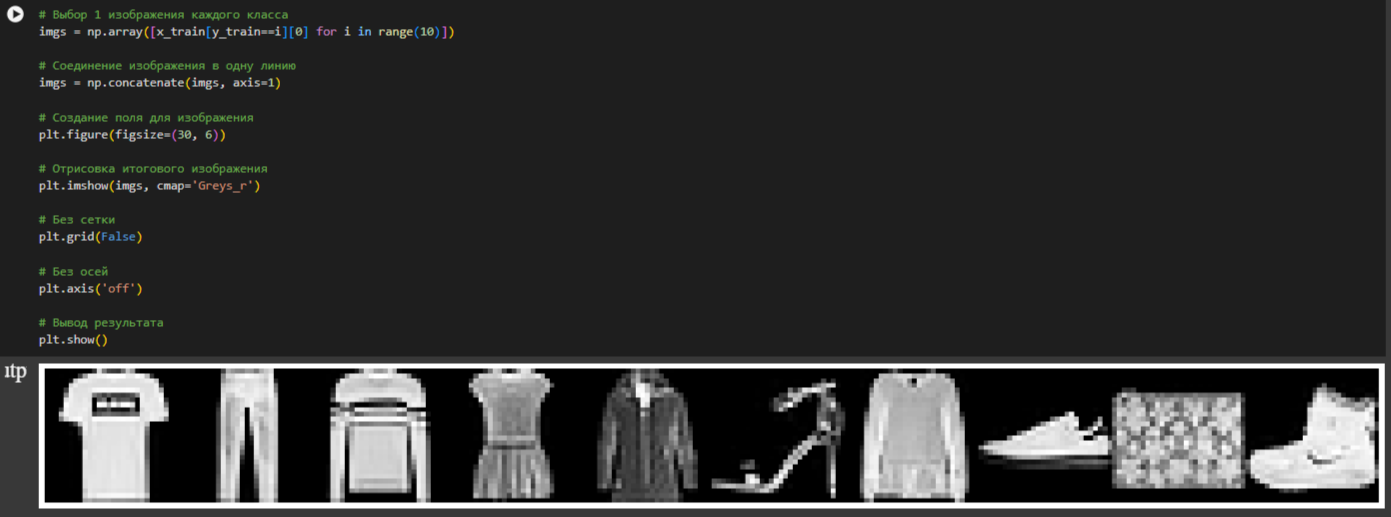


Рисунок 1.3 – Проверка содержимого массива изображений

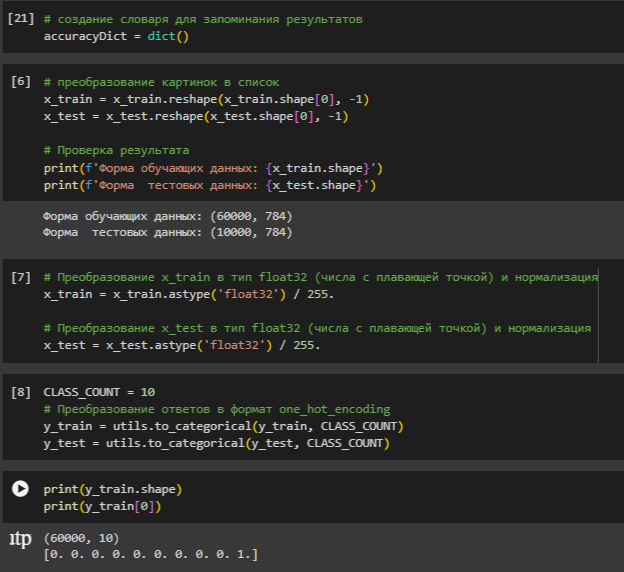


Рисунок 1.4 – Разделение и преобразование данных

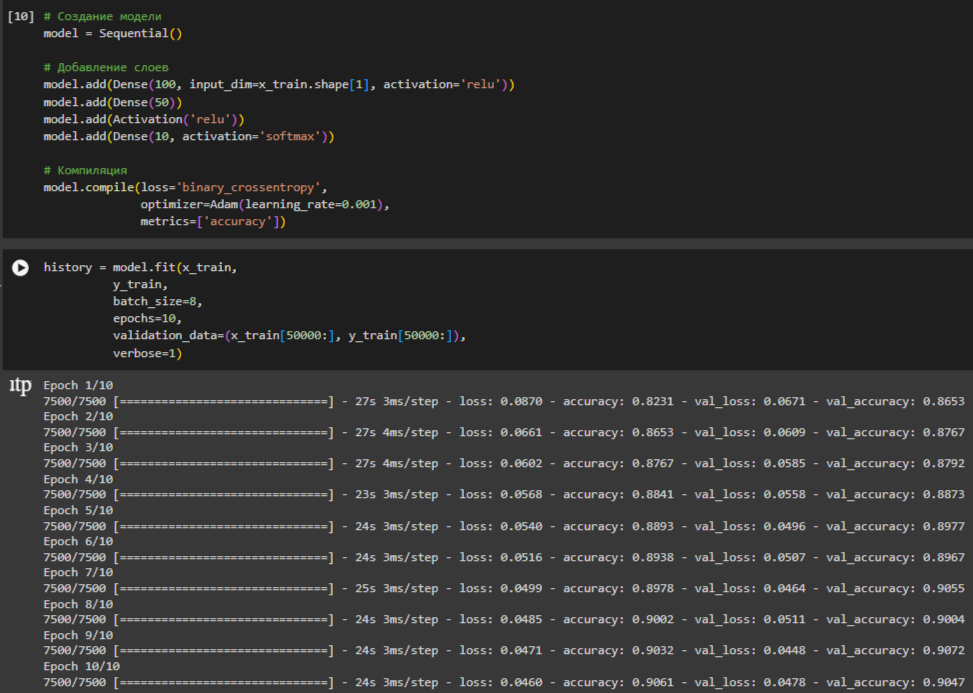


Рисунок 1.5 – Создание и обучение модели

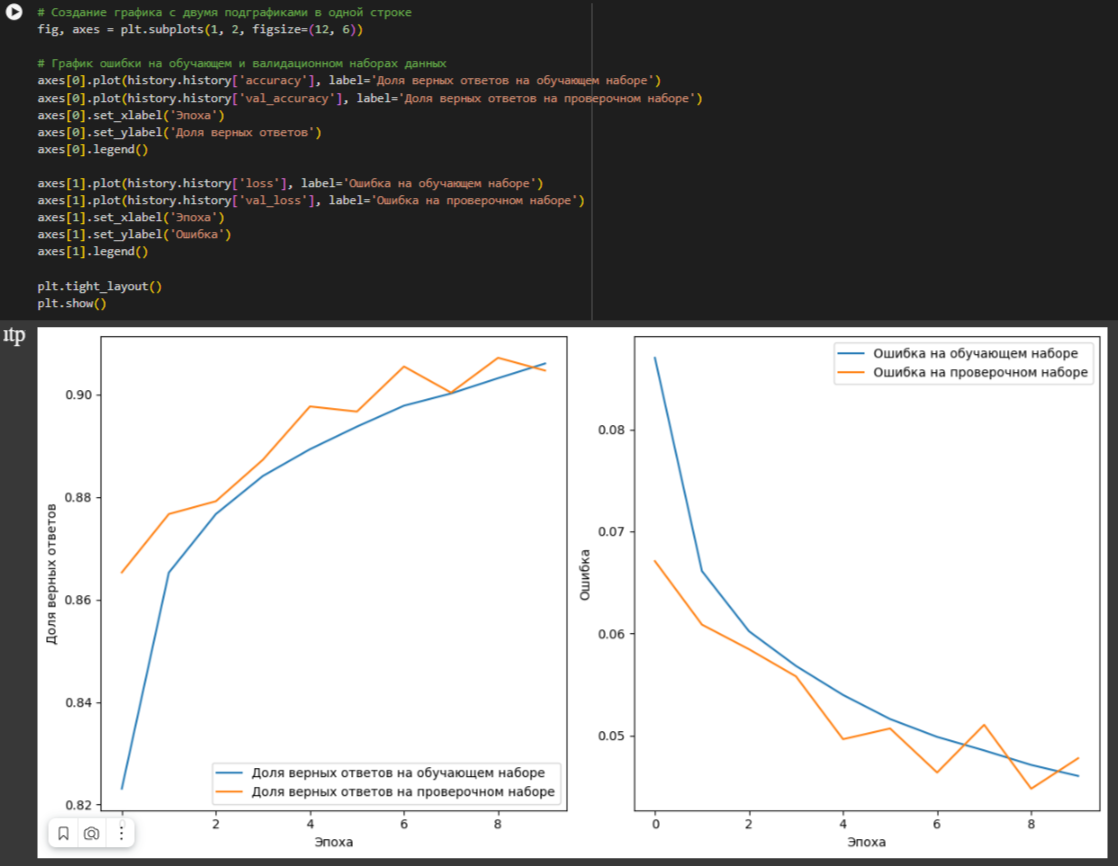


Рисунок 1.6 – Построение графика

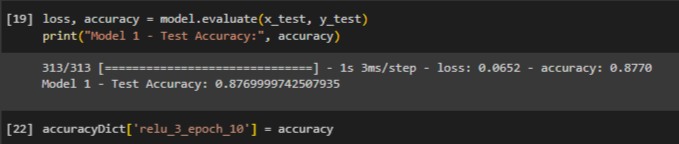


Рисунок 1.7 – Определение точности модели на тестовой выборке

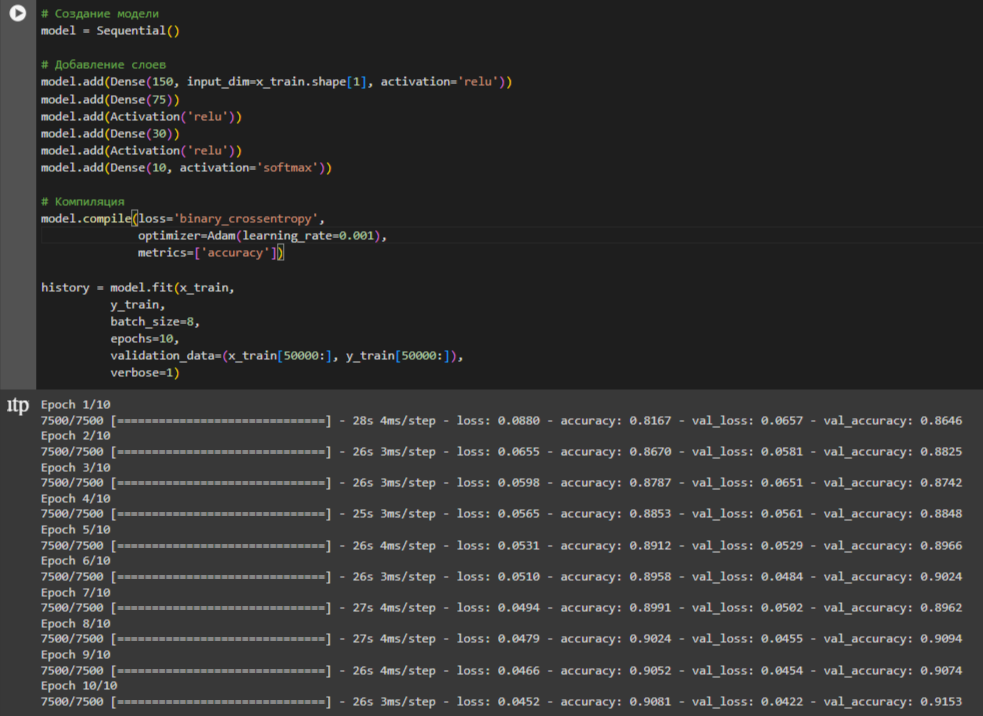


Рисунок 1.8 – Создание и обучение модели

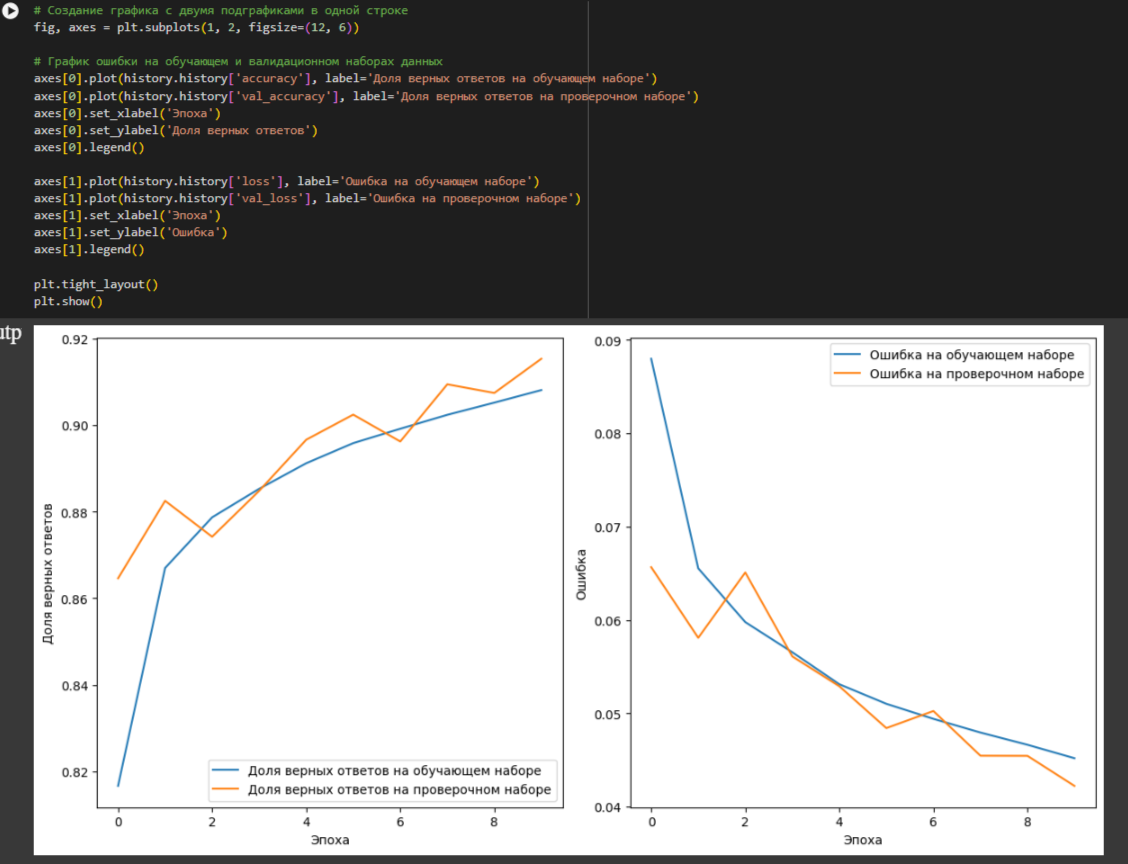


Рисунок 1.9 – Построение графика

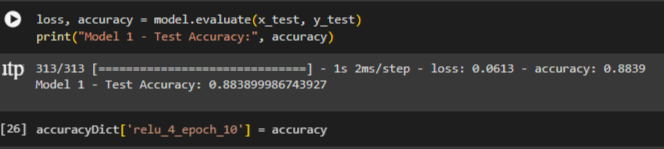


Рисунок 1.10 – Определение точности модели на тестовой выборке

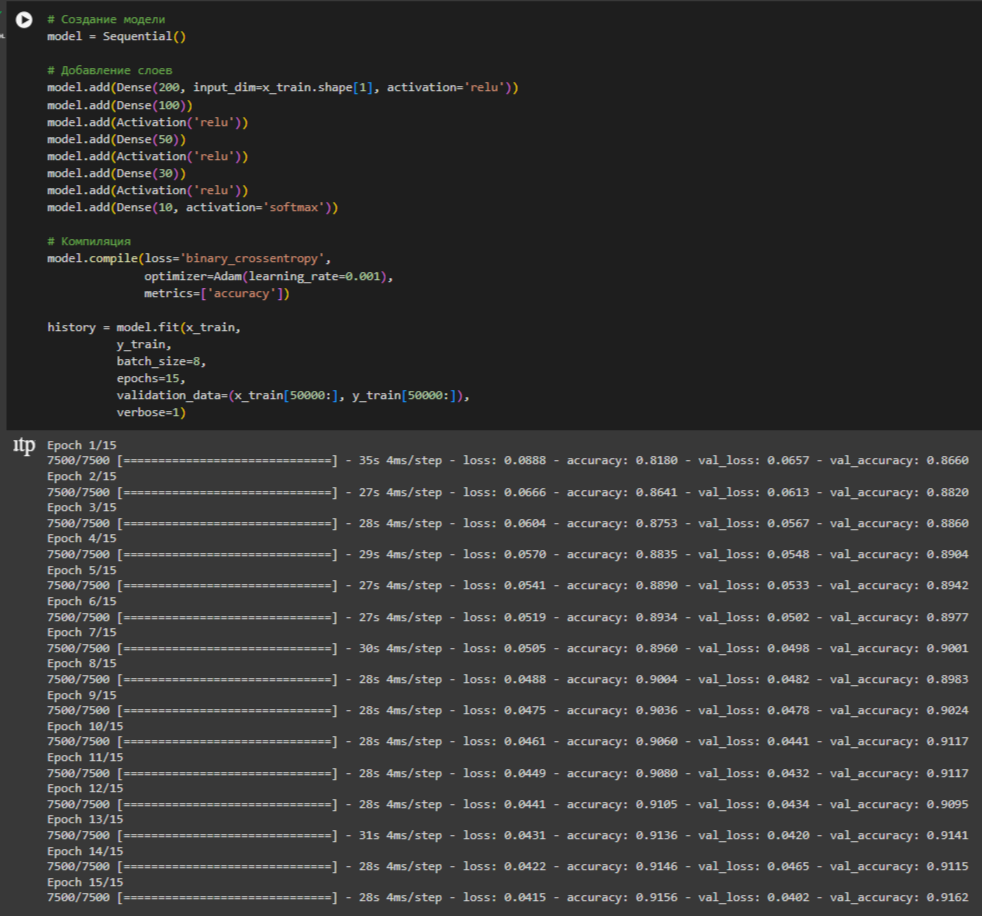


Рисунок 1.11 – Создание и обучение модели

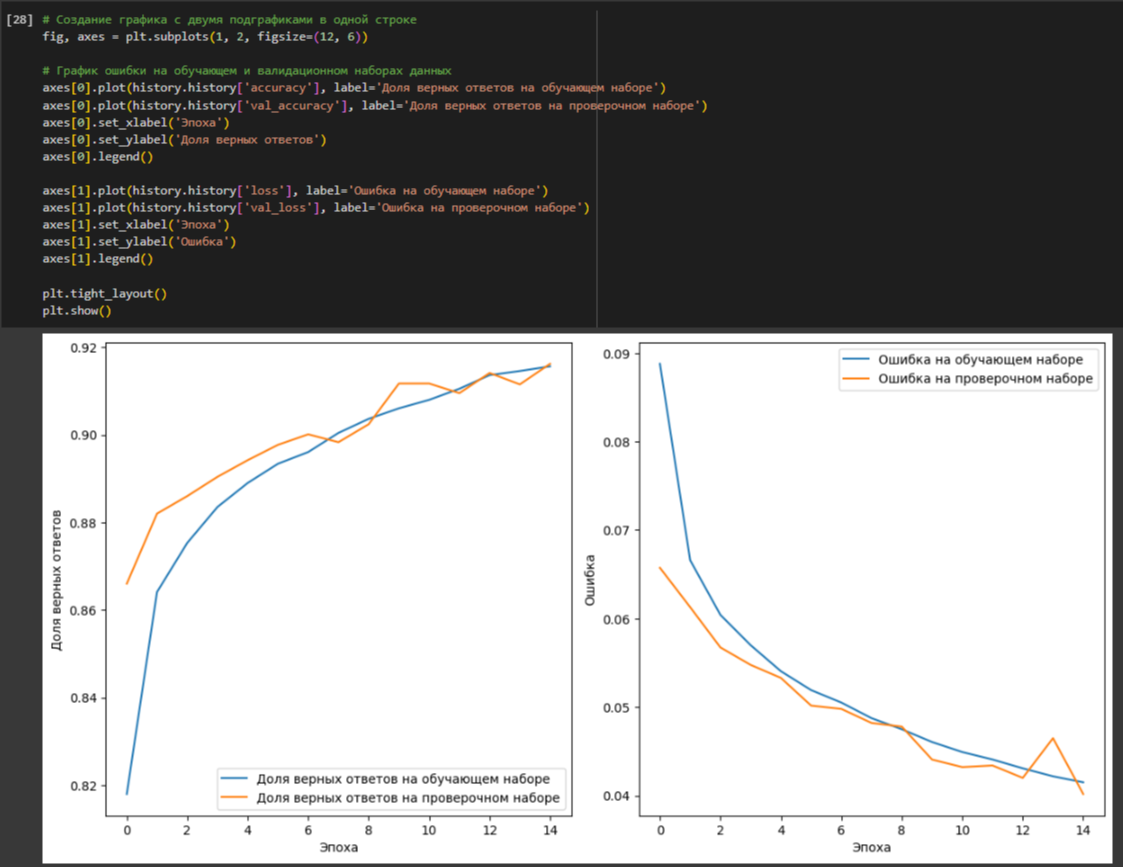


Рисунок 1.12 – Построение графика

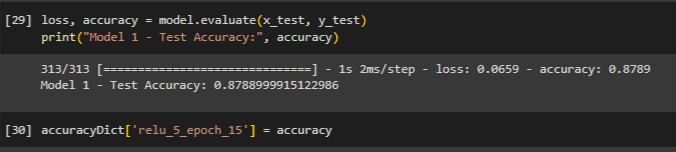


Рисунок 1.13 – Определение точности модели на тестовой выборке

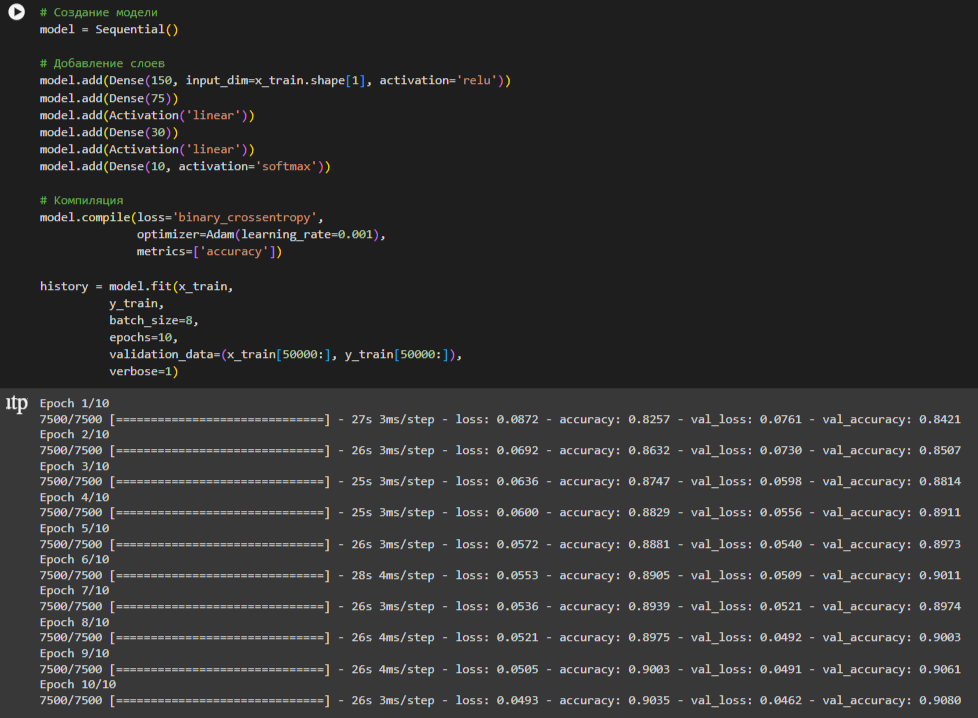


Рисунок 1.14 – Создание и обучение модели

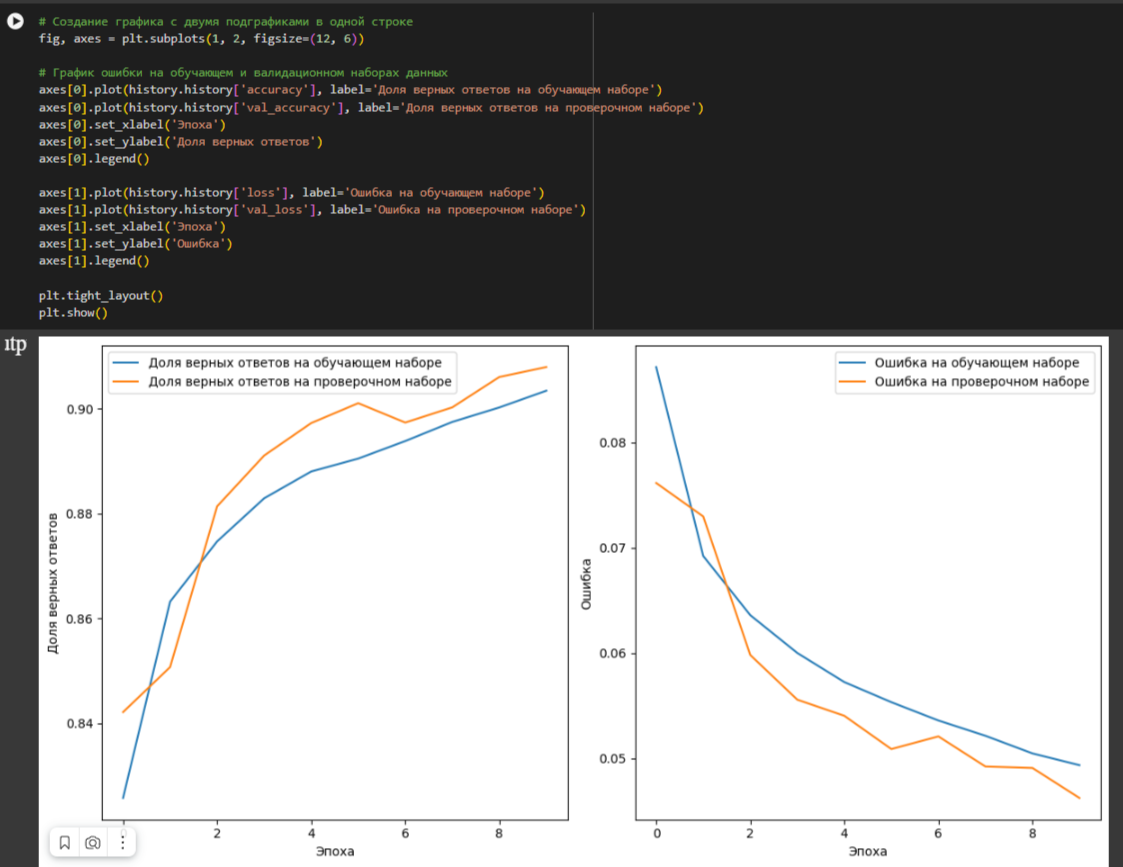


Рисунок 1.15 – Построение графика

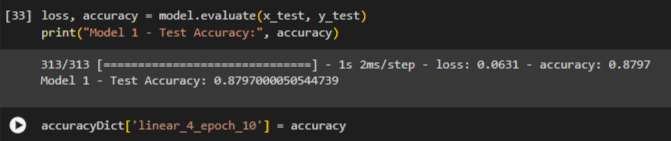


Рисунок 1.16 – Определение точности модели на тестовой выборке

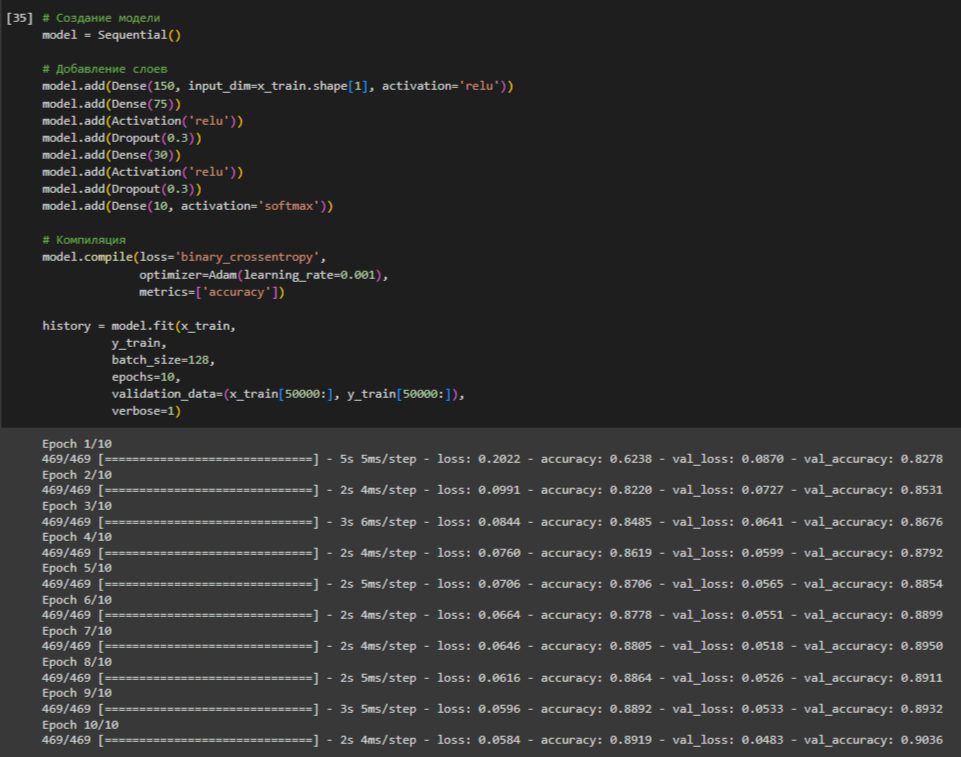


Рисунок 1.17 – Создание и обучение модели

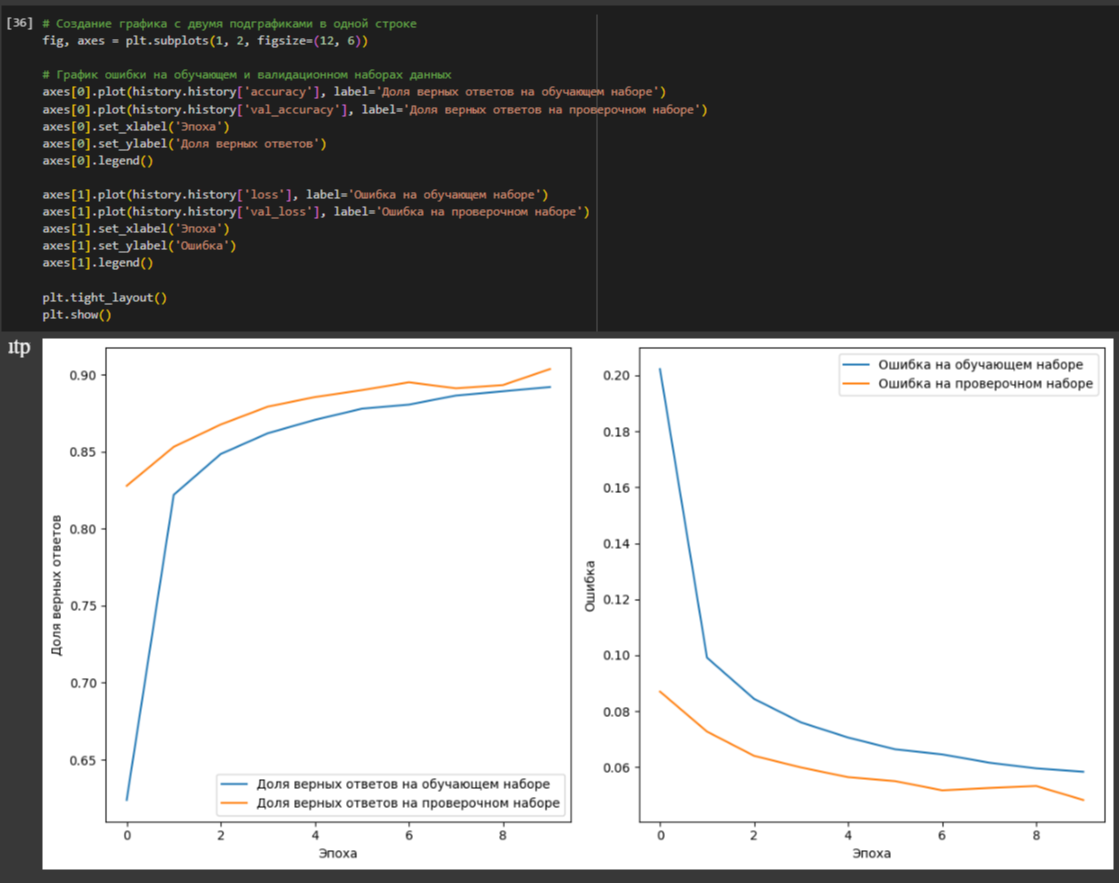


Рисунок 1.18 – Построение графика

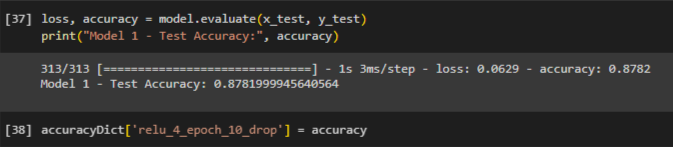


Рисунок 1.19 – Определение точности модели на тестовой выборке

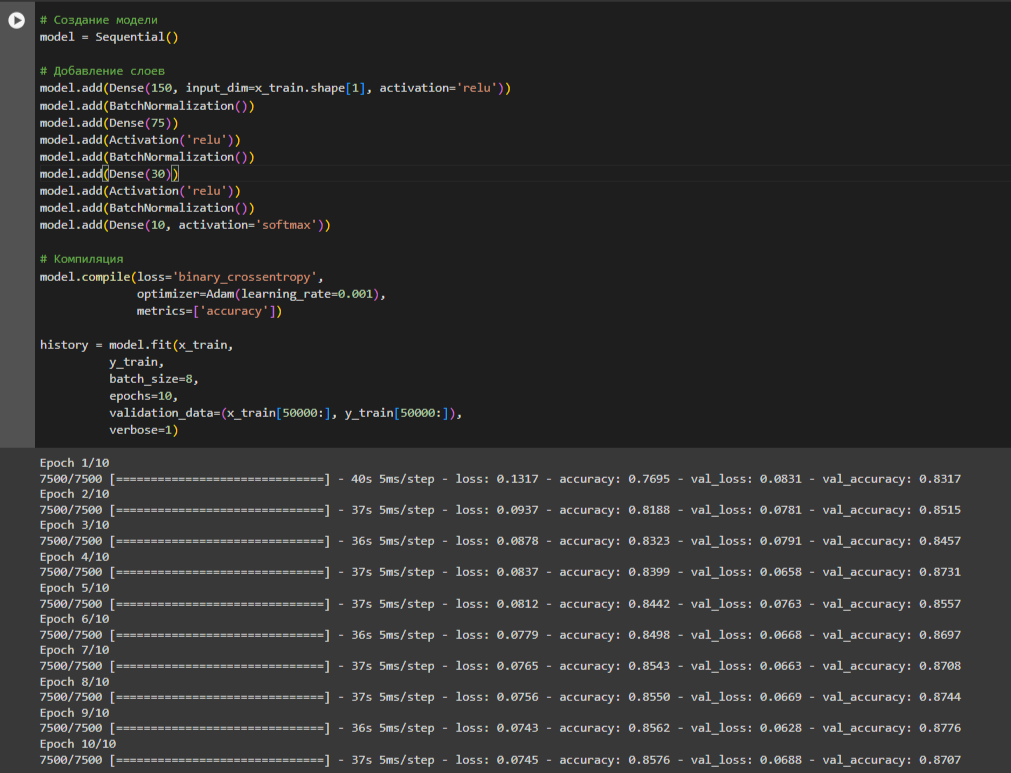


Рисунок 1.20 – Создание и обучение модели

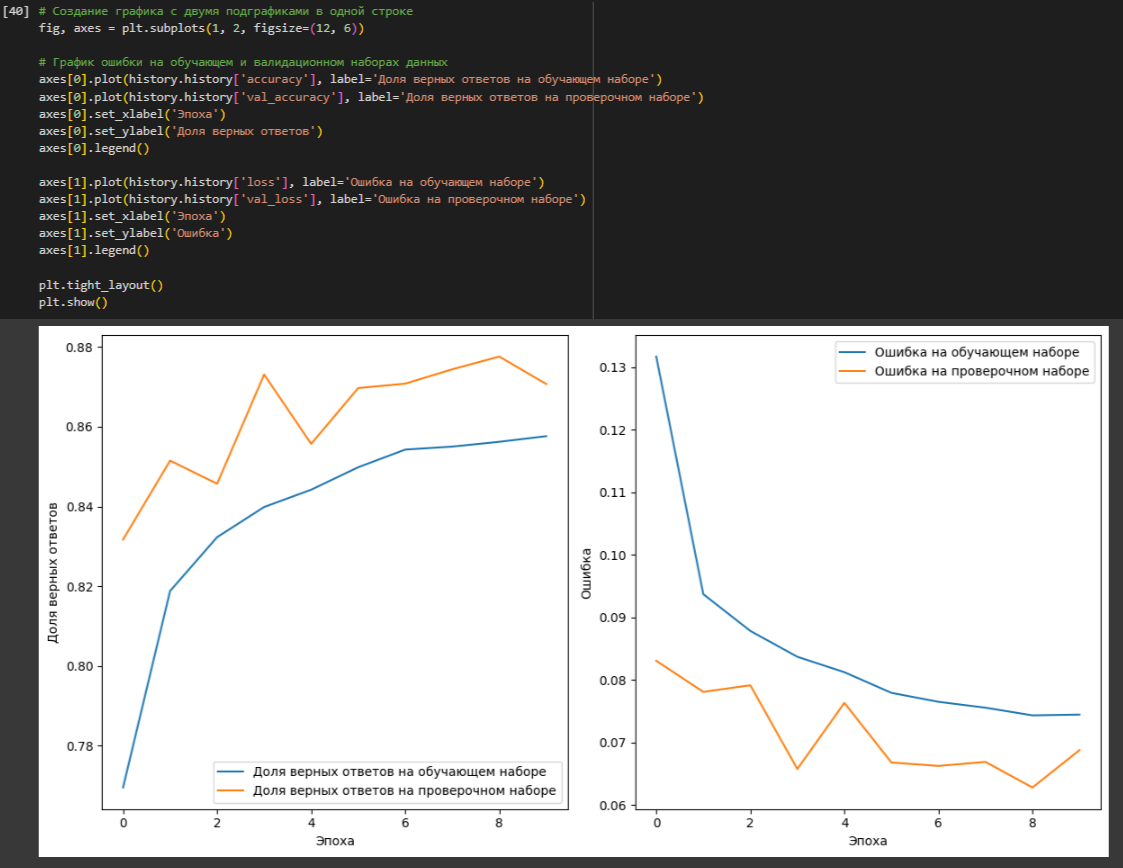


Рисунок 1.21 – Построение графика

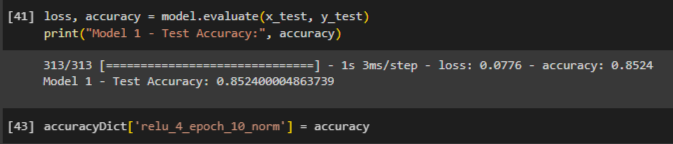


Рисунок 1.22 – Определение точности модели на тестовой выборке

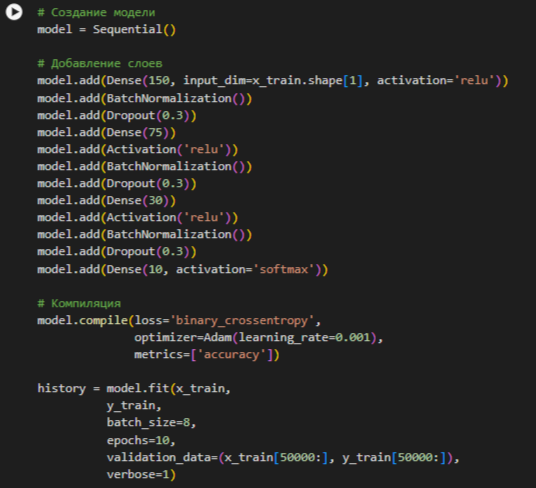


Рисунок 1.23 – Создание и обучение модели

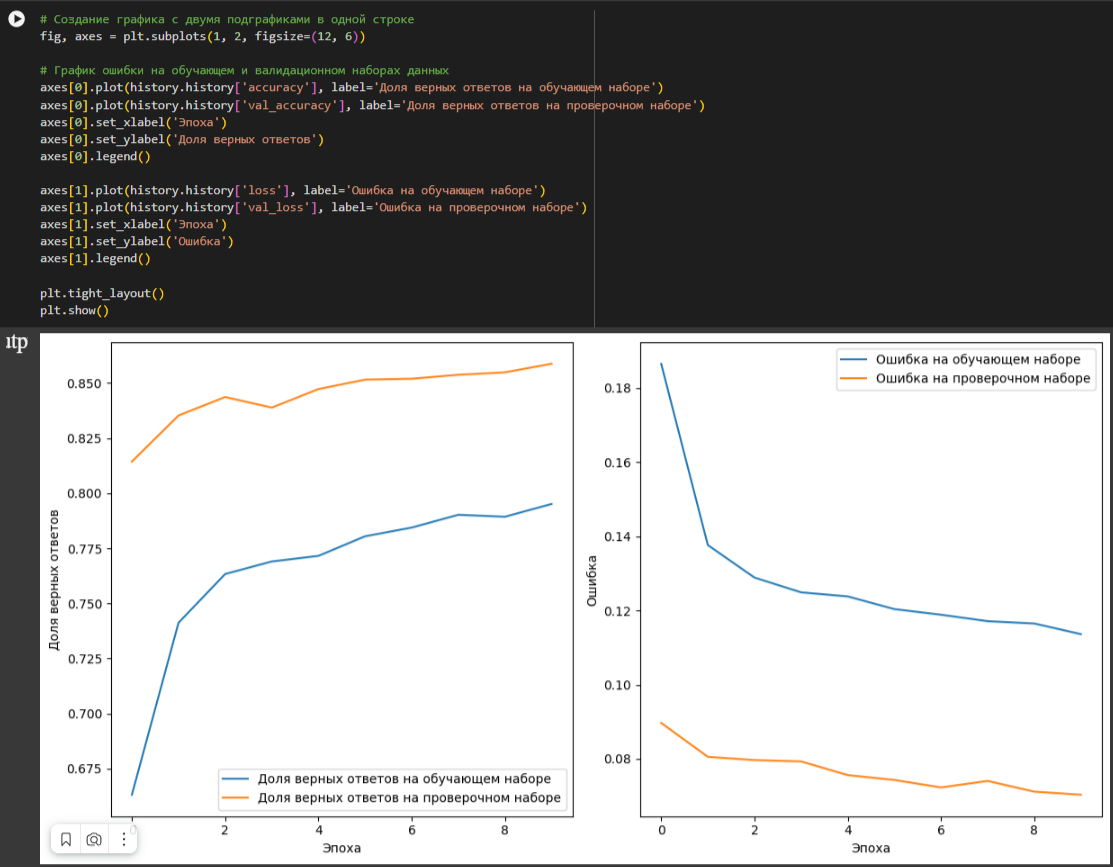


Рисунок 1.24 – Построение графика

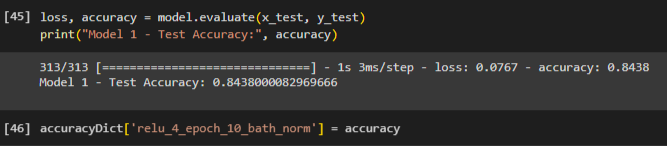


Рисунок 1.25 – Определение точности модели на тестовой выборке

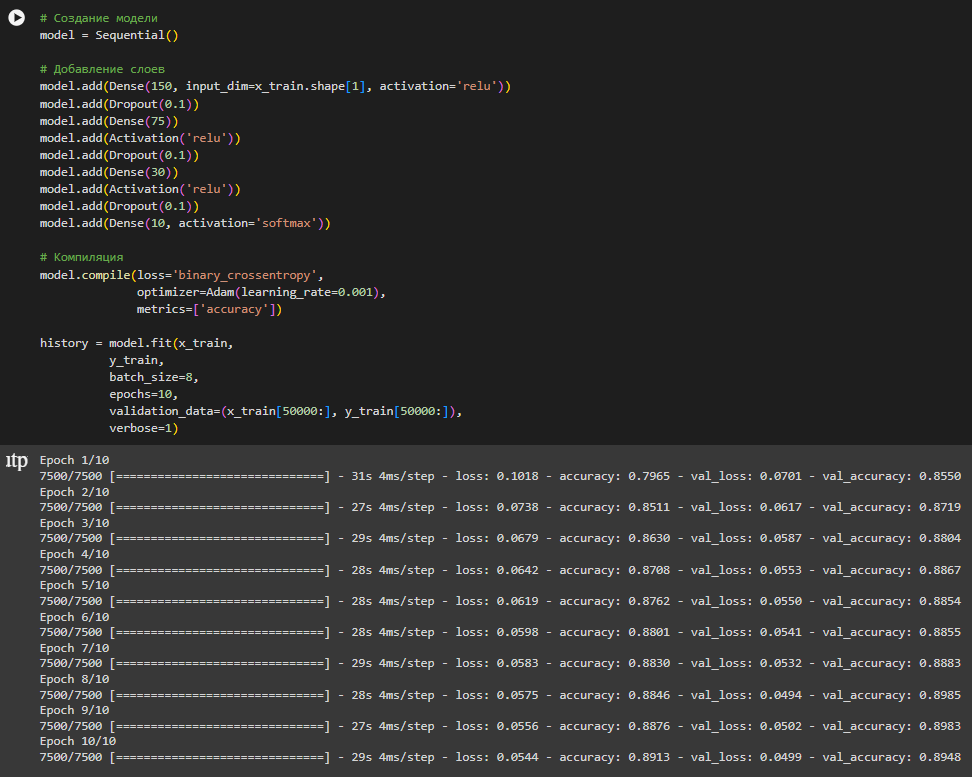


Рисунок 1.26 – Создание и обучение модели

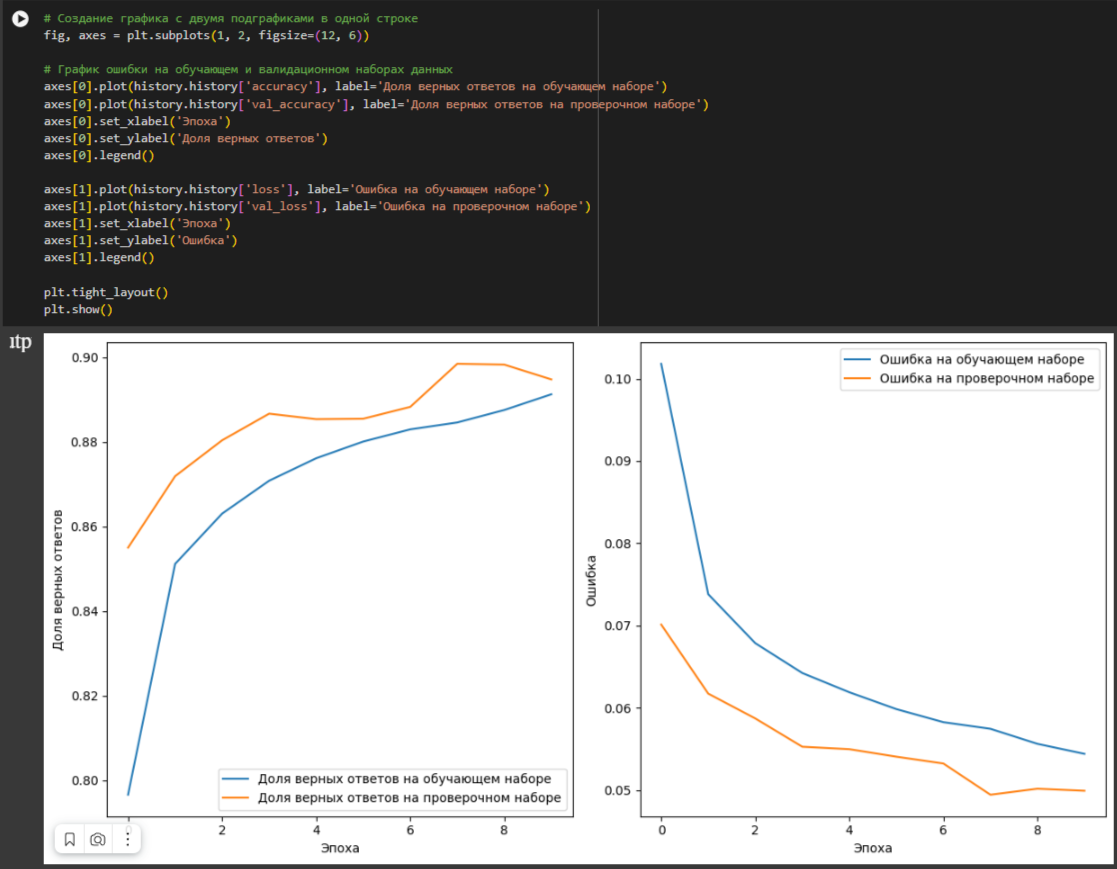


Рисунок 1.27 – Построение графика

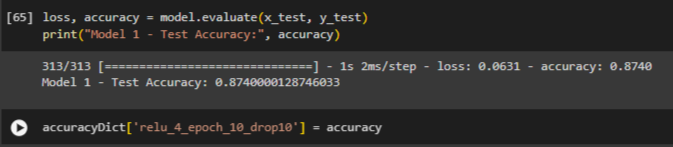


Рисунок 1.28 – Определение точности модели на тестовой выборке

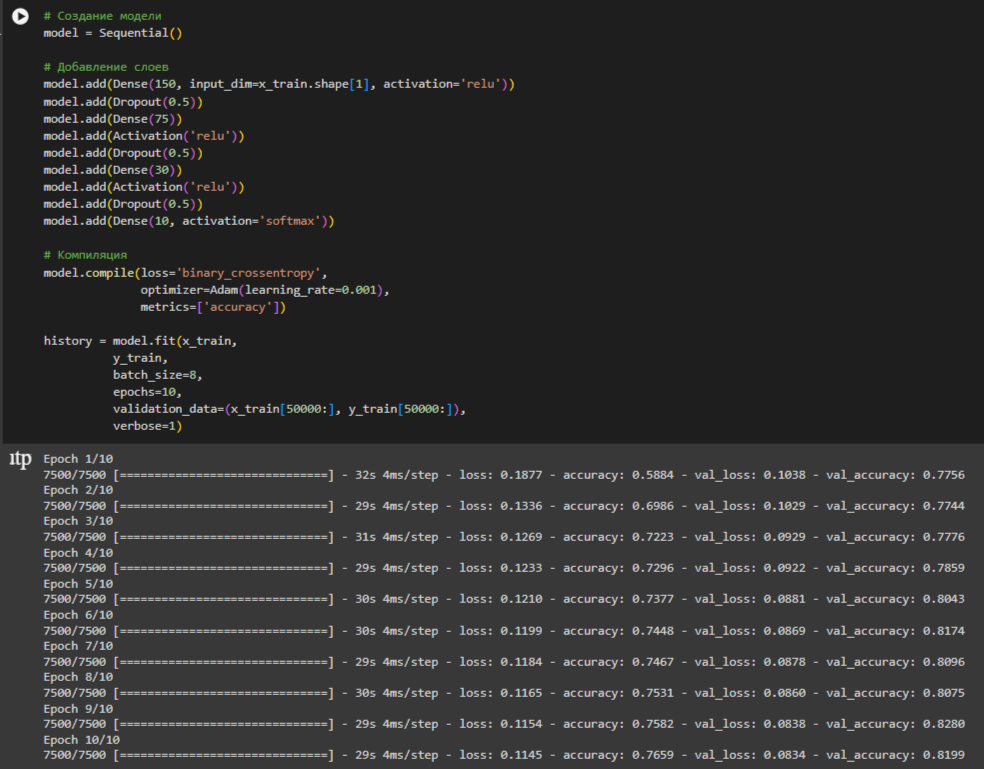


Рисунок 1.29 – Создание и обучение модели

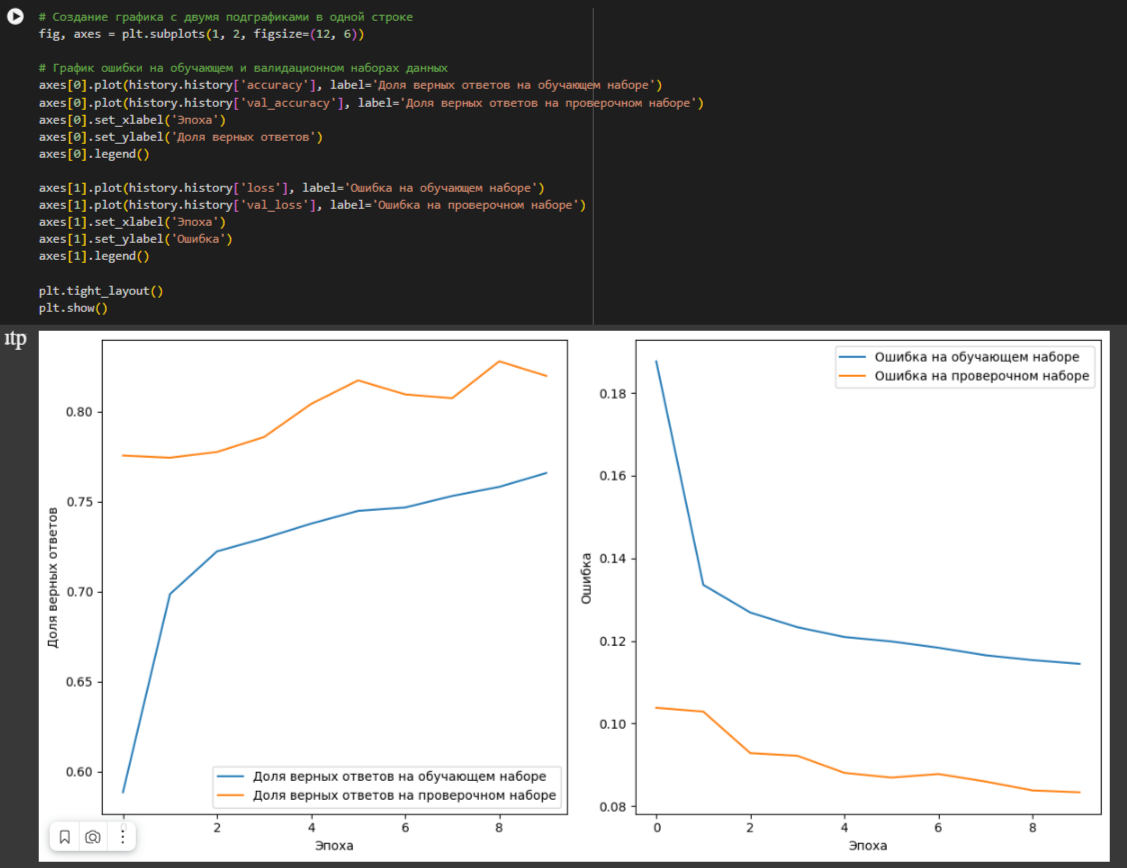


Рисунок 1.30 – Построение графика

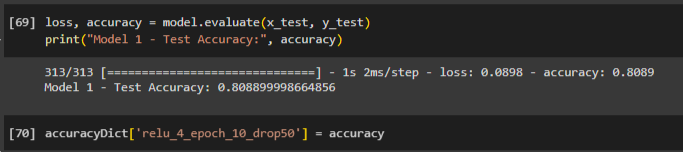


Рисунок 1.31 – Определение точности модели на тестовой выборке

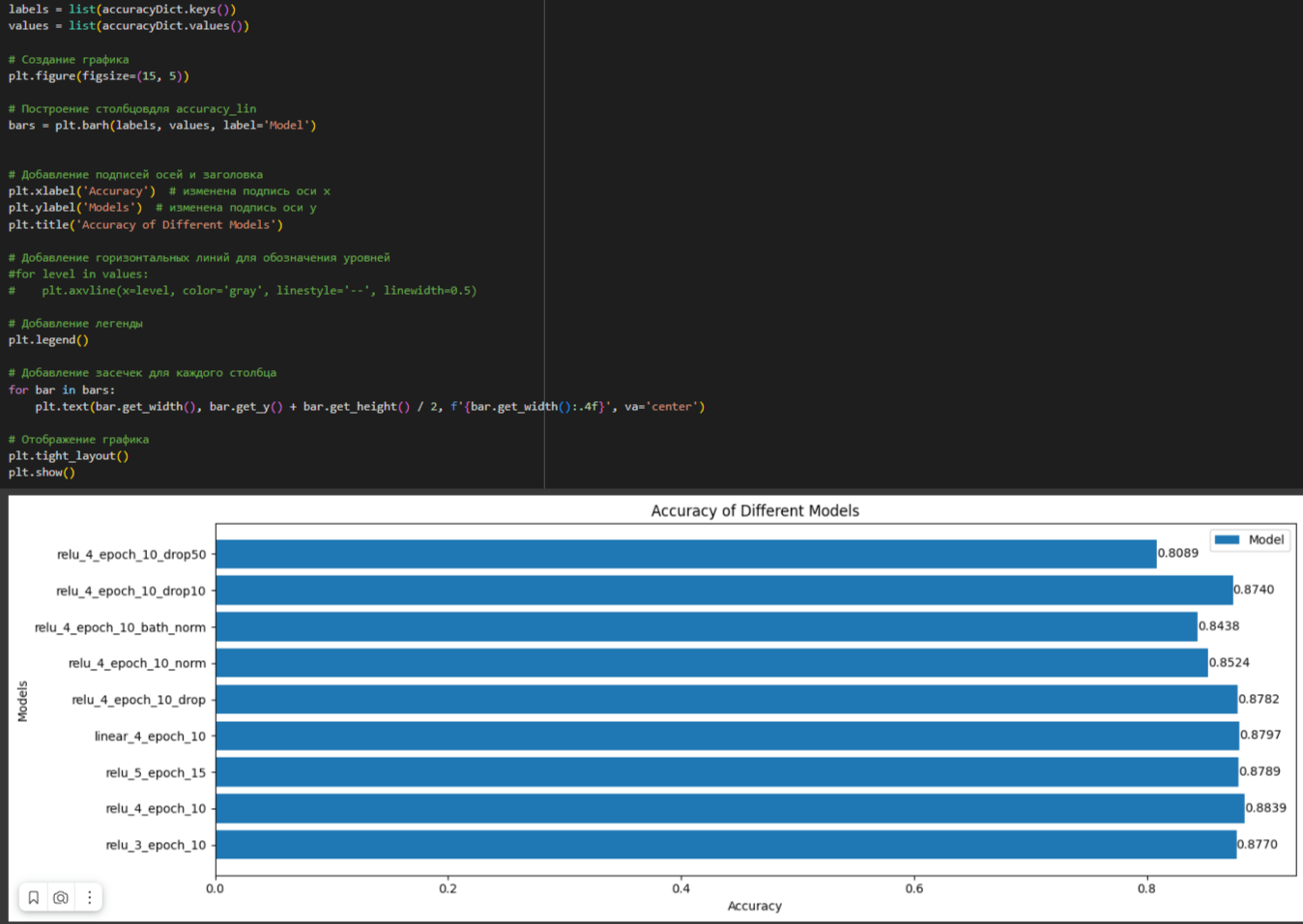


Рисунок 1.32 – Построение графика

**Задача 2.** Используя модуль datasets библиотеки sklearn, загрузите базу вин (.load\_wine()).

Используя шаблон ноутбука, выполните загрузку, подготовку и предобработку данных. Обязательное условие: разделение данных на три выборки осуществляется по шаблону (изменять параметры подготовки данных запрещается).

Проведите серию экспериментов и добейтесь максимальной точности классификации на тестовой выборке выше 94%. С помощью метода .summary() зафиксируйте количество параметров созданной вами нейронной сети.

**Решение.**

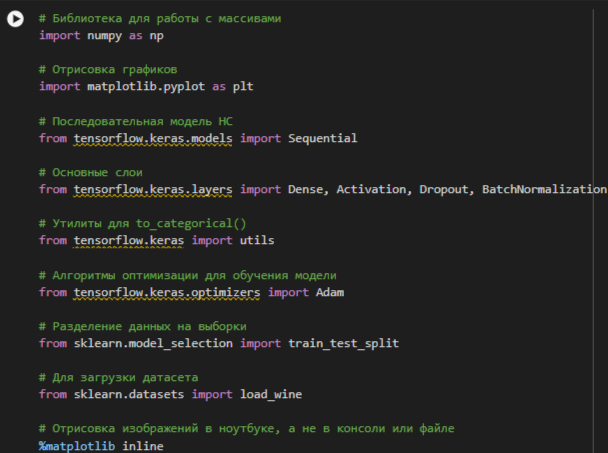


Рисунок 2.1 – Подключение библиотек

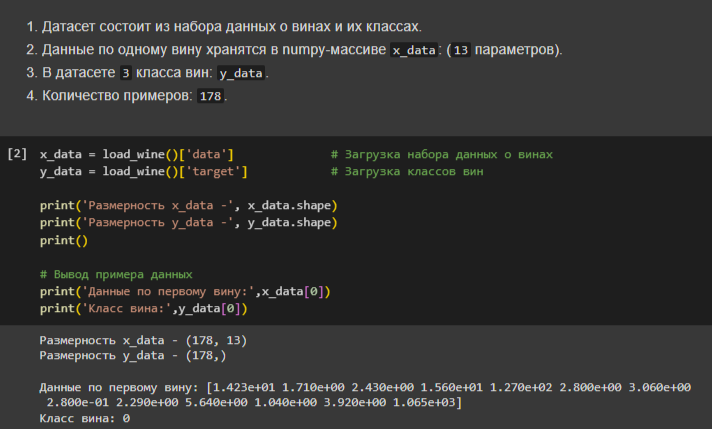


Рисунок 2.2 – Загрузка данных

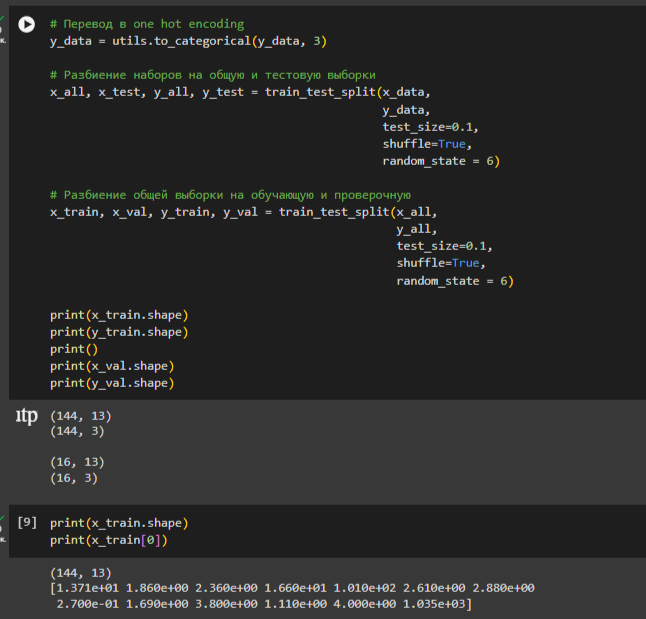


Рисунок 2.3 – Распределение и преобразование данных

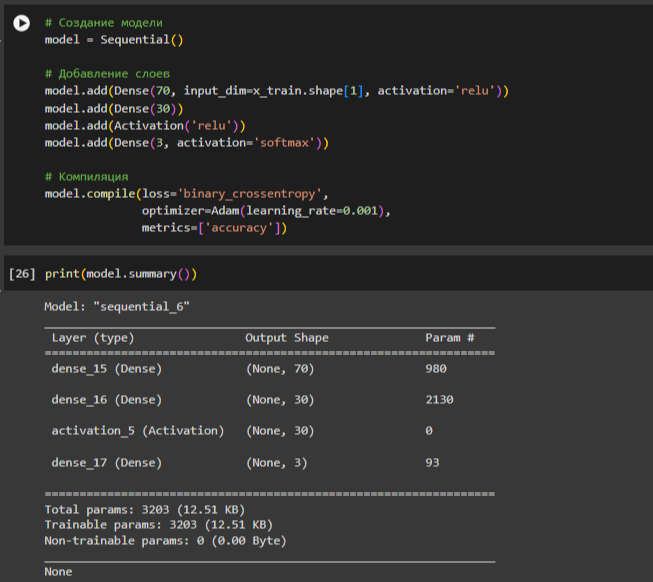


Рисунок 2.5 – Создание и обучение модели

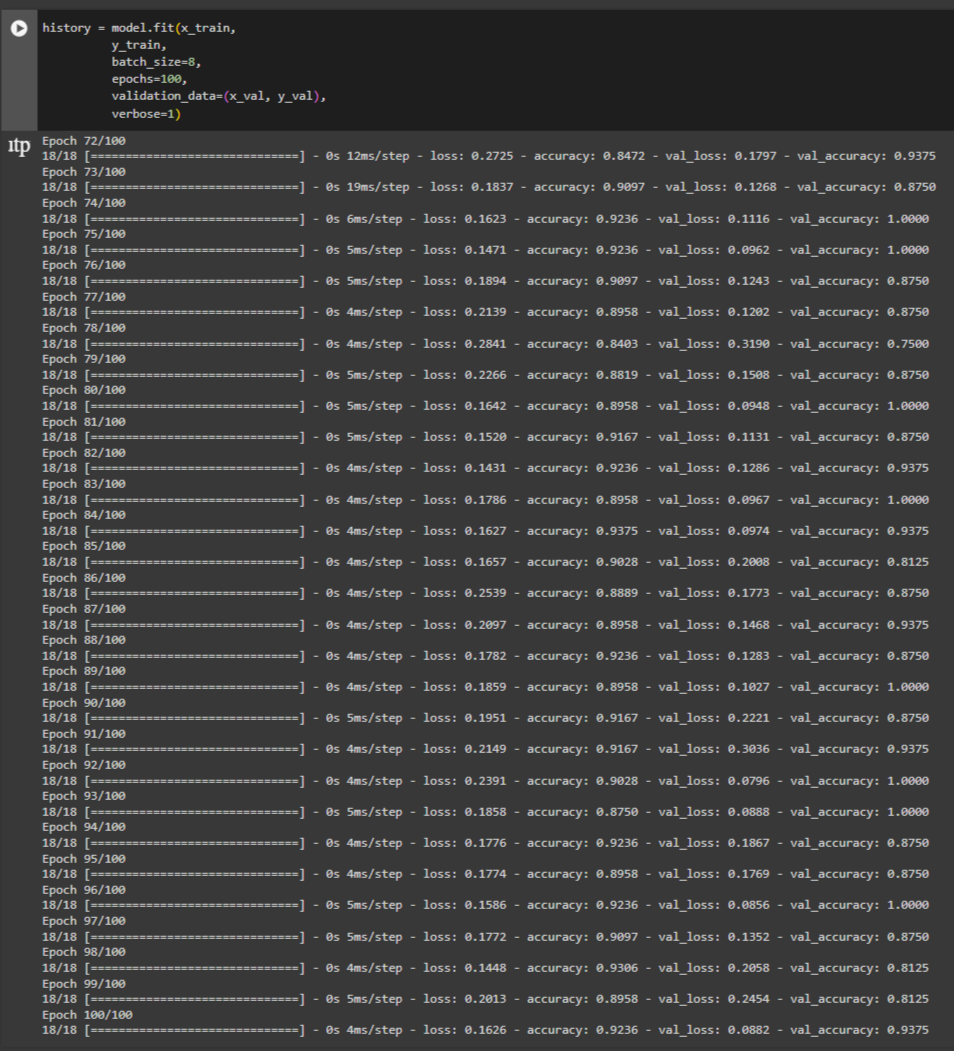


Рисунок 2.6 – Обучение модели

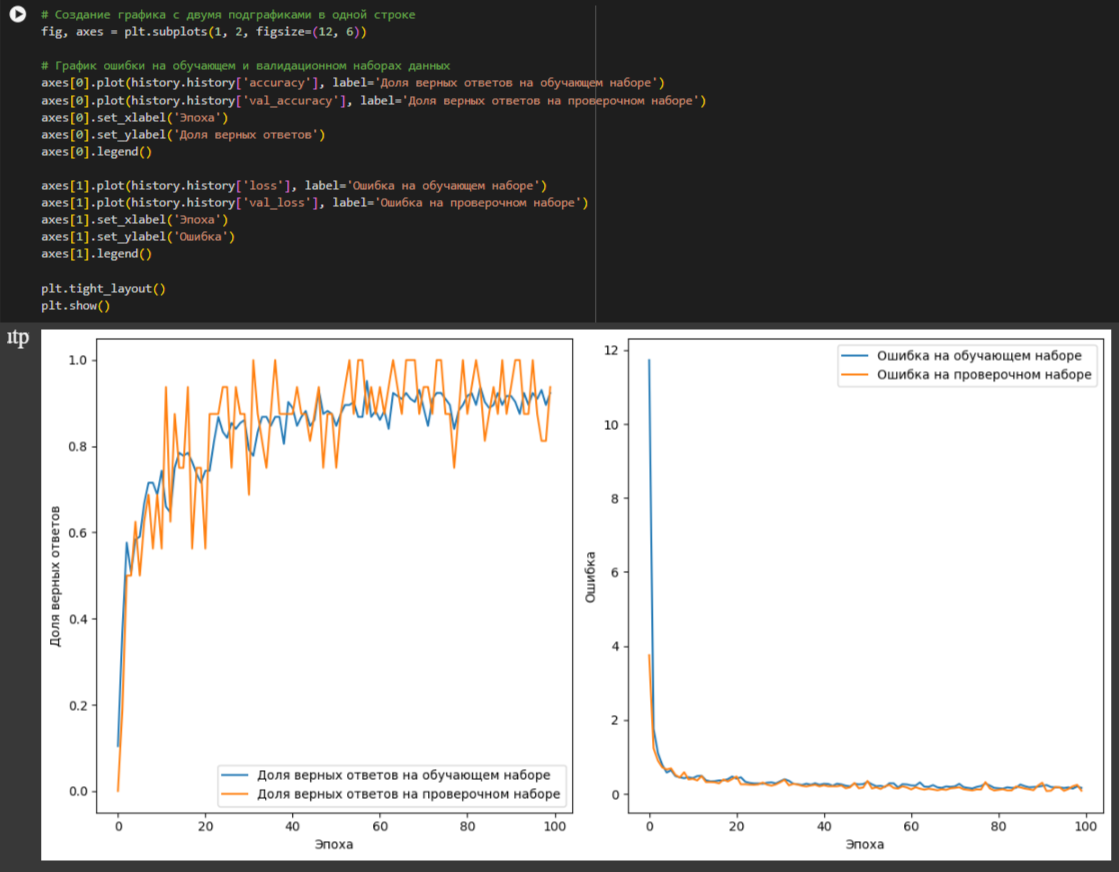


Рисунок 2.7 – Вывод графиков точности и ошибки

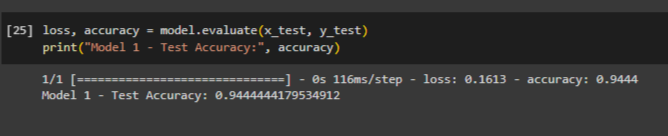


Рисунок 2.7 – Вывод точности и ошибки на тестовой выборке

**Задача 3.** Используя базу "Пассажиры автобуса", подготовьте данные для обучения нейронной сети, классифицирующей изображение на два класса:

* входящий пассажир
* выходящий пассажир

Добейтесь точности работы модели на проверочной выборке не ниже 85%. Ссылка на датасет:

https://storage.yandexcloud.net/aiueducation/Content/base/l4/bus.zip

**Решение.**

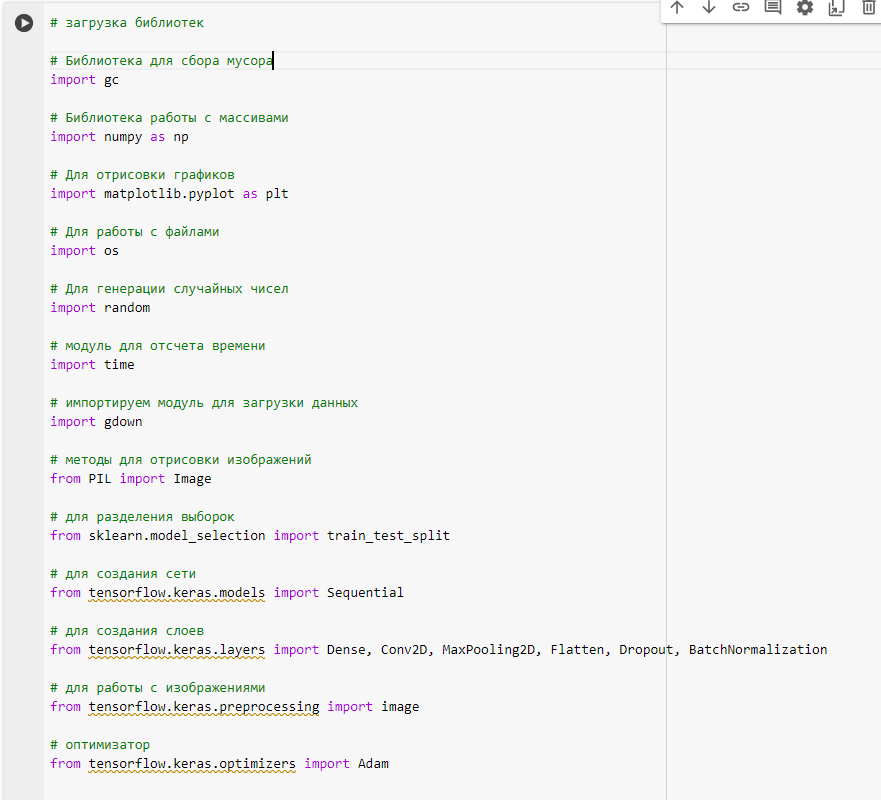


Рисунок 3.1 – Подключение библиотек

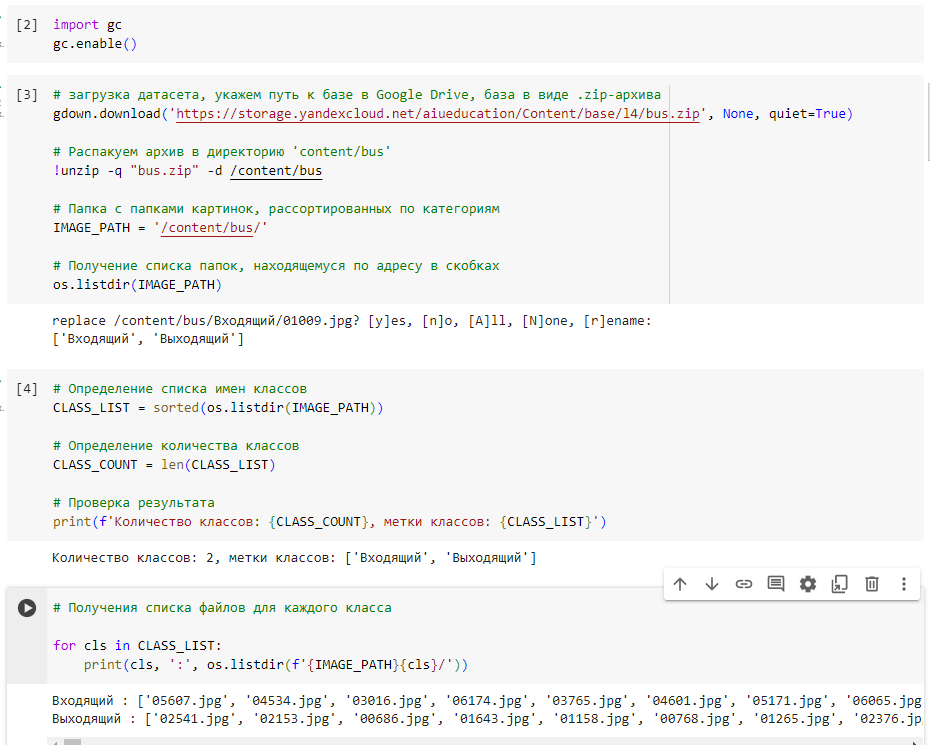


Рисунок 3.2 – Загрузка данных для обучения



Рисунок 3.2 – Вывод полученных данных

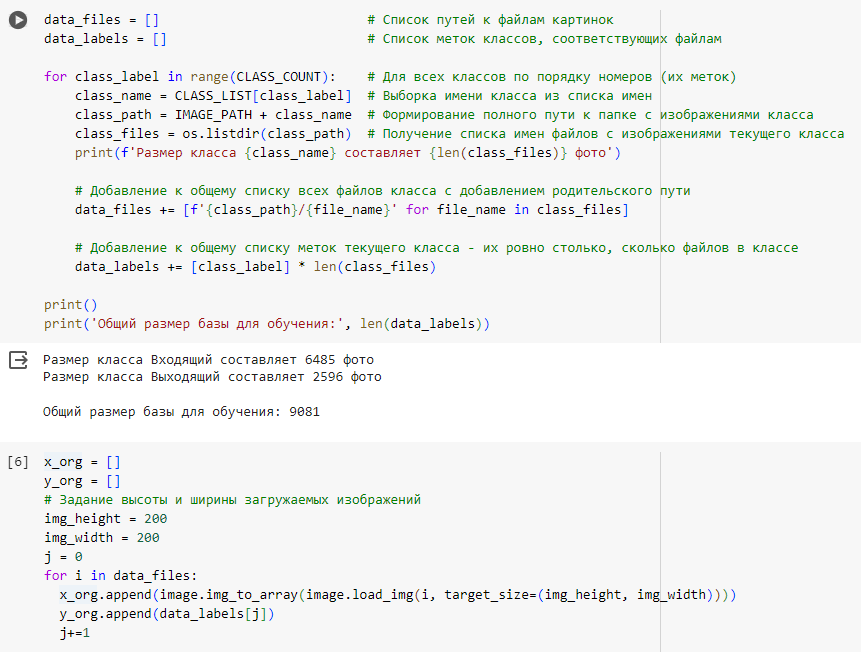
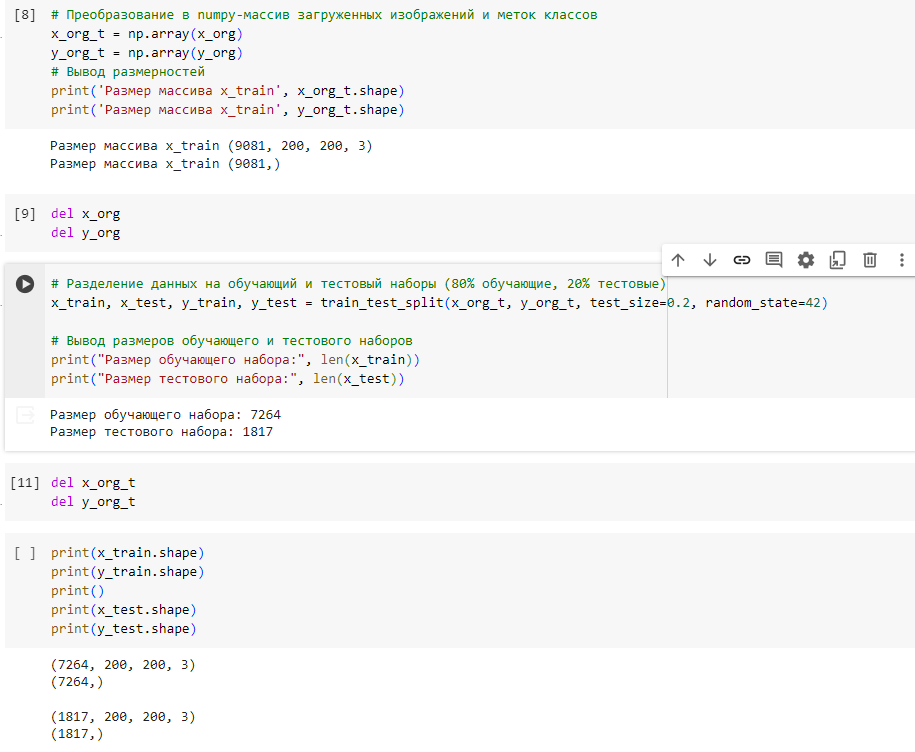


Рисунок 3.3 – Загрузка данных



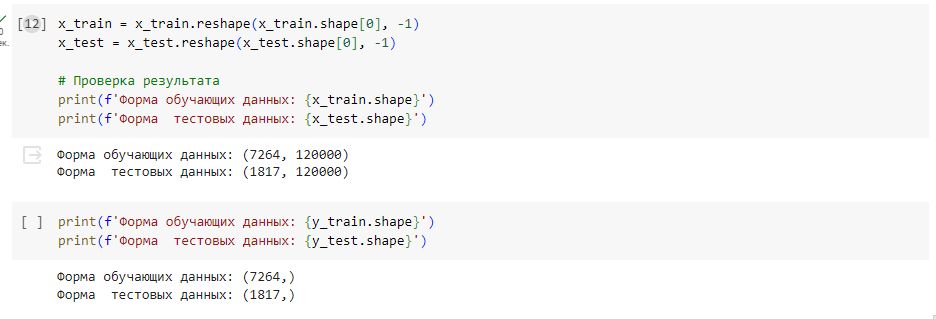


Рисунок 3.4 – Преобразование и разделение данных

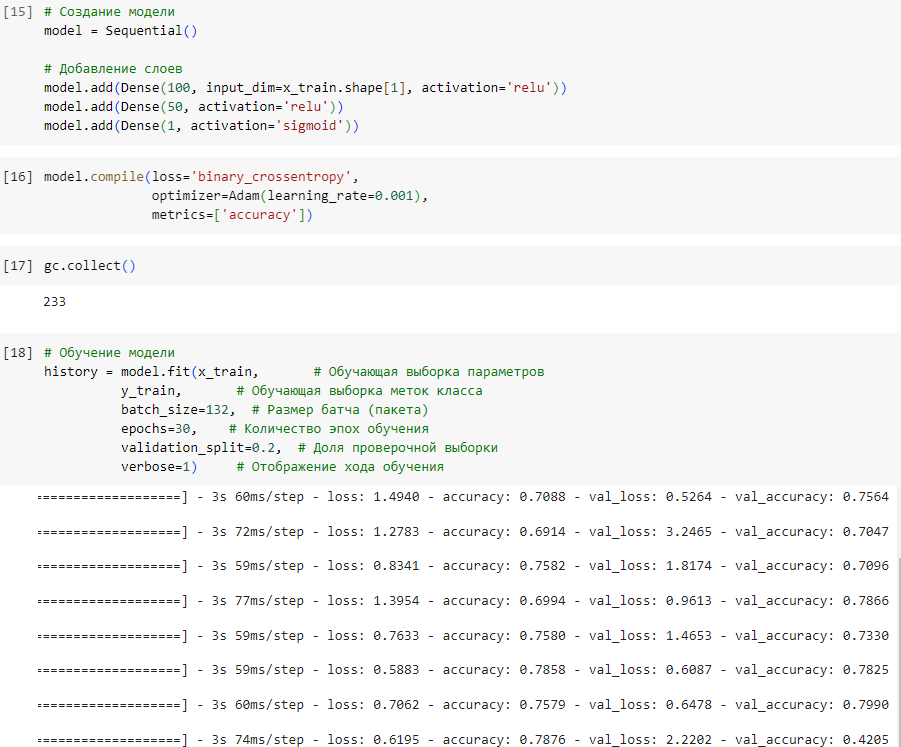


Рисунок 3.5 – Создание и обучение модели



Рисунок 3.6 – Визуализация точности

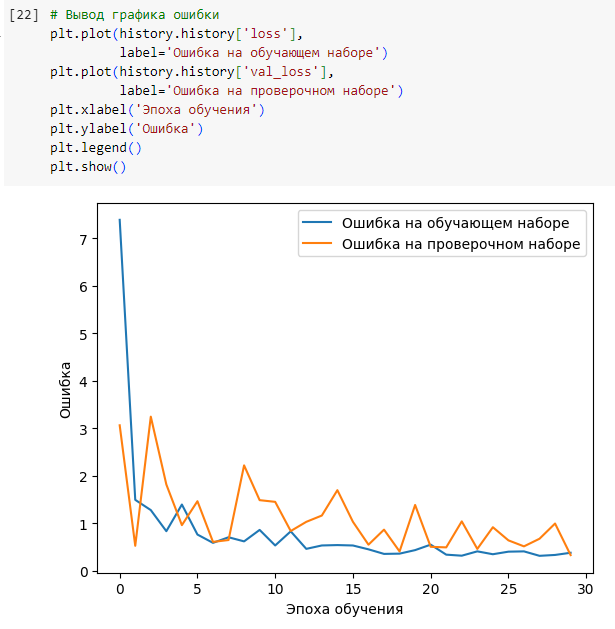


Рисунок 3.7 – Визуализация ошибки

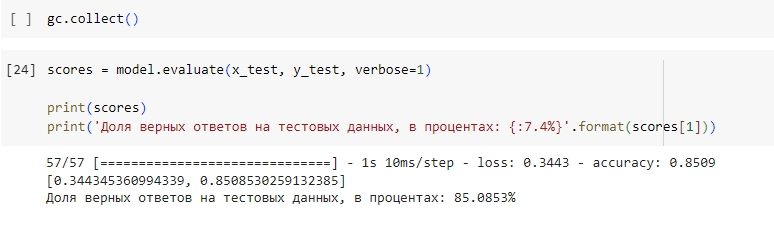


Рисунок 3.8 – Проверка точности на тестовой выборке

**Выводы.**  В ходе выполнения лабораторной работы была найдена наиболее точная архитектура нейронной сети, путем изменения модели, а именно изменением числа нейронов, количества слоев, в том числе использования слоев Dropout и BatchNormalization. При анализе результатов, можно сделать вывод, что наиболее точной оказалась модель с функцией активации ReLu, 4-мя слоями без слоев Dropout и BatchNormalization.

Было выполнено разбиение данных на три выборки согласно шаблону, проведены серии экспериментов, которые позволили добиться точности свыше 94%.

Используя подготовленную базу изображений, была создана и обучена нейронная сеть, распознающая две категории изображений: входящие и выходящие пассажиры, с точностью распознавания более 85%.