|  |  |
| --- | --- |
|  | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ Информатика и системы управления \_

КАФЕДРА Системы обработки информации и управления \_

**ОТЧЕТ ПО УЧЕБНОЙ ПРАКТИКЕ**

Студент Попов Илья Андреевич\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*фамилия, имя, отчество*

Группа ИУ5-81Б \_

Тип практики Преддипломная \_

Название предприятия МГТУ им. Н.Э.Баумана\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент **\_\_\_\_\_\_\_\_28.05.2021 \_\_Попов И. А.\_\_\_\_**

*подпись, дата фамилия, и.о.*

Научный руководитель **\_\_\_\_\_\_\_\_28.05.2021 \_\_Сёмкин А.П.\_\_\_**

*подпись, дата фамилия, и.о.*

Руководитель практики **\_\_\_\_\_\_\_\_28.05.2021 \_\_Балдин А.В.\_\_\_**

*подпись, дата фамилия, и.о.*

Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_

*2021 г.*

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное

учреждение высшего образования

«Московский государственный технический университет имени Н.Э. Баумана

(национальный исследовательский университет)»

(МГТУ им. Н.Э. Баумана)

Кафедра «Системы обработки информации и управления» (ИУ5)

**ЗАДАНИЕ**

**на прохождение производственной практики**

на предприятии \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_НУК ИУ МГТУ им. Н.Э.Баумана \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент Попов Илья Андреевич, ИУ5-81Б

(фамилия, имя, отчество; инициалы; индекс группы)

Во время прохождения преддипломной практики студент должен:

1. Исследовать возможность определения местоположения на территории главном здании МГТУ им. Баумана с помощью анализа Wi-Fi сетей.

2. Разработать модель, с помощью которой будет возможно определить местоположение на основе Wi-Fi сетей.

Дата выдачи задания « \_7 \_ » \_\_\_\_мая\_\_\_\_ 2021 г.

Студент **\_\_\_\_\_\_\_\_28.05.2021 \_\_Попов И. А.\_\_\_\_**

*подпись, дата фамилия, и.о.*

Научный руководитель **\_\_\_\_\_\_\_\_28.05.2021 \_\_Сёмкин А.П.\_\_\_**

*подпись, дата фамилия, и.о.*

Руководитель практики **\_\_\_\_\_\_\_\_28.05.2021 \_\_Балдин А.В.\_\_\_**

*подпись, дата фамилия, и.о.*

**Оглавление**

[Введение 4](#_Toc73102953)

[Основная часть 5](#_Toc73102954)

[Разработка приложения для сбора данных 5](#_Toc73102955)

[Обработка данных 7](#_Toc73102956)

[Определение местоположения 8](#_Toc73102957)

[Оценка точности моделей 9](#_Toc73102958)

[Заключение 10](#_Toc73102959)

[Список использованных источников 10](#_Toc73102960)

# **Введение**

В время прохождения учебной преддипломной практики необходимо достичь следующих целей:

* Исследовать возможность определения местоположения на территории ГЗ МГТУ им. Баумана с помощью анализа Wi-Fi сетей.
* Разработать модель, с помощью которой будет возможно определить местоположение на основе Wi-Fi сетей.

Для достижения поставленных целей необходимо решить следующие задачи:

* Разработать мобильное приложение для сбора данных о доступных Wi-Fi сетях в ГЗ МГТУ им. Баумана.
* Отсеять ненадёжные точки доступа.
* Разработать алгоритм для определения местоположения
* Разработать модель для проверки точности работы алгоритма.

# **Основная часть**

## **Разработка приложения для сбора данных**

Приложение для сбора данных о сетях Wi-Fi разработано для работы на устройствах под управлением операционной системы Android версии не ниже 6.0.0 на языке Java. Выбор языка обусловлен тем, что опыт разработки на нём уже имеется.

Функциональные требования приложения:

* Сканирование сетей Wi-Fi;
* Сохранение видимых устройством сетей с запоминанием их: MAC-адреса, SSID, этажа и части помещения, в котором сеть видна, силы сигнала, точного времени сохранения.
* Возможность выбора местоположения, в котором ведётся запись данных.
* Просмотр данных о нескольких видимых точках доступа в данный момент.
* Уведомление о количестве записанных сетях.

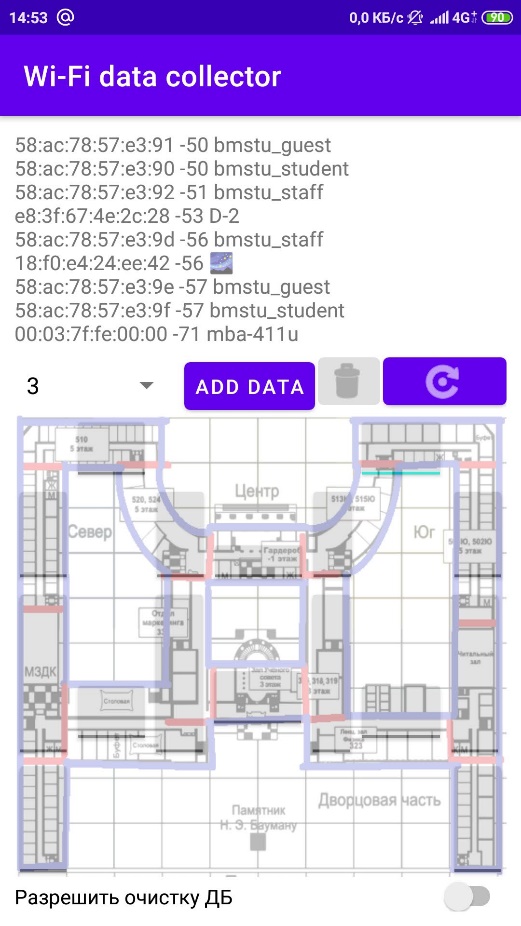


Рисунок 1. Итоговый вид интерфейса приложения

В верней части расположено поле, в котором отображаются сети с самыми сильными сигналами. Ниже расположены элементы управления:

* Выпадающий список для выбора этажа, на котором производится сканирование сетей.
* Кнопка «add data», по нажатии которой происходит запись данных в БД
* Кнопка (неактивная) со значком мусорной корзины предназначена для очистки БД.
* Кнопка с заворачивающейся стрелкой нужна для ручного обновления сканирования сетей.
* На фоне карты ГЗ расположены 16 залипающих кнопок, которые расположены над будущими зонами позиционирования.
* В самом низу переключатель, предохраняющий от случайного нажатия на кнопку очистки данных.

В смартфонах Android сканирование сетей происходит раз в 3-7 секунд, этого времени хватает чтобы можно было перейти в новую зону позиционирования, но, для большей точности данных была размещена кнопка для ручного обновления результатов сканирования. Уведомления о результате записи сетей выглядит следующим образом:

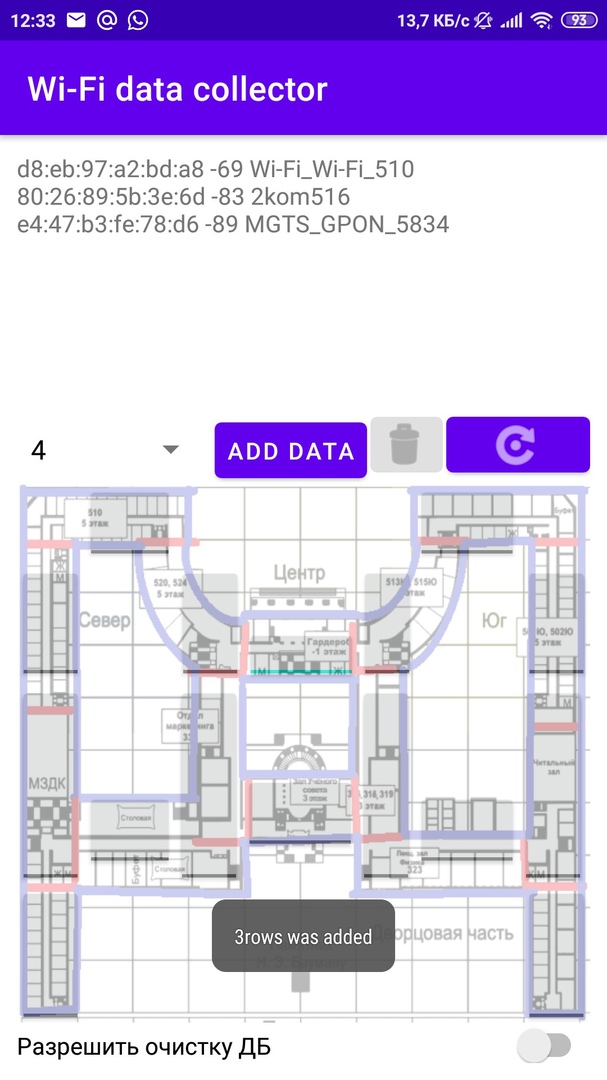


Рисунок 2. Пример уведомления о количестве записанных строк

Также при попытке записи данных без выбора зоны записи не происходит, а появляется предупреждение.

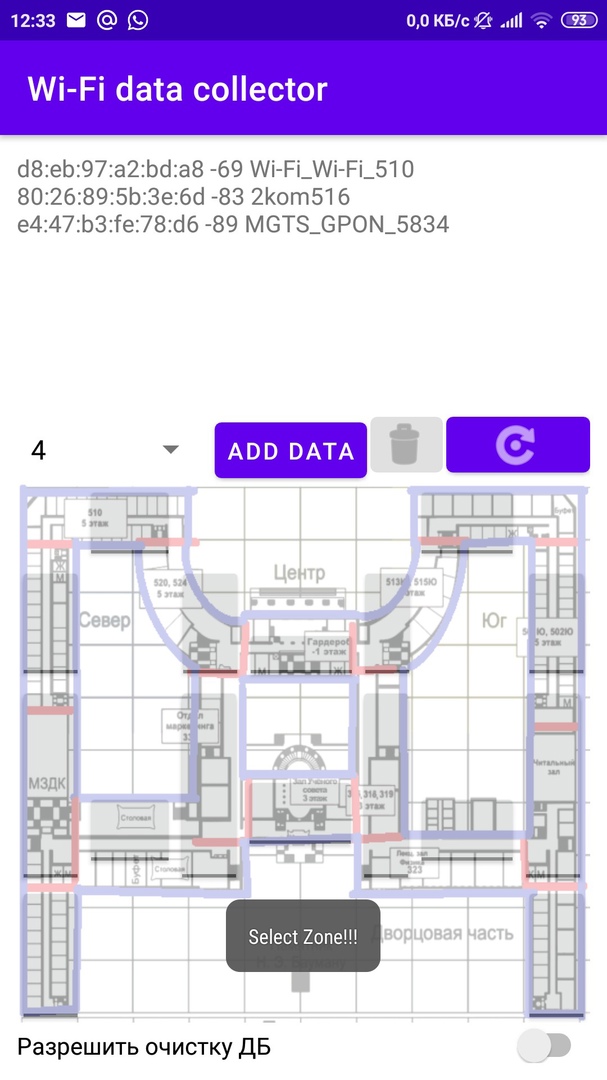


Рисунок 3. Уведомление при попытке записи без выбранной зоны.

Сбор данных происходил в 2 дня, в каждый из которых измерения проводились примерно по часу. Данные записывались примерно каждые 2-5 метров, таким образом на каждую зону приходилось по 3-8 сканирований.

Во второй день сканирование производилось немного чаще, чем в первый и также, были собраны данные с некоторых мест, где были допущены ошибки с указанием позиции сбора данных.

## **Обработка данных**

Было принято решение вести обработку данных с на языке Python, с использованием среды Jupyter Notebook, для большей быстроты разработки.

Для обработки и визуализации данных были использованы следующие библиотеки:

* Numpy;
* Pandas;
* Matplotlib.

По тогу сбора данных имелось 337 сетей и 650 точек доступа.

Критерий, по которому производилась отсеивание точек доступа – SSID.

Из списка длинной 337 SSID, были вычеркнуты сети, которые предположительно, базировались на мобильных устройствах других людей, а также бытовой офисной техникой, такой как МФУ. Сети, название которых содержало номер кабинета\аудитории или кафедры, а также сети со скрытым SSID, было решено оставить как надёжные. таким образом список сетей сократился до 122.

## **Определение местоположения**

В качестве кандидата на роль алгоритма рассматривалась нейросеть, однако, для её корректной работы катастрофически не хватало данных (по 3-8 образцов на зону для каждого этажа).

Для итогового алгоритма была сформирована таблица, которая содержит самый сильный сигнал от точки доступа и место, где этот сигнал был найден. Телефон, при сканировании уже располагает пойманный точки доступа в порядке убывания силы сигналов. Поэтому, для определения местоположения достаточно по очереди попытаться найти каждую из видимых сетей в «опорной таблице». Поиск идёт до первого найденного совпадения.

Минусы этой модели в том, что если в месте, которое мы пытаемся определить, нет достаточно сильных надёжных точек доступа, а ближайшая надёжная точка находится в соседней зоне или этаже, то место будет определено неверно.

Плюс состоит в простоте и быстроте работы и, несмотря на минусы, неплохой точности в 80% по этажам и 75% по зонам. 25% неточности по зонам скорее всего из попыток определить местоположение рядом с другой зоной, а так как почти у каждой зоны есть сосед, то это получается примерно по 12% от территории хоны с каждой из её сторон.

Также была попытка использовать алгоритм, при котором поиск по опорной таблице не останавливался на первом совпадении, а формировался список совпадающих сетей, при этом важность ответа, полученного по таблице исходила из силы полученного сигнала от каждой точки доступа. Сила сигнала нормировалась от 0 до 1, где 1 имел самый сильный полученный сигнал, а 0 самый слабый. Данная модель показала результат определения этажей на 1% лучше, чем простой поиск по опорной таблице до первого совпадения, поэтому, из-за его сложности было принято решение в пользу простого поиска.

## **Оценка точности моделей**

В качестве наборов данных, по которым производилась оценка были данные сканирования, полученные за второй день (так как сканирование велось с меньшими интервалами), которые использовались как «обучающие», и тестовые данные из набора сканирования первого дня.

Для проверки была сделана функция, принимающая на вход набор сетей, полученных за одно сканирование (такой набор формируется исходя из времени, в которое данные были записаны, то есть функция получает на вход тот же набор данных, что получил бы телефон в реальных условиях). Функция ищет каждую из сетей в опорной таблице, до первого совпадения и выдаёт на выход зону и этаж, которые удалось определить и те, что были в данных на входе. Получалось 4 значения, которые должны были быть попарно равны. В случае неравенства одной или обеих пар, делался вывод что это место определено неправильно.

Данная функция применялась по всем наборам сетей из тестового набора, после чего выводилась статистика правильности определения зон и этажей.

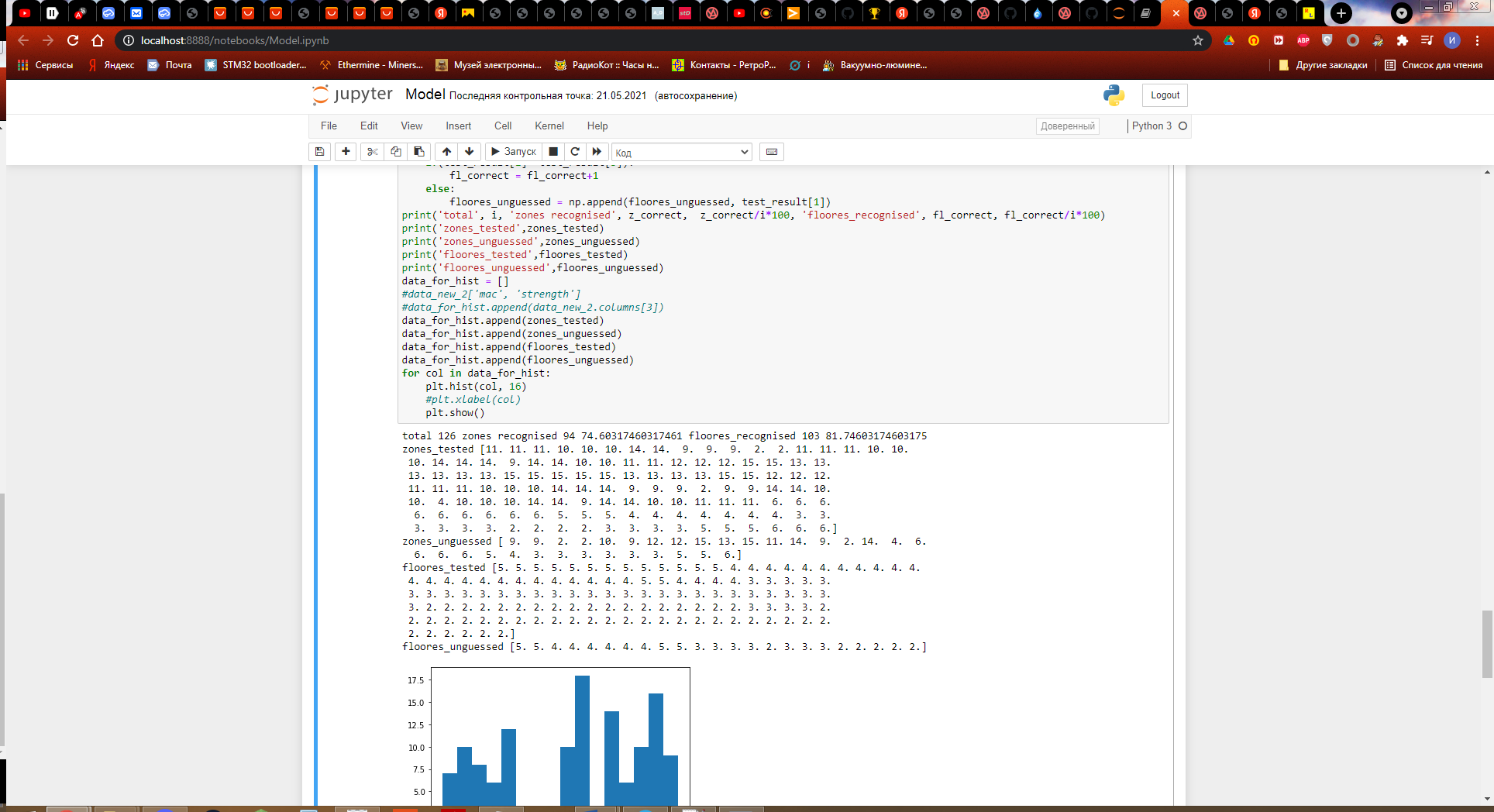


Рисунок 4. Вывод результата тестирования модели по опорной таблице (74,6% угаданных зон, 81,7% угаданных этажей на 126 примерах)

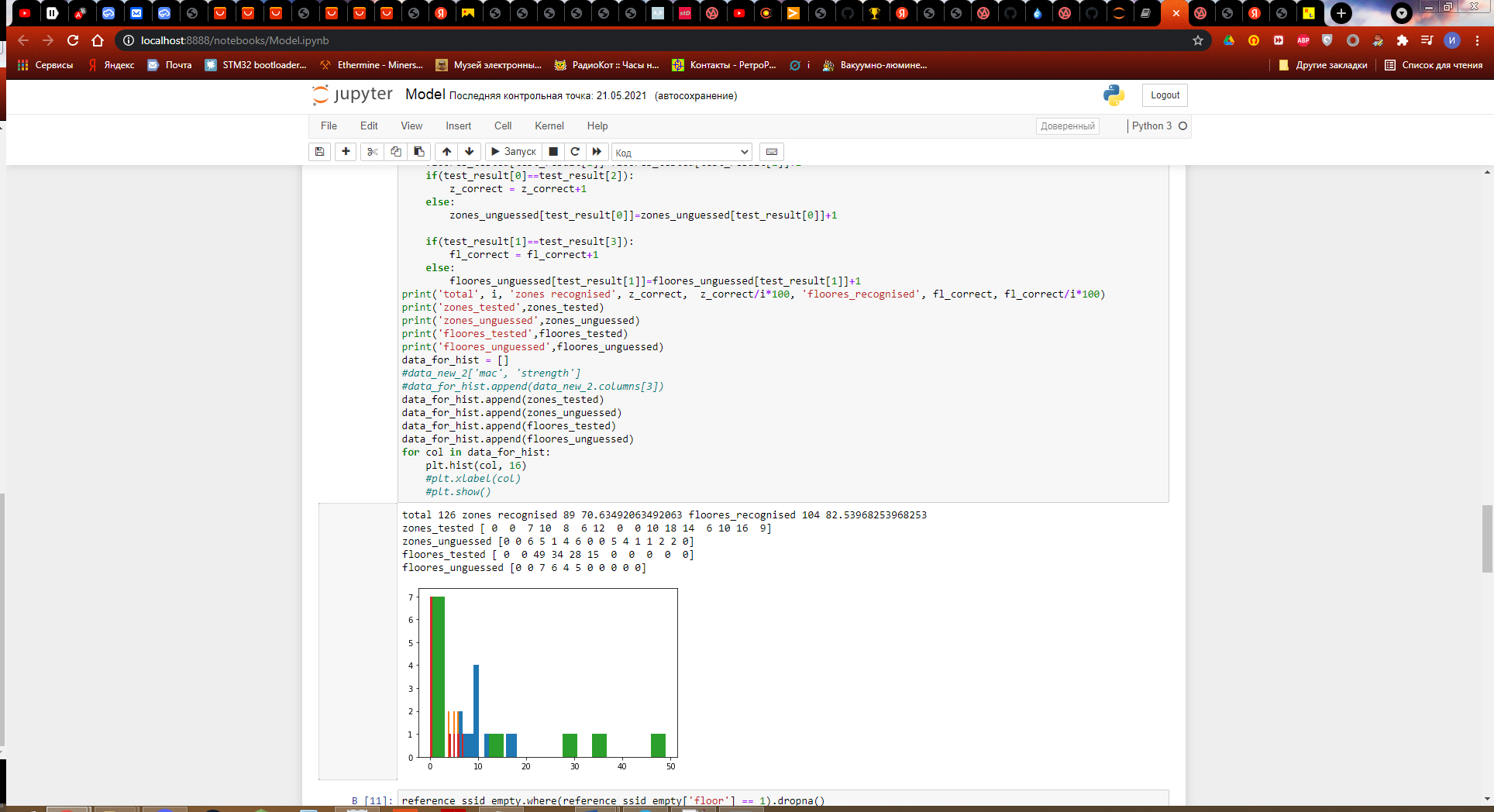


Рисунок 5. Вывод результата тестирования сложной модели (70,6% угаданных зон, 82,5% угаданных этажей на 126 примерах)

Также тестировались варианты этих моделей на данных с отсутствующими сетями со скрытым SSID. Результаты их тестирования показали немного более высокие результаты (на картинах как раз пример работы с таким набором данных).

# **Заключение**

В ходе выполнения преддипломной практики были достигнуты все поставленные цели и задачи.

Доказана возможность определения местоположения внутри главного здания МГТУ им. Баумана с помощью анализа Wi-Fi сетей с приемлемой точностью.

Разработана модель, производящая определение местоположения пользователя, которая будет реализована в ВКРБ в системе навигации по ГЗ МГТУ им. Баумана с помощью анализа Wi-Fi сетей.

# **Список использованных источников**

1. Антонов А.И. Лекции по курсу Беспроводные сети;
2. Гапанюк Ю.Е. Лекции по курсу Технологии машинного обучения;