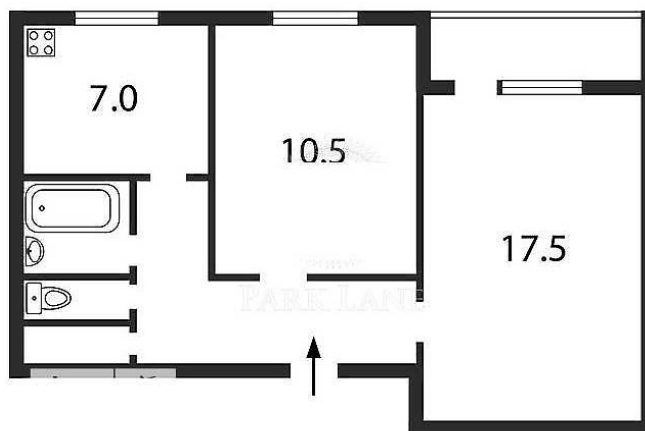
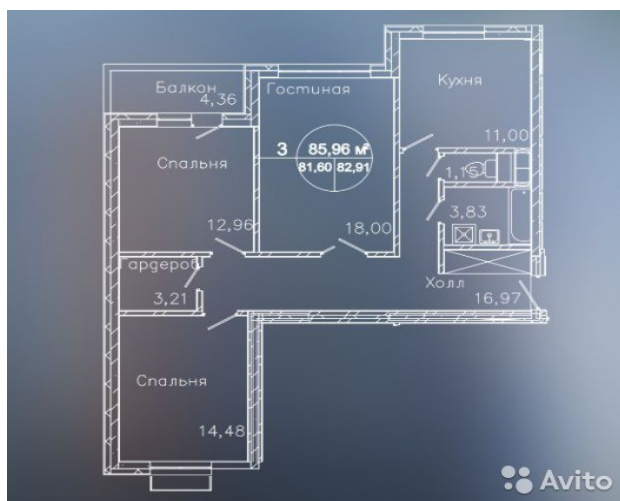




Тестовое задание “Планировки”

В рамках данного задания Вам предлагается реализовать классификатор изображений на два класса: “планировка” и “другое” (все остальное).

Планировками мы называем чертежи планов квартир (в т.ч. сканированные и т.п.), отрендеренные 3D-планы квартир и т.д. Например:



Больше примеров:

<https://cdn.riastatic.com/photos/dom/photo/3957/395757/39575749/39575749f.jpg>

<https://cdn.riastatic.com/photos/dom/photo/3872/387268/38726893/38726893f.jpg>

<https://cdn.riastatic.com/photos/dom/photo/3969/396949/39694956/39694956f.jpg>

<https://cdn.riastatic.com/photos/dom/photo/3953/395364/39536459/39536459f.jpg>

<https://cdn.riastatic.com/photos/dom/photo/3944/394422/39442227/39442227f.jpg>

<https://cdn.riastatic.com/photos/dom/photo/3765/376506/37650631/37650631f.jpg>

<https://39.img.avito.st/640x480/1232011539.jpg>

<https://cdn.riastatic.com/photos/dom/photo/3921/392164/39216449/39216449f.jpg>

<https://cdn.riastatic.com/photos/dom/photo/3816/381681/38168160/38168160f.jpg>

Долгое время для решения данной задачи в рамках нашей системы использовалась эвристика, которая учитывала долю белого цвета на изображении. Тут стоит отметить, что примеры выше не дают правильного представления о распределении всех планировок по видам в реальном мире. На самом деле в объявлениях Украины среди изображений планировок квартир преобладают черно-белые чертежи, аналогичные тому, который приведен в примерах на первой странице. Нетрудно заметить, что такое изображений действительно имеет большую долю пикселей белого цвета. Такая эвристика при приемлемой полноте (recall) давала хорошую точность (precision, см. [precision and recall](#)). Но поскольку и полнота для нас является важной характеристикой, имеется большое пространство для совершенствования такого классификатора.

Понимая сложность данного задания, мы нормально отнесемся к решению, которое в своей алгоритмической части будет реализовывать подход, уже описанный выше. Доведение известного алгоритма «до ума» (в рабочее состояние и с приемлемыми результатами) в свою очередь тоже является хорошей задачей. Но в то же время поощряется поиск своих решений, при чем описанную выше эвристику можно использовать для сравнительной проверки качества Вашего алгоритма.

Входные данные

В качестве входного набора данных предлагается использовать изображения из рабочей базы данных поиска объявлений ЛУН в Украине. В прикрепленном файле *images_urls.csv* содержится экспорт 1 млн. изображений, представленных адресами из хранилища ЛУН (при этом некоторые из адресов со временем могут становиться недоступными).

Задание

1. разработать и реализовать алгоритм классификации, который будет давать ответ на вопрос: “это изображение — планировка, или нет?” Дать аргументированное обоснование того, что данный алгоритм может быть использован в системе, подобной ЛУН (форма и способ аргументации произвольный: числа, графики, красивые слова,...);
2. классифицировать весь входной набор данных;
3. предоставить возможность в веб-интерфейсе в табличном виде с постраничным выводом просматривать результаты классификации по всем изображениям подряд (визуализация результатов п.2). В столбцах таблицы вывести: image_id, само изображение (с помощью HTML-тега), определенный класс: планировка, или нет;
4. реализовать веб-интерфейс, который позволит загружать изображение с компьютера и классифицировать его, показывая на странице результат классификации и само изображение.

От исполнителя данного задания ожидается комплексное решение данной задачи, в которой можно выделить такие подзадачи:

1. алгоритмическая;
2. архитектурная («красота» Вашего решения);
3. обработка данных (парсинг входа, работа с БД, HTTP, большими объемами данных, изображениями и т.д.);
4. написание web-интерфейса.

Тут стоит заметить, что алгоритмическая часть не является превалирующей, и она не играет главную роль в оценке Вашего решения. Выполненное задание будет оцениваться как комплексное решение (см. выше), при чем важным требованием является наличие рабочего решения. Если у Вас придуман мощный алгоритм, но что-то не работает, или чего-то нет — задание считается не выполненным.

Технические ограничения

Реализация на Python.

Любые вопросы по задаче приветствуются. Ждем Ваших вопросов.