

ARRAYINDEXOUTOFBOUNDSEXCEPTIONS

Кейс 2. Система определения предлагаемой заработной платы

Номинации:

1. Лучшая концепция решения
2. Лучшее решение на коде

Капитан:
Козлов Михаил
+79049438688

Идея проекта

Основной идеей проекта является создание сервиса способного определить, какая заработная плата должна быть у рабочего на основании анализа опыта работы, навыков, места работы, города и других характеристик. Наш сервис специализируется на оказании качественной, быстрой и технологичной помощи как работодателям в установке наиболее удачной оплаты труда, поиске сотрудников, так и работникам в оценке заработной платы, на которую они могут рассчитывать.

МЫ ТАКИЕ РАЗНЫЕ, НО МЫ ОДНА КОМАНДА!

В ходе разработки проекта мы поделились на две группы. Первая группа занималась сбором информации и взаимодействием с сайтом hh.ru, а также нормализацией и обработкой полученных массивов данных. Вторая группа занималась проектированием прототипа и переводом обработанных данных к виду, подходящему для ввода в модель.

Parsing информации с сайта hh.ru

В ходе нашей работы был написан parser на языке python, обращающийся к сайту через id вакансии или по времени создания, что позволило нам собрать наиболее полный массив данных. Во время тестирования решения мы столкнулись с медленным протоколом взаимодействия с сайтом, тогда было принято решение использовать в нашем парсере мультитредовость и асинхронность запросов, что позволило ускорить программу в несколько раз.

Оригинальность и инновационность

В процессе работы командой, занимающейся ML, была поставлена задача разбиения множества всех скилов на минимальное число непересекающихся подмножеств таких, что никакие два элемента лежащие в одном из подмножеств не лежат оба в одном из исходных множеств. Задача была решена с помощью продвинутого алгоритма, основанного на теории эволюции Genetic algorithm.

Обработка данных

Для начала нужно было “просеять” данные, так как некоторые признаки очевидно коррелировали (например: [specializations] и [specializations.name], и т.д.). Затем нужно было сократить количество признаков, чтобы модель не переобучалась, и <паша напиши что ты сделал и вставь скрин, в слайд “картинки говна”>.

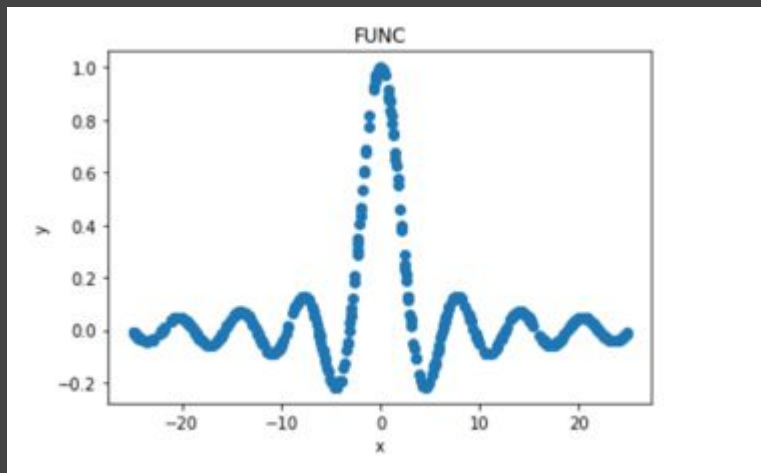
Модель

<Паша напиши поподробней говна, можешь на 2 слайда
(тогда создай его)>

Картинки зовна

Обучение нейронной сети

Function to predict



Our result

