kaggle   Search kaggle   Q   **Competitions**   **Datasets**   **Kernels**   **Discussion**   **Jobs**

Job Salary Prediction

Predict the salary of any UK job ad based on its contents

$6,000 · 289 teams · 5 years ago

Overview   Data   Discussion   Leaderboard   Rules
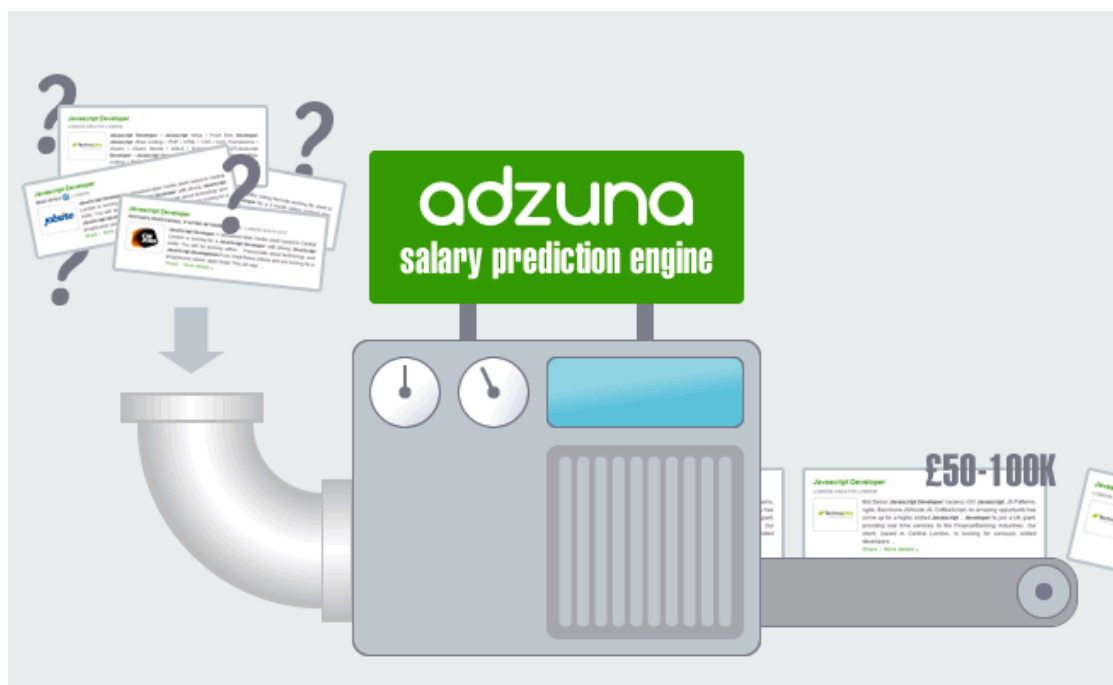
Overview

Description

Evaluation

Prizes

About Adzuna

Timeline

Winners

This competition was launched under the Kaggle Startup Program. If you're a startup with a predictive modelling challenge, please apply!

Adzuna wants to build a prediction engine for the salary of any UK job ad, so they can make huge improvements in the experience of users searching for jobs, and help employers and jobseekers figure out the market worth of different positions. At the moment, approximately half of the UK job ads they index have a salary publicly displayed. They need your help to bring more transparency to this important market.

Adzuna has a large dataset (hundreds of thousands of records), which is mostly unstructured text, with a few structured data fields. These can be in a number of different formats because of the hundreds of different sources of records. Adzuna needs the help of the Kaggle community to figure out the best techniques to apply to this data set to build a highly accurate predictive model for new ads. You will build, train and test your salary prediction engines against a wide field of competitors.

As an added perk, Adzuna intends to implement their chosen model on their website, both in the UK and worldwide. You will have the satisfaction of seeing your work implemented in production and change the way people search for jobs in the future.

Successful models will incorporate some analysis of the impact of including different keywords or phrases, as well as making use of the structured data fields like location, hours or company. Some of the structured data shown (such as category) is 'inferred' by Adzuna's own processes, based on where an ad came from or its contents, and may not be "correct" but is representative of the real data.

You will be provided with a training data set on which to build your model, which will include all variables including salary. A second data set will be used to provide feedback on the public leaderboard. After approximately 6 weeks, Kaggle will release a final data set that does not include the salary field to participants, who will then be required to submit their salary predictions against each job for evaluation.

Leaderboard >

1 lazylearner

2 Vlado Boza

3 LR

4 dejavu

5 VintageToasterMonthly

6 kg

7 arnaudsj

8 Daniel Velkov

34 discussion topics >

Visualizing individual RandomForestRegressor ...
2 replies · 3 months ago

Large scale text mining in R
20 replies · 7 months ago

I am looking for predictor results from test data...
0 replies · 9 months ago

How to add crossvalidation to scikit RandomFo...
6 replies · a year ago

Looking for a local teammate in SiliconValley/S...
1 reply · 4 years ago

Launch
5 years ago

Close
5 years ago

289	**336**
Teams	Competitors

Points **This competition awarded standard ranking points**

Tiers **This competition counted towards tiers**

Tags   mae   small

Our Team   Terms   Privacy   Contact/Support