Recommendation Engine For Online Programming Platform

04.12.2019

Video URL:

Dhaval Suthar

Prajakta Sumbe

Ravi Kiran

Veena Iyer

Final Project: INFO 6105 Data Science Eng Methods

OVERVIEW

Competitive programming is a mind sport usually held over the Internet or a local network, involving participants trying to program according to provided specifications. The aim of competitive programming is to write source code of computer programs which are able to solve given problems. Major companies hire from top coding platforms. The planning, workforce, time and money which goes into recruiting is cut to half by the competitive coding platforms.

GOALS

1. To build a model that can predict the number of attempts required by user to solve a problem based on problem complexity and the user's profile as to the number of problems solved, total number of submissions completed, user contribution to the online platform, location of the user, rating of the user and the rank of the user.

2. Recommending the questions that a programmer should solve given his/her current expertise is a big challenge for Online Programming Platforms but is an essential task to judge a programmer's expertise in that particular area which will help companies in their hiring process.

USE CASES

1. Online platforms: Can use the predicted attempts to evaluate the expertise of the user and suggest questions appropriately.

2.Hiring Company: Know the expertise level of candidates and hire an appropriate candidate who fit their expectations.

Machine Learning Models

1. Random Forest
2. Regression
3. Decision Tree

DATA

We will work with the data wherein the features are as below:

user_id	problem_id	level_type	attempts_range	submission_count	problem_solved	contribution	country	follower_count	max_rating	rating	rank
user_1	prob_918	E	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_2990	F	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_1358	D	2	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_4278	A	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_1868	A	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_2872	A	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_948	E	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_4386	E	2	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_1981	A	2	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_4550	C	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_1911	B	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_4930	E	2	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_522	A	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_655	D	2	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_1279	C	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_70	D	2	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_6304	A	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_6173	B	1	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_5115	C	3	84	73	10	Bangladesh	120	502.007	499.713	advanced
user_1	prob_4864	A	3	84	73	10	Bangladesh	120	502.007	499.713	advanced

170	user_1000	prob_1689	A	3	259	235	0	India	41	371.273	336.583	intermediate
171	user_1000	prob_1899	C	2	259	235	0	India	41	371.273	336.583	intermediate
172	user_1000	prob_4886	C	2	259	235	0	India	41	371.273	336.583	intermediate
173	user_1000	prob_6434	A	2	259	235	0	India	41	371.273	336.583	intermediate
174	user_1000	prob_3508	A	1	259	235	0	India	41	371.273	336.583	intermediate
175	user_1000	prob_3209	B	1	259	235	0	India	41	371.273	336.583	intermediate
176	user_1000	prob_5585	A	1	259	235	0	India	41	371.273	336.583	intermediate
177	user_1000	prob_5801	B	1	259	235	0	India	41	371.273	336.583	intermediate
178	user_1000	prob_3334	B	3	259	235	0	India	41	371.273	336.583	intermediate
179	user_1000	prob_757	C	1	259	235	0	India	41	371.273	336.583	intermediate
180	user_1000	prob_1705	B	2	259	235	0	India	41	371.273	336.583	intermediate
181	user_1000	prob_4672	A	2	259	235	0	India	41	371.273	336.583	intermediate
182	user_1000	prob_6004	B	3	259	235	0	India	41	371.273	336.583	intermediate
183	user_1000	prob_1394	B	3	259	235	0	India	41	371.273	336.583	intermediate
184	user_1000	prob_3024	B	1	259	235	0	India	41	371.273	336.583	intermediate
185	user_1000	prob_3453	C	1	259	235	0	India	41	371.273	336.583	intermediate
186	user_1000	prob_5890	B	1	259	235	0	India	41	371.273	336.583	intermediate
187	user_1000	prob_713	A	2	259	235	0	India	41	371.273	336.583	intermediate

PROCESS OUTLINE

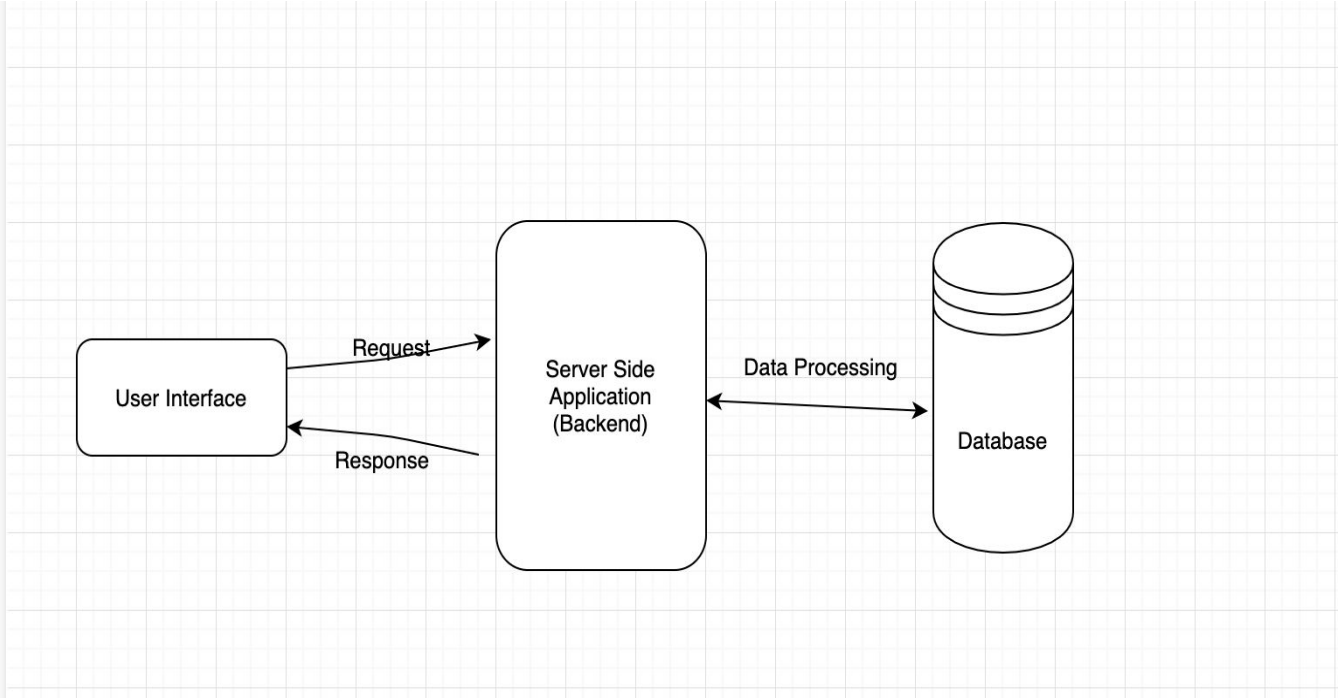
1. Data Preprocessing
   - ➔ Preparing final dataset by joining three subsets of data
   - ➔ Data Cleaning, handling missing values.
2. Exploratory Data Analysis.
3. Study supervised approaches and select the best model for prediction.
4. Design a pipeline and system to implement this approach.
5. Deploy the model.

MILESTONES

Timeframe	Delivery
Day 1-2	Topic Selection
Day 3-5	Data Gathering and Preprocessing
Day 6-7	Exploratory Data Analysis
Day 8-12	Model Building, Training, Selection
Day 13-14	Deployment of model
Day 15	Documentation

Deployment Details

1. **Language**: Python, Java
2. **Container**: Docker
3. **Web Platform**: Heroku



USER INTERFACE DESIGN PLAN

1. Have hiring company enter the expertise level as input and return list of candidates who fulfill that level.