

Rec-user: Recommending Users for GitHub Projects

Raunak Sinha
Akarsha Sehwal
Hridoy Sankar Dutta

Problem Statement

- **Github** - collaborative social network for programmers.
- **Problem Statement:** *Recommending GitHub Users for GitHub Projects based on activities in GitHub and StackOverflow*
- **Solution:** Creating user features based on activities performed in GitHub and StackOverflow and recommending the matched GitHub Projects for them.

Dataset Description

1. We had a list of common GitHub ID's and StackOverflow ID's.
2. From **GitHub API**, we extracted the languages of the repositories a particular user has contributed to.
3. For **StackOverflow(SO) Dump**, we extracted the following features:
 - a. Reputation of a particular user
 - b. Number of Questions with 0 answers
 - c. Number of Questions with atleast 2 answers
 - d. Score of a user (Total Upvotes-Downvotes count for all his/her questions/answers)

You gain reputation when:

- question is voted up: +5
- answer is voted up: +10
- answer is marked "accepted": +15 (+2 to acceptor)
- suggested edit is accepted: +2 (up to +1000 total per user)
- bounty awarded to your answer: + full bounty amount
- one of your answers is awarded a bounty automatically: + half of the bounty amount ([see more details about how bounties work](#))
- site association bonus: +100 on each site (awarded a maximum of one time per site)
- example you contributed to is voted up: +5
- proposed change is approved: +2
- first time an answer that cites documentation you contributed to is upvoted: +5

If you are an experienced Stack Exchange network user with 200 or more reputation on at least one site, you will receive a starting +100 reputation bonus to get you past basic new user restrictions. This will happen automatically on all current Stack Exchange sites where you have an account, and on any other Stack Exchange sites at the time you log in.

You lose reputation when:

- your question is voted down: -2
- your answer is voted down: -2
- you vote down an answer: -1
- you place a bounty on a question: - full bounty amount
- one of your posts receives 6 spam or offensive flags: -100

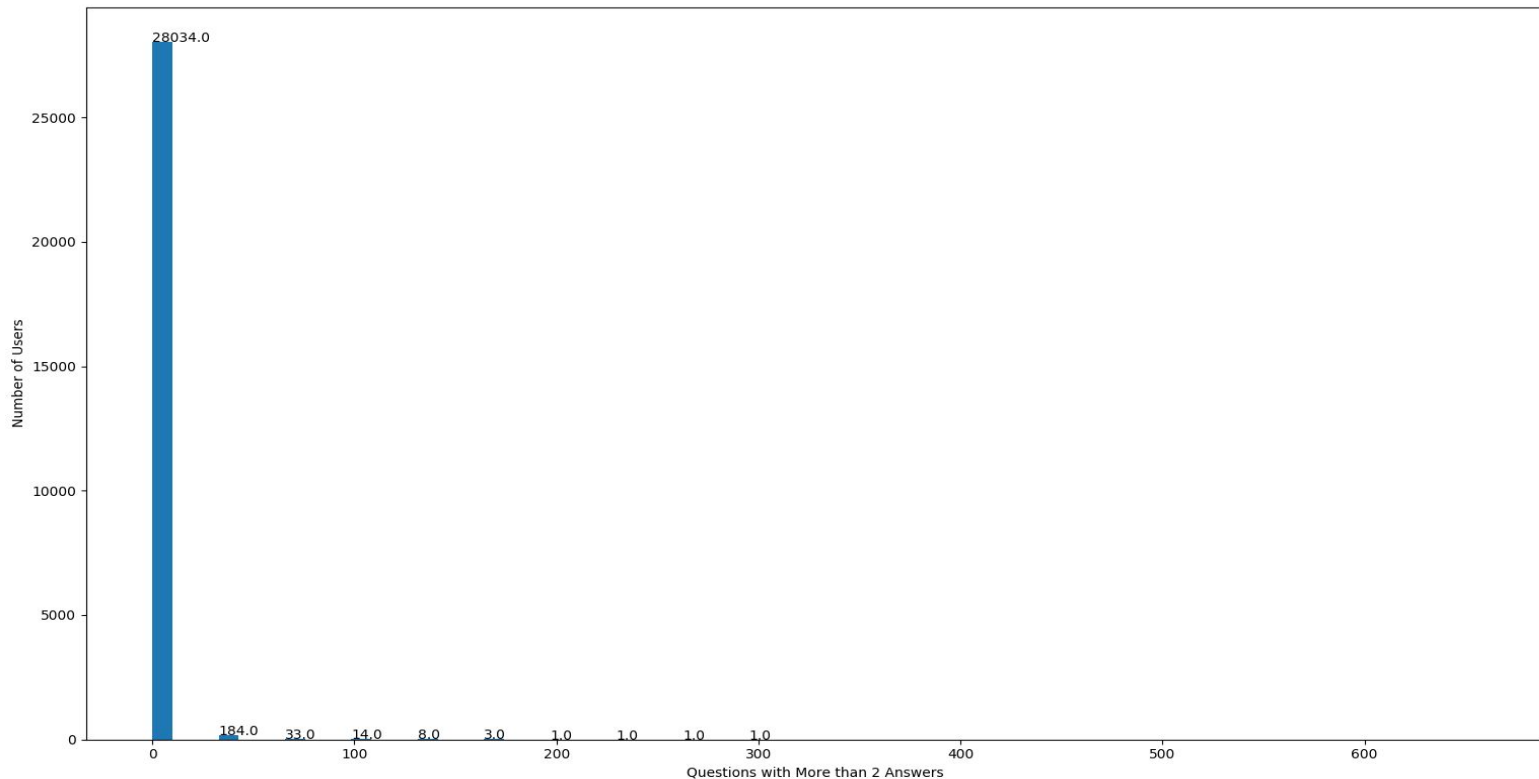
All users start with one reputation point, and reputation can never drop below 1. Accepting your own answer does not increase your reputation. Deleted posts do not affect reputation, for voters, authors or anyone else involved, in [most cases](#). If a user reverses a vote, the corresponding reputation loss or gain will be reversed as well. Vote reversal as a result of voting fraud will also return lost or gained reputation.

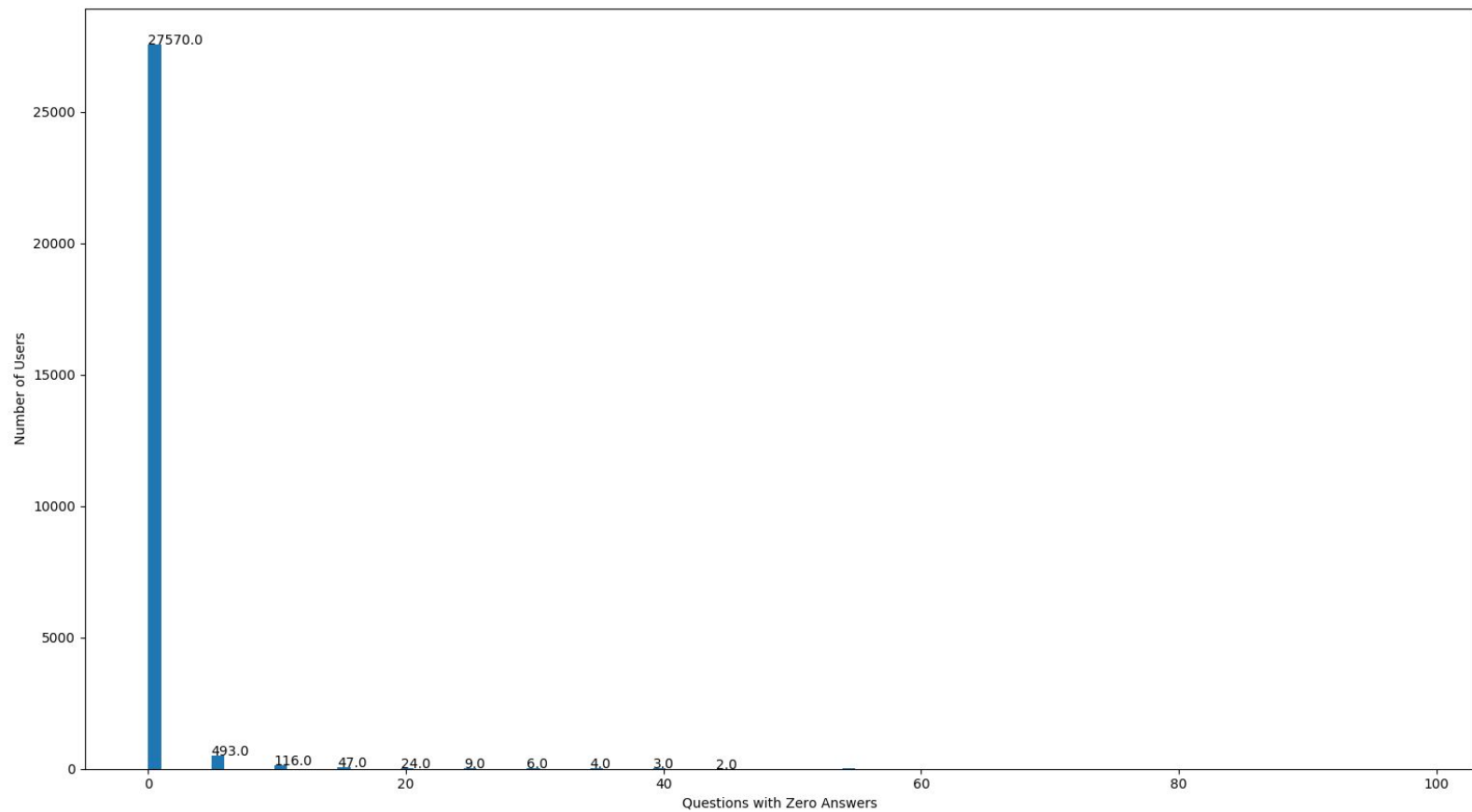
At the high end of this reputation spectrum there is little difference between users with high reputation and ♦ moderators. That is intentional. We don't run this site. The community does.

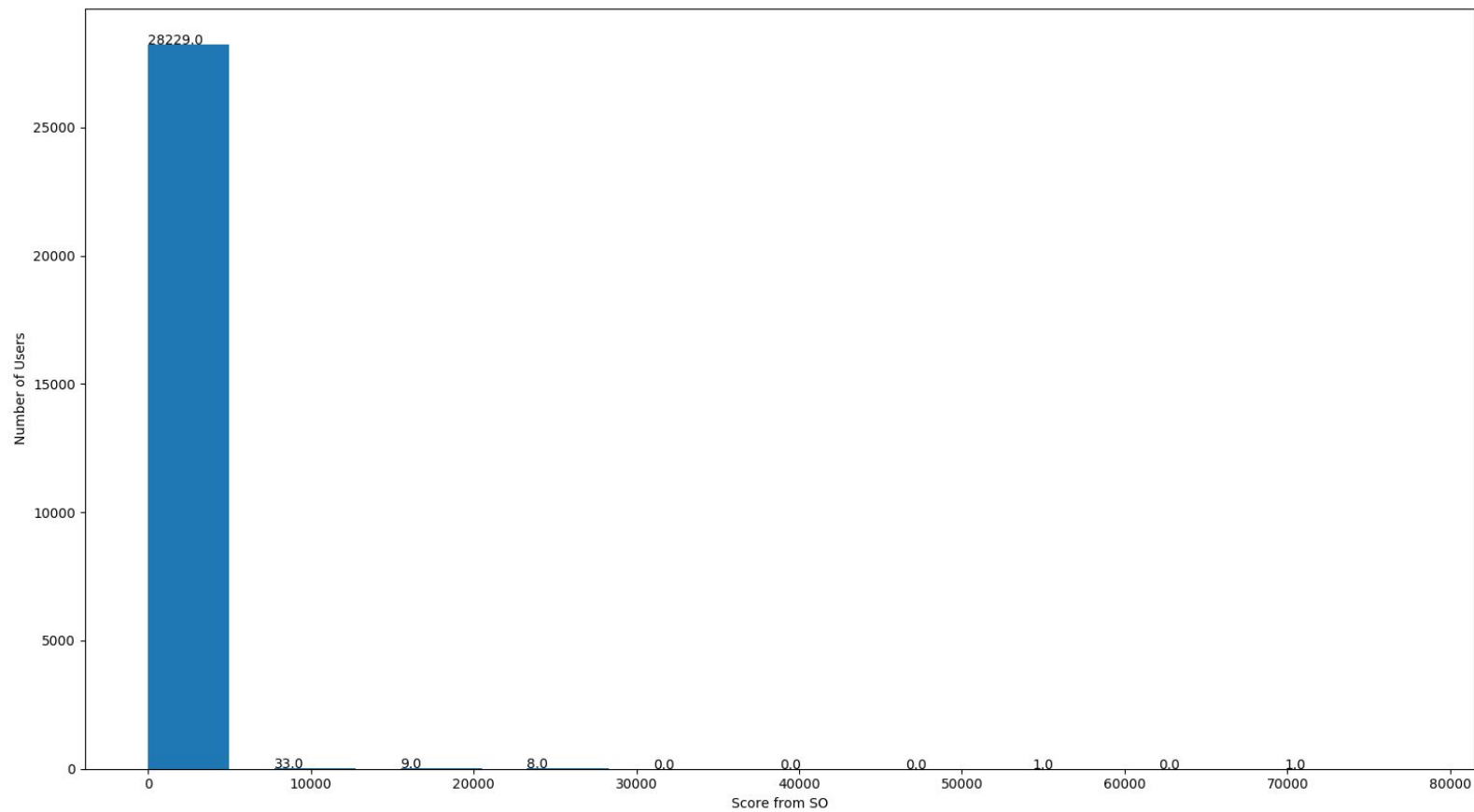
Dataset Summary

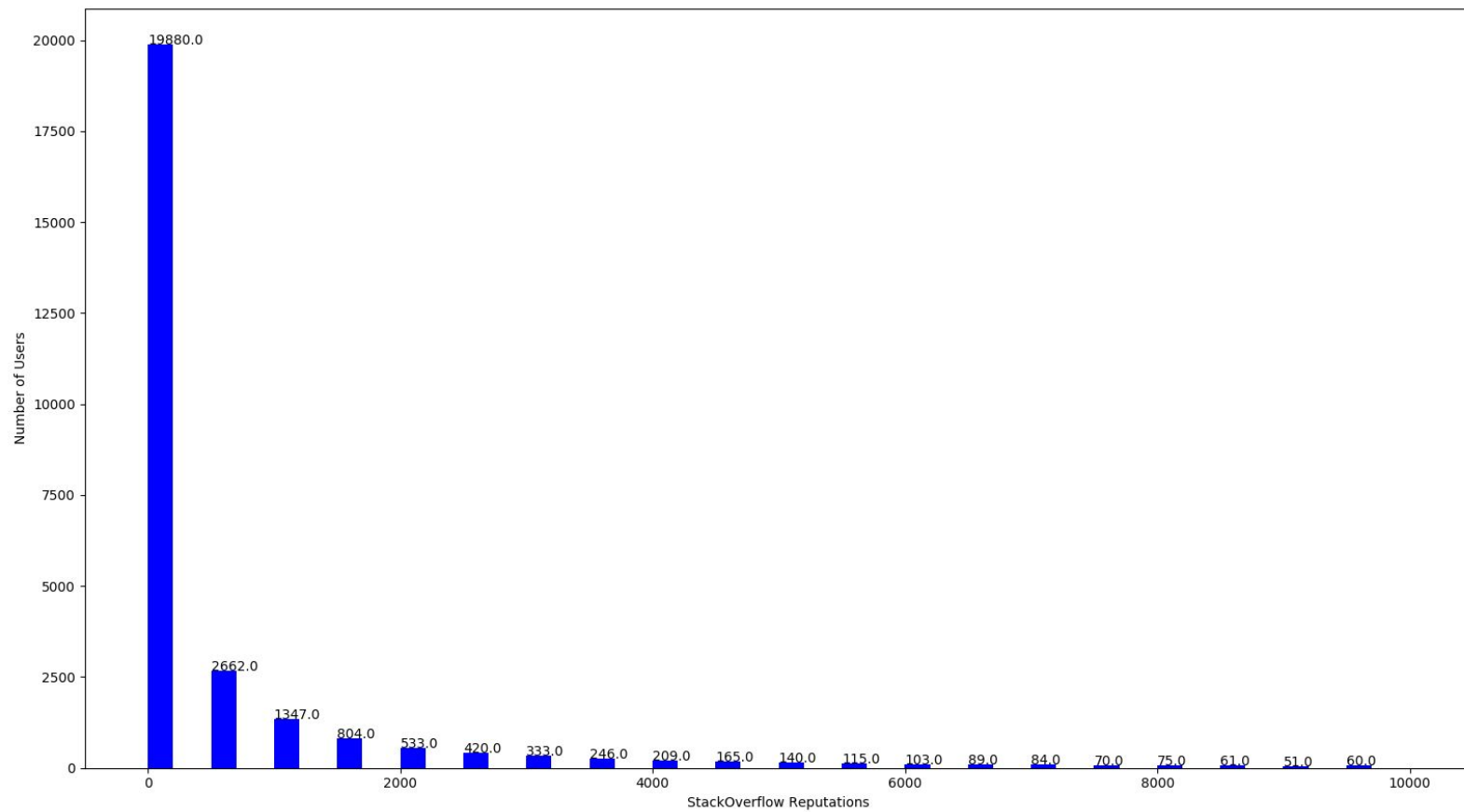
GitHub & StackOverflow Common Users	1936
StackOverflow Users Features Extracted	4
GitHub Repositories Features Extracted	256
Github User Features Extracted	256

Data Analysis from StackOverflow

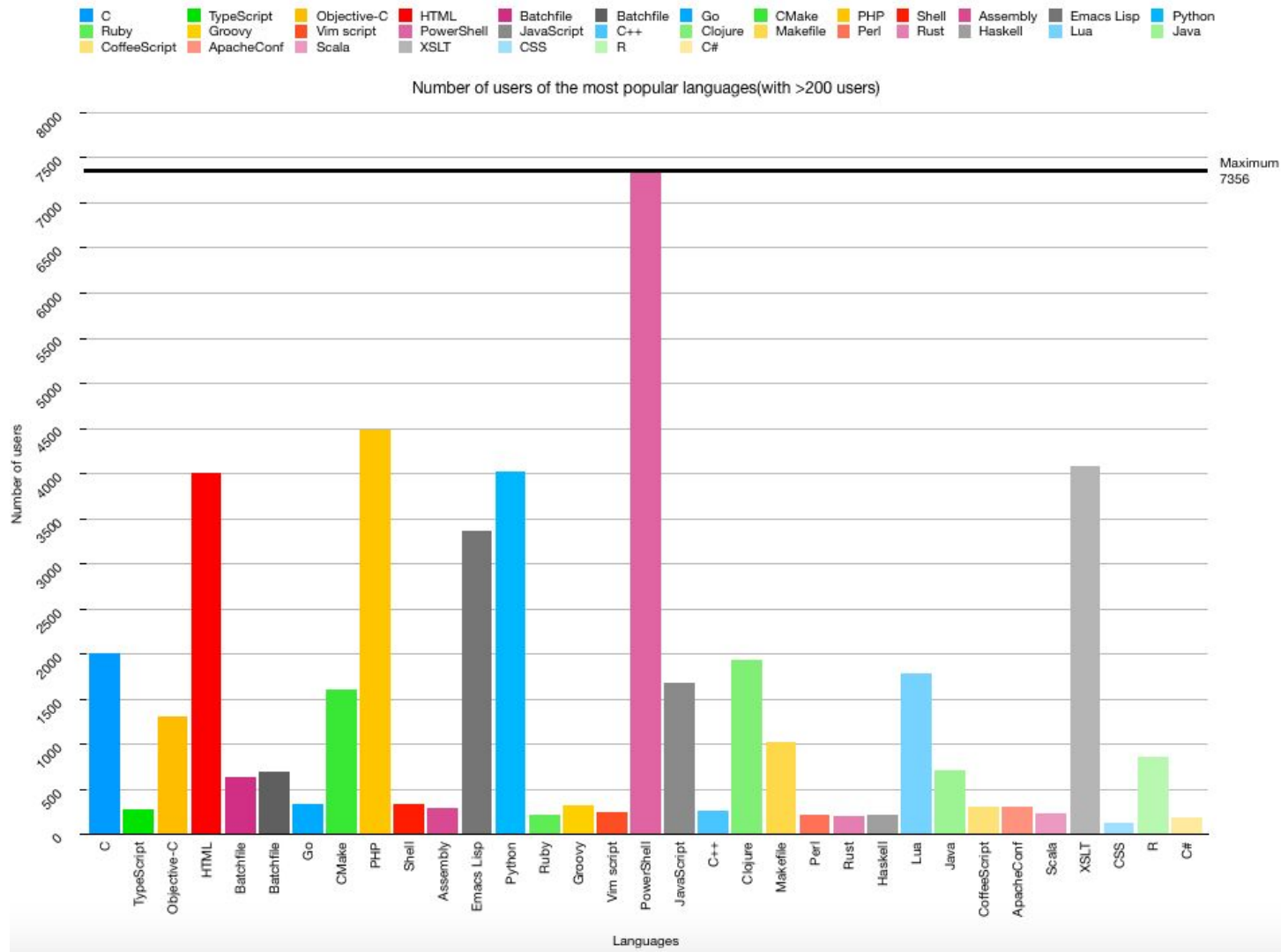








Data Analysis from GitHub



User-Item Recommendation using **LightFM**

<i>C</i>	<i>C++</i>	<i>Python</i>	<i>Java</i>	<i>Perl</i>	
5	4	1	8	1	<i>User1</i>
0	1	4	0	7	<i>User2</i>
6	5	3	0	2	<i>User3</i>

<i>C</i>	<i>C++</i>	<i>Python</i>	<i>Java</i>	<i>Perl</i>	
1	4	16	8	1	<i>Repo1</i>
2	6	11	4	7	<i>Repo2</i>
3	7	3	4	1	<i>Repo3</i>

<i>User1</i>	<i>User2</i>	<i>User3</i>	<i>User4</i>	<i>User5</i>	
4	14	5	8	7	<i>Repo1</i>
3	4	3	8	3	<i>Repo2</i>
4	3	7	7	8	<i>Repo3</i>

- A hybrid matrix factorisation model for linear combinations of their latent factors to recommend the best possible match user for a particular user.
- A collaborative filtering model, users and items are represented as latent vectors (embeddings)
- Works with three sets of information, **user metadata**, **repository metadata** and **interactions** between them.