**Codechef Report**

ML Team 6 - Mandar Mhaske, Priyanka shah, Sanik Malepati

Mentor: Sai Likhith K

**Summary:**

There are 3 metrics in the codechef forum such as category, number of posts and tags.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Site** | **discuss.codechef.com** | **No.of posts** | **Stemaway.com** | **No. of posts** |
| **Category:** | Editorial  Uncategorized  Codechef-DSA-learners  General  Announcement  Help  Tutorial  Site Feedback  Meta  FAQ  WICP- CP Basic session  Alibabacloud  SRI-B learning & Developme loo nt | 4455  2568  369  20315  166  4743  206  66  57  99  6  136  1 | STEM-away conversations  Friendly Competitions  STEM-away Awards  June 2020 Piazza  Announcements (July 2020)  FAQ  ML summer 2020  ML-Team 6-July2020  BI-Summer 2020  UX-Summer 2020  FS-Summer 2020  STEMcasts Webinars  Virtual internship  1-click Resume  Virtual Internship showcase  Community Contributions  About US  Support  Fall 2020 (New visitors)  Summer 2020 archives | 5  8  9  34  51  26  51  4  65  37  47  0  0  0  0  1  0  9  6  10 |
| **Tags:** | Ad-hoc  Array  Basic-math  Bfs  Binary-search  C-plus-plus  Cakewalk  Challenge  Combinatorics  Data-structure  Dfs  Dynamic-programming  Easy  Easy-medium  Editorial  Editorials  Fft  Gcd  Geometry  Hard  Hashing  Help  Implementation  Keteki  Likecs  Long-challenge  Long-contest  Math  Maths  Matrix-expo  Medium  Medium-hard  Number-theory  Observations  Prefix-sum  Probability  Segment-tree  Simple  Simple-math  Sorting  String  Strings  Taran\_1407  Tmwilliamlin  Tree  Video-editorials  Zio  zioeditorials |  | Teamwork  Bioinformatics  Team-building  Edustem  Full-stack-design  Mentoring  Communication  Us-research  Stemway-pathway  Ux-design  Bls  Learning-resources-ml  Advanced-biology  Gcp  Project-management  Python  Iot  Webinar  Advanced  Discourse  Ecc  Fireside-chat  Intermediate  Self-assessments-open  Stemway-blog  Teams-july2020  Django  Hardware engineering  R-trainings  Soft-skills |  |

**Navigation and Display**

* Users can filter through topics by:
  + Categories
  + Tags
  + Latest and top posts
  + Events using calendar
* For each topic:
  + The category and tags are shown at the top of the post and at the sticky top bar.
  + Categories are represented by colored squares
  + Tags are represented by gray squares
* Visually, it has a similar look to the STEM-Away forum.
  + We suspect that Codechef forum is built off Discourse, like STEM-Away

**Detailed metrics**

* **Tags System**
  + Popular Tags: “algorithm”, “challenge”, “problem”, “mathematics” and “contest”
  + Are Posts Properly Tagged: Yes, most of the time posts are tagged appropriately
  + Are Tags Related to Keywords: Yes, tags are related to keywords (see popular tags for examples)
* **Categories**
  + Popular Categories: “General”, “Help”, and “Editorial”, a large number of forums are in the category “Uncategorized” (~2.5k forums)
  + Are Most of the Posts In a Reasonable Category: Yes, most of the posts are in an appropriate category
* **Titles**
  + Are Most Titles Informative: No, most titles are not informative and used a lot of abbreviations in the title. User will not figure out the project or problem based on the title.
* **Content**
  + Popular Topics: challenge, contest, data structure
  + Keywords: code, problem, Editorial
* **Comments**
  + What People Comment About: Programming, providing coding help, answering coding questions, asking specific questions relating to the original post’s topic such as programming language, frameworks, help, and projects.
  + Are Comments Relevant: Yes, the comments are relevant to the forum and relevant for the purpose of training a topic recommendation and forum classification for STEM-Away because the forum classified posts based on the programming languages and frameworks related to the post.

**Description:**

According to the above table, there are a very few similarities between codechef and stem away. There are no subcategories in the codechef forum.

CodeChef is an extension of Unacademy which is a learning website so most of the forums are related to helping students with their specific problems, while Stem-Away is an internship website so most of the forums are related to informing interns about new tasks or events that are coming up.

**Conclusion**

After looking at the metrics and description, **we conclude that the forum is not suitable for our project** because the ML algorithm on this type of data will make the data set noisier and just lower the accuracy.