Final project - discren the topics of problem, and analyze interview problems in leetcode, interviewbit, codechef

Abstract

Using MySQL, API scraping, Selenium, and the BERT model, we are able to handle data from three coding practice websites, including Interviewbit, Codechef, and Leetcode. With the BERT model, we analyze the topics of problems, which helps users identify the topics of problems they may be unfamiliar with, allowing them to strengthen their problem-solving skills in related areas. We also conduct simple analyses on the three websites to determine which problem topics appear more frequently in interviews. The purpose of this project is to assist programmers in identifying specific problem topics and determining the topics of unknown problems.
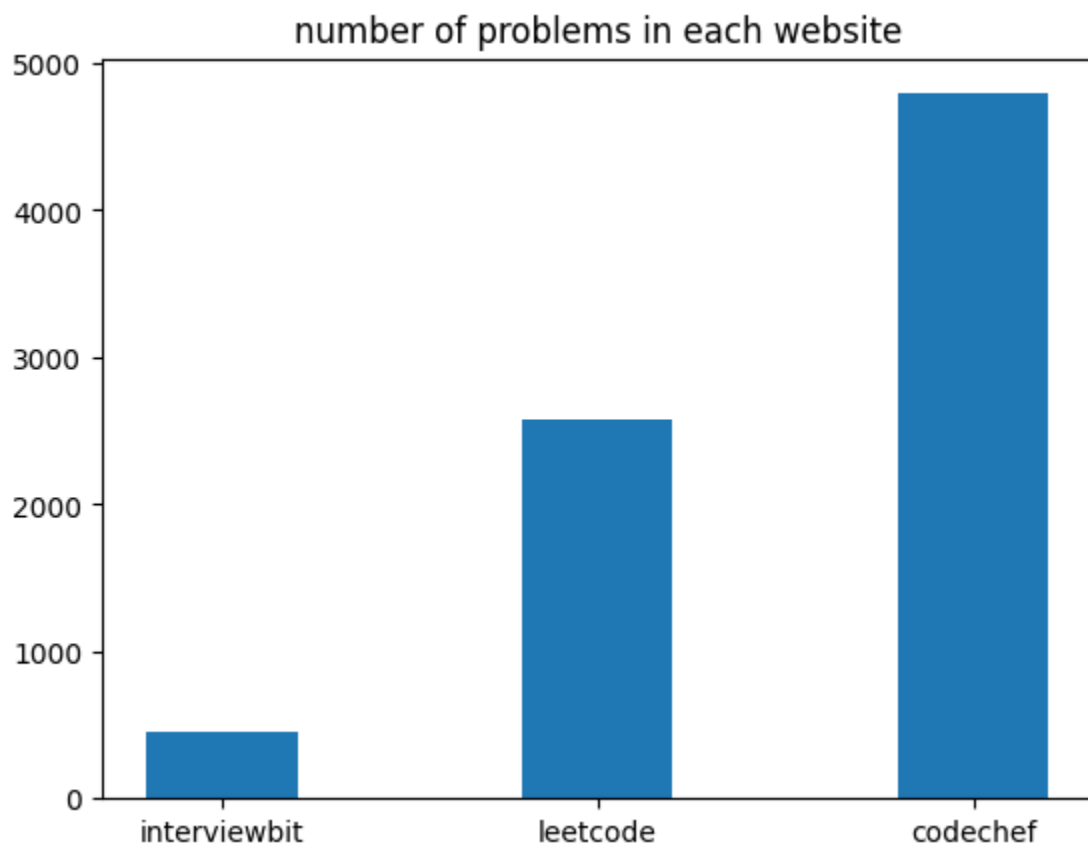
Motivation

Software engineering has become a highly competitive field worldwide, and many people are interested in making a career change to a software engineering-related position. Therefore, it is essential to enhance coding skills. This motivation led us to explore the differences between three coding practice websites and use them as a basis to determine problem topics using the BERT model. Additionally, we created an integrated search function to enable users to access coding problems across all three websites. We firmly believe that this project will be of great value to programmers.

Data pre-processing

We first use APIs, Selenium, and regression expressions to retrieve problem descriptions, topics, company tags, and accuracy rates for each problem from three coding practice websites: Leetcode, Interviewbit, and Codechef. After obtaining the data, we store it in MySQL and use SQL commands and Python to conduct further investigation and modeling.
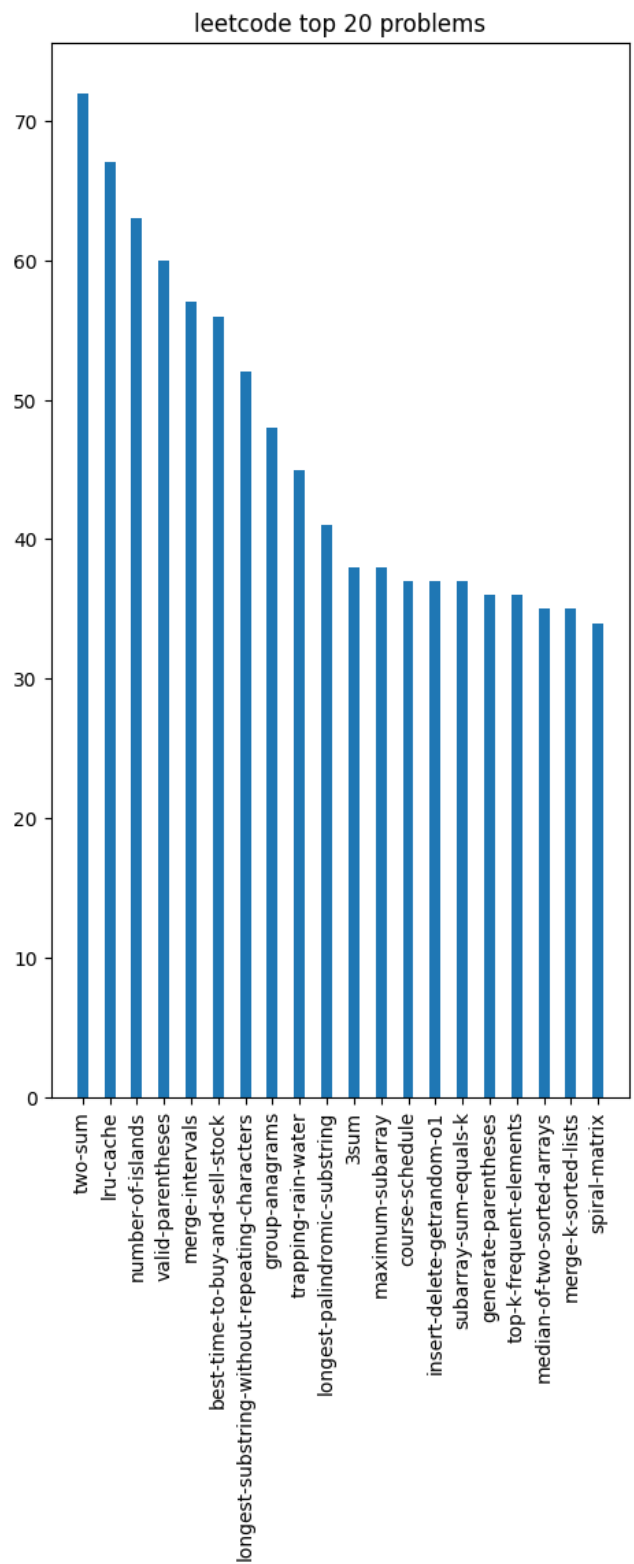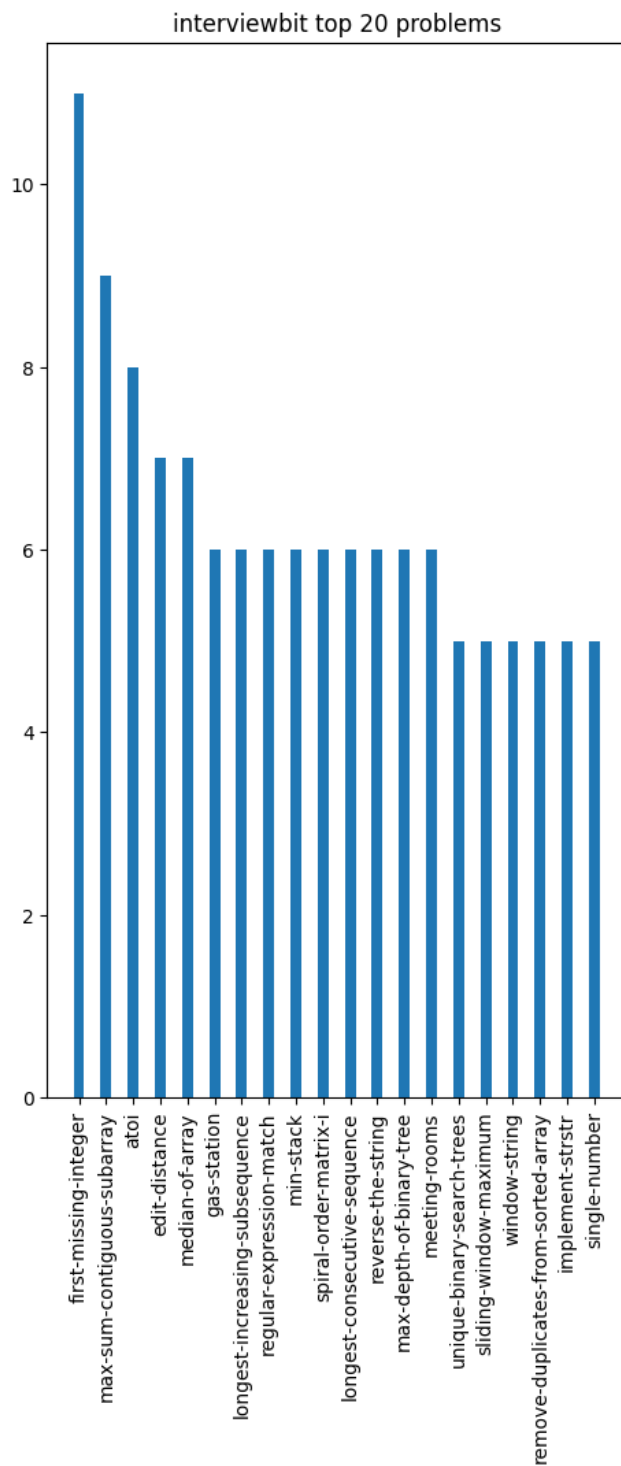
The difference among three websites

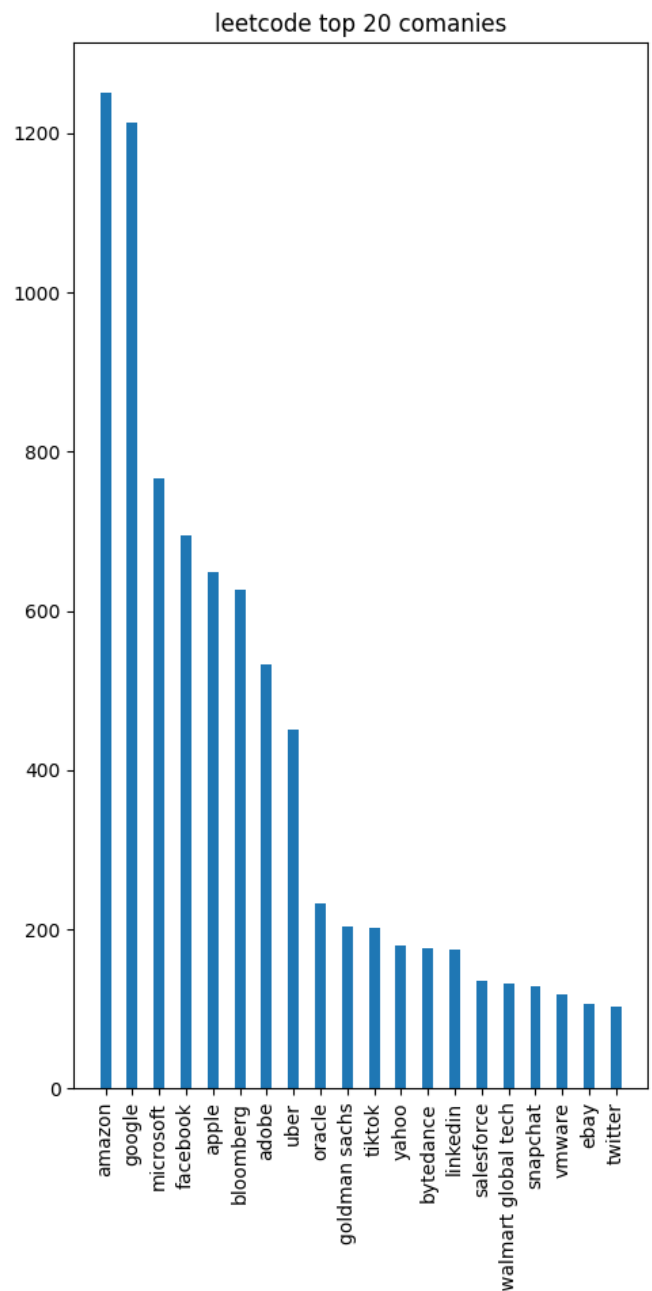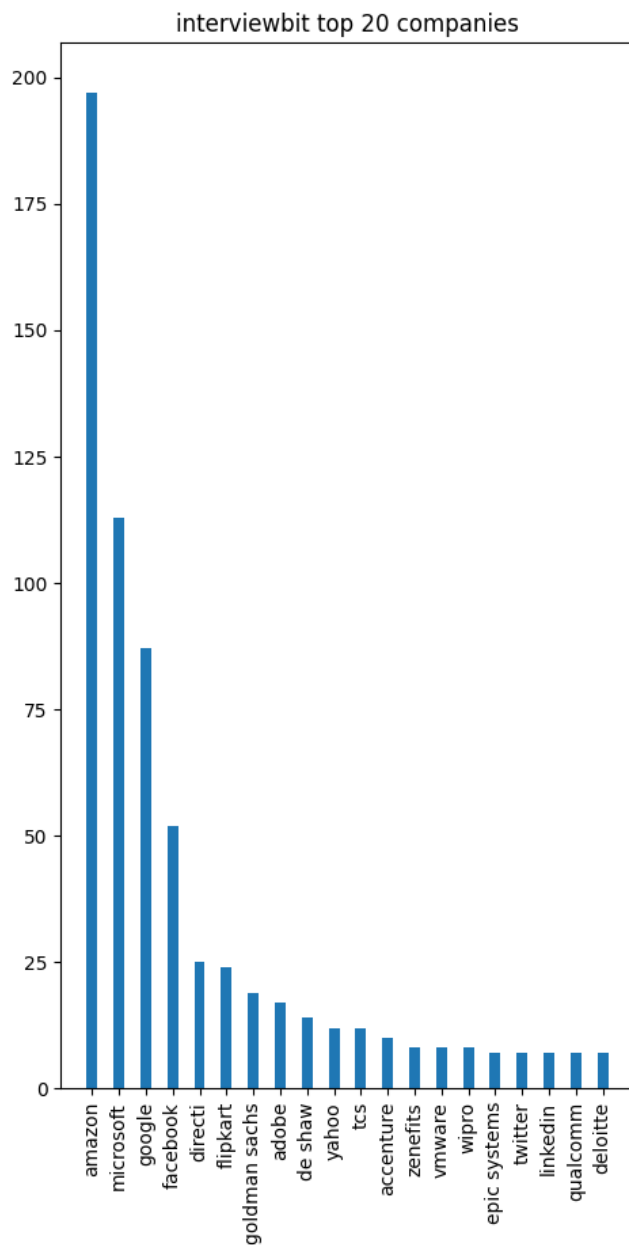### Numer of problems in three websites

number of problems in each website

The three websites have different numbers of problems, with Codechef having the most and Interviewbit having the least.

Topics of questions that companies use most often in three websites

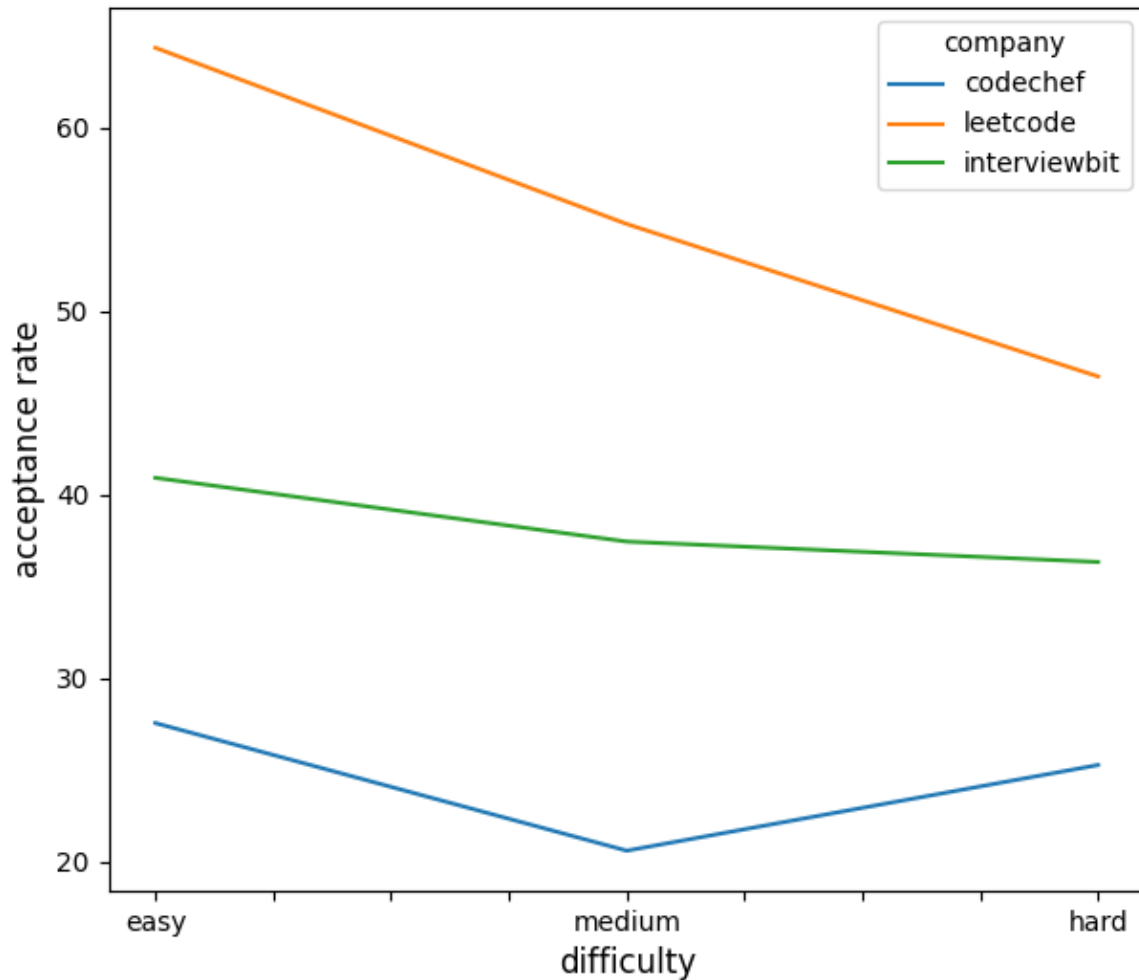interviewbit top 20 problems | leetcode top 20 problems

In this section, we will recommend that interviewees begin with these problems, as they are the most commonly adopted problems among the majority of companies.

The most commom companies in three websites

interviewbit top 20 companies   leetcode top 20 comanies

Since Codechef does not have company tags for each problem, we conducted company-related analyses using Interviewbit and Leetcode. Based on the plot above, Amazon, Microsoft, and Google are the top three most common companies on Interviewbit and Leetcode. Furthermore, it is clear that multiple companies appear in both images, and most of these companies belong to the information technology industry."
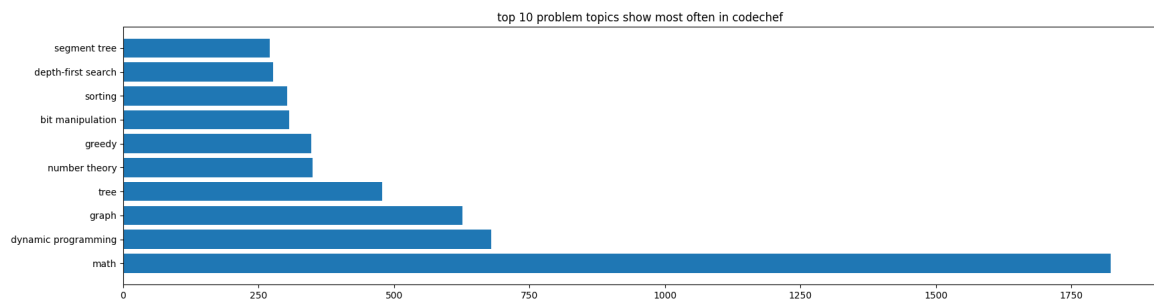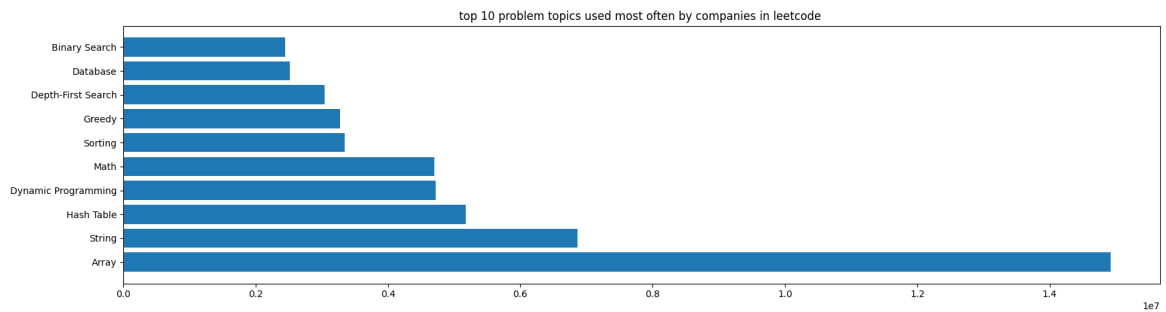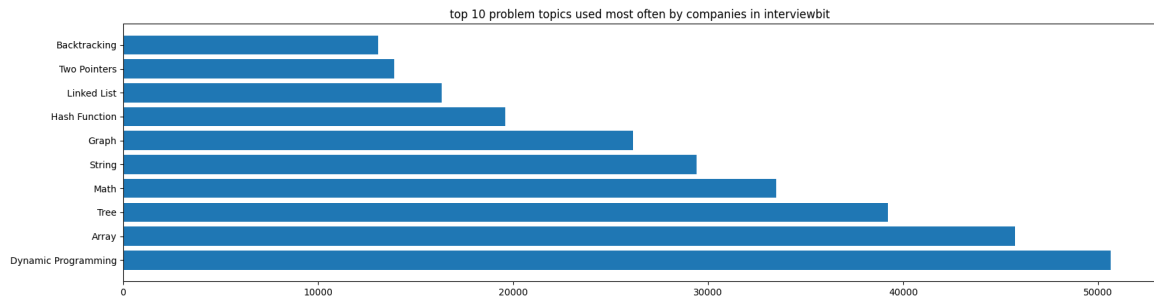
The accuracy rate in three difficulty levels among three websites

In this section, it is interesting to note that Codechef has the lowest mean accuracy rate for problems at the medium level. This may be due to the fact that Codechef has multiple difficulty levels, including some unknown levels such as orange and eee. As a result, the mean accuracy rate on Codechef may not be accurate.

We would recommend that beginners learning to code start by practicing easy-level problems on Leetcode, as it has the highest accuracy rate.

Top 10 most common topics that companies adopt in interviewbit and leetcode data, top 10 commom topics in codechef.

top 10 problem topics used most often by companies in interviewbit

top 10 problem topics used most often by companies in leetcode

top 10 problem topics show most often in codechef

Based on the plot above, the most common interview problems include dynamic programming, arrays, and strings. From this, we can infer that companies in the information technology industry prefer to select interview problems in these three areas. Therefore, interviewees should have experience in these topics.