The shortest superstring problem is the following: Given a finite alphabet \( \Sigma \), and a set of \( n \) strings \( \{s_1, s_2, \ldots, s_n\} \) over this alphabet, construct the shortest string \( s \) that contains each \( s_i \) as a substring.

Example: strings are ATCG, CGT, TAAT, CGC, the string TAATCGCGT contains all these strings as substrings and is of length 9 – is this shortest possible??

1. Prove this problem is NP-hard.

2. Give a \( O(\log n) \) approximation factor algorithm for this problem

3. (Optional-hard): Can you find a constant factor approximation algorithm for this problem?