Shannon Fano Algorithm

- Introduction
    - The Shannon-Fano algorithm is a greedy algorithm for lossless data compression. It is named after Claude Shannon and Robert Fano, who developed it independently in the 1940s.
    - This is a greedy algorithm that assigns variable-length codes to symbols based on their probabilities of occurrence in a message.
- Category
    - The Shannon-Fano algorithm can be classified as a recursive algorithm. It operates by repeatedly splitting the set of symbols into two subgroups, until each subset consists of just one symbol.
    - The algorithm calculates the probability of each symbol in the subset at each step of the recursion and then assigns a binary code to each symbol depending on its probability. The binary code is constructed by appending a "0" or "1" to the code assigned to the symbol in the previous step.
- General description
    - The purpose of the algorithm is to encode a message in a way that requires the least amount of space possible, while assigning variable-length codes to symbols in a given message in such a way that the most frequent symbols are assigned shorter codes, and the least frequent symbols are assigned longer codes.
- Steps in the algorithm
    - Here is a step-by-step walkthrough of the Shannon-Fano algorithm:
        - i. Calculate the probability of each symbol in the message.
        - ii. Sort the symbols in decreasing order of probability.
        - iii. Divide the set of symbols into two subsets such that the sum of the probabilities of the symbols in each subset is as close as possible to half the total probability.
        - iv. Assign a '0' to all symbols in the first subset and a '1' to all symbols in the second subset.
        - v. Repeat steps 3-4 recursively for each group until each group contains only one symbol.
        - vi. If either subset contains more than one symbol, repeat steps 2-3 recursively on each subset until each subset contains only one symbol.
        - vii. Concatenate the codes assigned to each symbol to form the compressed message.
- The time and space complexity of the algorithm
    - The time complexity of the Shannon-Fano algorithm is O(n log n), where n is the number of symbols in the message. This is because the algorithm involves sorting the symbols, which takes O(n log n) time.
    - The space complexity of the algorithm is O(n), as it requires storing the probability and code for each symbol in the message.