BASIC RECOMMENDATION ENGINE TASK - DOCUMENTATION

Approach/Methodology:

The Dataset consist of job roles and their associated skills. Converted this data to simple dataframe and then for faster lookup and use converted this to dictionary and skills to sets(for easy comparison).

Used MultiLabelBinarizer to transform skills and binary vectors to compute the similarity based on this binary vectors based on input role binary vector.

Used Jaccard Similarity for computing similarity. Formula for calculating Jaccard Similarity is,

$$J(A, B) = |A \text{ intersection } B| / |A \text{ union } B|$$

Where

Intersection is common elements between A and B

Union is unique elements between A and B

This formula measures how similar two sets of skills are by comparing their intersection dividing by union.

Final Recommendation is given by calculating jaccard similarity of input role to the each role of data. And finally top three roles are recommended by sorting based on the jaccard similarity score.

Why I chose Jaccard Similarity Technique:

I used Jaccard Similarity because it is good option for categorical data. It is best suited because skills are unique and separate values. It handles this type of data very well.

On other hand Cosine Similarity focuses more on the frequency of data, means it focuses on how frequent is data. But in this data skills are either present or absent. Frequency is not a issue here.

And lastly Higher Jaccard Similarity means there is higher overlapping skills between two roles. And best recommends best three closest roles.

This approach ensures a simple yet effective recommendation engine for job roles based on required skills.

Thank You!