Objective :

Implement a user-friendly search functionality using a provided sample dataset of 5000 user queries.
when a user initiates a specific search, the system should perform the following actions:

1) Analyze the user's query.
2) Retrieve relevant results from the dataset.
3) Rank the results based on relevance.
4) Display the top 3/5 relevant results to the user.

The solution should handle:

- Semantic Understanding
- Misspelled Words Handling
- Multi-Word Query
- No Results

Bonus objective :

Binary Classification - Do an exploratory analysis of the dataset provided, decide on feature selection, preprocessing before training a model to classify as class ‘0’ or class ‘1’.

- training_set.csv - To be used as training and validation set - 3910 records, 57 features, 1 output
- test_set.csv (without Ground Truth) - 691 records, 57 features

Submission should include:

- Readme file - explaining any relevant thought process as well as the general approach for the task
- Model performance analysis on validation set in terms of various risks
- A script that generates/prints the performance of model as in step 2 for a validation set
- Model predictions for the test dataset.
- A notebook/script showcasing the EDA/Feature selection and preprocessing steps.
- A list of dependencies/libraries & their versions to run the code.

Note:

- Candidate should split the training data in training and validation set with ratio of 4:1 to evaluate performance of the model on validation set.
- Any classification model can be used, as deemed appropriate for the task. The candidate is free to explain the model selection process also if any.
- The scope of this assignment is not to solely evaluate the accuracy of the model, rather to review the complete process and solution approach followed.

Timelines

The assignment should be submitted within 2-3 days of receiving it. Extension can be permitted if necessary and should be asked for, before the deadline is over.