HomeWork-6 Report Template

1. Data Preparation
   a. Data Selection
      i. Description of the QREL-based document selection process
         1. The QREL-based document selection process involves selecting documents based on relevance judgments provided in the QREL file. Each query in the QREL file is associated with relevant documents along with their relevance scores. These relevance scores indicate the degree of relevance of each document to its corresponding query.
         2. During the document selection process, only the relevant documents, as indicated by the QREL file, are included for further analysis or model training.
      ii. Explanation of non-relevant document inclusion
         1. Non-relevant documents, i.e., those not included in the QREL file or marked as irrelevant in the file, are typically excluded from the document selection process. Including non-relevant documents can introduce noise and adversely affect the performance of retrieval models or machine learning algorithms trained on this data.
   b. Data Splitting
      i. Methodology for splitting data into training and testing queries
         1. Utilized a random splitting approach to divide the queries into training and testing sets. Specifically, it shuffles the unique query IDs and then selects the first 20 for training (train_queries) and the remaining for testing (test_queries). This ensures that both sets have a diverse representation of queries. Subsequently, the code filters the DataFrame based on these selected query IDs to obtain separate training and testing datasets.

2. Feature Extraction
   a. Document-Query IR Features
      i. Detailed explanation of the feature extraction process
         1. The feature extraction process involves computing various scores we used for HW1 but running it for all the documents.
      ii. List of IR models and features used (e.g., BM25, Language Models)
         1. Okapi
         2. BM25
         3. ES
         4. TF-IDF
         5. LMLaplacian
         6. LMJM
      iii. Approach to handling documents outside the top 1000 rankings
         1. Reran the HW1 scoring for all the documents

b. Additional Features
 i. Description of any additional features used (e.g., document length, PageRank)
  1. Did not use additional feature

3. Machine Learning Model
 a. Training the Model
  i. Description of the learning algorithm(s) used
   1. Logistic Regression
   2. Random Forest
   3. Neural Network

 b. Model Testing and Evaluation
  i. Methodology for testing the model on the 5 testing queries
   1. Once the model is trained on the training data, it is applied to the testing queries by making predictions on the corresponding documents. The predicted relevance scores are then evaluated against the true relevance judgments (QREL scores) using appropriate evaluation metrics.
  ii. Approach for evaluating the model on the 20 training queries
   1. The training queries are used for model training, and the evaluation is conducted on the same set of queries after training. This allows assessing how well the model generalizes to unseen data (testing queries) while also evaluating its performance on data it was trained on.
  iii. Description of treceval application and results interpretation
   1.

4. Results and Analysis
 a. Testing Performance
  i. Presentation and analysis of results from testing queries

```
1.  Error due to 5
2.
3.  Queryid (Num):        5
4.  Total number of documents over all queries
5.     Retrieved:      5000
6.     Relevant:        266
7.     Rel_ret:         250
8.  Interpolated Recall - Precision Averages:
9.     at 0.00          0.9111
10.    at 0.10          0.6402
11.    at 0.20          0.5800
12.    at 0.30          0.5172
```

```
13.    at 0.40       0.5007
14.    at 0.50       0.4650
15.    at 0.60       0.3929
16.    at 0.70       0.3524
17.    at 0.80       0.2801
18.    at 0.90       0.1062
19.    at 1.00       0.0206
20. Average precision (non-interpolated) for all rel
    docs(averaged over queries)
21.                  0.4059
22. Precision:
23.  At    5 docs:   0.5600
24.  At   10 docs:   0.5000
25.  At   15 docs:   0.4667
26.  At   20 docs:   0.4600
27.  At   30 docs:   0.4400
28.  At  100 docs:   0.3260
29.  At  200 docs:   0.2070
30.  At  500 docs:   0.0964
31.  At 1000 docs:   0.0500
32. R-Precision (precision after R (= num_rel for a query)
    docs retrieved):
33.    Exact:        0.4442
34.
```

    ii.   Detailed treceval results and interpretation
- 1. Precision at various document cutoffs ranged from 5% to 56%, indicating the proportion of relevant documents retrieved.
- 2. Average precision (non-interpolated) was 40.59%, suggesting the model's overall effectiveness in retrieving relevant documents.
- 3. R-Precision was 44.42%, showing good performance in retrieving relevant documents early in the retrieval list.

b. Training Performance
    i.   Discussion of the model's performance on training data
- 1. Precision at different document cutoffs ranged from 5% to 44%, with lower values compared to testing performance.
- 2. Average precision (non-interpolated) was 26.83%, indicating slightly lower effectiveness compared to testing.

3. R-Precision was 29.30%, indicating relatively good performance in retrieving relevant documents early in the retrieval list.
ii. Comparative analysis of training vs. testing performance
1. The model exhibited better performance on testing queries compared to training queries, which shows the model is generalized.

5. Extra Credit (If Attempted)
a. Description of the extra credit task(s) undertaken. Show the results (if any)
i. Trained a Random Forest Model
ii. Trained a Neural Network Model.