

Report 6

1. Data Preparation

- Documents are filtered based on their presence in the QREL file, which is read_relevanc() function. This file is parsed to map query IDs to their respective documents and their relevance labels, ensuring that only assessed documents are considered for further processing. For each document noted in the QREL, corresponding scores from various IR models such as Elasticsearch, BM25, TF-IDF, and Language Models and etc are gathered. This is done by reading scores from pre-generated files (e.g., read_es_scores(), read_bm_scores(), etc.), ensuring that each document is represented by a comprehensive set of features.
- Non-relevant documents are sourced from previous IR runs (as from HW1), particularly those not listed in the top results by the retrieval models but are essential for providing a broader context and preventing the model from being biased towards only relevant documents. Approximately 1000 non-relevant documents per query are included to enrich the training and testing datasets. This step is crucial for training the model to distinguish between relevant and non-relevant documents effectively.

2. Feature Extraction

- Scores from different IR models are read from pre-processed output files through functions like read_es_scores(), read_bm_scores() etc. Each function parses a specific file and maps each document to its score for a particular query.
- A DataFrame is constructed where each row represents a query-document pair, and each column corresponds to a feature derived from the IR models' scores. This DataFrame df is central to both the training and testing phases of the machine learning model.
- If a document does not have a score from an IR run (because it is not in the top 1000), a default score of 0 is assigned. This approach assumes that documents outside the top 1000 are less relevant, but not entirely irrelevant.
- This default scoring is applied consistently across all IR models to maintain uniformity in feature representation, ensuring that the absence of a score does not distort the model's learning process.

3. Machine Learning Model

- Linear regression provides a straightforward and interpretable model where the relationship between features (IR scores) and the target variable (document relevance) is modeled as a linear combination of the input features.
- Linear regression was implemented using the LinearRegression class from sklearn.linear_model, ensuring a quick and efficient training process. This model

was trained using features extracted from various IR scores and additional document characteristics, with the relevance judgment as the target variable.

- Using the previously constructed DataFrame, data pertaining to the 5 testing queries is isolated. This subset of data is used to assess the model's performance in predicting document relevance.
- The trained model is then used to predict relevance scores for each document in the testing dataset. These scores represent the model's assessment of how relevant each document is to its respective query.
- Documents are ranked based on the predicted relevance scores for each query. This ranking is crucial as it simulates how documents would be presented to users in a real-world scenario.

4. Results and Analysis

Queryid (Num): 5

Total number of documents over all queries

Retrieved: 5000

Relevant: 147

Rel_ret: 147

Interpolated Recall - Precision Averages:

at 0.00 0.7512

at 0.10 0.5335

at 0.20 0.4778

at 0.30 0.4636

at 0.40 0.4207

at 0.50 0.4049

at 0.60 0.3334

at 0.70 0.2504

at 0.80 0.1583

at 0.90 0.1139

at 1.00 0.0766

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.3263

Precision:

At 5 docs: 0.3600

At 10 docs: 0.4400

At 15 docs: 0.3867

At 20 docs: 0.3700

At 30 docs: 0.3267

At 100 docs: 0.1760

At 200 docs: 0.1020

At 500 docs: 0.0524

At 1000 docs: 0.0294

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.3634

Queryid (Num): 100
Total number of documents over all queries
Retrieved: 1000
Relevant: 31
Rel_ret: 31
Interpolated Recall - Precision Averages:
at 0.00 1.0000
at 0.10 0.8333
at 0.20 0.6000
at 0.30 0.4615
at 0.40 0.4483
at 0.50 0.3913
at 0.60 0.3800
at 0.70 0.2973
at 0.80 0.1295
at 0.90 0.0735
at 1.00 0.0529
Average precision (non-interpolated) for all rel docs(averaged over queries)
0.4220
Precision:
At 5 docs: 0.8000
At 10 docs: 0.6000
At 15 docs: 0.6000
At 20 docs: 0.4500
At 30 docs: 0.4333
At 100 docs: 0.2300
At 200 docs: 0.1250
At 500 docs: 0.0580
At 1000 docs: 0.0310
R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.4194

Queryid (Num): 54
Total number of documents over all queries
Retrieved: 1000
Relevant: 51
Rel_ret: 51
Interpolated Recall - Precision Averages:
at 0.00 1.0000
at 0.10 0.6923
at 0.20 0.6000
at 0.30 0.5484
at 0.40 0.4800
at 0.50 0.3881
at 0.60 0.3523
at 0.70 0.2836
at 0.80 0.2320
at 0.90 0.1716

at 1.00 0.0576
Average precision (non-interpolated) for all rel docs(averaged over queries)
0.4261
Precision:
At 5 docs: 0.8000
At 10 docs: 0.6000
At 15 docs: 0.6000
At 20 docs: 0.6000
At 30 docs: 0.5333
At 100 docs: 0.3200
At 200 docs: 0.2100
At 500 docs: 0.0980
At 1000 docs: 0.0510
R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.4706

Queryid (Num): 56
Total number of documents over all queries
Retrieved: 1000
Relevant: 167
Rel_ret: 167
Interpolated Recall - Precision Averages:
at 0.00 1.0000
at 0.10 1.0000
at 0.20 0.9722
at 0.30 0.9123
at 0.40 0.8481
at 0.50 0.7500
at 0.60 0.6824
at 0.70 0.6467
at 0.80 0.5019
at 0.90 0.2581
at 1.00 0.2581
Average precision (non-interpolated) for all rel docs(averaged over queries)
0.7195

Precision:
At 5 docs: 1.0000
At 10 docs: 1.0000
At 15 docs: 1.0000
At 20 docs: 1.0000
At 30 docs: 0.9667
At 100 docs: 0.7900
At 200 docs: 0.6050
At 500 docs: 0.2940
At 1000 docs: 0.1670
R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.6707

Queryid (Num): 58

Total number of documents over all queries

Retrieved: 1000

Relevant: 70

Rel_ret: 69

Interpolated Recall - Precision Averages:

at 0.00 1.0000

at 0.10 1.0000

at 0.20 1.0000

at 0.30 1.0000

at 0.40 0.9655

at 0.50 0.8800

at 0.60 0.8800

at 0.70 0.8167

at 0.80 0.5490

at 0.90 0.5164

at 1.00 0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.8149

Precision:

At 5 docs: 1.0000

At 10 docs: 1.0000

At 15 docs: 1.0000

At 20 docs: 1.0000

At 30 docs: 0.9333

At 100 docs: 0.5500

At 200 docs: 0.3300

At 500 docs: 0.1380

At 1000 docs: 0.0690

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.7714

Queryid (Num): 59

Total number of documents over all queries

Retrieved: 1000

Relevant: 291

Rel_ret: 270

Interpolated Recall - Precision Averages:

at 0.00 1.0000

at 0.10 0.3964

at 0.20 0.3964

at 0.30 0.3964

at 0.40 0.3964

at 0.50 0.3964

at 0.60 0.3964

at 0.70 0.3250

at 0.80 0.3094

at 0.90 0.2821

at 1.00 0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.3186

Precision:

At 5 docs: 0.6000

At 10 docs: 0.5000

At 15 docs: 0.3333

At 20 docs: 0.3500

At 30 docs: 0.4333

At 100 docs: 0.3100

At 200 docs: 0.2950

At 500 docs: 0.3700

At 1000 docs: 0.2700

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.3608

Queryid (Num): 60

Total number of documents over all queries

Retrieved: 1000

Relevant: 7

Rel_ret: 7

Interpolated Recall - Precision Averages:

at 0.00 1.0000

at 0.10 1.0000

at 0.20 0.2857

at 0.30 0.0349

at 0.40 0.0349

at 0.50 0.0202

at 0.60 0.0202

at 0.70 0.0202

at 0.80 0.0202

at 0.90 0.0202

at 1.00 0.0202

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.1983

Precision:

At 5 docs: 0.2000

At 10 docs: 0.2000

At 15 docs: 0.1333

At 20 docs: 0.1000

At 30 docs: 0.0667

At 100 docs: 0.0300

At 200 docs: 0.0150

At 500 docs: 0.0140

At 1000 docs: 0.0070

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.2857

Queryid (Num): 61

Total number of documents over all queries

Retrieved: 1000

Relevant: 24
Rel_ret: 20
Interpolated Recall - Precision Averages:
at 0.00 0.3750
at 0.10 0.3750
at 0.20 0.2963
at 0.30 0.2963
at 0.40 0.2245
at 0.50 0.1600
at 0.60 0.1079
at 0.70 0.1000
at 0.80 0.0426
at 0.90 0.0000
at 1.00 0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.1731

Precision:
At 5 docs: 0.2000
At 10 docs: 0.3000
At 15 docs: 0.2667
At 20 docs: 0.2500
At 30 docs: 0.2667
At 100 docs: 0.1300
At 200 docs: 0.0850
At 500 docs: 0.0400
At 1000 docs: 0.0200

R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.2500

Queryid (Num): 62
Total number of documents over all queries
Retrieved: 1000
Relevant: 115
Rel_ret: 115

Interpolated Recall - Precision Averages:
at 0.00 0.4545
at 0.10 0.4464
at 0.20 0.4464
at 0.30 0.4259
at 0.40 0.4259
at 0.50 0.3602
at 0.60 0.3286
at 0.70 0.2485
at 0.80 0.1870
at 0.90 0.1669
at 1.00 0.1274

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.3159

Precision:

At 5 docs: 0.4000
At 10 docs: 0.4000
At 15 docs: 0.4000
At 20 docs: 0.3500
At 30 docs: 0.3333
At 100 docs: 0.4100
At 200 docs: 0.3250
At 500 docs: 0.1840
At 1000 docs: 0.1150

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.4174

Queryid (Num): 63

Total number of documents over all queries

Retrieved: 1000

Relevant: 1

Rel_ret: 1

Interpolated Recall - Precision Averages:

at 0.00 0.0098
at 0.10 0.0098
at 0.20 0.0098
at 0.30 0.0098
at 0.40 0.0098
at 0.50 0.0098
at 0.60 0.0098
at 0.70 0.0098
at 0.80 0.0098
at 0.90 0.0098
at 1.00 0.0098

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.0098

Precision:

At 5 docs: 0.0000
At 10 docs: 0.0000
At 15 docs: 0.0000
At 20 docs: 0.0000
At 30 docs: 0.0000
At 100 docs: 0.0000
At 200 docs: 0.0050
At 500 docs: 0.0020
At 1000 docs: 0.0010

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.0000

Queryid (Num): 64

Total number of documents over all queries

Retrieved: 1000

Relevant: 123

Rel_ret: 114

Interpolated Recall - Precision Averages:

at 0.00	0.6667
at 0.10	0.2714
at 0.20	0.1406
at 0.30	0.1250
at 0.40	0.1208
at 0.50	0.1163
at 0.60	0.1163
at 0.70	0.1163
at 0.80	0.1163
at 0.90	0.1161
at 1.00	0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.1348

Precision:

At 5 docs:	0.4000
At 10 docs:	0.4000
At 15 docs:	0.3333
At 20 docs:	0.3000
At 30 docs:	0.2000
At 100 docs:	0.1900
At 200 docs:	0.1350
At 500 docs:	0.1060
At 1000 docs:	0.1140

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.1545

Queryid (Num): 68

Total number of documents over all queries

Retrieved: 1000

Relevant: 13

Rel_ret: 13

Interpolated Recall - Precision Averages:

at 0.00	1.0000
at 0.10	0.3333
at 0.20	0.0845
at 0.30	0.0845
at 0.40	0.0845
at 0.50	0.0761
at 0.60	0.0656
at 0.70	0.0204
at 0.80	0.0204
at 0.90	0.0136
at 1.00	0.0136

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.1443

Precision:

At 5 docs:	0.2000
At 10 docs:	0.2000

At 15 docs: 0.1333
At 20 docs: 0.1000
At 30 docs: 0.0667
At 100 docs: 0.0700
At 200 docs: 0.0450
At 500 docs: 0.0180
At 1000 docs: 0.0130

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.1538

Queryid (Num): 71

Total number of documents over all queries

Retrieved: 1000

Relevant: 132

Rel_ret: 125

Interpolated Recall - Precision Averages:

at 0.00 1.0000
at 0.10 0.9474
at 0.20 0.8182
at 0.30 0.6308
at 0.40 0.3462
at 0.50 0.2012
at 0.60 0.1641
at 0.70 0.1583
at 0.80 0.1386
at 0.90 0.1288
at 1.00 0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.4002

Precision:

At 5 docs: 0.8000
At 10 docs: 0.9000
At 15 docs: 0.9333
At 20 docs: 0.9000
At 30 docs: 0.8000
At 100 docs: 0.5200
At 200 docs: 0.2750
At 500 docs: 0.1580
At 1000 docs: 0.1250

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.3939

Queryid (Num): 80

Total number of documents over all queries

Retrieved: 1000

Relevant: 5

Rel_ret: 5

Interpolated Recall - Precision Averages:

at 0.00 0.0125

at 0.10	0.0125
at 0.20	0.0125
at 0.30	0.0060
at 0.40	0.0060
at 0.50	0.0060
at 0.60	0.0060
at 0.70	0.0060
at 0.80	0.0060
at 0.90	0.0060
at 1.00	0.0060

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.0061

Precision:

At 5 docs:	0.0000
At 10 docs:	0.0000
At 15 docs:	0.0000
At 20 docs:	0.0000
At 30 docs:	0.0000
At 100 docs:	0.0100
At 200 docs:	0.0050
At 500 docs:	0.0020
At 1000 docs:	0.0050

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact:	0.0000
--------	--------

Queryid (Num): 85

Total number of documents over all queries

Retrieved:	1000
------------	------

Relevant:	393
-----------	-----

Rel_ret:	361
----------	-----

Interpolated Recall - Precision Averages:

at 0.00	1.0000
at 0.10	0.6066
at 0.20	0.6000
at 0.30	0.6000
at 0.40	0.6000
at 0.50	0.6000
at 0.60	0.5983
at 0.70	0.5983
at 0.80	0.5844
at 0.90	0.4102
at 1.00	0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.5127

Precision:

At 5 docs:	0.4000
At 10 docs:	0.5000
At 15 docs:	0.6000
At 20 docs:	0.5500

At 30 docs: 0.4667

At 100 docs: 0.5700

At 200 docs: 0.5700

At 500 docs: 0.5860

At 1000 docs: 0.3610

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.5903

Queryid (Num): 87

Total number of documents over all queries

Retrieved: 1000

Relevant: 47

Rel_ret: 45

Interpolated Recall - Precision Averages:

at 0.00 0.2712

at 0.10 0.2712

at 0.20 0.2712

at 0.30 0.2712

at 0.40 0.2593

at 0.50 0.2553

at 0.60 0.2267

at 0.70 0.2267

at 0.80 0.2267

at 0.90 0.0808

at 1.00 0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)

0.1898

Precision:

At 5 docs: 0.0000

At 10 docs: 0.0000

At 15 docs: 0.0000

At 20 docs: 0.0000

At 30 docs: 0.0667

At 100 docs: 0.2400

At 200 docs: 0.1950

At 500 docs: 0.0840

At 1000 docs: 0.0450

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.1915

Queryid (Num): 89

Total number of documents over all queries

Retrieved: 1000

Relevant: 26

Rel_ret: 25

Interpolated Recall - Precision Averages:

at 0.00 1.0000

at 0.10 0.3158

at 0.20 0.3158

at 0.30	0.2553
at 0.40	0.2553
at 0.50	0.2540
at 0.60	0.2540
at 0.70	0.1439
at 0.80	0.0775
at 0.90	0.0313
at 1.00	0.0000

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.2180

Precision:

At 5 docs:	0.2000
At 10 docs:	0.2000
At 15 docs:	0.1333
At 20 docs:	0.3000
At 30 docs:	0.2333
At 100 docs:	0.1700
At 200 docs:	0.0950
At 500 docs:	0.0440
At 1000 docs:	0.0250

R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.2692

Queryid (Num): 91

Total number of documents over all queries

Retrieved: 1000

Relevant: 1

Rel_ret: 1

Interpolated Recall - Precision Averages:

at 0.00	0.0016
at 0.10	0.0016
at 0.20	0.0016
at 0.30	0.0016
at 0.40	0.0016
at 0.50	0.0016
at 0.60	0.0016
at 0.70	0.0016
at 0.80	0.0016
at 0.90	0.0016
at 1.00	0.0016

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.0016

Precision:

At 5 docs:	0.0000
At 10 docs:	0.0000
At 15 docs:	0.0000
At 20 docs:	0.0000
At 30 docs:	0.0000
At 100 docs:	0.0000

At 200 docs: 0.0000
At 500 docs: 0.0000
At 1000 docs: 0.0010
R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.0000

Queryid (Num): 93
Total number of documents over all queries
Retrieved: 1000
Relevant: 94
Rel_ret: 94
Interpolated Recall - Precision Averages:
at 0.00 1.0000
at 0.10 1.0000
at 0.20 0.9200
at 0.30 0.9118
at 0.40 0.8837
at 0.50 0.7966
at 0.60 0.1913
at 0.70 0.1913
at 0.80 0.1913
at 0.90 0.1860
at 1.00 0.1025
Average precision (non-interpolated) for all rel docs(averaged over queries)
0.5523

Precision:
At 5 docs: 1.0000
At 10 docs: 1.0000
At 15 docs: 0.9333
At 20 docs: 0.9000
At 30 docs: 0.9000
At 100 docs: 0.4700
At 200 docs: 0.2500
At 500 docs: 0.1820
At 1000 docs: 0.0940
R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.5000

Queryid (Num): 98
Total number of documents over all queries
Retrieved: 1000
Relevant: 9
Rel_ret: 9
Interpolated Recall - Precision Averages:
at 0.00 0.6667
at 0.10 0.6667
at 0.20 0.6667
at 0.30 0.6000
at 0.40 0.5000

at 0.50	0.5000
at 0.60	0.5000
at 0.70	0.5000
at 0.80	0.5000
at 0.90	0.1875
at 1.00	0.1875

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.4715

Precision:

At 5 docs:	0.6000
At 10 docs:	0.4000
At 15 docs:	0.4667
At 20 docs:	0.4000
At 30 docs:	0.2667
At 100 docs:	0.0900
At 200 docs:	0.0450
At 500 docs:	0.0180
At 1000 docs:	0.0090

R-Precision (precision after R (= num_rel for a query) docs retrieved):

Exact: 0.4444

Queryid (Num): 99

Total number of documents over all queries

Retrieved: 1000

Relevant: 85

Rel_ret: 85

Interpolated Recall - Precision Averages:

at 0.00	0.5625
at 0.10	0.5625
at 0.20	0.4286
at 0.30	0.3611
at 0.40	0.3434
at 0.50	0.2775
at 0.60	0.2684
at 0.70	0.2500
at 0.80	0.1739
at 0.90	0.1338
at 1.00	0.1016

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.2933

Precision:

At 5 docs:	0.2000
At 10 docs:	0.3000
At 15 docs:	0.5333
At 20 docs:	0.5000
At 30 docs:	0.4667
At 100 docs:	0.3400
At 200 docs:	0.2600
At 500 docs:	0.1460

At 1000 docs: 0.0850
R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.3412
Error due to 20

Queryid (Num): 20
Total number of documents over all queries
Retrieved: 20000
Relevant: 1685
Rel_ret: 1608

Interpolated Recall - Precision Averages:

at 0.00	0.7010
at 0.10	0.5371
at 0.20	0.4433
at 0.30	0.3966
at 0.40	0.3617
at 0.50	0.3220
at 0.60	0.2775
at 0.70	0.2480
at 0.80	0.2009
at 0.90	0.1397
at 1.00	0.0469

Average precision (non-interpolated) for all rel docs(averaged over queries)
0.3161

Precision:

At 5 docs:	0.4400
At 10 docs:	0.4250
At 15 docs:	0.4200
At 20 docs:	0.4025
At 30 docs:	0.3717
At 100 docs:	0.2720
At 200 docs:	0.1935
At 500 docs:	0.1271
At 1000 docs:	0.0804

R-Precision (precision after R (= num_rel for a query) docs retrieved):
Exact: 0.3342

- The evaluation of our machine learning model's performance on the testing queries, using trec_eval, reveals a range of outcomes that reflect both the strengths and limitations of our approach. For instance, in the detailed trec_eval results for query 85, the average precision was recorded at 0.5127, indicating a moderate level of effectiveness in identifying relevant documents across various levels of recall. This precision level, coupled with the precision metrics at different document cut-offs—such as a precision of 0.4000 at the top 5 documents and 0.5700 at the top 100—suggests that while our model performs reasonably well at surfacing relevant documents initially, there is room for improvement in maintaining this precision across a larger set of retrieved documents. The R-Precision value of 0.5903 further underscores this point, indicating that our model is

capable of retrieving a significant proportion of relevant documents, but the precision drops as more documents are pulled into the result set.

- In contrast, for query 91, the performance dramatically drops with an average precision of just 0.0016, demonstrating the model's difficulty with this particular query, where even at higher retrieval thresholds, the interpolated recall-precision averages remained consistently low, barely rising above zero. This stark variance in performance across different queries highlights the challenges in creating a universally effective IR model and underscores the need for targeted improvements in both our feature engineering and model training approaches, especially to enhance the handling of queries that currently yield poor results.