

## TF-IDF vs BM25 Experiment Result

Nama : Muhammad Faisal Adi Soesatyo

NPM : 1906293184

No	Scoring Regimes	RBP, $p = .8$	DCG	AP
1.	<b>TF-IDF</b>	0.6172251405 195845	12.8592269 31024727	0.49196440 25103477
2.	<b>BM25, <math>k1 = 1.35, b = .75</math></b>	0.6196988970 370321	12.9004571 15683992	0.49666612 34590155
3.	<b>BM25, <math>k1 = 1.5, b = .75</math></b>	0.6238950198 274699	12.9474494 18461526	0.50084601 86189042
4.	<b>BM25, <math>k1 = 1.75, b = .75</math></b>	0.6365672035 687002	13.0306235 92553203	0.50789537 36510926
5.	<b>BM25, <math>k1 = 1.35, b = .65</math></b>	0.6212782774 266953	12.8939767 38409647	0.49637975 824408376
6.	<b>BM25, <math>k1 = 1.5, b = .5</math></b>	0.6205046652 944286	12.9263204 1450316	0.49919835 677982927
7.	<b>BM25, <math>k1 = 1.75, b = .35</math></b>	0.6221414546 009594	12.9428726 50741046	0.50299454 87887156

Legend:

- : highest score (overall) → the best
- : lowest score (overall) → the worst

Pada tabel di atas dapat terlihat bahwa secara keseluruhan hasil dari BM25 mengungguli TF-IDF. Dengan catatan nilai dari  $k1$  dan  $b$  pada BM25 masing-masing  $1.2 \leq k1 \leq 2$  dan  $b \leq .75$ . Pada hasil BM25 dapat terlihat bahwa,  $k1$  dan  $b$  masing-masing memiliki peran yang besar dalam ketiga metrik di atas. Contoh pada nomor 3 dan 6, meskipun memiliki nilai  $k1$  (tf scaling) yang sama, namun nilai  $b$  (doc length norm) pada nomor 3 lebih besar dari nomor 6, menghasilkan score yang berbeda. Semakin besar gap pada nilai  $b$ , maka score yang dihasilkan pun semakin berbeda jauh (contoh pada nomor 4 dan 7). Berlaku pada yang memiliki nilai  $b$  yang sama, contoh nomor 2, 3 dan 4. Semakin besar selisih dari nilai  $k1$ , maka semakin besar pula selisih pada metric scorenya.

Makna dari masing-masing skor pada tabel metric score di atas adalah. Semakin tinggi scorenya, menandakan bahwa algoritma BM25 mengembalikan dokumen yang semakin relevan dengan query. Sebaliknya, apabila semakin rendah scorenya, maka algoritma BM25 mengembalikan dokumen yang semakin tidak relevan dengan query. Semakin tinggi nilai  $k_1$  (term frequency scaling), menandakan jumlah masing-masing term frequency semakin mendapat leverage (seolah-olah menjadi banyak). Semakin tinggi nilai  $b$  (document length normalisation), menandakan ukuran dari dokumen yang dioperasikan semakin mendekati realita. Dengan demikian masuk akal apabila semakin tinggi nilai  $k_1$  dan  $b$  pada BM25 maka metric score yang muncul akan semakin besar a.k.a hasil yang dikembalikan semakin baik/relevan terhadap query.