|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Avg. Precision@10 | Avg. Recall@10 | Avg. nDCG@10 | Avg. Coverage@10 | Avg. Diversity@10 |
| Audio-based(cosine, mfcc\_stats) |  |  |  |  | 4.81123 |
| Audio-based(cosine,  Blf-correlation) |  |  |  |  | 4.89424 |
| Audio-based(cosine, ivec 256) |  |  |  |  | 4.90361 |
| Audio-based(cosine,  musicnn) |  |  |  |  | 4.7048 |
| Text-based(cosine, tf-idf) |  |  |  |  | 4.97433 |
| Text-based(cosine,  word2vec) |  |  |  |  | 4.8492 |
| Text-based(cosine,  Bert) |  |  |  |  | 4.84481 |
| Text-based(Random-Baseline) |  |  |  |  |  |
| Video-Based(cosine,vgg19) |  |  |  |  | 4.97154 |
| Early-Fusion |  |  |  |  |  |
| Late-Fusion |  |  |  |  |  |