|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Avg. Precision@10 | Avg. Recall@10 | Avg. nDCG@10 | Avg. Coverage@10 | Avg. Diversity@10 |
| Audio-based(cosine, mfcc\_stats) | 0.428874356 | 0.001762932 |  | 0.038229897 | 4.81123 |
| Audio-based(cosine,  Blf-correlation) | 0.409848410 | 0.001744662 |  | 0.0382780904142002 | 4.89424 |
| Audio-based(cosine, ivec 256) | 0.434363853 | 0.001946139 |  | 0.03822989749338124 | 4.90361 |
| Audio-based(cosine,  musicnn) | 0.481252477 | 0.002228871 |  | 0.03839898698204262 | 4.7048 |
| Text-based(cosine, tf-idf) | 0.379369798 | 0.001416642 |  | 0.03772497653356882 | 4.97433 |
| Text-based(cosine,  word2vec) | 0.385087198 | 0.001694450 |  | 0.03785609165768803, | 4.8492 |
| Text-based(cosine,  Bert) | 0.419193421 | 0.002046413 |  | 0.0343963508707004 | 4.84481 |
| Random-Baseline |  |  |  |  |  |
| Video-Based(cosine,vgg19) | 0.327831170 | 0.000955034 |  | 0.037394347 | 4.97154 |
| Videobased(cosine,resnet) | 0.334132567 | 0.000969763 |  | 0.037849994 |  |
| Videobased(cosine,incp) | 0.336903111 | 0.000969458 |  | 0.037822241 |  |
| Early-Fusion |  |  |  |  |  |
| Late-Fusion |  |  |  |  |  |