|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Avg. Precision@10 | Avg. Recall@10 | Avg. nDCG@10 | Avg. Coverage@10 | Avg. Diversity@10 |
| Audio-based(cosine, mfcc\_stats) | 0.4289 | 0.0018 | 0.7577 | 0.0382 | 4.8112 |
| Audio-based(cosine,  Blf-correlation) | 0.4098 | 0.0017 | 0.6653 | 0.0382 | 4.8942 |
| Audio-based(cosine, ivec 256) | 0.4344 | 0.0019 | 0.6587 | 0.0382 | 4.9036 |
| Audio-based(cosine,  musicnn) | 0.4813 | 0.0022 | 0.7137 | 0.0383 | 4.7048 |
| Text-based(cosine, tf-idf) | 0.3794 | 0.0014 | 0.6820 | 0.0377 | 4.9743 |
| Text-based(cosine,  word2vec) | 0.3851 | 0.0017 | 0.6639 | 0.0378 | 4.8492 |
| Text-based(cosine,  Bert) | 0.4192 | 0.0020 | 0.6820 | 0.0343 | 4.8448 |
| Random-Baseline | 0.3274 | 0.0010 |  | 0.0373 |  |
| Video-Based(cosine,vgg19) | 0.3278 | 0.0010 | 0.7176 | 0.0373 | 4.9715 |
| Early-Fusion | 0.3202 | 0.0009 |  | 0.0373 |  |
| Late-Fusion | 0.3301 | 0.0009 |  | 0.0373 |  |