

## BAB 6

### KESIMPULAN DAN SARAN

#### 6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan pada tugas akhir ini, diambil beberapa kesimpulan sebagai berikut:

1. Algoritma distribusi normal multivariat dapat digunakan dalam pembuatan sistem rekomendasi pada data implisit dengan memanfaatkan sifat distribusi normal yang ada pada setiap baris interaksi. Setiap baris interaksi ini memiliki sifat distribusi normal, sehingga seluruh baris interaksi membentuk distribusi normal multivariat.
2. Pada data implisit, terdapat *item-item* yang tidak berinteraksi dengan pengguna sehingga terdapat *missing value* atau *item* yang memiliki nol interaksi dengan pengguna. Dengan menggunakan distribusi normal multivariat, interaksi yang nol atau kosong tersebut dapat diprediksi nilai interaksinya dan nilai interaksinya ini dapat diurutkan dari nilai interaksi terbesar ke yang terkecil untuk dibuatkan rekomendasi *item-item* dengan nilai interaksi tertinggi untuk pengguna.
3. Dari hasil penelitian pada tugas akhir ini menggunakan data MovieLens, didapatkan hasil bahwa akurasi dari pemodelan distribusi normal multivariat memiliki hasil  $\text{precision}@k$  dan  $\text{nDCG}@k$  yang kurang baik dengan nilai  $\text{precision}@k$  dan  $\text{nDCG}@k$  berkisar 0,2 hingga 0,3 pada metode pembagian data *train test split*. Lalu dengan memperbesar rasio interaksi yang dihilangkan ( $s$ ) menjadi 0,7 dapat meningkatkan hasil evaluasi  $\text{precision}@k$  dan  $\text{nDCG}$  menjadi senilai 0,3 hingga 0,5. Tetapi dengan menaikkan nilai  $s$  atau rasio interaksi yang dihilangkan memiliki resiko *overfitting* model, karena hasil prediksi memiliki peluang yang lebih besar untuk masuk ke *top k*. Untuk itu dilakukan pengujian menggunakan Kfold dengan  $K = 5$  dan menggunakan rasio  $s = 0,5$  untuk menghindari *overfitting*. Hasilnya dengan menggunakan Kfold dengan nilai  $s$  yang sama, akurasi  $\text{precision}@k$  dan  $\text{nDCG}@k$  sedikit meningkat. KFold disini mengurangi ketidakstabilan data dengan menguji sebanyak 5 data berbeda, sehingga dapat menghasilkan akurasi yang lebih baik. Dengan hasil evaluasi model pada rasio  $s = 0,7$ , nilai  $\text{precision}@$  dan  $\text{nDG}@k$  tersebut dapat dibilang kurang memberikan rekomendasi yang cukup relevan. Walaupun hasilnya kurang baik, hasil prediksi masih bisa digunakan untuk sistem rekomendasi karena pengguna bisa terbuka untuk mengeksplorasi berbagai pilihan film walaupun kurang sesuai dengan minat pengguna dan memiliki toleransi yang terhadap rekomendasi yang kurang presisi.

#### 6.2 Saran

Berdasarkan penelitian yang telah dilakukan pada skripsi ini, diambil beberapa saran untuk dijadikan bahan perbaikan di kemudian hari, yaitu sebagai berikut:

1. Model distribusi normal multivariat yang dipakai pada penelitian ini bisa menggunakan formula yang sudah dimodifikasi, yaitu formula yang memprediksi setiap interaksi dengan berbasis yang berinteraksi atau yang tidak berinteraksi. Dengan formula yang sudah dimodifikasi ini diharapkan akan mendapatkan hasil yang lebih baik, tetapi waktu *runtime* kode untuk

memprediksi akan lebih lama dari memakai formula yang biasa, dikarenakan matriks kovarian akan berbeda terus apabila beda *item* pada satu pengguna, sehingga membutuhkan waktu lebih untuk membentuk matriks kovarian baru setiap *item*-nya. Untuk penelitian selanjutnya dapat mencoba formula yang dimodifikasi itu untuk melihat perbedaan hasil rekomendasi dengan menggunakan formula biasa.

2. Dapat mencoba nilai K yang lain pada Kfold untuk melihat hasil model untuk nilai K yang berbeda. Pada penelitian ini tidak dilakukan karena keterbatasan waktu karena Kfold memakan waktu yang tidak sedikit.
3. Dapat membuat *website* yang dapat melakukan prediksi dengan melakukan *input* data manual dan dapat mengeluarkan hasil prediksi berupa rekomendasi film. Pada penelitian ini tidak diimplementasikan karena keterbatasan waktu dan apabila diimplementasikan, proses pembuatan rekomendasi akan membutuhkan waktu yang cukup lama untuk menampilkan hasil rekomendasi, sehingga menjadi tidak efektif.

## DAFTAR PUSTAKA

- [1] Isinkaye, F., Folajimi, Y., dan Ojokoh, B. (2015) Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, **16**, 261–273.
- [2] Chalmers, D. (2018) *Probability Statistics 1*, 1st edition. University Printing House, United Kingdom.
- [3] Kashikar, A. S. (2020) Multivariate normal distribution. Academia.
- [4] Markus Viljanen, T. P. (2020) New recommendation algorithm for implicit data motivated by the multivariate normal distribution. arXiv.
- [5] Müller, A. C. dan Guido, S. (2016) *Introduction to Machine Learning with Python*. O'Reilly Media, Sebastopol.
- [6] Francesco Ricci, B. S., Lior Rokach (2015) *Recommender Systems: Introduction and Challenges*, 2nd edition. Springer, London.
- [7] Gentle, J. E. (2007) *Matrix Algebra: Theory, computations, and applications in statistics*, 1st edition. Springer Science, New York.