

ABSTRAK

Penelitian ini dilatarbelakangi oleh kesulitan penonton dalam menemukan film yang sesuai dengan selera mereka di tengah banyaknya film yang diproduksi. Rating film yang ada saat ini sering kali hanya didasarkan pada penilaian langsung oleh penonton tanpa mempertimbangkan faktor seperti *genre*, kategori usia penonton, dan deskripsi cerita film. Penelitian ini bertujuan untuk memprediksi rating film menggunakan pendekatan *Neural Factorization Machines* (NFM). Metode penelitian meliputi *data preparation* yang mencakup penggabungan file dataset, pemetaan kategori usia, pembersihan data, konversi teks menjadi huruf kecil, penghapusan *regular expression*, penghapusan teks selain bahasa Inggris, tokenisasi, *lemmatizing*, *word embedding*, *one-hot encoding*, dan *label encoding*. Proses modeling dilakukan dengan membangun model NFM yang terdiri dari *input fitur*, *embedding layer*, *bi-interaction layer*, *hidden layers*, dan *prediction score*. Evaluasi model dilakukan dengan mengatur *hyperparameter* yaitu *epoch* dan *batch size* untuk mengoptimalkan kinerja model. Penelitian ini dilakukan dengan 9 kali pengujian menggunakan kombinasi *epoch* (30, 50, dan 100) dan *batch size* (64, 128, dan 256). Hasil evaluasi menunjukkan bahwa nilai MSE terendah yang berarti terbaik pada data *training* adalah 1,181 dengan *batch size* 256 dan *epoch* 100, serta pada data *validation* adalah 1,230 dengan *batch size* 256 dan *epoch* 100. Namun, pada data test, konfigurasi dengan *batch size* 128 dan *epoch* 50 memberikan MSE terbaik sebesar 1,280. Meskipun model menunjukkan performa terbaik pada data *training* dan *validation* dengan konfigurasi *batch size* 256 dan *epoch* 100, grafik evaluasi mengindikasikan adanya *overfitting*. Temuan ini menunjukkan bahwa model NFM mampu dalam memprediksi *rating* film berdasarkan *genre*, kategori usia penonton, dan deskripsi cerita film.

Kata kunci: *Neural Factorization Machines* (NFM), prediksi, *rating* film, *epoch*, *batch size*, *Mean Squared Error* (MSE)

ABSTRACT

This study is motivated by the difficulty viewers face in finding films that match their preferences amid the large volume of film productions. Current film ratings are often based solely on direct viewer feedback without considering factors such as genre, age category of viewers, and film descriptions. This research aims to predict film ratings using the Neural Factorization Machines (NFM) approach. The research methodology includes data preparation involving dataset merging, age category mapping, data cleaning, text lowercasing, removal of regular expressions, removal of non-English text, tokenization, lemmatization, word embedding, one-hot encoding, and label encoding. The modeling process involves constructing an NFM model consisting of input features, embedding layers, bi-interaction layers, hidden layers, and prediction scores. Model evaluation is conducted by tuning hyperparameters, specifically epoch and batch size, to optimize model performance. The study includes 9 tests using combinations of epochs (30, 50, and 100) and batch sizes (64, 128, and 256). Evaluation results show that the lowest MSE, indicating the best performance, for the training data is 1.181 with a batch size of 256 and epoch 100, and for the validation data is 1.230 with a batch size of 256 and epoch 100. However, for the test data, the configuration with a batch size of 128 and epoch 50 yields the best MSE of 1.280. Although the model shows the best performance on training and validation data with the configuration of batch size 256 and epoch 100, the evaluation graph indicates overfitting. The findings suggest that the NFM model is capable of predicting film ratings based on genre, age category of viewers, and film descriptions.

Keywords: *Neural Factorization Machines (NFM), prediction, movie ratings, epoch, batch size, Mean Squared Error (MSE)*