

1. Pendahuluan

1.1. Latar Belakang

Pengguna media digital mengharapkan pengalaman yang personal dan rekomendasi yang relevan[1]. Sistem pemberi rekomendasi merupakan alat yang dapat memenuhi kebutuhan tersebut. Kecerdasan buatan dalam sistem pemberi rekomendasi dan search engine sangat membantu dalam memanifestasikan informasi dari data yang tak terbatas[2]. Beberapa sistem pemberi rekomendasi yang populer seperti Netflix, Apple TV, dan Disney+ membuat inisiatif untuk mengetahui apa yang diinginkan pengguna dan memberikan rekomendasi produk yang relevan[3].

Penggunaan sistem pemberi rekomendasi mampu menyelesaikan masalah tentang kelebihan informasi yang didapat pengguna dengan memberikan hanya *item* yang sesuai dengan minat dan *behaviour* pengguna [4]. Saat ini telah dibuat berbagai macam metode pada sistem pemberi rekomendasi, yang dapat memanfaatkan collaborative filtering, content-based filtering atau hybrid filtering [5].

Collaborative filtering adalah metode yang umum digunakan pada sistem pemberi rekomendasi, dengan menggunakan data rating user lain yang memiliki preferensi yang mirip [6]. Pada user-item matrix yang digunakan pada collaborative filtering menjadi sangat besar dan tersebar dimana-mana atau biasa disebut *sparsity problem* dan *cold start problem* disebabkan oleh data yang kurang pada *user*[7].

Content-Based Filtering adalah metode yang menggunakan data deskripsi pada *item* untuk merekomendasikan *item* yang mirip dengan apa yang *user* inginkan [8]. Karena content-based dibuat berdasarkan data *item*, maka content-based tidak memerlukan data user lain, hal ini dapat mempermudah penskalaan di sejumlah *user* [9] dan dapat menyelesaikan masalah *cold start problem* pada user.

Untuk menyelesaikan permasalahan yang terdapat pada kedua metode tersebut dapat dilakukan dengan cara mengkombinasikan kedua metode tersebut atau biasa disebut *hybrid filtering* [7]. Dengan *Hybrid filtering* sistem pemberi rekomendasi akan dapat teroptimasi dengan memanfaatkan kelebihan dan menutupi kelemahan dari kedua metode tersebut.

Dalam tugas akhir ini, digunakan algoritma Restricted Boltzmann Machine (RBM) karena efektif dalam melakukan encoding distribusi dan lebih efisien secara komputasi. Selain itu, aktivasi pada lapisan tersembunyi dapat digunakan sebagai fitur untuk meningkatkan kinerja model[10]. RBM dimanfaatkan dalam membangun model dengan memanfaatkan nilai dari masing-masing film pada dataset *MovieLens*.

Dan pada content-based filtering menggunakan Word2Vec dikarenakan pada *dataset* film akan memanfaatkan data deskripsi dari masing-masing film. Word2Vec dapat digunakan untuk membuat

word embedding pada data deskripsi film yang berkualitas tinggi, kemudian algoritma ini dapat dilatih pada *corpora* yang besar dengan sangat cepat, dan juga dapat digunakan untuk membuat neural network yang mampu memahami bahasa natural [11].

1.2. Perumusan Masalah

Rumusan masalah pada tugas akhir ini adalah:

1. Bagaimana pengimplementasian metode hybrid dengan Word2vec dan Restricted Boltzmann Machine dapat memberikan rekomendasi pada domain film ?
2. Bagaimana kualitas *ranking* pada metode hybrid dengan Word2vec dan Restricted Boltzmann Machine ?

1.3. Tujuan

Tujuan dari tugas akhir ini adalah:

1. Membuat sistem pemberi rekomendasi dengan menggunakan metode hybrid dengan Word2vec dan Restricted Boltzmann Machine yang dapat membantu memberi saran film kepada pengguna.
2. Mengukur kualitas dari *ranking* film yang disarankan dari penggunaan metode hybrid dengan Word2vec dan Restricted Boltzmann Machine.

1.4. Batasan Masalah

Batasan masalah pada tugas akhir ini adalah:

1. Dataset yang digunakan tidak lebih dari 100.000 data film dan *user rating* pada film.
2. Sumber dataset yang digunakan hanya menggunakan dataset dari *MovieLens*.