p-ISSN: 2301-5373
e-ISSN: 2654-5101

# Implementasi Variational Autoencoder dan Regularized Singular Value Decomposition pada Sistem Rekomendasi Film

**Ida Bagus Gede Basudewa Weda[a1], Ngurah Agus Sanjaya ER[a2], Cokorda Rai Adi Pramartha[a3], Gst. Ayu Vida Mastrika Giri[a4]**
[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana, Jl. Raya Kampus Unud, Jimbaran, 80361, Badung-Bali
[1]weda.2208561077@student.unud.ac.id
[2]agus_sanjaya@unud.ac.id
[3]cokorda@unud.ac.id
[4]vida@unud.ac.id

**ABSTRAK**

Melimpahnya pilihan film yang tersedia menimbulkan tantangan bagi penonton dalam menemukan konten yang sesuai dengan preferensi mereka. Penelitian ini mengusulkan sistem rekomendasi film berbasis hybrid ensemble yang menggabungkan Variational Autoencoder (VAE) dengan KL annealing dan Regularized Singular Value Decomposition (RSVD) menggunakan pendekatan weighted average ensemble pada dataset MovieLens 100K. Evaluasi dilakukan menggunakan 10-fold cross-validation, pengujian final pada hold-out test set, dan uji signifikansi statistik melalui paired t-test. Hasil penelitian menunjukkan bahwa hybrid ensemble dengan bobot alpha = 0,25 untuk VAE dan beta = 0,75 untuk RSVD mencapai RMSE sebesar 0,9307 ± 0,0043 dan MAE sebesar 0,7338 ± 0,0034 pada cross-validation, serta RMSE sebesar 0,9365 dan MAE sebesar 0,7401 pada test set, mengungguli RSVD standalone (RMSE 0,9474) dan VAE standalone (RMSE 1,0926) secara konsisten. Seluruh perbedaan performa dikonfirmasi signifikan secara statistik melalui paired t-test dengan p-value jauh di bawah 0,05.

**Kata Kunci :** Variational Autoencoder, Regularized Singular Value Decomposition, Sistem Rekomendasi Film, Hybrid Ensemble, Collaborative Filtering, KL Annealing

**ABSTRACT**

*The abundance of available films presents a challenge for viewers in finding content that matches their preferences. This study proposes a hybrid ensemble movie recommendation system combining a Variational Autoencoder (VAE) with KL annealing and Regularized Singular Value Decomposition (RSVD) using a weighted average approach on the MovieLens 100K dataset. Evaluation is conducted using 10-fold cross-validation, final testing on a hold-out test set, and statistical significance testing via paired t-test. Results show that the hybrid ensemble with weights alpha = 0.25 for VAE and beta = 0.75 for RSVD achieves an RMSE of 0.9307 ± 0.0043 and MAE of 0.7338 ± 0.0034 on cross-validation, and RMSE of 0.9365 and MAE of 0.7401 on the test set, consistently outperforming standalone RSVD (RMSE 0.9474) and standalone VAE (RMSE 1.0926). All performance differences are confirmed statistically significant by paired t-test with p-values well below 0.05.*

**Keywords :** *Variational Autoencoder, Regularized Singular Value Decomposition, Movie Recommendation System, Hybrid Ensemble, Collaborative Filtering, KL Annealing*

## 1. PENDAHULUAN

Industri perfilman kembali tumbuh signifikan pascapandemi COVID-19, ditandai dengan kenaikan produksi film dan pendapatan *global box office* pada periode 2022–2023 [1]. Melimpahnya konten

film yang tersedia memberikan banyak pilihan kepada penonton, namun sekaligus menimbulkan tantangan untuk menemukan film yang sesuai dengan preferensi mereka. Sistem rekomendasi hadir sebagai solusi atas tantangan ini, yaitu sistem yang berfungsi memperkirakan dan menyajikan informasi yang sesuai dengan preferensi seorang pengguna [2].

Salah satu pendekatan sistem rekomendasi yang dominan adalah *collaborative filtering* (CF), yang memberikan rekomendasi berdasarkan kesamaan preferensi antar pengguna. Tantangan utama CF adalah *data sparsity*, di mana matriks pengguna-item memiliki banyak entri kosong karena pengguna hanya berinteraksi dengan sebagian kecil item yang tersedia. Berbagai penelitian telah mengusulkan faktorisasi matriks berbasis SVD untuk mengatasi masalah ini. Rajput et al. [3] menggabungkan *standard autoencoder* dengan SVD pada sistem rekomendasi multi-kriteria, sementara Wibisono et al. [4] menerapkan SVD dengan *item-based collaborative filtering*. Namun, seluruh penelitian tersebut menggunakan *standard autoencoder* dan SVD tanpa regularisasi, sehingga terdapat gap penelitian yang belum dieksplorasi.

Penelitian ini mengajukan pendekatan *hybrid ensemble* yang menggabungkan *Variational Autoencoder* (VAE) dengan KL *annealing* dan *Regularized Singular Value Decomposition* (RSVD). VAE dipilih karena sifat Bayesian-nya memungkinkan penanganan data *sparse* yang lebih tepat melalui representasi laten sebagai distribusi probabilistik, sehingga menghasilkan *latent space* yang lebih terstruktur dibandingkan *standard autoencoder* [5], [6]. Namun, pelatihan VAE pada data CF berisiko mengalami *posterior collapse*, yaitu kondisi ketika model mengabaikan ruang laten sehingga representasi laten menjadi tidak informatif [7]. Untuk mengatasinya, diterapkan KL *annealing* yang menaikkan bobot β pada KL divergence secara bertahap agar model terlebih dahulu mempelajari rekonstruksi sebelum tekanan regularisasi diberikan secara penuh [6]. Di sisi lain, RSVD mengatasi kelemahan SVD standar yang rentan *overfitting* pada data *sparse* dengan menambahkan penalti regularisasi pada parameter laten [8]. Kombinasi VAE dan RSVD yang belum pernah dieksplorasi sebelumnya ini merupakan kontribusi utama penelitian ini. Dataset MovieLens 100K dipilih sebagai *benchmark* karena merupakan salah satu dataset yang paling banyak digunakan dalam penelitian CF [9].

## 2. METODE PENELITIAN

### 2.1 Dataset dan Preprocessing

Dataset MovieLens 100K diperoleh dari situs grouplens.org dan terdiri dari 100.000 data rating dari 943 pengguna terhadap 1.682 film pada skala 1 hingga 5, dengan tingkat *sparsity* sekitar 93,7%. Data rating tersimpan dalam berkas u.data berformat *tab-separated* dengan empat kolom (user_id, item_id, rating, dan timestamp), sedangkan metadata film termasuk judul dan genre tersimpan dalam berkas u.item dengan 19 kolom genre biner; kolom timestamp tidak digunakan karena model tidak memodelkan dimensi waktu. Data dibagi menggunakan rasio 70% data latih, 10% data validasi, dan 20% data uji dengan strategi *user-level stratified split* untuk menjamin representasi setiap pengguna di ketiga subset. Rating dinormalisasi ke rentang [0, 1] sebelum pelatihan, kemudian dikembalikan ke skala aslinya saat evaluasi.

### 2.2 Variational Autoencoder (VAE)

VAE adalah model generatif yang memetakan setiap pengguna ke distribusi Gaussian di ruang laten dengan parameter rata-rata μ dan log-varians $\log(\sigma^2)$ [6]. Fungsi objektif yang dioptimalkan adalah *Evidence Lower Bound* (ELBO):

$$L = E[q(z|x) \log p(x|z)] - \beta \cdot KL(q(z|x) \,//\, p(z))$$

Suku pertama mengukur kualitas rekonstruksi, sementara suku kedua adalah regularisasi KL divergence yang mendorong distribusi laten mendekati prior Gaussian standar N(0,1). Untuk mencegah *posterior collapse*, nilai β dinaikkan secara bertahap dari 0 hingga 1 selama pelatihan melalui mekanisme KL *annealing*. Arsitektur encoder menggunakan lapisan *hidden* [1024, 512] dengan dropout_rate = 0,1. Proses sampling menggunakan teknik *reparameterization trick* dengan formulasi:

p-ISSN: 2301-5373
e-ISSN: 2654-5101

$$z = \mu + \sigma \cdot \varepsilon, \quad \varepsilon \sim N(0,1)$$

sehingga gradien dapat mengalir melalui operasi sampling selama proses pelatihan.

### 2.3 Regularized Singular Value Decomposition (RSVD)

RSVD memodelkan rating pengguna-item menggunakan dekomposisi matriks dengan penambahan *bias terms*. Prediksi rating diformulasikan sebagai:

$$\hat{r}_{ui} = \mu + b_u + b_i + U_u \cdot \Sigma \cdot V_i^T$$

di mana μ adalah rata-rata global, $b_u$ dan $b_i$ adalah bias pengguna dan item, sedangkan U, Σ, dan V adalah faktor laten [10]. Parameter diperbarui menggunakan *stochastic gradient descent* (SGD) dengan fungsi objektif yang meminimalkan *regularized squared error*:

$$L = \Sigma_{ui} (r_{ui} - \hat{r}_{ui})^2 + \lambda(||U||^2 + ||\Sigma||^2 + ||V||^2 + ||b_u||^2 + ||b_i||^2)$$

Penalti regularisasi λ mengurangi risiko *overfitting* dan meningkatkan stabilitas model pada data *sparse* [8].

### 2.4 Hybrid Ensemble dan Hyperparameter Tuning

Prediksi akhir diperoleh dari kombinasi linear:

$$\hat{r}hybrid = \alpha \cdot \hat{r}VAE + (1-\alpha) \cdot \hat{r}RSVD$$

Nilai α optimal ditentukan pada data validasi dengan mencari nilai α ∈ [0, 1] (dengan langkah 0,01) yang menghasilkan RMSE minimum. *Hyperparameter tuning* untuk VAE dan RSVD dilakukan menggunakan framework Optuna dengan strategi *Tree-structured Parzen Estimator* (TPE). Proses tuning dilaksanakan pada data latih untuk menghindari *data leakage*. Ruang pencarian VAE meliputi: latent_dim ∈ {50, 100, 200}, learning_rate ∈ [1e-4, 1e-2], batch_size ∈ {32, 64, 128}, dropout_rate ∈ [0,0, 0,5], β ∈ [0,001, 1,0], dan hidden_dims dari beberapa pilihan arsitektur. Ruang pencarian RSVD meliputi: n_factors ∈ {10, 20, 50, 100}, learning_rate ∈ [1e-4, 1e-2], dan λ_reg ∈ [1e-5, 1e-1].

### 2.5 Evaluasi Model

Performa model diukur menggunakan RMSE dan MAE. Evaluasi dilakukan melalui dua tahap: (1) 10-fold cross-validation pada gabungan data latih dan validasi untuk mengukur stabilitas model [11], dan (2) pengujian final pada *hold-out test set* yang tidak pernah dilihat model selama pelatihan. Signifikansi statistik perbedaan performa antar model diuji menggunakan *paired t-test* pada distribusi metrik per fold, dengan α = 0,05 sebagai threshold signifikansi [12]. Pendekatan ini mengikuti rekomendasi Wainer & Cawley [13] bahwa untuk kebanyakan aplikasi praktis, evaluasi dengan *hold-out* yang dikombinasikan dengan cross-validation sudah memadai tanpa memerlukan *nested cross-validation*.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Hyperparameter Tuning

Proses *hyperparameter tuning* menggunakan Optuna menghasilkan konfigurasi terbaik sebagai berikut. Konfigurasi VAE: latent_dim = 100, learning_rate = 0,001, batch_size = 64, dropout_rate = 0,1, β = 0,01, dan arsitektur encoder [1024, 512]; konfigurasi RSVD: n_factors = 10, learning_rate = 0,005, dan λ_reg = 0,001. Nilai *latent_dim* = 100 dengan arsitektur encoder [1024, 512] memberikan kapasitas representasi yang memadai sambil mengendalikan *overfitting* melalui dropout. Nilai β = 0,01 yang kecil menunjukkan bahwa tekanan regularisasi KL divergence yang rendah lebih optimal, konsisten dengan manfaat KL *annealing*. Pada RSVD, n_factors = 10 yang relatif kecil mencerminkan bahwa kompleksitas faktor laten yang terlalu tinggi tidak selalu menghasilkan performa lebih baik pada dataset *sparse*.

### 3.2 Hasil 10-Fold Cross-Validation

Tabel 1 menyajikan ringkasan statistik 10-fold cross-validation untuk seluruh metrik. Nilai alpha ensemble optimal berkisar antara 0,23 hingga 0,26 dengan rata-rata 0,24 ± 0,01 di seluruh fold, menunjukkan konsistensi bobot yang tinggi. Hybrid ensemble secara konsisten mencapai RMSE dan MAE terkecil di seluruh 10 fold.

**Tabel 1. Ringkasan Statistik 10-Fold Cross-Validation (RMSE dan MAE)**

| Model | RMSE (Mean ± Std) | MAE (Mean ± Std) |
|---|---|---|
| VAE Standalone | 1,0926 ± 0,0080 | 0,8447 ± 0,0071 |
| RSVD Standalone | 0,9474 ± 0,0040 | 0,7464 ± 0,0028 |
| **Hybrid Ensemble** | **0,9307 ± 0,0043** | **0,7338 ± 0,0034** |

Standar deviasi RMSE hybrid (0,0043) yang lebih rendah daripada VAE standalone (0,0080) menunjukkan penggabungan kedua model juga meningkatkan stabilitas prediksi antar fold, dengan RMSE hybrid yang konsisten pada kisaran 0,9217–0,9368.

### 3.3 Hasil Uji Signifikansi Statistik (Paired t-Test)

Tabel 2 menyajikan hasil *paired t-test* pada distribusi RMSE dan MAE per fold. Seluruh pasangan perbandingan menunjukkan *t-statistic* yang sangat besar secara absolut dan *p-value* yang jauh di bawah threshold α = 0,05, mengonfirmasi bahwa perbedaan performa bukan disebabkan oleh variasi acak.

**Tabel 2. Hasil Paired t-Test Antar Model (RMSE dan MAE)**

| Perbandingan | t (RMSE) | p (RMSE) | t (MAE) | p (MAE) | Signifikan? |
|---|---|---|---|---|---|
| Hybrid vs VAE | -77,48 | $5{,}03\times10^{-14}$ | -67,11 | $1{,}83\times10^{-13}$ | Ya (α=0,05) |
| Hybrid vs RSVD | -20,25 | $8{,}12\times10^{-9}$ | -17,09 | $3{,}62\times10^{-8}$ | Ya (α=0,05) |
| RSVD vs VAE | -52,05 | $1{,}79\times10^{-12}$ | -42,74 | $1{,}05\times10^{-11}$ | Ya (α=0,05) |

### 3.4 Hasil Pengujian Final pada Test Set

Hasil evaluasi pada *hold-out test set* yang tidak pernah dilihat model selama proses pelatihan ditunjukkan pada Tabel 3.

**Tabel 3. Hasil Pengujian pada Test Set**

| Model | RMSE | MAE | Δ RMSE vs Hybrid | Δ MAE vs Hybrid |
|---|---|---|---|---|
| VAE Standalone | 1,0876 | 0,8446 | +16,13% | +14,12% |
| RSVD Standalone | 0,9506 | 0,7508 | +1,51% | +1,45% |
| **Hybrid Ensemble (α=0,25)** | **0,9365** | **0,7401** | — | — |

Pada test set, hybrid ensemble mencapai RMSE 0,9365, mengungguli RSVD standalone (0,9506) dan VAE standalone (1,0876). Konsistensi antara hasil cross-validation (0,9307) dan test set (0,9365) menunjukkan generalisasi yang baik, dan performa ini kompetitif terhadap baseline *collaborative filtering* Sami et al. [9] pada dataset yang sama (RMSE 0,912–0,943) meskipun metodologinya tidak sepenuhnya identik.

### 3.5 Contoh Hasil Rekomendasi

Sebagai pelengkap metrik agregat di atas, ditampilkan contoh keluaran rekomendasi bagi satu pengguna. Pengguna dengan user_id 655 dipilih karena memiliki rating terbanyak pada data uji (148 rating). Sistem memprediksi rating seluruh film yang belum dinilai pengguna, mengurutkannya dari prediksi tertinggi, lalu menyajikan film teratas sebagai rekomendasi. Akurasi prediksi diverifikasi dahulu pada film yang telah dinilai (Tabel 4), kemudian sepuluh film teratas yang belum ditonton disajikan sebagai rekomendasi (Tabel 5).

p-ISSN: 2301-5373
e-ISSN: 2654-5101

**Tabel 4. Perbandingan Rating Asli dan Prediksi (Pengguna 655)**

| Judul Film | Rating Asli | Prediksi | Selisih |
|---|---|---|---|
| Belle de jour (1967) | 5 | 3,35 | 1,65 |
| The Fugitive (1993) | 5 | 3,35 | 1,65 |
| Citizen Kane (1941) | 4 | 3,93 | 0,07 |
| Big Night (1996) | 4 | 3,34 | 0,66 |
| Mighty Aphrodite (1995) | 3 | 2,90 | 0,10 |
| The White Balloon (1995) | 3 | 2,64 | 0,36 |
| Get Shorty (1995) | 2 | 3,19 | 1,19 |
| Seven (Se7en) (1995) | 2 | 3,15 | 1,15 |
| Amateur (1994) | 1 | 2,87 | 1,87 |

Tabel 4 membandingkan rating asli dan prediksi *hybrid* untuk pengguna contoh di berbagai level rating. Prediksi sangat akurat pada rentang menengah (misalnya Citizen Kane: rating 4, prediksi 3,93; Mighty Aphrodite: rating 3, prediksi 2,90), namun pada rating ekstrem menyusut ke nilai tengah: film dirating 5 diprediksi sekitar 3,35 dan film dirating 1 sekitar 2,87. Pola regresi ke nilai tengah ini lazim pada model berbasis minimisasi galat kuadrat di data *sparse*. Meski demikian, rata-rata selisih absolut untuk pengguna ini hanya 0,56, lebih rendah daripada MAE global 0,7401, menandakan *error* besar terkonsentrasi pada rating ekstrem sementara mayoritas prediksi tetap dekat nilai sebenarnya.

**Tabel 5. Sepuluh Film Rekomendasi Teratas untuk Pengguna 655**

| Peringkat | Judul Film | Genre | Prediksi Rating |
|---|---|---|---|
| 1 | Wallace & Gromit: The Best of Aardman Animation (1996) | Animation | 3,96 |
| 2 | The Wrong Trousers (1993) | Animation, Comedy | 3,80 |
| 3 | High Noon (1952) | Western | 3,80 |
| 4 | The Sting (1973) | Comedy, Crime | 3,78 |
| 5 | A Close Shave (1995) | Animation, Comedy, Thriller | 3,76 |
| 6 | The Bridge on the River Kwai (1957) | Drama, War | 3,76 |
| 7 | Paths of Glory (1957) | Drama, War | 3,73 |
| 8 | The 39 Steps (1935) | Thriller | 3,72 |
| 9 | The Maltese Falcon (1941) | Film-Noir, Mystery | 3,70 |
| 10 | Apocalypse Now (1979) | Drama, War | 3,68 |

Sepuluh film dengan prediksi *hybrid* tertinggi yang belum pernah dinilai pengguna 655 disajikan pada Tabel 5 sebagai daftar rekomendasi. Rekomendasi yang dihasilkan didominasi film klasik yang diakui secara kritis dengan genre yang mencakup drama, perang, *thriller*, dan *film-noir*. Kecenderungan ini sebagian besar dipengaruhi oleh dominasi kontribusi RSVD (75%) yang memodelkan bias item secara eksplisit, sehingga film yang secara umum bernilai tinggi cenderung

menempati peringkat teratas. Komposisi tersebut juga cukup selaras dengan preferensi pengguna 655 yang memberi rating tinggi pada sejumlah film klasik seperti The Godfather dan Annie Hall.

### 3.6 Analisis Kompleksitas Komputasi

Selain akurasi, biaya komputasi merupakan pertimbangan penting dalam menilai kelayakan pendekatan *hybrid*. Analisis berikut meninjau kompleksitas VAE, RSVD, dan *hybrid ensemble* dari tiga sisi, yaitu ukuran model, kompleksitas asimtotik pelatihan dan inferensi, serta pengukuran empiris waktu dan memori.

Dari sisi ukuran, VAE jauh lebih besar daripada RSVD. Dengan arsitektur *encoder* [1024, 512] dan *latent_dim* 100 pada dimensi masukan-keluaran 1.682, VAE memuat sekitar 4,65 juta parameter terlatih, terdiri atas 2,35 juta parameter pada *encoder* dan 2,30 juta pada *decoder*. Sebaliknya, RSVD hanya memuat sekitar 28,9 ribu parameter yang mencakup matriks faktor U (943×10), V (1.682×10), diagonal Σ, serta *bias* pengguna dan item. Dengan demikian jumlah parameter VAE sekitar 161 kali lebih banyak dibandingkan RSVD, sebagaimana dirangkum pada Tabel 6.

**Tabel 6. Perbandingan Ukuran Model**

| Model | Jumlah Parameter | Rincian |
|---|---|---|
| VAE | ± 4,65 juta | Encoder 1682→1024→512→100; Decoder 100→512→1024→1682 |
| RSVD | ± 28,9 ribu | U (943×10), V (1682×10), Σ (diagonal 10), b_u (943), b_i (1682), μ |

Meskipun jumlah parameternya paling sedikit, pelatihan RSVD justru terukur lebih lama daripada VAE (29,63 detik berbanding 21,92 detik; lihat Tabel 8), sejalan dengan karakteristik kompleksitas berikut. Pelatihan VAE berjalan pada kompleksitas O(E·N·C), dengan E jumlah *epoch*, N jumlah pengguna, dan C biaya satu *forward-backward pass* yang didominasi perkalian matriks lapisan terlebar (1.682×1.024) dan dijalankan secara tervektorisasi per *batch* pada TensorFlow. Pelatihan RSVD mengoptimalkan parameter melalui *stochastic gradient descent* pada rating teramati sehingga secara algoritmik berkompleksitas O(E·|R|·k), dengan |R| jumlah rating dan k faktor laten; namun implementasi pada penelitian ini menelusuri seluruh sel matriks pengguna-item (m×n ≈ 1,59 juta sel) di setiap *epoch* melalui *loop* bersarang Python, sehingga waktu tempuh nyatanya lebih dekat ke O(E·m·n).

Pada tahap inferensi, VAE cukup menjalankan satu *forward pass* deterministik (*encoder* → z_mean → *decoder*) berkompleksitas O(N·C), sedangkan RSVD menghasilkan satu prediksi pengguna-item dalam O(k) atau rekonstruksi penuh dalam satu operasi matriks. *Hybrid ensemble* hanya menambahkan satu penjumlahan berbobot atas kedua prediksi, sehingga *overhead* penggabungan bersifat konstan, O(1) per prediksi, dan praktis dapat diabaikan. Ringkasan kompleksitas asimtotik disajikan pada Tabel 7.

**Tabel 7. Kompleksitas Asimtotik Pelatihan dan Inferensi**

| Proses | VAE | RSVD | Hybrid Ensemble |
|---|---|---|---|
| Pelatihan | O(E·N·C) | O(E·\|R\|·k) | Penjumlahan biaya VAE dan RSVD |
| Inferensi | O(N·C) | O(k) per prediksi | +O(1) per prediksi |

Untuk melengkapi analisis teoretis di atas, waktu pelatihan dan waktu inferensi untuk memproduksi prediksi seluruh pengguna diukur secara langsung pada perangkat dengan prosesor AMD Ryzen 5 7500F (6 inti) dan RAM 32 GB 6000 MT/s; seluruh komputasi dijalankan pada CPU, dengan hasil dirangkum pada Tabel 8. Penggunaan memori puncak proses selama pengukuran tercatat sebesar 672,5 MB, mencakup TensorFlow, kedua model, dan data.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

**Tabel 8. Hasil Pengukuran Empiris Waktu Pelatihan dan Inferensi**

| Model | Waktu Pelatihan | Waktu Inferensi |
|---|---|---|
| VAE Standalone | 21,92 s | 307,25 ms |
| RSVD Standalone | 29,63 s | 5,84 ms |
| Hybrid Ensemble | 51,54 s | 326,11 ms |

Secara keseluruhan, peningkatan akurasi *hybrid ensemble* diperoleh dengan biaya komputasi tambahan yang terbatas. Beban utama terletak pada tahap pelatihan karena dua model harus dilatih dan disimpan alih-alih satu; namun pada tahap inferensi — yang paling relevan bagi penggunaan nyata — penggabungan hanya berupa penjumlahan berbobot dengan overhead terukur sekitar 13 milidetik, sehingga tidak menambah beban yang berarti. Dengan kata lain, perbaikan RMSE sebesar 1,51% terhadap RSVD *standalone* dan 16,13% terhadap VAE *standalone* (Tabel 3) dicapai tanpa mengorbankan efisiensi inferensi, sehingga pendekatan *hybrid* tetap layak diterapkan selama sumber daya pelatihan tersedia.

### 3.7 Diskusi

Dominansi RSVD (beta = 0,75) menunjukkan kestabilan prediksinya lebih diandalkan pada dataset ini, kemungkinan karena *sparsity* tinggi (93,7%) membatasi kemampuan VAE membentuk representasi laten yang stabil saat interaksi per pengguna sangat sedikit. Meskipun demikian, kontribusi VAE (alpha = 0,25) secara konsisten meningkatkan performa *hybrid* di atas RSVD *standalone* di seluruh 10 fold, dan keunggulan ini terkonfirmasi oleh *paired t-test* ($|t| = 20{,}25$, $p = 8{,}12\times10^{-9}$) yang menolak hipotesis nol bahwa performa keduanya setara.

Stabilitas bobot alpha (0,23–0,26; std = 0,01) menunjukkan bobot *ensemble* tidak bersifat *data-dependent* secara berlebihan sehingga model diharapkan berperilaku serupa pada data baru. Nilai $\beta = 0{,}01$ yang sangat kecil juga menandakan mekanisme KL *annealing* berhasil menjaga tekanan regularisasi pada tingkat optimal: cukup untuk mencegah *posterior collapse* tanpa menghambat kemampuan rekonstruksi model.

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, evaluasi hanya menggunakan metrik berbasis *rating prediction error* (RMSE dan MAE), sementara kualitas rekomendasi dalam konteks nyata juga diukur melalui *ranking-based metrics* seperti NDCG dan Precision@K yang belum dievaluasi. Kedua, eksperimen hanya dilakukan pada satu dataset (MovieLens 100K), sehingga generalisasi ke dataset lain dengan karakteristik berbeda belum dapat dikonfirmasi. Ketiga, pendekatan *weighted average ensemble* yang digunakan bersifat sederhana; eksplorasi teknik ensemble yang lebih canggih seperti *stacking* atau *learned ensemble weights* berpotensi memberikan peningkatan performa lebih lanjut.

## 4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi film berbasis *hybrid ensemble* yang menggabungkan VAE dengan KL *annealing* dan RSVD pada dataset MovieLens 100K. Dengan bobot optimal alpha = 0,25 (VAE) dan beta = 0,75 (RSVD), model mencapai RMSE 0,9307 ± 0,0043 dan MAE 0,7338 ± 0,0034 pada 10-fold *cross-validation* serta RMSE 0,9365 pada *test set*, secara konsisten mengungguli RSVD dan VAE *standalone* dengan perbedaan yang signifikan secara statistik (*paired t-test*, p jauh di bawah 0,05). Selain menurunkan galat prediksi, sistem juga mampu menghasilkan daftar rekomendasi terurut yang relevan sebagaimana diilustrasikan pada Sub-bab 3.5. Penelitian selanjutnya disarankan mengeksplorasi *ranking-based evaluation metrics* seperti NDCG dan Precision@K, pengujian pada dataset yang lebih besar, serta integrasi informasi kontekstual seperti metadata film dan demografis pengguna untuk meningkatkan relevansi rekomendasi.

**DAFTAR PUSTAKA**

[1] J. Hancock et al., "Global film industry report 2024," Industry Analysis Report, 2024.

[2] A. Nugroho and S. Rahayu, "Sistem rekomendasi: Tinjauan dan perkembangan," Jurnal Teknologi Informasi, 2020.

[3] D. S. Rajput et al., "Autoencoder dan SVD untuk sistem rekomendasi multi-kriteria," Jurnal Sistem Informasi, 2024.

[4] A. Wibisono et al., "Item-based collaborative filtering dengan SVD untuk reduksi dimensi matriks rating," Jurnal Sistem Informasi, 2021.

[5] J. Bobadilla et al., "Deep variational models for collaborative filtering-based recommender systems," Neural Computing and Applications, 2021.

[6] D. Liang et al., "Variational autoencoders for collaborative filtering: A survey," ACM Computing Surveys, 2024.

[7] Y. Wang et al., "Posterior collapse in variational autoencoder collaborative filtering," in Proceedings of AAAI 2025, 2025.

[8] I. K. G. D. Putra and I. W. Santiyasa, "Implementasi regularized singular value decomposition dalam sistem rekomendasi buku collaborative filtering," JuTISI, Universitas Udayana, 2025.

[9] A. Sami et al., "A deep learning based hybrid recommendation model for internet users," Scientific Reports, 2024.

[10] H. I. Alshbanat et al., "Singular value decomposition with bias terms and stochastic gradient descent for matrix factorization in recommendation systems," Information Fusion, 2025.

[11] M. Z. Abedin et al., "K-fold cross-validation for model performance evaluation," Scientific Reports, 2026.

[12] Rahmani et al., "Paired t-test untuk uji signifikansi dalam penelitian komputasional," JPION, 2025.

[13] J. Wainer and G. Cawley, "Nested cross-validation when selecting classifiers is overzealous for most practical applications," Expert Systems with Applications, 2021.