

BAB IV HASIL DAN PEMBAHASAN

Sebelum membahas hasil imputasi dengan metode mean, max, min, dan KNN-Imputer perlu dipahami karakteristik data yang digunakan dalam penelitian ini. Dataset yang digunakan mencakup peringkat pengguna terhadap item dalam konteks sistem rekomendasi. Data ini bersifat sparse, dengan sejumlah besar nilai yang hilang karena beberapa pengguna hanya memberikan peringkat pada sebagian kecil movie. Untuk sample uji coba saya menggunakan 2500 data dari data rating yang terdiri dari 50 UserId dan 50 MovieID . Tabel berikut ini adalah data yang akan di jadikan contoh

Tabel 4.1 Pivot Tabel Sampel Rating (rata-rata)

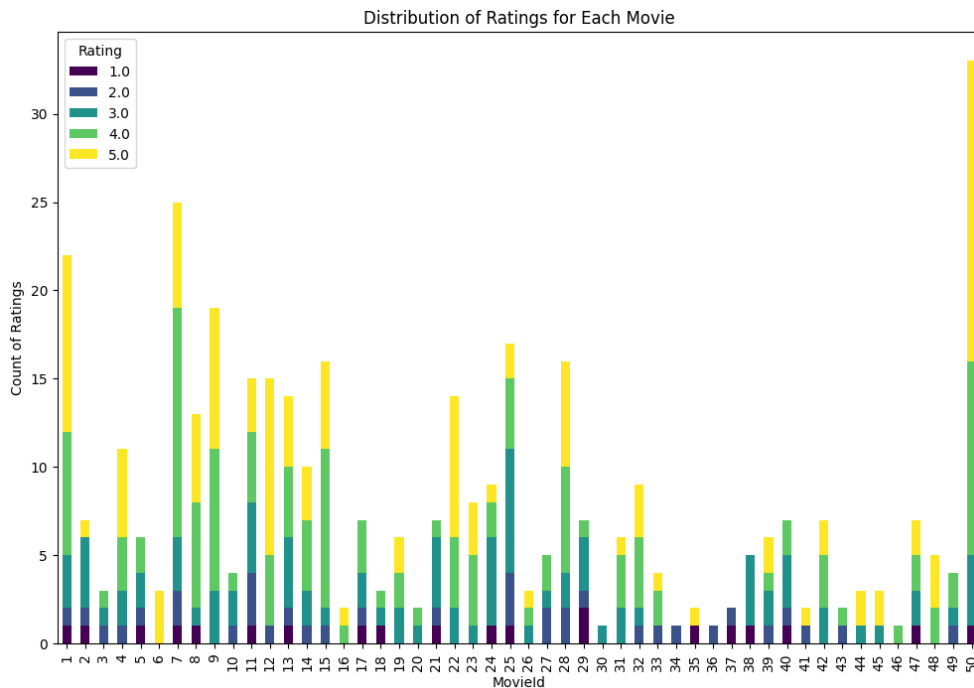
movieId	1	2	3	4	5	...	46	47	48	49	50
userId											
1	5.0	3.0	4.0	3.0	3.0	...	4.0	4.0	5.0	3.0	5.0
2	4.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	5.0
3	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	5.0
5	4.0	3.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	4.0

Dan setelah di hitung diketahui bahwa jumlah sampel data yang kosong dan data yang sudah ada nilai rating nya adalah

- Jumlah nilai yang kosong: 2100
- Jumlah nilai yang tidak kosong: 400
- Jumlah nilai Keseluruhan: 2500
-

Tabel 4.2 Jumlah Rating Sample Data

movieId	1	2	3	4	5	...	46	47	48	49	50
Rating											
1	1	1	0	0	1	...	0	1	0	0	1
2	1	1	1	1	1	...	0	0	0	1	0
3	3	4	1	2	2	...	0	2	0	1	4
4	7	0	1	3	2	...	1	2	2	2	11
5	10	1	0	5	0	...	0	2	3	0	17
Jumlah	22	7	3	11	6	...	1	7	5	4	33



Gambar 4.1 Gambar sebaran rating pada sample data rating

Dari data pada tabel 4.2 dan pada gambar 4.1 di ketahui bahwa tidak semua user memberi rating pada semua movie seperti contoh user melakukan rating pada movie id 1 hanya berjumlah 22 rating sedangkan jumlah data user berjumlah 50 user Id

4.1. Imputation dengan Nilai Rata-Rata

Metode imputasi nilai yang pertama dievaluasi dalam penelitian ini adalah metode mean imputation. Pendekatan ini melibatkan penggantian nilai yang hilang dengan nilai rata-rata dari movieId. Proses ini dijalankan pada setiap nilai rating movieId yang hilang. Hasil Imputation dapat dilihat dalam tabel berikut

Tabel 4.3 Imputation Metode Mean (rata-rata)

movieId	1	2	3	4	5	...	46	47	48	49	50
userId											
1	5	3	4	3	3	...	4	4	5	3	5
2	4	3	3	4	3	...	4	4	5	3	5
3	4	3	3	4	3	...	4	4	5	3	4
4	4	3	3	4	3	...	4	4	5	3	5
5	4	3	3	4	3	...	4	4	5	3	4

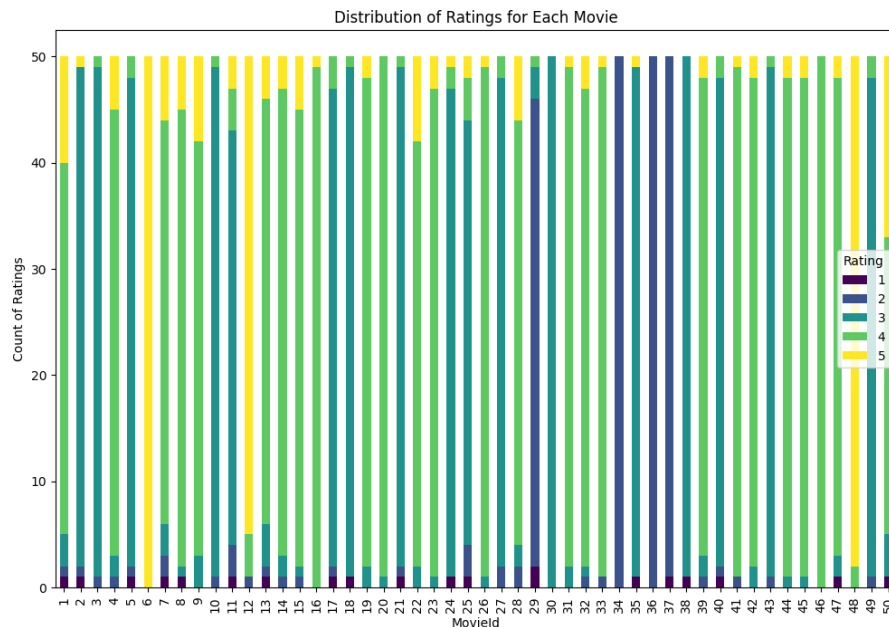
Hasil imputasi dengan metode mean menunjukkan peningkatan signifikan dalam mengurangi jumlah nilai yang hilang. Dengan hasil sebagai berikut

- Jumlah nilai yang hilang sebelum imputasi: 2100
- Jumlah nilai yang hilang setelah imputasi: 0

Seluruh data yang kosong sudah terisi dengan nilai rata-rata dari setiap row dataset dan jika di lihat dari tabel dan grafik sebaran rating berikut ini terlihat bahwa semua movie sudah terrating

Tabel 4.4 Jumlah Rating Mean Imputation

movieId	1	2	3	4	5	...	46	47	48	49	50
rating											
1	1	1	0	0	1	...	0	1	0	0	1
2	1	1	1	1	1	...	0	0	0	1	0
3	3	47	48	2	46	...	0	2	0	47	4
4	35	0	1	42	2	...	50	45	2	2	28
5	10	1	0	5	0	...	0	2	48	0	17
Jumlah	50	50	50	50	50	0	50	50	50	50	50



Gambar 4.2 Sebaran Rating Mean Imputation

Pada tabel 4.4 dan gambar 4.2 grafik menunjukkan bahwa nilai rating sudah terpenuhi dari 50 user untuk 50 movie dan rata-rata pada setiap rating movie paling banyak nilai rating 3,4 dan 5.

Kemudain akan dilakukan pengecekan similarity pada tabel imputation mean (rata-rata). Berikut hasil dari pengecekan similarity

Tabel 4.5 Similarity Imputation Mean

	0	1	2	3	...	46	47	48	49
0	0.000000	0.968228	0.969706	0.968202	...	0.969706	0.966923	0.939474	0.963993
1	0.968228	0.000000	0.996983	0.997055	...	0.996983	0.993976	0.961151	0.993976
2	0.969706	0.996983	0.000000	0.998575	...	1.000.000	0.996984	0.970662	0.996984
3	0.968202	0.997055	0.998575	0.000000	...	0.998575	0.995691	0.965630	0.995691
4	0.962771	0.991599	0.994599	0.992920	...	0.994599	0.991905	0.963586	0.991905

Pada tabel 4.5 diketahui bahwa nilai kedekatan pada setiap nilai rating mencapai rata-rata 0.98

Kemudian Evaluasi kinerja imputasi dilakukan menggunakan Root Mean Square Error (RMSE). Metrik ini digunakan untuk mengevaluasi seberapa baik metode Mean Imputation mampu merekonstruksi nilai yang hilang pada dataset MovieLens.

Hasil dari evaluasi kinerja dengan metrik RMSE adalah sebagai berikut:

Evaluating RMSE of algorithm SVD on 5 split(s).							
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std
RMSE (testset)	0.3630	0.4154	0.4311	0.4710	0.3647	0.4091	0.0411
Fit time	0.02	0.03	0.02	0.02	0.02	0.02	0.00
Test time	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average RMSE: 0.4090736222124957							

4.2. Imputation dengan Nilai Tertinggi

Metode imputasi nilai maksimum melibatkan penggantian nilai yang hilang dengan nilai maksimum dari seluruh dataset. Pendekatan ini diterapkan pada setiap nilai yang hilang dalam data movieId.

Tabel 4.6 Imputation Metode Max (tertinggi)

movieId	1	2	3	4	5	...	46	47	48	49	50
userId											
1	5	3	4	3	3	...	4	4	5	3	5
2	4	5	4	5	4	...	4	5	5	4	5
3	5	5	4	5	4	...	4	5	5	4	5
4	5	5	4	5	4	...	4	5	5	4	5
5	4	3	4	5	4	...	4	5	5	4	4

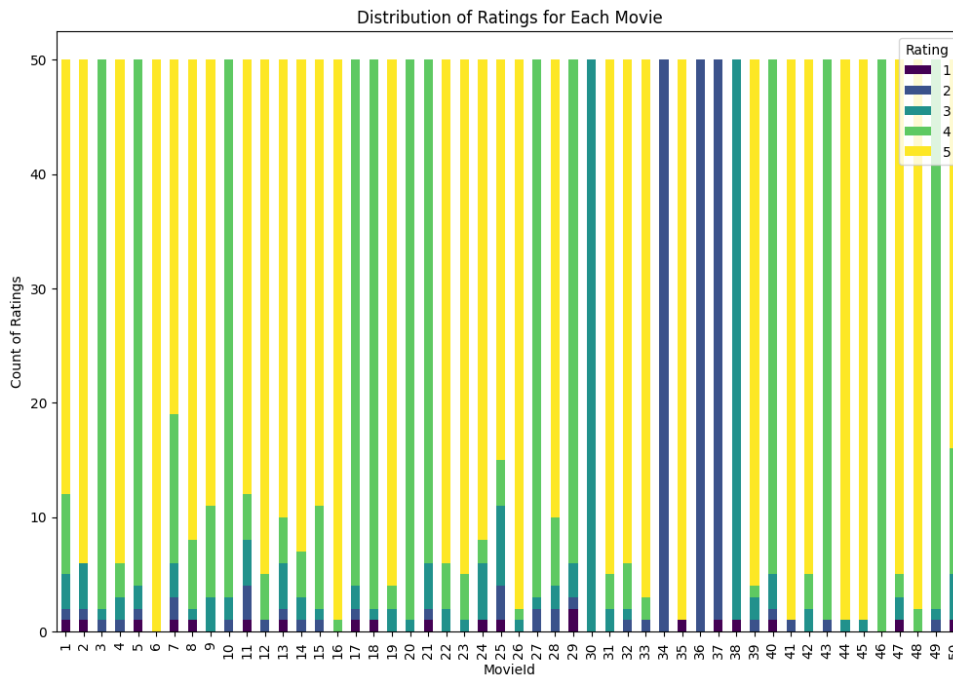
Hasil imputasi dengan metode max juga menunjukkan peningkatan signifikan dalam mengurangi jumlah nilai yang hilang. Dengan hasil sebagai berikut

- Jumlah nilai yang hilang sebelum imputasi: 2100
- Jumlah nilai yang hilang setelah imputasi: 0

Seluruh data yang kosong sudah terisi dengan nilai tertinggi dari setiap row dataset dan jika di lihat dari tabel dan grafik sebaran rating berikut ini terlihat bahwa semua movie sudah terrating

Tabel 4.7 Jumlah Rating Max Imputation

movieId	1	2	3	4	5	...	46	47	48	49	50
rating											
1	1	1	0	0	1	...	0	1	0	0	1
2	1	1	1	1	1	...	0	0	0	1	0
3	3	4	1	2	2	...	0	2	0	1	4
4	7	0	48	3	46	...	50	2	2	48	11
5	38	44	0	44	0	...	0	45	48	0	34
Jumlah	50	50	50	50	50	0	50	50	50	50	50



Gambar 4.3 Sebaran Rating Max Imputation

Pada tabel 4.7 dan gambar 4.3 grafik menunjukkan bahwa nilai rating sudah terpenuhi dari 50 user untuk 50 movie dan rata-rata pada setiap rating movie paling banyak nilai rating tertinggi yaitu 4 dan 5.

Kemudain akan dilakukan pengecekan similarity pada tabel imputation max. Berikut hasil dari pengecekan similarity

Tabel 4.8 Similarity Imputation Max

	0	1	2	3	...	47	48	49	50
0	0.000000	0.945219	0.956130	0.957977	...	0.956130	0.950285	0.916589	0.946761
1	0.945219	0.000000	0.994539	0.993884	...	0.994539	0.989046	0.950331	0.989046
2	0.956130	0.994539	0.000000	0.999526	...	1.000.000	0.995219	0.951887	0.995219
3	0.957977	0.993884	0.999526	0.000000	...	0.999526	0.994653	0.952671	0.994653
4	0.954565	0.990751	0.994566	0.993907	...	0.994566	0.990063	0.962867	0.989052

Pada tabel 4.8 diketahui bahwa nilai kedekatan pada setiap nilai rating mencapai rata-rata 0.95

Kemudian Evaluasi kinerja imputasi dilakukan menggunakan Root Mean Square Error (RMSE). Metrik ini digunakan untuk mengevaluasi seberapa baik metode Max Imputation mampu merekonstruksi nilai yang hilang pada dataset MovieLens. Hasil dari evaluasi kinerja dengan metrik RMSE adalah sebagai berikut:

```

Evaluating RMSE of algorithm SVD on 5 split(s) .

RMSE (testset)      Fold 1  Fold 2  Fold 3  Fold 4  Fold 5  Mean  Std
Fit time           0.02   0.02   0.02   0.02   0.02   0.02  0.00
Test time          0.00   0.11   0.00   0.00   0.00   0.02  0.04
Average RMSE: 0.5261958549773049
    
```

4.3.Imputation dengan Nilai Terendah

Metode imputasi nilai minimum melibatkan penggantian nilai yang hilang dengan nilai minimum dari seluruh dataset. Pendekatan ini digunakan untuk setiap nilai yang hilang pada data movieId.

Tabel 4.9 Imputation Metode Min (terendah)

movieId	1	2	3	4	5	...	46	47	48	49	50
userId											
1	5	3	4	3	3	...	4	4	5	3	5
2	4	1	2	2	1	...	4	1	4	2	5
3	1	1	2	2	1	...	4	1	4	2	1
4	1	1	2	2	1	...	4	1	4	2	5
5	4	3	2	2	1	...	4	1	4	2	4

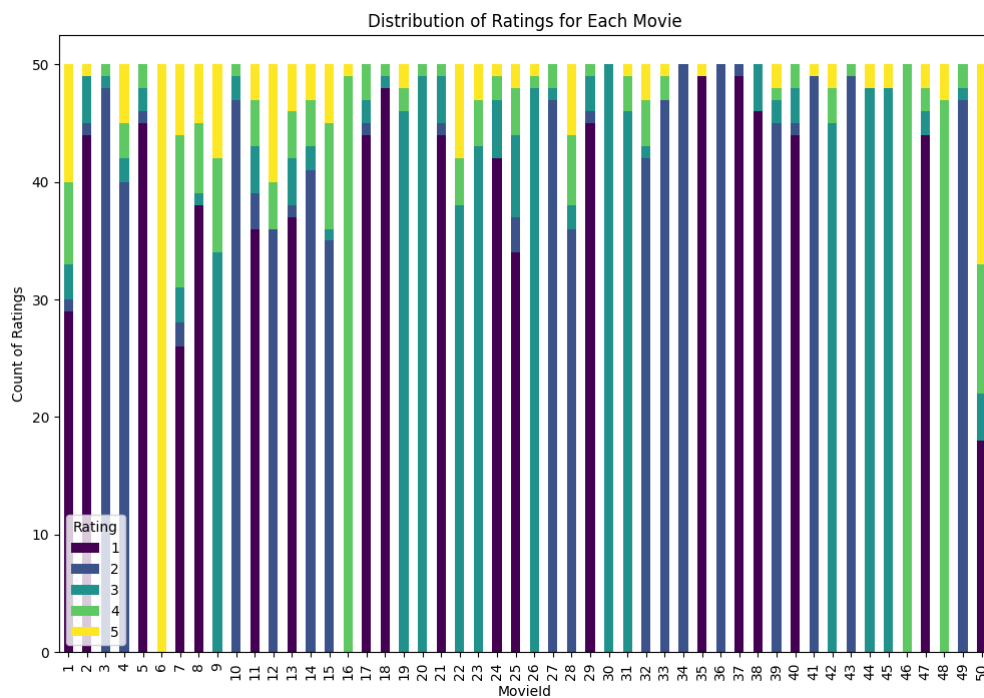
Hasil imputasi dengan metode min juga menunjukkan peningkatan signifikan dalam mengurangi jumlah nilai yang hilang. Dengan hasil sebagai berikut

- Jumlah nilai yang hilang sebelum imputasi: 2100
- Jumlah nilai yang hilang setelah imputasi: 0

Seluruh data yang kosong sudah terisi dengan nilai tertinggi dari setiap row dataset dan jika di lihat dari tabel dan grafik sebaran rating berikut ini terlihat bahwa semua movie sudah terrating

Tabel 4.10 Jumlah Rating Max Imputation

movieId	1	2	3	4	5	...	46	47	48	49	50
rating											
1	29	44	0	0	45	...	0	44	0	0	18
2	1	1	48	40	1	...	0	0	0	47	0
3	3	4	1	2	2	...	0	2	0	1	4
4	7	0	1	3	2	...	50	2	47	2	11
5	10	1	0	5	0	...	0	2	3	0	17
Jumlah	50	50	50	50	50	...	50	50	50	50	50



Gambar 4.4 Sebaran Rating Min Imputation

Pada tabel 4.7 dan gambar 4.4 grafik menunjukkan bahwa nilai rating sudah terpenuhi dari 50 user untuk 50 movie dan rata-rata pada setiap rating movie paling banyak nilai rating terendah yaitu 1 dan 2.

Kemudain akan dilakukan pengecekan similarity pada tabel imputation min. Berikut hasil dari pengecekan similarity

Tabel 4.11 Similarity Imputation Max

	0	1	2	3	...	47	48	49	50
0	0.000000	0.939957	0.910815	0.904805	...	0.910815	0.919197	0.925031	0.910085
1	0.939957	0.000000	0.929243	0.938596	...	0.929243	0.952863	0.899220	0.926856
2	0.910815	0.929243	0.000000	0.956880	...	1.000.000	0.983259	0.935953	0.998157
3	0.904805	0.938596	0.956880	0.000000	...	0.956880	0.983160	0.911850	0.954820
4	0.904716	0.914736	0.909587	0.903586	...	0.909587	0.921179	0.869762	0.906800

Pada tabel 4.11 diketahui bahwa nilai kedekatan pada setiap nilai rating mencapai rata-rata 0.93

Kemudian Evaluasi kinerja imputasi dilakukan menggunakan Root Mean Square Error (RMSE). Metrik ini digunakan untuk mengevaluasi seberapa baik metode Min Imputation mampu merekonstruksi nilai yang hilang pada dataset MovieLens.

Hasil dari evaluasi kinerja dengan metrik RMSE adalah sebagai berikut:

Evaluating RMSE of algorithm SVD on 5 split(s).							
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std
RMSE (testset)	0.7077	0.7686	0.7686	0.7646	0.8382	0.7695	0.0414
Fit time	0.02	0.02	0.02	0.02	0.02	0.02	0.00
Test time	0.01	0.00	0.01	0.00	0.00	0.00	0.00
Average RMSE: 0.7695348350061556							

4.4.Imputation dengan KNN-Imputer

Metode K-Nearest Neighbors (KNN-Imputer) digunakan untuk menggantikan nilai yang hilang dengan menggunakan nilai dari tetangga terdekat. Proses ini melibatkan perhitungan jarak antara data yang hilang dengan data yang memiliki nilai yang sudah diketahui. Hasilnya adalah nilai imputasi yang didasarkan pada nilai dari K tetangga terdekat.

Tabel 4.12 Imputation Metode KNN-i

movieId	1	2	3	4	5	...	46	47	48	49	50
userId											
1	5	3	4	3	3	...	4	4	5	3	5
2	4	2	4	5	3	...	4	3	4	3	5
3	4	3	3	4	3	...	4	4	5	3	4
4	4	2	2	4	3	...	4	5	4	4	5
5	4	3	2	5	3	...	4	3	4	4	4

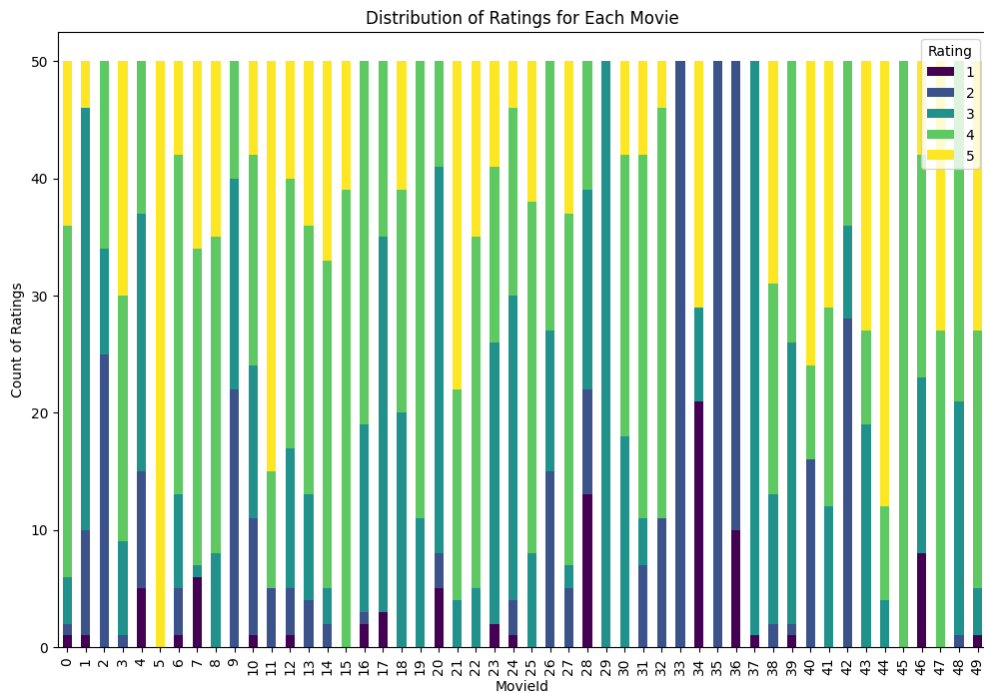
Hasil imputasi dengan metode KNN-i juga menunjukkan peningkatan signifikan dalam mengurangi jumlah nilai yang hilang. Dengan hasil sebagai berikut

- Jumlah nilai yang hilang sebelum imputasi: 2100
- Jumlah nilai yang hilang setelah imputasi: 0

Seluruh data yang kosong sudah terisi dengan nilai tertinggi dari setiap row dataset dan jika di lihat dari tabel dan grafik sebaran rating berikut ini terlihat bahwa semua movie sudah terrating

Tabel 4.13 Jumlah Rating KNN-Imputation

movieId	1	2	3	4	5	...	46	47	48	49	50
rating											
1	1	1	0	0	5	...	0	8	0	0	1
2	1	9	25	1	10	...	0	0	0	1	0
3	4	36	9	8	22	...	0	15	0	20	4
4	30	0	16	21	13	...	50	19	27	29	22
5	14	4	0	20	0	...	0	8	23	0	23
Jumlah	50	50	50	50	50	...	50	50	50	50	50



Gambar 4.5 Sebaran Rating KNN-i Imputation

Pada tabel 4.13 dan gambar 4.5 grafik menunjukkan bahwa nilai rating sudah terpenuhi dari 50 user untuk 50 movie dan nilai rating terbagi rata pada setiap rating movie.

Kemudain akan dilakukan pengecekan similarity pada tabel imputation KNN-I. Berikut hasil dari pengecekan similarity

Tabel 4.14 Similarity Imputation KNN-i

	0	1	2	3	...	47	48	49	50
0	0.000000	0.936127	0.969706	0.946491	...	0.969706	0.945702	0.938497	0.931178
1	0.936127	0.000000	0.977074	0.976032	...	0.977074	0.945791	0.909887	0.936847
2	0.969706	0.977074	0.000000	0.977736	...	1.000.000	0.965408	0.951109	0.965834
3	0.946491	0.976032	0.977736	0.000000	...	0.977736	0.956895	0.921206	0.953795
4	0.924109	0.967242	0.970402	0.955796	...	0.970402	0.974392	0.906254	0.953487

Pada tabel 4.14 diketahui bahwa nilai kedekatan pada setiap nilai rating mencapai rata-rata 0.97

Kemudian Evaluasi kinerja imputasi dilakukan menggunakan Root Mean Square Error (RMSE). Metrik ini digunakan untuk mengevaluasi seberapa

baik metode KNN-i Imputation mampu merekonstruksi nilai yang hilang pada dataset MovieLens.

Hasil dari evaluasi kinerja dengan metrik RMSE adalah sebagai berikut:

Evaluating RMSE of algorithm SVD on 5 split(s).							
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std
RMSE (testset)	0.8011	0.8454	0.8077	0.8424	0.8054	0.8204	0.0193
Fit time	0.03	0.02	0.02	0.02	0.02	0.02	0.00
Test time	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average RMSE: 0.8204036605991119							

4.5. Hasil Evaluasi RMSE

Setelah dilakukan imputation dan hitung nilai similarity dari ke empat data table imputation rating movie, dilakukan pengujian performa RMSE dengan hasil seperti pada data table 4.15

Tabel 4.15 Hasil Evaluasi RMSE

	Tanpa Inputation	Mean	Max	Min	KNN-i
RMSE	1.38468	0.40907	0.52619	0.76953	0.82040

Hasil pengujian didapatkan bahwa nilai yang unggul adalah dari metode imputation Mean dengan nilai RMSE = 0.40907 merupakan nilai terkecil dibandingkan dengan lainnya, yang menunjukkan nilai prediksi yang dihasilkan lebih baik. Selain itu di dukung dengan nilai similarity yang paling tinggi yaitu 0.99. Berdasarkan serangkaian hasil evaluasi dapat disimpulkan bahwa metode imputation khususnya adalah Mean dapat menyelesaikan masalah sparsity pada Collaborative Filtering dan mampu meningkatkan kualitas rekomendasi.