

BAB II TINJAUAN PUSTAKA

2.1 Missing Value Imputaion

Menggunakan metode statistic atau pembelajaran mesin untuk memperkirakan data observasi yang dipilih untuk menggantikan nilai yang kosong. Missing value imputation menganalisis pola di mana data yang hilang sebagai keluaran data model klasifikasi[15]. Penelitian ini menggunakan teknik imputation yaitu mean, min, max, dan KNN Imputer[16]. Algoritma KNN imputasi menggunakan observasi yang mirip dengan observasi yang memiliki nilai serupa yaitu dengan menentukan nilai K atau banyaknya observasi terdekat yang akan digunakan. Persamaan KNN imputer menggunakan Persamaan 1 [17].

$$d(x_a, x_b) = \sqrt{\sum_{m_j = 1}^m (x_{aj} - x_{bj})^2} \quad (1)$$

Penelitian ini menggunakan python library dalam menerapkan teknik imputation. Python library memiliki pendekatan terhadap missing value dengan mengganti data yang hilang, dengan beberapa metode pendekatan nilai "mean", metode pendekatan nilai maksimal "max", dan pendekatan nilai minimal "min"[18].

2.2 Cosin Similarity

Cosine Similarity merupakan metode yang bekerja dengan cara menghitung tingkat kesamaan (similarity) antar dua buah objek. Pada penelitian ini akan digunakan untuk membandingkan kemiripan antar user. Perhitungan Cosin Similarity menggunakan Persamaan 2.

$$(d_j, q) = \frac{\sum_{i=1}^t (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t w_{ij}^2 \sum_{i=1}^t w_{iq}^2}} \quad (2)$$

2.3 Evaluasi Performa

Penelitian ini menggunakan Root Mean Squared Error (RMSE) untuk melakukan evaluasi terhadap kinerja Collaborative Filtering dari pengimplementasian teknik imputation dan cosin similarity[19]. Root Mean Squared Error (RMSE) merupakan evaluasi kinerja untuk menghitung nilai rata-rata yang mempunyai selisih kuadrat antara nilai aktual dan nilai rating prediksi[20][21]. RMSE menghitung nilai pangkat yang di-root dari hasil RMSE merupakan matriks yang mempunyai matriks hasil yang sangat tinggi ketika kesalahan yang sangat tinggi tidak diinginkan[22]. Perhitungan RMSE menggunakan persamaan 3[23].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (|p_i - q_i|)} \quad (3)$$

2.4 Penelitian Terkait

Penelitian-penelitian tedahulu yang terkait dan mendasari penelitian ini dapat dilihat pada table 2.1 berikut

Tabel 2.1 Penelitian Terkai

N o	Peneliti, Tahun	Judul Penelitian	Dataset	Metode	Hasil penelitian
1	Mohamed, Marwa Hussien Khafagy, Mohamed Helmy Elbeh, Heba Abdalla, (2019)	Sparsity and cold start recommendati on system challenges solved by hybrid feedback,	Last.Fm (Song Dataset)	Item-based Collaborati ve Filtering	Hasil Penelitian mampu untuk merekomendasi kan item menggunakan preferensi pengguna tidak hanya item yang di nilai namun juga menggunkana tag item pengguna dan jumlah hitungan pemutaran lagu
2	Alhijawi , Bushra Al- Naymat, Ghazi Obeid, Nadim Awajan, Arafat (2019)	Mitigating the Effect of Data Sparsity: A Case Study on Collaborative Filtering Recommende r System	MovieLe ns dan HetRec 2011	UIP Matrix	Hasil Penelitian menunjukkan jarak tertinggi antara hasil akurasi menggunakan Matrik UIP mencapai [0,58%-15,1%] dan [1,8%- 17,5%].
3	Peng Yu (2019)	Merging attribute characteristics in	MoviLen s	Item-based Collaborati ve Filtering	Hasil penelitian ini menghasilkan bahwa algoritma

		collaborative filtering to alleviate data sparsity and cold start			yang ditingkatkan jauh lebih baik daripada algoritma Item-based collaborative filtering
--	--	---	--	--	---

Berdasarkan penelitian di atas, maka dapat disimpulkan bahwa dalam mengatasi sparsity dilakukan dengan beberapa metode seperti metrik UIP. Oleh karena itu penelitian ini akan mengatasi sparsity dengan metode imputation dan melakukan pengujian atau evaluasi terhadap imputation yang akan dilakukan.