

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan dari penelitian yang sudah dilaksanakan dengan judul “Analisis Prediksi *The Movie Database Rating Film* dengan *Ensemble Learning* Menggunakan Metode *Random Forest Regression*” adalah sebagai berikut:

1. Penggunaan *M-Estimate Encoding* dalam pengubahan bentuk *categorical data* ke bentuk *numerical data* pada *attribute cast1, cast2, cast3, director1, genre1, genre2, pro_com1* dan *pro_com2* memberikan solusi pada data-data yang jarang sekali muncul ataupun pada data yang tidak terdaftar pada *training data* sehingga nilai bobot setiap data tidak terlalu tinggi atau bias. Karena dilakukan perhitungan *smoothing values* dengan menetapkan nilai *M*.
2. Dari hasil proses *Feature Selection* dengan *Pearson Coefficient Correlation* diperoleh 4 *attributes* yaitu *attribute director1* dengan perolehan nilai *correlation* tertinggi yaitu sebesar 0.83, disusul dengan *attribute cast3* dengan nilai 0.82. Pada posisi ketiga ada *attribute cast2* dengan nilai 0.8 dan diposisi keempat adalah *attribute cast1* dengan nilai 0.76.
3. Hasil dari pencarian *Grid Search* didapatkan model dengan parameter yang optimal dan lebih akurat yaitu dengan penyesuaian parameter *max_depth* dengan nilai 80, *min_samples_leaf* dengan nilai 4, *min_samples_split* dengan nilai 3, dan parameter *n_estimator* dengan nilai

800 *trees*.

4. Hasil pengujian dan pengevaluasian performa model *Random Forest Regression* dengan metrik R^2 Score sebesar 86% untuk jumlah *training dataset* sebanyak 4907 *rows*. Diperoleh nilai model dengan metrik MSE sebesar 12%. Untuk peroleh nilai dari metrik RMSE adalah sebesar 35%. Dan untuk peroleh metrik *regression* dari MAE adalah sebesar 22%. Dan perolehan dari pengevaluasian performa dengan teknik *10-Fold Cross Validation* didapatkan rata-rata nilai adalah sebesar 85%.

