

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka peneliti memperoleh beberapa kesimpulan dalam penelitian sebagai berikut :

1. Tahapan analisis gaji data *science* berdasarkan pengaruh variabel pengalaman dan keahlian kerja menggunakan algoritma XGBoost melalui 6 tahapan. Tahapan pertama adalah pengumpulan data yang di download melalui platform *Kaggle*, setelah itu tahap berikutnya adalah preprocessing yaitu membersihkan nilai yang hilang dan menghapus data duplikat yang bertujuan mempermudah analisis, tahapan ketiga adalah *Exploratory Data Analysis* yang bertujuan untuk mengetahui pola, distribusi dan visualisasi data secara lengkap, tahapan selanjutnya yaitu pembagian data dengan rasio 80/20 untuk mencegah *overfitting* saat melakukan analisis, tahapan kelima adalah penggunaan model *Extreme Gradient Boosting*, tahapan terakhir adalah pengujian dan evaluasi model.
2. Dari total data 14.838 data yang dikumpulkan, terdapat 7.301 data yang digunakan untuk melakukan pelatihan atau sebesar 80% dalam membangun model. Dan 1.826 data yang digunakan untuk menguji performa model *machine learning* atau sebesar 20%.
3. Hasil penerapan *Extreme Gradient Boosting* dalam menganalisis gaji data *science* berdasarkan pengaruh variabel pengalaman dan keahlian kerja cukup baik karena menggunakan pembagian data dengan rasio 80/20.
4. Hasil *preprocessing* sudah baik karena tidak terdapat *missing value* dan data duplikat sebanyak 5.711 data sudah dihapus.
5. Hasil dari evaluasi matrix mendapatkan nilai *Mean Squared Error* (MSE) sebesar 5246971904.0 atau 5,25.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran sebagai berikut :

1. Dalam penelitian ini hanya menganalisis variabel pengalaman dan keahlian kerja untuk melihat pengaruhnya terhadap besaran gaji data science, untuk penelitian selanjutnya diharapkan dapat menambah variabel baru seperti jenis pekerjaan atau ukuran perusahaan.
2. Dalam penelitian ini hanya menggunakan satu matriks evaluasi yaitu *mean squared error* dengan hasil rata-rata error masih tergolong besar, untuk penelitian selanjutnya diharapkan dapat menambahkan *matrix* evaluasi dan memperbaiki hasil *mse* untuk memperoleh hasil lebih baik.
3. Dalam penelitian ini terbatas hanya menggunakan metode *Extreme Gradient Boosting*. Diharapkan untuk penelitian selanjutnya, dapat menggunakan metode regresi lainnya selain metode yang digunakan untuk analisis ini, seperti *LightGBM* dan *Multi-Layer Perceptron Regressor* (MLPRegressor), atau metode lain.
4. Dalam penelitian ini masih menggunakan tahapan *preprocessing* yang *simple*, diharapkan penelitian selanjutnya menggunakan *preprocessing* yang lebih kompleks untuk keperluan analisis lebih baik.