

## **BAB V**

### **SIMPULAN DAN SARAN**

#### **5.1 Simpulan**

Penelitian ini berhasil melakukan analisis komparatif antara dua algoritma pemodelan prediktif, yaitu Linear Regression dan XGBoost, dalam memprediksi gaji pemain bisbol berdasarkan data statistik dan atribut pemain. Linear Regression, sebagai model linier klasik, menunjukkan  $R^2$  sebesar 0.653 pada data uji dan rata-rata  $R^2$  cross-validation sebesar 0.559, dengan RMSE sebesar 3.991.861. Meskipun model ini memiliki keunggulan dalam hal sederhana, cepat, dan mudah diinterpretasikan, hasil evaluasi menunjukkan bahwa Linear Regression memiliki keterbatasan utama, yaitu ketidakmampuannya dalam menangkap hubungan non-linear dan interaksi antar fitur. Hal ini menyebabkan penurunan performa saat dihadapkan pada data yang kompleks seperti statistik pemain yang memiliki banyak variabel saling berkorelasi.

Sebaliknya, XGBoost, sebagai model berbasis ensemble yang lebih canggih, menunjukkan performa yang jauh lebih unggul dengan  $R^2$  sebesar 0.7836 pada data uji, rata-rata  $R^2$  cross-validation sebesar 0.7444, dan RMSE sebesar 2.652.046. Keunggulan utama XGBoost terletak pada kemampuannya dalam menangani non-linearitas, mengelola missing value, serta melakukan regularisasi dan pemilihan fitur secara otomatis. XGBoost juga lebih tahan terhadap outlier dan dapat menyesuaikan bobot antar fitur untuk memaksimalkan akurasi prediksi. Dengan demikian, model ini terbukti mampu menangkap pola hubungan kompleks

antar variabel yang tidak dapat ditangani oleh Linear Regression, menjadikannya baseline model yang lebih kuat untuk pengembangan sistem prediksi gaji pemain baseball.

Nilai  $R^2$  yang belum mencapai angka mendekati 1 dalam penelitian ini dapat dijelaskan oleh beberapa keterbatasan. Pertama, cakupan data terbatas hanya pada musim 2010–2024, sehingga tren jangka panjang belum tergambar secara utuh. Kedua, fitur yang digunakan masih terbatas pada statistik performa dasar seperti Salary Differences, Age, Hits, Runs, RBI, Walks, Strikeout, Stolen Bases, dan OPS, tanpa memasukkan variabel strategis seperti posisi bermain, performa dalam situasi krusial, kondisi cedera, atau pengalaman profesional. Ketiga, hanya dua algoritma yang diuji tanpa eksplorasi ke model alternatif lain seperti Random Forest, LightGBM, atau neural network. Terakhir, faktor eksternal non-statistik yang berperan besar dalam negosiasi kontrak, seperti agen pemain, citra publik, serta permintaan pasar, belum diikutsertakan dalam model.

Penelitian ini juga menghasilkan sebuah aplikasi prediksi berbasis web menggunakan Streamlit yang memungkinkan pengguna menginput data statistik dan secara langsung memperoleh hasil prediksi gaji. Aplikasi ini membuka peluang penerapan nyata di bidang analisis olahraga dan manajemen pemain profesional.

Secara keseluruhan, perbandingan antara Linear Regression dan XGBoost memberikan kontribusi penting dalam memahami trade-off antara interpretabilitas dan akurasi model prediksi. Linear Regression unggul dari sisi kemudahan interpretasi dan efisiensi, namun kurang tangguh untuk data dengan pola kompleks. Sebaliknya, XGBoost memiliki keunggulan dalam menangkap pola tersembunyi

dan menghasilkan prediksi yang lebih akurat dan stabil. Oleh karena itu, untuk konteks prediksi gaji pemain bisbol yang melibatkan banyak variabel dan kemungkinan non-linearitas, XGBoost lebih layak dijadikan pilihan utama dalam pengembangan sistem berbasis data yang aplikatif dan canggih.

## 5.2 Saran

Untuk penelitian lanjutan, demi meningkatkan akurasi model dalam menjelaskan anomali kontrak gaji yang sangat tinggi, perlu dilakukan pengumpulan dan analisis data tambahan. Beberapa faktor eksternal yang dapat mempengaruhi nilai kontrak pemain antara lain:

1. Peran agen pemain: Agen yang berpengaruh dan memiliki reputasi baik dapat meningkatkan nilai tawar pemain dalam negosiasi kontrak.
2. Daya tarik pasar (marketability): Faktor non-statistik seperti popularitas di media, kepemimpinan di tim, dan citra publik turut berkontribusi terhadap nilai kontrak meskipun sulit diukur secara langsung.
3. Permintaan pasar: Ketidakseimbangan antara permintaan tim terhadap posisi tertentu dan ketersediaan pemain dengan spesifikasi yang dibutuhkan dapat menyebabkan peningkatan kontrak secara signifikan.

Penelitian lanjutan disarankan untuk memfokuskan analisis pada pemain yang sudah memasuki masa free agency serta melakukan studi khusus terhadap

25% kontrak pemain dengan nilai tertinggi, guna memperoleh pemahaman lebih mendalam terhadap faktor-faktor penentu gaji yang tergolong ekstrem (outlier).

Selain itu, pada sisi metode, disarankan untuk mengeksplorasi lebih lanjut berbagai variasi parameter XGBoost seperti `max_depth`, `learning_rate`, `subsample`, dan `colsample_bytree`. Penggunaan grid search atau randomized search dengan rentang parameter yang lebih luas, serta kombinasi dengan teknik Bayesian optimization, dapat meningkatkan performa model secara signifikan.

Metode ensemble lain seperti Random Forest dan LightGBM, atau bahkan pendekatan deep learning, dapat diuji untuk melihat apakah terdapat peningkatan akurasi dibandingkan XGBoost. Penggunaan cross-validation dengan jumlah fold lebih banyak (misalnya 10-fold atau stratified K-fold) juga dapat memberikan estimasi performa yang lebih stabil. Terakhir, peningkatan kualitas dan keberagaman dataset—termasuk menambahkan data dari liga atau musim yang berbeda—akan memperkaya generalisasi model, sehingga hasil prediksi lebih robust dan relevan untuk pengambilan keputusan di dunia nyata, khususnya dalam evaluasi dan negosiasi gaji pemain baseball.