

## Perbandingan Metode *Adaptive Boosting* dan *Extreme Gradient Boosting* Untuk Prediksi Hasil Pertandingan Liga Spanyol

Muhammad Rohid Saputro<sup>1</sup>, Umi Mahdiyah<sup>2</sup>, Daniel Swanjaya<sup>3</sup>

Universitas Nusantara PGRI Kediri<sup>1,2,3</sup>

[rohidiunior@gmail.com](mailto:rohidiunior@gmail.com)<sup>1</sup>, [umimahdiyah@gmail.com](mailto:umimahdiyah@gmail.com)<sup>2</sup>, [daniel@unpkediri.ac.id](mailto:daniel@unpkediri.ac.id)<sup>3</sup>

### Abstrak

Sepakbola merupakan olahraga yang paling terkenal di seluruh dunia dengan hampir 4 miliar pengagum dari berbagai belahan bumi. Negara besar di Eropa memiliki kompetisi sepakbola yang terstruktur dan memiliki tingkatan kompetisi yang lengkap. Algoritma *AdaBoost* dan *XGBoost* merupakan metode *machine learning* yang dapat digunakan untuk mengatasi suatu permasalahan yang berhubungan dengan deret dan situasi peramalan. Perlunya mengetahui prediksi kemenangan tim pertandingan sepak bola Liga Spanyol selalu menjadi pembahasan yang tidak pernah dilewatkan oleh penggemar sepak bola, oleh karena itu prediksi sangat berguna untuk para penggemar sepakbola dan pelatih tim sepak bola dapat mengantisipasi suatu kejadian yang mendatang. Misalnya, penggemar ataupun pelatih tim sepak bola Liga Spanyol dapat memperkirakan kemenangan tim pada masa mendatang. Data yang digunakan menggunakan dataset statistik pertandingan 2 tingkat teratas Liga Spanyol selama 4 musim yaitu pada musim 2018/2019 sampai musim 2021/2022 yang didapat dari [www.football-data.co.uk](http://www.football-data.co.uk). Pengujian menggunakan metode *AdaBoost* memperoleh tingkat akurasi sebesar 64,02%, dan metode *XGBoost* memperoleh tingkat akurasi sebesar 61,79%. Dari hasil pengujian yang telah dilakukan pada dataset Liga Spanyol musim 2018/2019 sampai 2021/2022, menunjukkan bahwa metode *AdaBoost* memperoleh hasil yang lebih baik jika dibandingkan dengan metode *XGBoost*.

**Kata Kunci** : *Adaptive Boosting*, *Data Mining*, *Extreme Gradient Boosting*, *Prediksi*.

### A. PENDAHULUAN

Sepak bola merupakan permainan yang dilakukan oleh 2 tim berbeda, dengan setiap tim diisi oleh 11 orang pemain. Tim yang berhasil mencetak gol lebih banyak akan keluar sebagai pemenang. Dikutip dari [bolasport.com](http://bolasport.com), sepak bola merupakan olahraga yang paling terkenal di seluruh dunia dengan hampir 4 miliar pengagum dari berbagai belahan bumi. Indonesia memiliki populasi penduduk sebanyak 275 juta jiwa dengan 77% penduduknya menyukai sepak bola.

Di setiap negara di seluruh penjuru dunia, sebagian besar memiliki kompetisi sepak bola. Di eropa sepak bola sudah sangat maju, banyak teknologi-teknologi baru yang berawal dari sana. Negara besar di eropa memiliki kompetisi sepak bola yang sangat terstruktur dan juga memiliki tingkatan kompetisi yang terbilang lengkap. Dari kasta pertama hingga beberapa kasta dibawahnya, setiap tim juga memiliki tim kelompok umur yang terbilang lengkap.

Liga Spanyol atau La Liga merupakan salah satu dari 5 Liga sepakbola terbaik di Eropa. La Liga dibentuk pada tahun 1927 oleh Seorang Direktur yang berasal dari Arenas Club de Getxo dan mempunyai nama Jose Maria Acha. Ketika pertama kali diselenggarakan yaitu pada tahun 1929, La Liga merupakan pertandingan sepak bola tertinggi yang ada di negara tersebut dengan tim yang bermain diantaranya adalah Real Madrid, Barcelona, Real Sociedad, Arenas Club de Getxo, Athletic Bilbao serta Real Union (Sulaiman, 2022).

Statistik menurut KBBI yaitu catatan angka-angka, data yang dikelompokkan sehingga dapat memberi informasi tertentu yang berkaitan dengan suatu topik. Dalam sepak bola, statistik berisi data tentang jalannya pertandingan, seperti skor pertandingan, jumlah tendangan yang telah dilakukan, jumlah pelanggaran yang telah dilakukan, jumlah kartu yang didapat, dan lain-lain. Dari data statistik tersebut dapat digunakan untuk memprediksi pertandingan sepak bola dengan menggunakan *machine learning*. *Machine learning* sebagai salah satu sub-bidang dari kecerdasan buatan, memiliki kemampuan belajar seperti manusia. *Machine learning* dapat digunakan untuk klafifikasi dan prediksi. Dalam *machine learning* terdapat proses pelatihan, pembelajaran, dan training sebagai ciri khasnya (Ahmad, 2017).

Pada penelitian ini penulis akan melakukan perbandingan dari 2 metode yaitu *Adaptive Boosting* (AdaBoost) dan *Extreme Gradient Boosting* (XGBoost) dengan data pertandingan sepak bola Liga Spanyol. AdaBoost dan XGBoost digunakan dalam *machine learning* untuk melakukan klasifikasi data secara *supervised learning*, di mana label ditentukan oleh atribut-atribut yang telah terklasifikasi dengan pengawasan. Sebagai hasilnya, kedua metode tersebut dapat diterapkan dalam melakukan analisis statistik liga sepak bola dan melakukan prediksi terkait kemenangan tim sepak bola di Liga Spanyol di masa depan. Semua metode ini juga dipilih untuk membandingkan keakuratan masing-masing metode tersebut.

## B. LANDASAN TEORI

### 1. Machine Learning

*Machine Learning* merupakan cabang ilmu komputer dengan fokus pada pengembangan sebuah sistem yang mampu belajar sendiri tanpa harus berulang kali di program oleh manusia. Namun, sebelum menghasilkan sebuah hasil data dari perilaku objek, *machine learning* membutuhkan data awal sebagai bahan yang akan dipelajari. *Machine learning* merupakan aplikasi dari disiplin ilmu *Artificial Intelligence* yang menggunakan teknik statistika untuk menghasilkan suatu model otomatis dari sekumpulan data, dengan tujuan memberikan komputer kemampuan untuk "belajar" (Rezkia, 2021).

### 2. Adaptive Boosting

*Adaptive Boosting* (AdaBoost) adalah salah satu varian dari algoritma *boosting* yang bekerja berdasarkan prinsip metode penjumlahan bertahap dimana banyak *weak learners* digunakan untuk mendapatkan *strong learners*. Algoritma AdaBoost telah berhasil diterapkan pada beberapa bidang karena prediksinya yang akurat dan sangat sederhana. Algoritma AdaBoost akan menghasilkan keputusan dari *weak learners* (Gultom, 2020). Algoritma AdaBoost memiliki beberapa kelebihan, salah satunya yaitu mampu mengurangi bias maupun perbedaan pada *weak classifier*. Selain itu, algoritma ini juga memiliki kemampuan generalisasi yang baik, output dari AdaBoost mengkonversi ke logaritma rasio terdekat.

### 3. Extreme Gradient Boosting

*Extreme Gradient Boosting* (XGBoost) adalah algoritma yang telah ditingkatkan berdasarkan *gradient boosting decision tree* dan mampu membangun *boosted trees* secara efisien dan parallel (Karo, 2020). Algoritma ini merupakan pengembangan dari *boosting ensemble* yang menggabungkan prediksi dari beberapa model lemah (*weak learner*) untuk membentuk model yang lebih kuat. *Weight* memiliki peran penting dalam algoritma ini, *weight* diberikan ke semua variabel independen dan kemudian dimasukkan ke *decision tree* untuk memprediksi hasil. *Weight* yang diprediksi salah akan dinaikkan dan kemudian dimasukkan ke *decision tree* kedua. Nilai prediksi dari setiap *decision tree* dijumlahkan untuk mendapatkan hasil akhir.

### 4. Confusion Matrix

*Confusion Matrix* merupakan salah satu metode untuk mengukur kinerja model yang telah dibuat. Nilai *confusion matrix* dapat dilihat pada tabel 1.

Tabel 1. *Confusion Matrix*

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP	FN
	Negatif	FP	TN

Berikut rumus untuk mengukur nilai akurasi, presisi, *recall* pada persamaan (2)(3)(4).

$$\text{Akurasi} = \left( \frac{TP+TN}{TP+TN+FP+FN} \right) \times 100\% \dots\dots\dots(2)$$

$$\text{Presisi} = \left( \frac{TP}{TP+FP} \right) \times 100\% \dots\dots\dots(3)$$

$$Recall = \left( \frac{TP}{TP+FN} \right) \times 100\% \dots\dots\dots(4)$$

$$F1-Score = \frac{2 \times recall \times precision}{recall+precision} \dots\dots\dots(5)$$

### C. METODE PENELITIAN

Penelitian ini dilakukan dengan beberapa tahapan pengerjaan, berikut tahapan-tahapan yang akan dilakukan :

#### 1. Studi Literatur

Pada tahap ini yaitu mempelajari teori dari jurnal terdahulu, buku, serta informasi yang berkaitan dengan metode AdaBoost dan XGBoost.

#### 2. Pengumpulan Data

Dalam penelitian ini dataset yang digunakan adalah data statistik pertandingan 2 tingkat teratas Liga Spanyol selama 4 musim yaitu pada musim 2018/2019 sampai musim 2021/2022 yang didapat dari [www.football-data.co.uk](http://www.football-data.co.uk). Dataset yang didapat terpisah setiap musim, dataset ini terdapat atribut sebanyak 105 atribut, namun hanya diambil atribut yang menjadi faktor utama dalam kemenangan sepak bola saja sehingga atribut yang digunakan sebanyak 16 atribut dan 1 label. Penjelasan setiap atribut dan label dapat dilihat pada tabel 2. Pada dataset ini terdapat 3347 total data dari seluruh pertandingan 2 tingkat teratas Liga Spanyol dari musim 2018/2019 sampai musim 2021/2022.

Tabel 2 Atribut Pada Dataset

Atribut	Keterangan	Deskripsi
<i>HomeTeam</i>	Atribut ( <i>Input</i> )	Nama dari tim kandang yang bermain.
<i>AwayTeam</i>	Atribut ( <i>Input</i> )	Nama dari tim tandang yang bermain.
HS	Atribut ( <i>Input</i> )	<i>Home Team Shots</i> . Jumlah tendangan kearah gawang oleh tim kandang.
AS	Atribut ( <i>Input</i> )	<i>Away Team Shots</i> . Jumlah tendangan kearah gawang oleh tim tandang.
HST	Atribut ( <i>Input</i> )	<i>Home Team Shots on Target</i> . Jumlah tendangan kearah gawang yang tepat sasaran oleh tim kandang.
AST	Atribut ( <i>Input</i> )	<i>Away Team Shots on Target</i> . Jumlah tendangan kearah gawang yang tepat sasaran oleh tim tandang.
HC	Atribut ( <i>Input</i> )	<i>Home Team Corners</i> . Jumlah tendangan pojok oleh tim kandang.
AC	Atribut ( <i>Input</i> )	<i>Away Team Corners</i> . Jumlah tendangan pojok oleh tim tandang.
HF	Atribut ( <i>Input</i> )	<i>Home Team Fouls Committed</i> . Jumlah pelanggaran yang dilakukan oleh tim kandang.
AF	Atribut ( <i>Input</i> )	<i>Away Team Fouls Committed</i> . Jumlah pelanggaran yang dilakukan oleh tim tandang.
HY	Atribut ( <i>Input</i> )	<i>Home Team Yellow Cards</i> . Jumlah kartu kuning yang didapat oleh tim kandang.
AY	Atribut ( <i>Input</i> )	<i>Away Team Yellow Cards</i> . Jumlah kartu kuning yang didapat oleh tim tandang.
HR	Atribut ( <i>Input</i> )	<i>Home Team Red Cards</i> . Jumlah kartu merah yang didapat oleh tim kandang.
AR	Atribut ( <i>Input</i> )	<i>Away Team Red Cards</i> . Jumlah kartu merah yang didapat oleh tim tandang.
HTHG	Atribut ( <i>Input</i> )	<i>Half Time Home Team Goals</i> . Skor tim kandang di babak pertama pada pertandingan.
HTAG	Atribut ( <i>Input</i> )	<i>Half Time Away Team Goals</i> . Skor tim tandang di babak pertama pada pertandingan.
FTR	Label ( <i>Output</i> )	<i>Full Time Result</i> . Hasil akhir pertandingan, memiliki 3 label

Atribut	Keterangan	Deskripsi
		yaitu A ( <i>Away</i> ) yang berarti tim tandang menang, D ( <i>Draw</i> ) yang berarti kedua tim seri, dan H ( <i>Home</i> ) yang berarti tim kandang menang.

### 3. Preprocessing Data

Pada tahap ini akan dilakukan *encoding* untuk mengubah data pada kolom *HomeTeam* dan *AwayTeam* supaya dapat dibaca oleh model. Metode *encoding* yang digunakan yaitu *Label Encoding*.

### 4. Pemodelan Metode

Setelah dilakukan *preprocessing*, data kemudian dimodelkan dengan menggunakan metode AdaBoost dan XGBoost menggunakan bantuan bahasa pemrograman *python* dengan *software* yang digunakan yaitu Google Colab untuk menulis dan menjalankan kode.

### 5. Evaluasi dan Validasi

Parameter yang digunakan untuk menganalisis kinerja model yaitu menggunakan *confusion matrix*. Dengan dasar dari *confusion matrix* kemudian dapat dilakukan perhitungan nilai akurasi, presisi, *recall*, dan *f1-score*.

### 6. Analisa Hasil

Pada tahap ini dapat disimpulkan bagaimana hasil perbandingan menggunakan metode AdaBoost dan XGBoost.

## D. HASIL DAN PEMBAHASAN

### 1. Analisis Data Set Penelitian

Dataset penelitian terlebih dahulu dilakukan proses input di google colab. Dataset tersebut di import dan ditampilkan di sistem seperti pada gambar 1.

```
dataset = pd.read_csv('Dataset/spanyol/tes_spanyol.csv')
dataset.head(10)
```

✓ 0.1s

	HomeTeam	AwayTeam	HS	AS	HST	AST	HC	AC	HF	AF	HY	AY	HR	AR	HTHG	HTAG	FTR
0	Betis	Levante	22.0	6.0	8.0	4.0	5.0	3.0	10.0	10.0	0.0	2.0	0.0	0.0	0.0	1.0	A
1	Girona	Valladolid	13.0	2.0	1.0	1.0	3.0	2.0	21.0	20.0	1.0	1.0	0.0	0.0	0.0	0.0	D
2	Barcelona	Alaves	25.0	3.0	9.0	0.0	7.0	1.0	6.0	13.0	0.0	2.0	0.0	0.0	0.0	0.0	H
3	Celta	Espanol	12.0	14.0	2.0	5.0	8.0	7.0	13.0	14.0	3.0	2.0	0.0	0.0	0.0	1.0	D
4	Villarreal	Sociedad	16.0	8.0	7.0	4.0	4.0	6.0	16.0	10.0	2.0	3.0	0.0	0.0	1.0	1.0	A
5	Eibar	Huesca	18.0	8.0	6.0	6.0	7.0	0.0	12.0	13.0	1.0	1.0	0.0	0.0	0.0	2.0	A
6	Real Madrid	Getafe	10.0	4.0	3.0	1.0	3.0	0.0	11.0	27.0	1.0	7.0	0.0	0.0	1.0	0.0	H
7	Vallecano	Sevilla	13.0	17.0	2.0	8.0	2.0	6.0	6.0	15.0	1.0	0.0	0.0	0.0	0.0	3.0	A
8	Ath Bilbao	Leganes	17.0	12.0	5.0	2.0	6.0	2.0	12.0	13.0	4.0	5.0	0.0	0.0	1.0	1.0	H
9	Valencia	Ath Madrid	13.0	9.0	4.0	3.0	4.0	10.0	10.0	15.0	2.0	3.0	0.0	0.0	0.0	1.0	D

Gambar 1. Dataset

Untuk Label pada penelitian ini terdapat pada kolom *Full Time Result (FTR)* yaitu *away (A)* merupakan hasil kemenangan untuk tim *Away* (tandang), *draw (D)* merupakan hasil seri, dan *home (H)* merupakan hasil kemenangan untuk tim *Home* (kandang).

### 2. Preprocessing Data

Pada kolom *HomeTeam* dan *AwayTeam* yang bertipe data *string* diubah menjadi *integer*, pada gambar 2. Proses *encoding* yang digunakan yaitu *Label Encoding* supaya nilai pada kolom berurutan.

```
le=LabelEncoder()
dataset['HomeTeam']=le.fit_transform(dataset['HomeTeam'])
dataset['AwayTeam']=le.fit_transform(dataset['AwayTeam'])
dataset['FTR']=le.fit_transform(dataset['FTR'])
```

Gambar 2. Encoding

Dataset dibagi menjadi data *training* (data latih) sebesar 80% dan data *testing* (data uji) sebesar 20% dari 3347 data yang digunakan, pembagian data dilakukan dengan *Train Test Split* seperti pada gambar 3.

```
X = dataset[dataset.columns[0:16]]
Y = dataset['FTR']
✓ 0.0s

# membagi dataset menjadi training dan testing
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=1)
✓ 0.0s
```

Gambar 3. *Split Data*

Data terlebih dahulu dipisah menjadi atribut dan label, Atribut ditandai dengan huruf X, label ditandai dengan huruf Y. Kemudian dilakukan pembagian data menjadi data latih dan data uji dengan fungsi *train\_test\_split()*. Atribut yang digunakan yaitu *HomeTeam*, *AwayTeam*, HS, AS, HST, AST, HC, AC, HF, AF, HY, AY, HR, AR, HTHG, HTAG. Sedangkan label yang digunakan yaitu FTR.

### 3. Pemodelan Metode

#### a. AdaBoost

Hasil prediksi yang dilakukan menggunakan algoritma AdaBoost mendapatkan tingkat akurasi sebesar 64,02%. Presentase tingkat akurasi dihasilkan dari menghitung data latih penelitian pada gambar 4.

```
abSpain = AdaBoostClassifier(random_state=10, learning_rate=0.20)
abSpain.fit(X_train, y_train)
abSpain_pred = abSpain.predict(X_test)
print(abSpain_pred) #hasil prediksi
print(y_test) #jawaban yang sebenarnya
print(metrics.accuracy_score(y_test, abSpain_pred)*100)
✓ 0.4s
```

```
[2 0 2 2 2 2 2 2 0 2 2 1 2 1 1 0 2 0 0 0 1 0 2 2 1 2 2 1 2 0 2 2 0 0 0 0 2
 2 2 2 0 0 1 2 1 2 2 2 2 1 2 0 1 0 1 2 1 2 2 0 2 0 2 1 2 2 1 0 0 1 2 2 0 2
 0 2 2 2 0 2 2 0 2 1 2 2 2 2 2 2 1 1 2 2 0 2 2 1 0 0 2 0 2 2 1 2 1 2 2 0 1
 0 2 2 2 2 2 1 2 2 1 2 2 2 0 0 2 1 2 2 1 2 0 2 1 0 1 1 2 0 1 2 1 2 2 2 0 2
 0 0 1 2 0 0 2 1 1 2 2 2 2 0 0 1 2 1 2 2 2 1 1 1 0 1 2 1 2 0 2 0 0 2 1 0 1
 1 0 0 0 2 2 0 0 0 1 1 2 0 1 2 2 0 2 2 1 2 1 0 2 1 1 2 0 2 1 0 2 0 0 2 2 0
 0 2 1 2 0 1 0 0 2 0 2 1 0 1 2 0 2 1 2 2 2 0 2 2 0 1 2 0 2 2 1 2 1 2 1 0 0
 1 1 1 1 0 2 0 2 2 0 2 0 1 0 1 2 2 0 2 2 1 1 2 2 1 1 2 1 2 2 2 2 2 1 0 1 1
 2 0 2 0 1 2 2 1 2 2 2 2 2 2 2 1 2 1 0 1 0 1 2 2 2 2 1 0 2 2 0 2 2 0 1
 1 0 2 1 0 1 0 2 2 1 2 2 2 1 1 2 2 2 1 0 0 1 2 2 1 2 2 0 0 1 2 1 2 2 2 0 1
 2 2 1 2 1 2 2 1 2 2 1 0 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 0 2 2 0
 2 2 2 2 2 0 0 1 0 0 2 1 2 0 0 1 1 0 0 2 2 0 2 2 2 0 2 2 2 1 2 2 0 2 2 1 2
 0 2 2 2 2 1 0 2 2 0 1 1 2 0 2 1 1 1 0 0 2 0 1 2 2 2 1 1 2 0 2 2 1 1 2 1 1
 2 2 0 2 2 2 1 0 0 0 0 2 2 2 2 0 2 2 1 0 0 2 1 1 0 0 1 1 0 2 1 2 0 0 1 1 1
 0 1 2 1 1 1 0 2 2 2 2 2 0 1 0 1 0 2 2 1 2 2 2 1 2 0 0 2 2 2 2 0 2 2 1 2 0
 0 0 1 0 2 2 2 0 0 1 0 0 0 0 0 0 0 2 0 0 1 0 2 1 0 0 2 2 2 1 2 2 0 1 2 2
 1 1 0 1 1 1 2 2 2 2 2 1 2 2 0 1 2 2 2 2 0 1 0 1 0 1 2 2 1 2 0 2 2 0 2 2
 1 2 2 2 0 1 0 2 0 0 1 2 2 2 2 2 1 2 2 2 1 2 2 1 0 2 1 2 0 2 2 2 2 0 0 2 2
 1 2 1 1]
```

```
1322 2
2372 1
2555 2
2097 2
1758 2
...
1751 1
2264 1
Name: FTR, Length: 670, dtype: int32
64.02985074626866
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Gambar 4. AdaBoost

Pada pengujian menggunakan algoritma AdaBoost, data train dimasukkan ke dalam model, parameter yang ditambahkan yaitu *learning\_rate* sebesar 0,20. Kemudian dilakukan prediksi dan menghasilkan akurasi sebesar 64,02%.

#### b. XGBoost

Hasil prediksi yang dilakukan menggunakan algoritma XGBoost mendapatkan tingkat akurasi sebesar 61,79%. Presentase tingkat akurasi dihasilkan dari menghitung data latihan penelitian pada gambar 5.

```

grSpain = xgb.XGBClassifier(random_state=10)
grSpain.fit(X_train, y_train)
grSpain_pred = grSpain.predict(X_test)
print("pred", grSpain_pred) #hasil prediksi
print("tes", y_test) #jawaban yang sebenarnya
print(metrics.accuracy_score(y_test, grSpain_pred) * 100)
✓ 0.8s

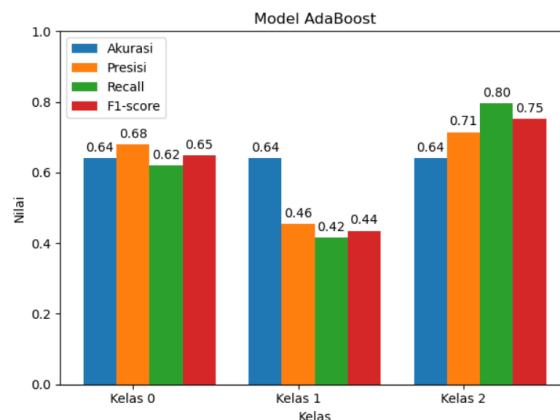
pred [2 1 2 0 2 2 2 2 2 0 2 2 1 2 1 1 0 2 0 0 0 0 2 2 2 0 2 2 0 2 0 2 1 0 2 0 2
2 2 2 0 0 1 2 1 2 2 0 2 1 2 0 1 2 1 1 0 2 1 1 2 2 1 1 2 2 1 1 2 2 1 0 0 1 2 2 2 2
0 2 2 2 0 0 2 0 2 1 1 2 2 2 2 1 1 1 2 2 1 2 2 1 2 0 2 0 1 2 1 2 2 2 1 0 1
0 2 2 2 2 2 1 2 2 2 2 2 0 0 2 1 1 1 1 2 1 2 1 0 1 1 2 0 1 1 2 1 2 2 2 0 2
1 0 1 2 0 1 1 0 2 2 1 2 2 0 0 0 1 1 2 2 2 1 1 2 1 1 2 1 2 0 2 0 0 1 1 0 1
1 0 0 0 2 1 0 0 0 1 0 2 0 2 2 2 0 2 2 1 2 1 0 2 1 2 2 0 1 1 0 2 0 0 2 2 0
0 1 0 2 0 2 1 1 2 0 2 2 0 0 2 0 2 1 2 2 2 0 2 2 0 1 2 0 2 2 2 2 1 2 1 0
0 1 2 0 0 2 0 2 1 2 2 2 1 0 2 2 2 0 2 2 2 1 2 2 0 2 2 1 1 1 1 2 2 1 0 2 1
2 0 1 0 0 1 1 2 2 2 2 2 2 2 2 0 2 2 0 0 1 0 0 2 2 1 2 0 1 1 2 2 2 2 1 0 1
1 1 2 1 0 1 0 2 1 1 1 2 2 0 1 1 2 2 0 0 0 1 2 2 1 2 2 0 0 1 2 0 2 2 1 0 1
1 1 1 2 1 2 2 2 1 2 1 0 2 0 1 1 2 2 2 2 1 2 2 2 1 1 2 2 2 1 2 2 1 0 2 2 1 1
2 0 1 2 1 0 1 0 0 2 0 2 0 1 0 2 0 0 2 2 0 2 0 2 0 1 2 1 1 2 2 0 2 1 2 1 1
1 2 2 1 2 1 0 2 2 0 1 0 1 1 1 1 2 2 0 0 1 0 1 2 2 2 1 1 2 0 2 2 0 1 2 1 1
1 2 0 2 2 2 0 0 2 0 0 2 2 2 2 0 2 2 1 0 0 2 1 1 0 0 1 0 0 2 1 2 1 0 0 2 1
0 1 2 1 1 1 0 2 2 1 2 2 0 1 0 0 0 0 2 0 2 2 2 1 2 1 0 2 1 1 2 0 2 2 2 1 0
0 0 1 0 2 2 2 0 0 0 0 0 0 0 0 1 0 2 0 1 0 0 2 0 0 2 1 2 2 0 2 2 0 2 1 2
1 2 0 0 1 0 2 2 2 2 2 1 2 1 0 1 2 2 1 2 0 0 0 1 0 1 2 2 1 1 0 2 2 1 2 2
1 2 2 1 0 0 1 2 0 1 0 2 2 2 2 2 0 2 2 2 1 2 2 0 0 1 1 2 1 2 2 2 0 0 2 1
2 2 0 2]
tes 1322 2
2372 1
2555 2
2097 2
1758 2
...
...
1751 1
2264 1
Name: FTR, Length: 670, dtype: int32
61.791044776119406
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

```

Gambar 5. XGBoost

Pada pengujian menggunakan algoritma XGBoost, data *train* dimasukkan ke dalam model, kemudiann dilakukan prediksi dan menghasilkan akurasi sebeasr 61,79%.

4. Evaluasi Model  
a. AdaBoost

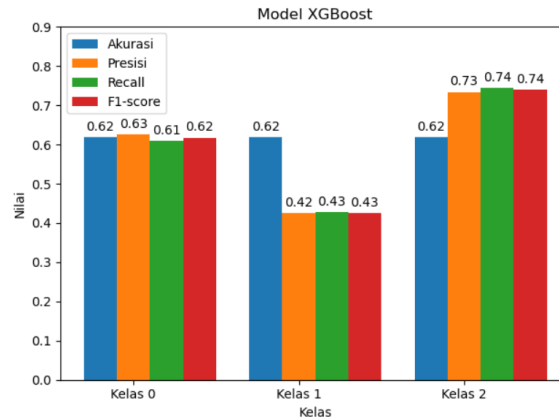


Gambar 6. Confusion Matrix AdaBoost

Data uji untuk kelas 0 berjumlah 192 data, kelas 1 berjumlah 185 data, kelas 2 berjumlah 293 data. Dari gambar 6 dapat diketahui kelas 1 menghasilkan nilai presisi, *recall*, dan *f1-score* paling rendah. Hal ini dikarenakan jumlah data dari kelas 1 lebih sedikit daripada jumlah data kelas 0 dan 2, sehingga kelas 1 kurang dikenali oleh model.

b. XGBoost





Gambar 7. *Confusion Matrix* XGBoost

Data uji untuk kelas 0 berjumlah 192 data, kelas 1 berjumlah 185 data, kelas 2 berjumlah 293 data. Dari gambar 7 dapat diketahui kelas 1 menghasilkan nilai presisi, *recall*, dan *f1-score* paling rendah. Hal ini dikarenakan jumlah data dari kelas 1 lebih sedikit daripada jumlah data kelas 0 dan 2, sehingga kelas 1 kurang dikenali oleh model.

### E. Kesimpulan dan Saran

Pada penelitian ini melakukan perbandingan metode AdaBoost dan XGBoost Dalam Prediksi Akurasi Pertandingan Liga Spanyol. Data yang digunakan yaitu data pertandingan 2 tingkat teratas Liga Spanyol musim 2018/2019 sampai 2021/2022. Berdasarkan implementasi dan pengujian yang telah dilakukan, dapat diambil kesimpulan bahwa hasil pengujian menggunakan metode AdaBoost memperoleh tingkat akurasi sebesar 64,02%, dan metode XGBoost memperoleh tingkat akurasi sebesar 61,79%. Dari hasil pengujian yang telah dilakukan pada dataset Liga Spanyol musim 2018/2019 sampai 2021/2022, menunjukkan bahwa metode AdaBoost memperoleh hasil yang lebih baik jika dibandingkan dengan metode XGBoost.

Berdasarkan kesimpulan, saran yang dapat diberikan untuk mengembangkan penelitian ini lebih lanjut yaitu penambahan dataset dari beberapa musim sebelum musim 2018/2019 atau setelah musim 2021/2022 agar akurasi yang didapat lebih tinggi.

### DAFTAR PUSTAKA

- Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine. *Jurnal Teknologi Indonesia*.
- Gultom, S. I. (2020). Implementasi Data Mining Menentukan Pola Hidup Sehat Bagi Pengguna KB. *Jurnal Informasi dan Teknologi Ilmiah (INTI)*, 298-304.
- Karo, I. M. (2020). Implementasi Metode XGBoost dan Feature Importance untuk Klasifikasi pada Kebakaran Hutan dan Lahan. 10-16.
- Prasetya, M. H. (2021, Januari 14). Retrieved from Bolasport.com: <https://www.bolasport.com/read/312509342/pssi-bentuk-tim-ad-hoc-dan-sebut-indonesia-negara-kedua-pecinta-sepak-bola-di-dunia>
- Rezki, S. M. (2021). Retrieved from <https://dqlab.id/mengenal-algoritma-machine-learning-hingga-cara-kerjanya>
- Sulaiman, F. (2022, Mei 28). Retrieved from wartaekonomi.co.id: <https://wartaekonomi.co.id/read417286/kisah-dan-sejarah-unik-di-balik-berdirinya-la-liga-spanyol#:~:text=sejarah%20liga%20Spanyol,-,Sejarah%20Liga%20Spanyol,sepak%20bola%20di%20negara%20Spanyol.>