

# Penerapan Model Regresi Logistik Biner dan Random Forest terhadap Prospek Atlet Muda pada Liga Basket DBL Tahun 2019\*

Marta Nur Muhammad<sup>1</sup>, Mohammad Masjkur<sup>1‡</sup>, Muhammad Nur Aidi<sup>1</sup>

<sup>1</sup>Department of Statistics, IPB University, Indonesia

‡corresponding author: masjkur@apps.ipb.ac.id

Copyright © 2023 Marta Nur Muhammad, Mohammad Masjkur, Muhammad Nur Aidi. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

Athletic performance is a key indicator of the success of athlete development within a sports discipline, including basketball. Effective development requires a competitive and professional platform for talent identification, such as the DBL East Java Series basketball league for senior high school students. A well-organized competition supports positive athlete development, enabling the evaluation of individual prospects through game statistics. Athlete prospects reflect future potential arising from present performance and are categorized according to the Indonesia Emas (PRIMA) Program Guidelines established by the Indonesian National Sports Committee in 2015, which define the Pratama class for athletes competing at national or regional levels. This study develops classification models to predict athlete prospects using match-level statistics. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied, while k-fold cross-validation is used to obtain robust model estimates. The findings show that all constructed models achieve strong predictive performance based on the Area Under the Curve (AUC). Furthermore, the variables points scored—representing scoring ability—and assists—representing ball-handling and playmaking ability—are identified as the most influential predictors of young athlete prospects in both the binary logistic regression and random forest models.

**Keywords:** basketball, athlete prospects, binary logistic regression, random forest, classification.

## 1. Pendahuluan

Cabang olahraga basket termasuk cabang olahraga yang bergengsi dan diminati di antara cabang olahraga bola besar lainnya. Menurut hasil survei yang dilakukan oleh Nielsen Sports pada tahun 2020, bola basket menjadi olahraga dengan peminat terbanyak ketiga di Indonesia dengan total 48% penggemar. Kondisi tersebut akan lebih baik ketika diikuti dengan upaya pembinaan minat dan bakat pada cabang olahraga bola basket di Indonesia sehingga mampu mengarahkan ke perkembangan prestasi nasional.

\* Received: Jun 2023; Reviewed: Aug 2023; Published: Dec 2023

Hasil Program Garuda Emas KONI pada 1998 menunjukkan bahwa pencapaian prestasi dalam olahraga sangat didukung oleh beberapa faktor, diantaranya adalah pembinaan prestasi. Secara garis besar ada beberapa faktor yang memengaruhi dalam peningkatan prestasi maksimal secara efektif. Adapun faktor-faktor tersebut antara lain: 1) Faktor internal pemain yang meliputi kemampuan individu; 2) Manajemen organisasi yang baik; 3) Program pembinaan; 4) Pemanfaatan sarana dan pra sarana pendukung. Penelitian menggunakan data statistik pertandingan di suatu kompetisi saat ini sedang populer. Menurut Ivankovic et al. (2010) jika data dari suatu pertandingan bola basket dikumpulkan, langkah selanjutnya adalah menerapkan sejumlah metode statistika yang memungkinkan dengan jaminan bahwa statistik yang diperoleh sesuai dengan kejadian di lapangan. Hal tersebut menunjukkan bahwa hasil pertandingan adalah cerminan kemampuan setiap atlet yang berlaga dan dapat dijadikan evaluasi serta pengembangan atlet di masa mendatang. Data pemain Indonesian Basketball League (IBL) 2022 menunjukkan bahwa terdapat 16 pemain (terbanyak daripada daerah lain) yang merupakan alumni Liga Basket DBL Seri Jawa Timur, namun hanya setengahnya yang merupakan pemain utama di tim. Hal tersebut menunjukkan bahwa kemampuan pemain kurang maksimal sehingga jarang mendapatkan menit bermain yang lebih lama.

Pembinaan olahraga nasional merupakan kegiatan berjenjang. Formula tentang pembinaan yang standar dibutuhkan di semua cabang olahraga termasuk sistem seleksi atlet. Sistem seleksi yang baik akan menghasilkan pembinaan yang efektif dengan atlet-atlet potensial dan konsisten. Seleksi atlet dilakukan secara terencana dengan kriteria tertentu yang selanjutnya akan dikelompokkan sesuai prospek atlet tersebut, yaitu kelas atau kelompok Pratama untuk tingkat kompetisi Nasional atau Regional (KONI 2015).

Model yang akan dibentuk dalam penelitian ini bertujuan membantu proses seleksi atlet muda dalam pengelompokan prospek berdasarkan peubah-peubah dari data statistik pertandingan di Liga Basket DBL East Java Series 2019 menggunakan metode regresi logistik biner dan random forest. Regresi logistik biner digunakan ketika peubah respon berupa data kualitatif dikotomi yang bernilai 1 untuk menyatakan keberadaaan karakteristik tertentu dan bernilai 0 untuk menyatakan ketidak beradaaan suatu karakteristik (Sepang et al. 2012). Metode ini merupakan bentuk khusus dari regresi sederhana non-linier untuk menganalisis hubungan antara satu peubah respon dengan beberapa peubah penjelas (Tranmer dan Elliot 2011; Shyntia et al. 2022)

Pemodelan terkait peubah respon dengan tipe kategorik terus mengalami perkembangan (Dito et al. 2019). Salah satu penerapan model tersebut adalah model klasifikasi random forest. Metode klasifikasi random forest memiliki beberapa keunggulan, yaitu dapat mengatasi masalah overfitting, kekar terhadap penciran dan dapat menghasilkan akurasi yang baik (Ali et al. 2012).

## 2. Metodologi

### 2.1 Data

Data yang digunakan dalam penelitian merupakan data statistik pertandingan setiap atlet Liga Basket DBL East Java Series tahun 2019 yang didapatkan melalui

situs resmi DBL Indonesia ([www dbl id/statistics](http://www dbl id/statistics)). Sebanyak 945 amatan (atlet) dari 32 tim (sekolah) di Jawa Timur akan digunakan dalam penelitian. Peubah respon yang akandigunakan adalah prospek atlet muda dengan Y=1 untuk atlet kelompok Pratama danlainnya untuk bukan atlet kelompok Pratama. Peubah penjelas yang digunakan sebanyak 15 peubah yang dapat dilihat pada Tabel 1.

Tabel 1 Rincian peubah yang digunakan.

Nama Peubah	Jenis	Keterangan
Prospek (Y)	Kategorik	1 = Atlet Kelompok Pratama, 0 = Bukan Atlet Kelompok Pratama
Point ( $X_1$ )	Numerik	Total poin selama kompetisi
Point per game ( $X_2$ )	Numerik	Rataan poin selama kompetisi
Assist ( $X_3$ )	Numerik	Total umpan selama kompetisi
Assist per game ( $X_4$ )	Numerik	Rataan umpan selama kompetisi
Block ( $X_5$ )	Numerik	Total blok selama kompetisi
Block per game ( $X_6$ )	Numerik	Rataan blok selama kompetisi
Steal ( $X_7$ )	Numerik	Total mencuri bola lawan selama kompetisi
Steal per game ( $X_8$ )	Numerik	Rataan mencuri bola lawan selama kompetisi
Rebound ( $X_9$ )	Numerik	Total mendapat bola pantul selama kompetisi
Rebound per game ( $X_{10}$ )	Numerik	Rataan mendapat bola pantul selama kompetisi
3-Point ( $X_{11}$ )	Numerik	Total tembakan 3-angka berhasil
3-Point percentage ( $X_{12}$ )	Numerik	Persentase tembakan 3-angka berhasil
Free-Throw ( $X_{13}$ )	Numerik	Total tembakan bebas berhasil
Free-Throw percentage ( $X_{14}$ )	Numerik	Persentase tembakan bebas berhasil

## 2.2 Metode Penelitian

Berikut adalah tahapan analisis yang digunakan dalam penelitian :

1. Praproses data dan melakukan pelabelan data berdasarkan sumber data pendukung sesuai peubah pada Tabel 1.
2. Eksplorasi data dengan melakukan analisis deskriptif pada peubah respon dan penjelas.
3. Melakukan penanganan data tidak seimbang menggunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE).
4. Menggunakan *k-fold cross validation* sebagai teknik validasi model dengan  $k=10$ . Dataset yang terdiri dari contoh sebanyak N akan dibagi menjadi k bagian yang berukuran sama. Nilai k biasanya berupa bilangan kecil seperti 5 atau 10 (Bramer

2007). Masing-masing dari k bagian akan dijadikan sebagai data uji untuk mengevaluasi model, sementara k – 1 bagian lainnya sebagai data latih untuk membangun model. Proses ini akan diulangi sebanyak k kali sehingga setiap bagian menjadi data uji sebanyak satu kali.

5. Memodelkan peubah penjelas dengan peubah respon menggunakan metode regresi logistik biner dengan persamaan sebagai berikut (Agresti 2007) :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)} \quad (1)$$

Fungsi dari  $\pi(x)$  didapatkan dari hasil transformasi berikut:

$$\text{logit}[\pi(x)] = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \quad (2)$$

- a. Melakukan uji multikolinearitas untuk melihat adanya hubungan linear antar peubah penjelas. Menurut O'brien (2007) menyatakan bahwa *Variance Inflation Factor* (VIF) adalah salah satu cara dalam mendeteksi adanya multikolinearitas.

### 3. Hasil dan Pembahasan

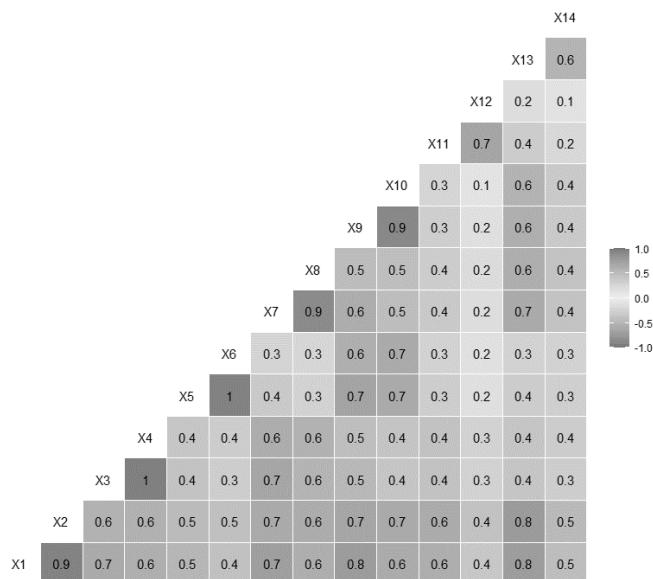
#### 3.1 Praproses dan Eksplorasi Data

Data yang digunakan pada penelitian terdiri dari 945 amatan atlet dengan peubah yang merupakan statistik individu dari setiap pertandingan. Hasil tahap praproses data tidak mendeteksi keberadaan data hilang (missing value) serta dilakukan proses pelabelan untuk megklasifikasikan prospek atlet untuk kelompok pratama berdasarkan ketentuan Pedoman Program Indonesia Emas (PRIMA) yang ditetapkan oleh Komite Olahraga Nasional Indonesia. Kriteria atlet dapat dikategorikan kelompok pratama ( $Y=1$ ) dengan setidaknya memenuhi salah satu diantara ketentuan berikut : 1) Peraih medali pada Pekan Olahraga Nasional (PON) atau PON Remaja; 2) Peraih medali kejuaraan resmi tingkat Nasional; 3) Kriteria atlet dari hasil talent spotting (atlet berbakat baik individu atau ataupun dari klub dan memiliki potensi setelah mendapatkan rekomendasi atau dinominasikan oleh Induk Cabang Olahraga). Berikut hasil statistik deskriptif dari peubah-peubah penjelas yang digunakan.

Tabel 2 Statistik deskriptif peubah penjelas

Peubah	Nilai Minimum	Nilai Maksimum	Rata-rata
Point ( $X_1$ )	0	97	10,54
Point per game ( $X_2$ )	0	16,60	2,33
Assist ( $X_3$ )	0	35	2,38
Assist per game ( $X_4$ )	0	5	0,50
Block ( $X_5$ )	0	15	0,62
Block per game ( $X_6$ )	0	2,50	0,14
Steal ( $X_7$ )	0	24	3,44
Steal per game ( $X_8$ )	0	5	0,79

Rebound ( $X_9$ )	0	91	9,43
Rebound per game ( $X_{10}$ )	0	15,2	2,19
3-Point ( $X_{11}$ )	0	13	0,48
3-Point percentage ( $X_{12}$ )	0	100	6,95
Free-Throw ( $X_{13}$ )	0	30	2,11
Free-Throw percentage ( $X_{14}$ )	0	100	25,14



Gambar 1 Matriks korelasi peubah penjelas

Matriks korelasi menunjukkan sebagian besar peubah memiliki korelasi yang cukup tinggi dan korelasi paling tinggi dimiliki oleh peubah yang berpasangan antara peubah berlabel angka ganjil dengan genap seperti pada peubah  $X_1$  dengan  $X_2$ , peubah  $X_3$  dengan  $X_4$ , peubah  $X_5$  dengan  $X_6$ , peubah  $X_7$  dengan  $X_8$ , peubah  $X_9$  dengan  $X_{10}$ , peubah  $X_{11}$  dengan  $X_{12}$ , dan peubah  $X_{13}$  dengan  $X_{14}$ . Tahapan selanjutnya akan dilakukan pemisahan peubah yang berpasangan tersebut menjadi dua kelompok peubah yang terpisah untuk menghindari eliminasi peubah yang terlalu banyak akibat korelasi yang tinggi sehingga antar peubah penjelas saling bebas. Data yang akan digunakan untuk pemodelan juga memiliki permasalahan ketidakseimbangan kelas antara kategori kelompok atlet pratama dan bukan atlet pratama dengan proporsi 4% banding 96%. Penanganan masalah tersebut dapat dilakukan dengan penerapan SMOTE sebelum memulai tahapan pemodelan. Penerapan algoritma SMOTE dilakukan dengan menambahkan sejumlah amatan minor. Berikut hasil perbandingan proporsi data sebelum dan setelah penerapan SMOTE.

**Tabel 3 Perbandingan data kategori kelompok atlet sebelum dan sesudah proses SMOTE**

Penanganan Data	Frekuensi		Proporsi	
	Atlet Pratama	Bukan Atlet Pratama	Atlet Pratama	Bukan Atlet Pratama
Pra-SMOTE	40	905	0.04	0.96
SMOTE	880	905	0.49	0.51

### 3.2 Regresi Logistik Biner

Model akan diterapkan pada dua kelompok peubah statistik pertandingan. Model pertama sebagai model yang merepresentasikan capaian individu atlet secara keseluruhan dengan peubah penjelas point (X1), assist (X3), block (X5), steal (X7), rebound (X9), 3-point (X11), dan free-throw (X13). Model kedua sebagai model yang merepresentasikan kekonsistenan capaian individu atlet di setiap pertandingan selama kompetisi berlangsung. Model kedua dengan peubah penjelas point per game (X2), assist per game (X4), block per game (X6), steal per game (X8), rebound per game (X10), 3-point percentage (X12) dan free-throw percentage (X14).

Hasil awal kedua model pada Tabel 4 setelah melalui uji parsial dan serentak menunjukkan peubah-peubah penjelas yang tidak signifikan dan berpotensi mengganggu penerapan model dan analisis sehingga tidak akan diikutsertakan dalam penerapan model selanjutnya. Peubah-peubah tersebut adalah block (X5), steal (X7), dan rebound (X9) pada model pertama, sedangkan pada model kedua yaitu block per game (X6), steal per game (X8), dan rebound per game (X10). Pembentukan model baru atau model lebih lanjut dilakukan agar diperoleh model yang lebih baik dengan mengeluarkan peubah yang tidak berpengaruh secara signifikan terhadap model.

**Tabel 4 Hasil awal dan uji multikolinearitas pemodelan regresi logistik biner**

Peubah	Dugaan Koefisien Regresi	Nilai-p	VIF
<b>Model A</b>			
Intersep	-3,329	<2x10 <sup>-16</sup>	-
Point (X <sub>1</sub> )	0,101	<2x10 <sup>-16</sup>	6,145
Assist (X <sub>3</sub> )	0,344	<2x10 <sup>-16</sup>	1,616
Block (X <sub>5</sub> )	-0,013	0,841	1,806
Steal (X <sub>7</sub> )	0,040	0,154	2,303
Rebound (X <sub>9</sub> )	0,008	0,568	3,163
3-Point (X <sub>11</sub> )	-0,373	2,87x10 <sup>-8</sup>	2,147
Free-Throw (X <sub>13</sub> )	-0,184	5,39x10 <sup>-14</sup>	3,026

Model B				
Intersep	-2,291	<2x10 <sup>-16</sup>	-	-
Point per game (X <sub>2</sub> )	0,344	6,32x10 <sup>-16</sup>	3,176	
Assist per game (X <sub>4</sub> )	2,036	<2x10 <sup>-16</sup>	1,505	
Block per game (X <sub>6</sub> )	0,642	0,114	1,664	
Steal per game (X <sub>8</sub> )	0,001	0,997	1,753	
Rebound per game (X <sub>10</sub> )	-0,104	0,064	2,622	
3-Point percentage (X <sub>12</sub> )	-0,028	1,60x10 <sup>-5</sup>	1,411	
Free-Throw percentage (X <sub>14</sub> )	-0,019	5,27x10 <sup>-8</sup>	1,776	

Pengujian parameter pada model lanjutan regresi logistik biner secara simultan dengan semua peubah penjelas pada masing-masing model menghasilkan nilai statistik uji G berturut-turut untuk Model A sebesar 1479,953 dan Model B sebesar 1072,199 hasil uji G kedua model menunjukkan bahwa nilai G lebih besar daripada nilai  $\chi^2(0,05,4) = 9,488$  sehingga tolak hipotesis nol yang artinya minimal terdapat satu peubah penjelas yang memengaruhi prospek atlet muda pada taraf nyata 5%. Pengujian parameter secara parsial pada kedua model lanjutan menggunakan uji Wald yang terdapat pada Tabel 5 menunjukkan bahwa semua peubah penjelas yang digunakan berpengaruh signifikan terhadap prospek atlet muda pada taraf nyata 5%.

Tabel 5 Hasil akhir pemodelan regresi logistik biner

Peubah	Dugaan Koefisien Regresi	Nilai Uji-G ( $\chi^2_{(\alpha, \text{db})}$ )	VIF	Rasio Odds
<b>Model A</b>				
Intersep	-3,396	<2x10 <sup>-16</sup>	-	0,034
Point (X <sub>1</sub> )	0,118	<2x10 <sup>-16</sup>	3,749	1,125
Assist (X <sub>3</sub> )	0,374	<2x10 <sup>-16</sup>	1,292	1,453
3-Point (X <sub>11</sub> )	-0,481	4,18x10 <sup>-14</sup>	1,622	0,618
Free-Throw (X <sub>13</sub> )	-0,197	2,52x10 <sup>-15</sup>	2,811	0,821
<b>Model B</b>				
Intersep	-2,312	<2x10 <sup>-16</sup>	-	0,099
Point per game (X <sub>2</sub> )	0,304	<2x10 <sup>-16</sup>	1,853	1,356
Assist per game (X <sub>4</sub> )	2,006	<2x10 <sup>-16</sup>	1,334	7,435
3-Point percentage (X <sub>12</sub> )	-0,022	0,001	(9,488)	1,279
Free-Throw percentage (X <sub>14</sub> )	-0,020	7,09x10 <sup>-8</sup>	1,649	0,980

Model Logit yang dihasilkan berdasarkan Tabel 5

$$\widehat{g}_1(x) = -3,396 + 0,118X_1 + 0,374X_3 - 0,481X_{11} - 0,197X_{13}, \text{ untuk Model A}$$

$$\widehat{g}_2(x) = -2,312 + 0,304X_2 + 2,006X_4 - 0,022X_{12} - 0,020X_{14}, \text{ untuk Model B}$$

Kejadian sukses pada penelitian ini adalah prospek atlet muda yang diklasifikasikan sebagai kelompok atlet pratama dan bukan kelompok pratama merupakan kejadian gagal. Interpretasi nilai rasio odds dilakukan terhadap semua peubah penjelas yang berpengaruh secara signifikan terhadap prospek atlet muda.

Interpretasi rasio odds hasil Tabel 5 untuk Model A menyatakan bahwa ketika total poin yang dicetak oleh atlet selama kompetisi bertambah satu poin maka kemungkinan atlet berprospek pratama adalah 1,125 kali dibandingkan dengan atlet yang perolehan poinnya konstan atau lebih rendah. Nilai rasio odds untuk peubah  $X_3$  dapat dinyatakan bahwa ketika total umpan yang berhasil dibuat oleh atlet selama kompetisi bertambah satu umpan maka kemungkinan atlet berprospek pratama adalah 1,453 kali dibandingkan dengan atlet yang kontribusi umpannya konstan atau lebih rendah. Nilai rasio odds untuk peubah  $X_{11}$  dapat dinyatakan bahwa ketika total tembakan tiga-angka yang berhasil dibuat oleh atlet selama kompetisi bertambah satu tembakan maka kemungkinan atlet berprospek pratama adalah 0,618 kali dibandingkan dengan atlet yang pencapaian tembakannya konstan atau lebih rendah. Nilai rasio odds untuk peubah  $X_{13}$  dapat dinyatakan bahwa ketika total tembakan bebas yang berhasil dibuat oleh atlet selama kompetisi bertambah satu tembakan maka kemungkinan atlet berprospek pratama adalah 0,821 kali dibandingkan dengan atlet yang pencapaian tembakannya konstan atau lebih rendah.

Interpretasi rasio odds hasil Tabel 5 untuk Model B menyatakan bahwa ketika poin yang dicetak oleh atlet pada setiap pertandingan meningkat maka kemungkinan atlet berprospek pratama adalah 1,356 kali dibandingkan dengan atlet yang perolehan poin di setiap pertandingannya konstan atau lebih rendah. Nilai rasio odds untuk peubah  $X_4$  dapat dinyatakan bahwa ketika umpan yang berhasil dibuat oleh atlet di setiap pertandingan meningkat maka kemungkinan atlet berprospek pratama adalah 7,435 kali dibandingkan dengan atlet yang kontribusi umpan di setiap pertandingannya konstan atau lebih rendah. Nilai rasio odds untuk peubah  $X_{12}$  dapat dinyatakan bahwa ketika tembakan tiga-angka yang berhasil dibuat oleh atlet di setiap pertandingan meningkat maka kemungkinan atlet berprospek pratama adalah 0,978 kali dibandingkan dengan atlet yang pencapaian tembakan tiga-angka di setiap pertandingannya konstan atau lebih rendah. Nilai rasio odds untuk peubah  $X_{14}$  dapat dinyatakan bahwa ketika tembakan bebas yang berhasil dibuat oleh atlet di setiap pertandingan bertambah maka kemungkinan atlet berprospek pratama adalah 0,980 kali dibandingkan dengan atlet yang pencapaian tembakan bebas di setiap pertandingannya konstan atau lebih rendah.

Tabel 6 Evaluasi kinerja model regresi logistik biner

Model Regresi	Sensitivitas	Spesifisitas	AUC
Model A	0,884	0,904	0,955
Model B	0,785	0,855	0,909

Tabel 6 menunjukkan hasil kinerja model regresi logistik biner. Ukuran kebaikan (AUC) dari kedua model memiliki nilai yang tidak kurang dari 0,9 sehingga dapat dikatakan bahwa semua model sangat baik dan terkemuka dalam kinerja

pengklasifikasian (Hosmer et al. 2013). Model A dikatakan lebih baik karena memiliki nilai ukuran kebaikan model yang lebih tinggi dibandingkan dengan model B.

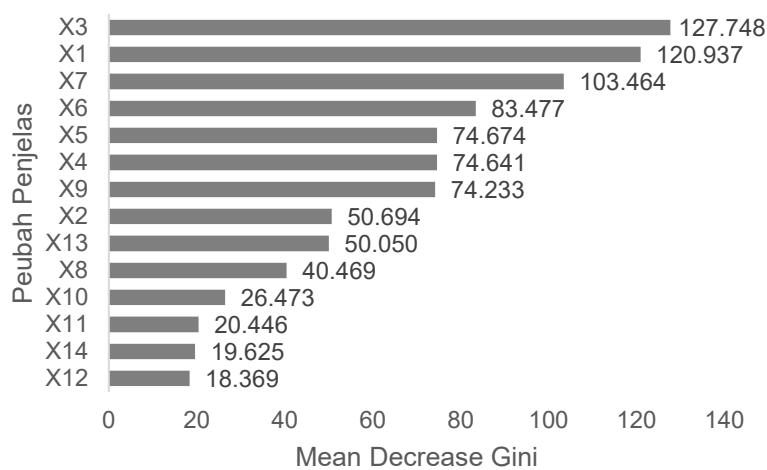
### 3.3 Random Forest

Tahapan pemodelan random forest dimulai dengan penentuan jumlah pohon yang dibangun untuk pemodelan yaitu sebanyak 500 pohon. Jumlah peubah penjelas  $m$  yang digunakan pada pemodelan klasifikasi random forest adalah sebanyak  $\lfloor \sqrt{m} \rfloor$  atau  $\lfloor \sqrt{14} \rfloor$ , hasilnya didapatkan empat peubah penjelas dari setiap proses pemilahan pada pohon yang dibangun (default). Nilai  $m$  lain yang akan digunakan yaitu  $m = 2, 8, 14$ . Hasil tuning parameter  $m$  pada Tabel 7 menunjukkan bahwa nilai  $m = 2$  memiliki nilai AUC terbaik dan akan digunakan sebagai nilai  $m$  untuk tahapan pemodelan selanjutnya.

Tabel 7 Hasil tuning parameter  $m$

Nilai $m$	AUC
2	0,9528
4	0,9312
8	0,9126
14	0,8936

Hasil akhir pemodelan random forest menggunakan teknik validasi  $\sim 10$ -fold cross validation dengan nilai  $m = 2$  didapatkan nilai AUC sebesar 0,944 sehingga model dapat dikatakan sangat baik. Peubah penting pada tahapan pemodelan dapat diketahui melalui nilai Mean Decrease Gini (MDG). Peubah penting merupakan peubah yang memberikan pengaruh besar terhadap kategori respon yaitu prospek atlet muda. Hasil pada Gambar 5 menunjukkan bahwa peubah assist (X3) dan point (X1) merupakan dua peubah terpenting dalam model karena memiliki nilai MDG terbesar.



Gambar 2 Tingkat kepentingan peubah model random forest

### 3.4 Evaluasi

Perbedaan regresi logistik biner dan random forest sebagai metode klasifikasi salah satunya yaitu regresi logistik biner merupakan pemodelan klasifikasi yang berbasis sebaran binomial, sedangkan random forest tidak berbasis sebaran. Metode random forest membangkitkan pohon keputusan hingga mencapai ukuran maksimum.

Regresi logistik biner memiliki beberapa asumsi yang harus terpenuhi untuk pembentukan model, seperti melakukan uji parameter dan multikolinearitas sehingga terdapat kemungkinan ada peubah yang tidak diikutsertakan dalam tahapan selanjutnya. Berdasarkan hal tersebut terdapat keterbatasan untuk melihat pengaruh seluruh peubah karena interpretasi rasio odds pada model regresi logistik biner tidak dapat dilakukan pada peubah yang tereliminasi. Sudut pandang lain terkait tingkat kepentingan peubah secara menyeluruh dapat diidentifikasi dengan metode random forest melalui perhitungan nilai Mean decrease Gini. Meskipun pada tahapan random forest hanya menggunakan sebanyak  $m$  acak peubah penjelas yang kurang dari  $p$  total peubah penjelas pada setiap proses pembentukan pohon keputusan, akan tetapi secara tidak langsung random forest tetap mengikutsertakan semua peubah penjelas tanpa terkecuali. Hal tersebut dilakukan untuk mengurangi kemungkinan korelasi yang besar antara setiap pohon yang ada. Hasil kedua metode juga menunjukkan beberapa hal yang dapat dikatakan sesuai. Nilai kebaikan dari model yang dibangun menggunakan metode regresi logistik biner dan random forest dapat dikatakan sangat baik dalam pengklasifikasian prospek atlet.

#### 4. Simpulan

Hasil penerapan metode regresi logistik biner dan random forest menunjukkan bahwa model yang dibangun dianggap sangat baik dalam kinerja klasifikasi prospek atlet berdasarkan ukuran kebaikan atau nilai Area Under Curve yang tidak kurang dari 0,9. Peubah assist ( $X_3$ ) dan point ( $X_1$ ) dapat dikatakan sebagai peubah yang paling memengaruhi prospek atlet muda. Hal tersebut dapat dilihat dari hasil interpretasi rasio odds yang menunjukkan bahwa atlet dengan pencapaian poin dan kontribusi umpan yang lebih banyak dibandingkan dengan atlet lain akan memiliki kecenderungan berprospek atlet pratama. Peubah assist ( $X_3$ ) dan point ( $X_1$ ) juga merupakan dua peubah terpenting dalam pemodelan random forest berdasarkan nilai Mean Decrease Gini. Berdasarkan hasil penelitian ini, atlet maupun pihak terkait seperti bagian kepelatihan atau induk cabang olahraga dapat lebih memperhatikan kemampuan dari aspek mencetak angka dan olah bola sebagai indikator utama dari hasil penelitian ini dengan catatan tetap mempertimbangkan peubah atau skill set lain sebagai aspek penunjang prospek atlet muda.

#### Daftar Pustaka

- Agresti A. 2007. *An Introduction to Categorical Data Analysis*. New Jersey (US): John Wiley and Sons.
- Ali J, Khan R, Ahmad N, Maqsood I. 2012. Random Forests and Decision Trees. *International Journal of Computer Science Issues*. 9(3) : 272-278.
- Breiman, L. 2001. Random Forests. *Machine Learning*. 45 : 5–32.
- Dito GA, Safitri A, Afendi FM, Anisa R, Salim A, Sartono B. 2019. Graphical User Interface (GUI) for the Least Absolute Shrinkage and Selection Operator (LASSO) Regression. IOP Conference Series: Earth and Environmental Science. 299(1): 012031. doi:10.1088/1755-1315/299/1/012031.

- Hosmer DW, Lemeshow S, Sturdivant RX. 2013. *Applied Logistic Regression*. New Jersey (US): John Wiley and Sons.
- [KONI] Komite Olahraga Nasional Indonesia. 2015. Kriteria Seleksi Calon Atlet, Pelatih, dan Manajer Program Indonesia Emas (PRIMA). Jakarta : KONI.
- Ivankovic Z, Rackovic M, Markoski B, Radosav D, Ivkovic M. 2010. Appliance of Neural Networks in Basketball Scouting. *Acta Polytechnica Hungarica*. 7(4) : 167- 180.
- O'brien RM. 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Uality & Quantity*. 41(5) : 673-690.
- Sepang F, Komalig H, Hatidja D. 2012. Penerapan Regresi Logistik untuk Menentukan Faktor-Faktor yang Memengaruhi Pemilihan Jenis Alat Kontrasespri di Kecamatan Modayag Barat. *Jurnal Mipa Unsrat*. 1(1):1-5.
- Shyntia, F. D., Kurnia, A. and Dito, G. A. (2022) Pendekatan Metode CHAID dan Regresi Logistik dalam Menganalisis Faktor Berpengaruh pada Kejadian Stunting di Provinsi Jawa Barat, *Xplore: Journal of Statistics*, 11(1), pp. 48–58. doi: 10.29244/xplore.v11i1.857.
- Tranmer M, Elliot M. 2011. *Binary Logistic Regression*. Manchester (UK) : Cathie Marsh Center.