

PREDIKSI MARKET VALUE PEMAIN SEPAK BOLA DI LIMA LIGA TOP EROPA MENGGUNAKAN *K-NEAREST NEIGHBOR*

Hirzi Firdausi^{1§}, I Wayan Sumarjaya², I Gusti Ayu Made Srinadi³

¹Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: hirzifirdausi@gmail.com]

²Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: sumarjaya@unud.ac.id]

³Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: srinadi@unud.ac.id]

[§]Corresponding Author

ABSTRACT

Every football club in competition at any country has ambition to be a champion. One of the important things to be a champion is the quality of player in the team. Therefore, when transfer time is opened, almost every club buys the required players. Busy transfer activity frequently occurs in five Europe top leagues. Nevertheless, a club frequently buys a player which has price beyond market value transfer (market value). The reason is that many clubs still don not know how to determine market value. The aim of this research is to predict the market value of football players in five Europe top leagues using K-nearest neighbor (KNN). Data which will be used are 26 football players who play in five Europe top leagues with club get into first tier league (promotion) in season of 2022/2023. The result is that the model which is used for prediction is a KNN model which has proportion of training data and test data at the ratio of 90:10 and parameter $K = 3$ since it has the lowest MAPE which is 10,45 %. From model selection we obtain MAPE value at 28,614 % for compared prediction of market value result with actual of market value.

Keywords: *football, market value, K-nearest neighbor (KNN)*

1. PENDAHULUAN

Sepak bola merupakan olahraga paling tenar di dunia (Majewski, 2014). Hampir setiap negara mempunyai kompetisi sepak bola sendiri yang dijadikan sebagai tontonan hiburan untuk masyarakat negara tersebut. Setiap klub dalam kompetisi di negara mana pun berambisi untuk menjadi juara. Menurut Devi (2014), hal yang paling penting untuk klub sepak bola untuk menjadi juara adalah pemain sepak bola itu sendiri. peluang untuk menjadi juara semakin besar, jika memiliki pemain-pemain hebat (Devi, 2014). Oleh karena itu, pada waktu transfer pemain dibuka, hampir setiap klub membeli pemain-pemain yang dibutuhkan.

Aktivitas transfer sibuk sering terjadi di lima liga top Eropa yaitu Premier League yang merupakan liga di negara Inggris, Serie A yang merupakan liga di negara Italia, Bundesliga yang merupakan liga di negara Jerman, Ligue 1 yang merupakan liga di negara Perancis, dan La Liga yang merupakan liga di negara Spanyol. Pemain yang mempunyai nilai tinggi kebanyakan berada pada posisi sebagai

penyerang. Namun, sebuah klub sering membeli pemain dengan harga jauh diatas pasaran nilai pasar transfer (*market value*).

Istilah yang digunakan dalam sepak bola untuk menentukan seberapa besar asset pemain adalah *market value* (Kaukab dan Falah, 2021). Pada banyak kasus, kualitas pemain yang dibeli belum sesuai dengan harga yang dibeli. Ini membuat rugi klub dari sisi finansial klub maupun untuk performa tim. Selain itu, banyak klub menjual pemain tidak sesuai dengan kualitas, sangat sering dijual di bawah harga yang sesuai dengan kualitasnya. Hal ini dikarenakan klub masih banyak yang tidak mengetahui cara menentukan *market value*. Untuk menghindari masalah tersebut, diperlukan sebuah cara agar klub dapat membeli pemain dengan harga yang sesuai kualitas pemain yang dibeli.

Sebelumnya, Adiwiyana dan Harymawan (2021) meneliti tentang faktor-faktor yang menentukan *market value* sepak bola professional di Indonesia. Penelitian tersebut menggunakan metode regresi linear berganda dan menggunakan data pemain dengan posisi

penyerang dan gelandang yang berkompetisi di liga Indonesia. Penelitian tersebut menyimpulkan bahwa jumlah goal, jumlah menit bermain dan jumlah *starting 11* seorang pemain tidak memengaruhi *market value*. Di lain pihak, penelitian yang dilakukan oleh Majewski (2016), tentang identifikasi faktor-faktor yang menentukan *market value* dari pemain-pemain sepak bola yang paling berharga, menyimpulkan bahwa jumlah *goal* dan *assists* berpengaruh kuat terhadap *market value* pesepak bola pada posisi penyerang.

Situs Transfermarkt menyediakan informasi tentang pemain sepak bola di berbagai liga seluruh dunia. Ketersediaan data yang ada pada situs tersebut, memungkinkan peneliti untuk memprediksi nilai pasar transfer pemain baru dengan menggunakan metode-metode statistika atau *machine learning*. Salah satunya adalah *k-nearest neighbors (KNN)*.

K-nearest neighbors (KNN) adalah metode klasifikasi yang menentukan sebuah objek berdasarkan data latih yang mempunyai jarak terdekat pada objek tersebut. Keunggulan metode *KNN* adalah membuat model sangat cepat, sederhana, dan efektif jika data pelatihan banyak (Mutrofin dkk. 2014). Selain itu banyak penelitian yang menggunakan metode *KNN*. Pada penelitian Alkatib et al. (2013) yang memprediksi harga saham memiliki nilai residual mendekati nol, bahkan beberapa nilai residualnya nol yang berarti data yang diprediksi dengan data yang sebenarnya mempunyai nilai yang sama. Selain *KNN* berfungsi untuk melakukan klasifikasi, *KNN* juga dapat digunakan melakukan prediksi untuk mendapatkan nilai eksak. Metode *KNN* untuk melakukan prediksi masih jarang digunakan pada penelitian-penelitian sebelumnya. Oleh karena itu, penulis melakukan penelitian prediksi *market value* pemain sepakbola dengan posisi sebagai penyerang yang bermain pada lima liga top Eropa yaitu liga Inggris, liga Italia, liga Spanyol, liga Perancis dan liga Jerman menggunakan *KNN*.

2. METODE PENELITIAN

Data pada penelitian merupakan data sekunder sebanyak 300 data hasil akhir pada musim 2021-2022 di Premier league, Serie A, Bundesliga, Ligue 1, dan La liga diperoleh dari laman web www.transfermarkt.com. Variabel-variabel pada penelitian ini adalah :

1. Market value

Market value merupakan variabel terikat .Data yang digunakan adalah market value pemain sepak bola di Premier league, Serie A, Bundesliga, Ligue 1, dan La liga pada musim 2021-2022.

2. Jumlah goal

Pada penelitian yang dilakukan oleh Majewski (2016) diperoleh bahwa jumlah goal yang dicetak berpengaruh positif terhadap market value. Jumlah goal merupakan variabel bebas.

3. Jumlah assist

Jumlah assist berpengaruh positif pada market value pada penelitian yang dilakukan oleh Adiwiyana dan Harymawan (2021). Jumlah assist merupakan variabel bebas.

4. Jumlah kartu kuning

Menurut penelitian yang dilakukan oleh Adiwiyana dan Harymawan (2021) menyimpulkan bahwa mempunyai pengaruh positif terhadap market value. Sehingga penelitian ini menggunakan jumlah kartu kuning sebagai variabel bebas.

5. Starting eleven

Starting eleven mempunyai pengaruh positif signifikan terhadap market value. Oleh karena itu, penelitian ini memerlukan starting eleven sebagai variabel bebas.

6. Umur

Menurut penelitian yang dilakukan oleh Adiwiyana dan Harymawan (2021), umur merugikan market value pesepak bola, yang mempunyai arti semakin tua pemain, semakin rendah market value pemain karena mereka tidak dalam usia yang produktif lagi. Oleh sebab itu, umur dijadikan variabel bebas.

Aplikasi yang digunakan yaitu Microsoft Excel dan bahasa pemrograman Python dengan Jupyter Notebook sebagai text editor. Langkah-langkah yang akan dilakukan dalam penelitian adalah sebagai berikut.

1. Mencari data market value, jumlah goal, umur, assist, starting eleven, dan kartu kuning di lima liga top Eropa.
2. Memisahkan variabel dependet yaitu market value dari dataset.
3. Membagi dataset menjadi data uji dan data latih. Karena menggunakan 10-fold cross

- validation dataset dibagi menjadi 10 partisi. Sembilan partisi akan digunakan sebagai data latih dan 1 partisi akan digunakan sebagai data uji. Pembagian proporsi data latih dan data uji dibuat untuk tiga skenario yaitu 70:30, 80:20, dan 90:10.
4. Data latih dibagi menjadi kelompok-kelompok data yang sesuai.
 5. Pada setiap skenario, model KNN dipilih dengan parameter K dengan nilai yaitu 3,4,5,6,7,8,9, dan 10. Nilai K yang kecil menyebabkan overfitting, Sedangkan sebaliknya, nilai K yang terlalu besar akan menyebabkan underfitting (Seruni, 2020). Oleh karena itu, K yang diambil adalah 3, 4, 5, 6, 7, 8, 9, dan 10.
 6. Prediksi data uji dengan model KNN dari data latih.
 7. Menghitung MAPE
 8. Ulangi langkah 3 sampai 7 hingga iterasi ke sepuluh karena menggunakan 10--fold cross validation.

9. Memilih model dengan MAPE terendah setelah melakukan 10 kali iterasi untuk membuat prediksi data baru pada setiap skenario.
10. Prediksi data baru. Data baru yang digunakan adalah data-data pemain di setiap lima liga top Eropa pada tahun 2022-2023 dengan tim yang baru dipromosikan atau masuk ke dalam liga tersebut.

3. HASIL DAN PEMBAHASAN

1.1. Data Penelitian

Data pada penelitian merupakan data sekunder sebanyak 300 data hasil akhir pada musim 2021-2022 di Premier league, Serie A, Bundesliga, Ligue 1, dan La liga diperoleh dari laman web www.transfermarkt.com.

1.2. Statistika Deskriptif

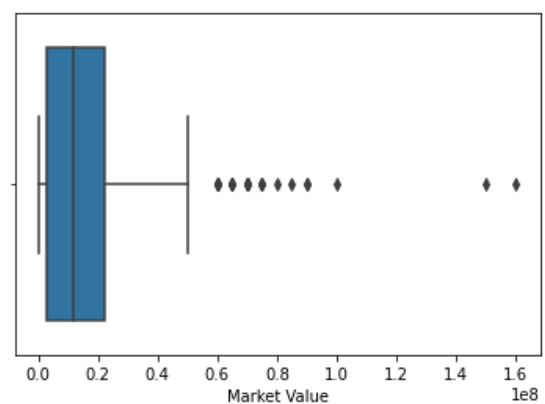
Ringkasan data *market value*, *goal*, *assist*, *yellow card*, *starting eleven* dan umur dapat dilihat pada tabel 3.1

Tabel 3.1. Ringkasan Dataset

	<i>Market value</i> (€)	<i>Goal</i>	<i>Assist</i>	<i>Yellow card</i>	<i>Starting eleven</i>	<i>Age</i>
Mean	18.143.170	9,016	4,296	3,573	23,126	25,723
Deviasi standar	22.547.720	7,420	3,919	2,730	11,374	4,574
Minimum	200.000	0	0	0	0	16
Q1	3.000.000	4	2	1	14,750	22
Q2	12.000.000	7	3	3	24	25
Q3	22.250.000	12	6	6	32	29
Maksimum	160.000.000	50	26	11	49	39

Pada Tabel 3.1, kolom *market value* menunjukkan rata-rata *market value* adalah 18.143.170 €. Nilai standar deviasi adalah 22.547.720 € lebih besar dari pada nilai rata-rata, ini berarti bahwa data yang tersebar bervariasi. *Market value* terendah adalah pemain yang bermain di liga perancis Jason mbock bernilai 200.000 € dan tertinggi adalah seorang pemain yang bermain di liga perancis, Kylian mbappe yang bernilai 160.000.000 €. Pemain tersebut juga menjadi pencetak assist terbanyak dengan jumlah 26. Jumlah goal terbanyak adalah 50 yang diraih oleh Robert lewandowski. Ini menjadikan Robert lewandowski menjadi pemenang golden shoe 2021—2022 di Eropa.

1.3. Deteksi Pencilan



Gambar 3.1. Pencilan

Pada Gambar 3.1, ditemukan banyak pencilan pada sebaran data *market value*. penentuan

pencilan diperoleh jika ada data yang melebihi perhitungan dari:

$$Q_3 + (1,5 \times (Q_3 - Q_1))$$

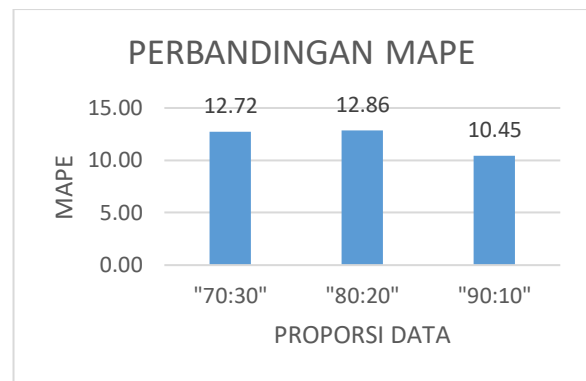
Dimana Q_3 adalah kuartil tiga dan Q_1 adalah kuartil satu. Terdapat 24 pencilan dengan nilai *market value* lebih dari 50.000.000 pada data ini. Pencilan-pencilan pada penelitian ini bukan menandakan bahwa terdapat kesalahan pada input data. Setelah dilihat pada dataset, data—data yang menjadi pencilan memiliki jumlah rata-rata *goal*, *assist*, kartu kuning, dan *starting eleven* lebih besar dari dataset. Sehingga dapat dikatakan bahwa outlier pada kasus ini memiliki arti khusus. Karena penelitian ini memiliki banyak pencilan dan pada variabel dependen tidak memiliki nilai yang mendekati 0 atau 0, pengukuran kinerja model pada penelitian ini menggunakan metode *mean absolute percentage error* (MAPE).

1.4. Data Latih dan Data Uji

Untuk membuat model mengenali pola pada data, diperlukan data uji dan data latih dari dataset. Oleh karena itu, dataset dibagi menjadi empat bagian yaitu data latih untuk variabel bebas, data latih untuk variabel dependent, data uji untuk variabel bebas, dan data uji untuk variabel dependent. Data—data tersebut diambil secara acak. Pada penelitian ini, terdapat tiga skenario dalam pembagian proporsi antara data uji dan data latih yaitu 70:30, 80:20, dan 90:10.

1.5. Evaluasi Model

Karena pada penelitian ini menggunakan *10-fold cross validation*, dataset dibagi menjadi 10 partisi. sembilan partisi digunakan sebagai data latih dan satu partisi digunakan sebagai data uji. Selanjutnya, melakukan pemodelan menggunakan *K nearest neighbor* dengan parameter nilai K pada KNN yaitu 3, 4, 5, 6, 7, 8, 9, dan 10 untuk mencari model KNN dengan parameter nilai K yang memiliki nilai rata-rata MAPE terkecil.



Gambar 3. 1 MAPE Proporsi Data 70:80, 80:20, dan 90:10

Dari tiga skenario pada Gambar 3.2 dapat disimpulkan bahwa nilai MAPE terkecil adalah 10,45% yang terdapat pada skenario 90 % data latih dan 10 % data uji dengan parameter $K = 3$. Sehingga model KNN dengan parameter $K = 3$ dan skenario data latih 90 % dan data uji 10 % dipilih sebagai model yang optimal untuk memprediksi.

1.6. Hasil Prediksi

Data yang digunakan untuk memprediksi adalah data pemain sepak bola posisi penyerang yang bermain di klub yang baru naik ke lima liga top Eropa pada musim 2022/2023 sebanyak 26 data.

Tabel 3. 1 Statistika Deskriptif Untuk Data Prediksi

	<i>Actual market value</i> (€)	<i>Market value</i> musim 2021-2022(€)	Prediksi <i>market value</i> (€)
Jumlah data	26	26	26
Mean	11.423.080	6.619.231	11.382.050
Deviasi standar	9.278.548	6.672.332	9.050.456
Mini-mum	1.500.000	1.000.000	2.666.667
Q1	4.825.000	1.850.000	4.391.667
Q2	8.000.000	3.500.000	8.500.000
Q3	17.500.000	15.250.000	15.250.000
Maksi-mum	32.000.000	30.000.000	30.000.000

Pada Tabel 3.2, nilai rata-rata *market value* pada musim 2021-2022 adalah € 6.672.332. Nilai rata-rata *market value* meningkat pada akhir musim 2022-2023 menjadi € 11.423.080. Ini berarti pada nilai *market value* saat awal musim pada kompetisi tahun 2022/2023 mengalami peningkatan hampir dua kali lipat dibandingkan musim 2021/2022. Hasil prediksi *market value* musim 2022/2023 dengan model KNN dengan parameter $k = 3$ memiliki nilai rata-rata € 11.382.050. Nilai MAPE Jika dibandingkan dengan nilai *market value* aktual 2022/2023 dengan nilai prediksi *market value* 2022/2023 adalah sebesar 28,614%. Nilai MAPE sebesar 28,614% dapat dikategorikan cukup.

4. KESIMPULAN DAN SARAN

Berdasarkan pembahasan yang sudah diuraikan, kesimpulan yang diperoleh adalah Model KNN yang paling optimal untuk memprediksi nilai *market value* pemain sepakbola posisi penyerang di lima liga top Eropa adalah model KNN dengan parameter $K = 3$ dan proporsi data uji dan data latih adalah 90:10 yang memiliki Nilai MAPE sebesar 10,45 % dan Nilai MAPE pada hasil prediksi nilai *market value* dibandingkan nilai aktual *market value* adalah 28,614 %. Nilai MAPE sebesar 28,614% dapat dikategorikan cukup. Saran yang diberikan oleh penulis untuk penelitian selanjutnya adalah memprediksi *market value* di lima liga top Eropa menggunakan metode lain dan menambahkan variabel lain.

DAFTAR PUSTAKA

- Adiwiyan, H.I & Harymawan, I. (2021). Factors that determine the market value of professional players. *Jurnal Dinamika Akutansi*, 13(1), 51-61.
- Alkhatib, K., Najadat, H., Hmeidi, I., & Shatnawi, M.K.A. (2013). Stock prediction using k-nearest neighbor (KNN) algorithm. *International Journal of Business, Humanities and Technology*, 2(3), 32-44.
- Devi, A.P.(2004). Akutansi untuk pemain sepak bola. *Jurnal Akutansi dan Keuangan Indonesia*, 1. 38-53. DOI: 10.21002/jaki.2004.04

- Han, J., Kamber, M. & Pei, J. (2012), *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Imandoust, S.D., & Bolandraftar, M.(2013). Application of k-nearest (KNN) approach for predicting economic events : theoretical background. *Int. Journal of Engineering Research and Applications*, 3(5), 605-610.
- Kaukab, M.E & Falah, Z.N. (2021). Football player market value: Apakah usia pemain berperan dalam penentuan harga pasar ?. *Jurnal penelitian dan pengabdian kepada masyarakat UNSIQ*, 9(1), 24-37.
- Koloğlu, Y., Birinci, H., Kanalmaz, S.I., & Özyilmaz, B.(2018). A multiple linear regression approach for estimating the market value of football players in forward position. *Arxiv:1807.01104*. <https://doi.org/10.48550/arXiv.1807.01104>
- Kurniawan, C.C., Rostianingsih, S., & Santoso, L.W.(2022). Penerapan metode knn-regresi dan multiplicative decomposition untuk prediksi data penjualan pada supermarket x. *Jurnal Infra*, 10(2),
- Majewski, S.(2016). Identification of determining market value of the most valuable players. *Central European Management Journal*, 3. 91-104.
- Yustanti, W. (2012). Algoritma k-nearest neighbour untuk memprediksi harga jual tanah. *Jurnal Matematika, Statistika & Komputasi*, 9(1), 57-68.