

Prediksi Hasil *FIFA World Cup Qatar 2022* Menggunakan *Machine Learning* dengan *Python*

Syahrul Zein, Gani Gunawan*

Prodi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ARTICLE INFO

Article history :

Received : 14/8/2022
Revised : 15/12/2022
Published : 21/12/2022



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Volume : 2
No. : 2
Halaman : 153-162
Terbitan : **Desember 2022**

ABSTRAK

Memprediksi hasil pertandingan adalah hal yang besar ketika selalu diungkapkan sebelum pertandingan dimulai, baik dari penggemar maupun analis. Untuk memprediksi kecocokan, diperlukan suatu teknologi yang dapat mengolah data input, analisis dan output yaitu Machine Learning. Machine Learning adalah sebuah sistem atau komputer untuk 'belajar' secara mandiri dan meningkatkan kemampuannya secara otomatis tanpa memerlukan instruksi pemrograman tertulis yang eksplisit. beberapa algoritma untuk memprediksi suatu kejadian yaitu Logistic Regression, K-Nearest Neighbors, Naïve-Bayes, Support Vector Machine, Neural Network dan Random Forest. Dalam penelitian ini, dengan bantuan Scikit-Learn on Python digunakan untuk mengukur akurasi FIFA World Cup 2006-2018 dan memprediksi hasil FIFA World Cup 2022. Algoritma SVM memiliki tingkat akurasi tertinggi pada tahun 2010 dan 2014. Artinya, jika dua tim yang sama bertemu di tahun 2010, jika bertemu lagi 4 tahun kemudian, mereka akan memiliki peluang yang sama untuk menang. Algoritma Neural Network memiliki tingkat akurasi tertinggi pada tahun 2006 dan 2018. Artinya jika dua tim yang sama bertemu pada tahun 2006, jika mereka bertemu lagi 12 tahun kemudian, mereka akan memiliki peluang yang sama untuk menang. Hasil prediksi menunjukkan bahwa Jerman berpeluang menjadi juara FIFA World Cup 2022.

Kata Kunci : Machine Learning; FIFA World Cup; Python.

ABSTRACT

Predicting the outcome of the match is a big thing when it's always expressed before the game starts, both from fans and analysts. To predict a match, we need a technology that can process input, analysis and output data, namely Machine Learning. Machine Learning is a system or computer to 'learn' independently and improve its capabilities automatically without the need for explicit written programming instructions. several algorithms to predict an event, namely Logistic Regression, K-Nearest Neighbors, Naïve-Bayes, Support Vector Machine, Neural Network and Random Forest. In this study, with the help of Scikit-Learn on Python, it was used to measure the accuracy of the FIFA World Cup 2006-2018 and predicts the results of the FIFA World Cup 2022. The SVM algorithm has the highest accuracy rate in 2010 and 2014. This means that if the same two teams meet in 2010, if they meet again 4 years later, they will have the same chance of winning. The Neural Network algorithm has the highest accuracy rate in 2006 and 2018. This means that if the same two teams met in 2006, if they met again 12 years later, they would have the same chance of winning. Prediction results show that Germany has a chance to win the FIFA World Cup 2022.

Keywords : Machine Learning; FIFA World Cup; Python.

A. Pendahuluan

Sepak bola merupakan olahraga yang memiliki penggemar terbesar di dunia, dan tidak diragukan lagi semua orang menyukai permainan yang indah ini. Selain itu, sepak bola tidak hanya sekedar olahraga, tetapi sepak bola menjadi sebuah industri yang mencakup beberapa aspek seperti: bisnis, hiburan dan perkembangan teknologi, serta pengetahuan sepak bola. Atas hal tersebut, banyak yang memikirkan bagaimana memprediksi hasil pertandingan sepak bola.

Prediksi hasil pertandingan merupakan sebuah hal yang besar. Prediksi ini selalu diutarakan sebelum pertandingan dimulai, baik dari penggemar maupun para analis. Terdapat beberapa faktor yang berpengaruh terhadap suatu prediksi, seperti jumlah kemenangan, imbang, jumlah gol, peringkat tim, dan faktor lainnya. Untuk memprediksi suatu pertandingan, diperlukan suatu teknologi yang dapat mengolah data masukan, analisis dan data keluaran. Teknologi tersebut dinamakan teknologi kecerdasan buatan atau Artificial Intelligence (AI). AI adalah simulasi dari kecerdasan yang dimiliki oleh manusia yang dimodelkan ke dalam mesin dan diprogram agar bisa berpikir seperti halnya manusia. Salah satu bagian dari AI adalah Machine Learning (ML). ML merupakan bagian dari AI yang memungkinkan sistem atau komputer untuk 'belajar' secara mandiri dan meningkatkan kemampuannya secara otomatis tanpa perlu instruksi pemrograman yang dituliskan secara eksplisit [1][2]. ML memiliki beberapa algoritma untuk memprediksi suatu kejadian, yaitu Logistic Regression, K-Nearest Neighbors, Naïve-Bayes, Support Vector Machine, Neural Network, Random Forest [3][4].

Pada penelitian ini, algoritma Logistic Regression, K-Nearest Neighbors, Naïve-Bayes, Support Vector Machine, Neural Network, Random Forest, dengan bantuan Scikit-Learn pada Python digunakan untuk mengukur keakuratan pada kompetisi FIFA World Cup 2006 s.d. 2018 dan memprediksi hasil dari FIFA World Cup Qatar 2022. Hal ini dilakukan karena belum ada penelitian yang melakukan prediksi terhadap kompetisi FIFA World Cup. Kompetisi ini merupakan kompetisi paling bergengsi di dunia dan diselenggarakan setiap 4 tahun sekali. Kompetisi ini juga membuat para penggemar sepak bola memiliki antusias yang sangat tinggi. Parameter yang digunakan adalah statistik kekuatan lini serang, lini tengah, lini belakang dan rata-rata kekuatan dari semua lini pada permainan eFootball 2022. Hasil yang didapatkan berupa peluang 2 tim ketika bertemu. Peluang yang dimaksud adalah peluang Home Win (H), Draw (D) dan Away Win (A).

Berdasarkan uraian latar belakang yang telah disampaikan, maka didapatkan beberapa rumusan permasalahan, yaitu sebagai berikut: (1) Bagaimana menyusun algoritma Machine Learning untuk memprediksi hasil pertandingan FIFA World Cup pada Python?; (2) Bagaimana hasil akurasi dari algoritma Machine Learning untuk hasil pertandingan FIFA World Cup sebelumnya?; (3) Bagaimana memprediksi hasil FIFA World Cup Qatar 2022 dengan algoritma Machine Learning?

Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb. (1) Menyusun algoritma Machine Learning untuk memprediksi hasil pertandingan FIFA World Cup pada Python; (2) Mengetahui hasil akurasi dari algoritma Machine untuk hasil pertandingan FIFA World Cup sebelumnya; (3) Memprediksi hasil FIFA World Cup Qatar 2022 dengan algoritma Machine Learning; (4) Penelitian ini tidak akan membahas proses perhitungan atau hasil secara manual, namun proses tersebut dilakukan pada Bahasa pemrograman Python dengan bantuan library Scikit-Learn.

B. Metode Penelitian

Pada penelitian ini data yang digunakan terdiri dari tiga jenis data, yaitu data kejadian masa lalu, statistik performa tim pada eFootball 2022 dan jadwal FIFA World Cup Qatar 2022. Data kejadian masa lalu merupakan data yang akan diolah oleh Machine Learning sebagai parameter atau variable independent dan tolak ukur keakuratan suatu algoritma. Statistik performa tim pada eFootball 2022 merupakan data performa setiap tim yang dirangkum dari konsol permainan eFootball 2022. Hal ini digunakan sebagai parameter, karena kurangnya ketersediaan statistik performa tim yang ada. Statistik yang digunakan untuk parameter adalah lini serang tuan rumah (*Home Attack*), lini tengah tuan rumah (*Home Midfielder*), lini belakang tuan rumah (*Home Defender*), rata-rata kekuatan dari seluruh lini tuan rumah (*Home Overate*), lini serang tamu (*Away Attack*), lini tengah tamu (*Away Midfielder*), lini belakang tamu (*Away Defender*), dan rata-rata kekuatan dari seluruh lini tamu (*Away Overate*). Jadwal FIFA World Cup Qatar 2022. Jadwal pertandingan FIFA World Cup Qatar

2022 merupakan data yang akan menjadi bahan prediksi oleh Machine Learning. Data berasal dari situs website resmi penyelenggara pertandingan, yaitu www.fifa.com [5].

Penelitian ini menggunakan 6 metode, yaitu Logistic Regression, K-Nearest Neighbors, Naïve-Bayes, Random Forest, Support Vector Machine dan Artificial Neural Network.

Logistic Regression

Logistic Regression (LR) merupakan algoritma yang digunakan untuk melakukan suatu prediksi peluang suatu kejadian dengan mencocokkan data pada fungsi logit [6]. Dalam LR, model tersebut dinamakan sigmoid function. Rumus sigmoid function pada penelitian ini adalah sebagai berikut:

$$y_{uji}^{(j)} = \frac{e^{\beta_0 + \sum_{i=1}^n \beta_{baru}^{(i)} x_{uji}^{(i)}}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_{baru}^{(i)} x_{uji}^{(i)}}} \tag{1}$$

K-Nearest Neighbors

K-Nearest Neighbors adalah algoritma untuk melakukan klasifikasi kepada objek berdasarkan data latih yang jaraknya paling dekat dengan objek tersebut dan tidak mempelajari cara mengkategorikan data, hanya mengingat data yang sudah ada [7]. Pada K-NN, huruf K adalah jumlah tetangga yang terdekat. Nilai k dipilih berjumlah ganjil untuk menghindari munculnya jarak yang sama. Untuk mencari k terbaik tergantung pada datasetnya sendiri atau tergantung terhadap data yang kita uji. Dalam penelitian ini, jumlah pertandingan yang akan diprediksi sebanyak 64. Karena nilai K harus berjumlah ganjil, maka nilai K yang paling optimum adalah 63. Untuk menemukan kedekatan terhadap tetangga, menggunakan persamaan Manhattan [8]. Persamaan Manhattan sebagai berikut:

$$d(x, y^{(i)}) = \sum_{i=1}^n |x_{test}^{(i)} - x_{latih}^{(i)}| \tag{2}$$

Dimana:

d(x,y) = jarak

xtest = Parameter yang akan diuji

xlatih = Parameter yang akan dilatih

i = banyaknya data

Setelah mengetahui jarak semua data latih terhadap data uji, selanjutnya mengurutkan jarak yang paling kecil ke paling besar. Karena K=63, maka diambil 63 data latih yang paling atas. Kemudian jumlahkan masing-masing berapa banyak output Home Win (2), Draw (1) dan Away Win (0). Setelah itu hitung peluang masing-masing output dengan persamaan dibawah ini:

Naïve-Bayes

Naïve-Bayes merupakan algoritma hasil pengembangan dari Teorema Bayes dimana untuk mengklasifikasi suatu kejadian perlu adanya asumsi bahwa efek dari nilai atribut bersifat independent atau bebas, sehingga keberadaan setiap variabel tidak berkaitan dengan nilai atribut yang lain. Pada penelitian ini, parameter yang digunakan bersifat numerik, sehingga harus diketahui terlebih dahulu jumlah kelas, rata-rata (mean) , simpangan baku (standar deviasi) dan nilai distribusi gaussian dari masing-masing parameter. Nilai jumlah dan peluang masing-masing kelas adalah

$$P(y_{test}^{(j)} | x_{test}) = P(x_{test} | y_{test}^{(j)}) \times P(y_{latih}^{(j)}) \tag{3}$$

Dimana :

P(y | x) = Peluang suatu kejadian ke-j

P(x | y) = peluang semua kelas jika kejadian ke-j

P(ylatih) = Peluang kejadian yang dilatih ke-j

P(ytest) = Peluang kejadian yang diuji ke-j

Random Forest

Random Forest merupakan algoritma ML dimana sistem bekerja mirip seperti manusia mengambil keputusan. Dalam Random Forest, terdapat beberapa algoritma untuk membentuk sebuah pohon keputusan, salah satunya Algoritma C4.5. Sebelum menentukan parameter sebagai akar, hal yang harus diketahui adalah nilai Entropy. Entropy adalah nilai informasi yang menyatakan ukuran ketidakpastian dari atribut dari suatu kumpulan objek data. Rumus untuk menentukan Entropy adalah sebagai berikut:

$$E(y_{latih}^{(j)}) = \sum_{j=1}^n -P(y_{latih}^{(j)}) * \log_2(P(y_{latih}^{(j)})) \quad (4)$$

Kemudian tentukan nilai Gain. Nilai Gain adalah ukuran efektifitas suatu atribut dalam mengklasifikasina data. Nilai Gain yang paling besar akan dipilih menjadi akar. Rumus menentukan nilai Gain adalah sebagai berikut:

$$G(x_{latih}^{(i)}) = E(y_{latih}^{(j)}) - \sum_{i=1}^n \frac{|P(x_{latih}^{(i)})|}{|S|} * E(x_{latih}^{(i)}) \quad (5)$$

Dimana :

S = Total Data

E(x) = Nilai Entropi pada parameter ke-i

P(x) = Nilai peluang pada parameter ke-i

G(x) = Nilai Gain pada Parameter ke-i

Parameter dengan nilai Gain paling besar akan dipilih sebagai akar. Ulangi proses tersebut sampai menghasilkan Node yang tidak dibagi node lain atau hasil akhir (Terminal Node). Apabila terdapat 2 jenis data pada parameter, dan salah satu $P(x_{latih}) = 1$, maka dapat dipastikan kelas bernilai 1 dan satunya lagi bernilai 0. Lakukan iterasi sampai parameter hanya mengeluarkan kelas 1 atau 0.

Support Vector Machine

Support Vector Machine merupakan algoritma ML untuk melakukan sebuah prediksi dengan prinsip dasar kasus klasifikasi yang secara linear dan dapat dipisahkan. Pada ruang berdimensi tinggi, akan dicari hyperplane yang dapat memaksimalkan jarak antar kelas data [9]. Jika terdapat beberapa parameter ($x_1, x_2, x_3, \dots, x_n$) maka akan terdapat beberapa fitur ($w_1, w_2, w_3, \dots, w_n$) sehingga untuk meminimalkan nilai persamaan hyperplane klasifikasi linear SVM adalah :

$$y_{latih}^{(i)} (b + \sum_{i=1}^n w^{(i)} x_{latih}^{(i)}) \geq 1 \quad (6)$$

Setelah memasukan parameter kepada persamaan diatas, maka terdapat beberapa persamaan dengan variabel w dan b. Selanjutnya akan dicari nilai w dan b dengan sistem eliminasi. Nilai w yang telah didapatkan kemudian substitusikan terhadap fungsi keputusan yaitu:

$$f(x) = w^{(i)} x + b \quad (7)$$

Setelah memasukan nilai w dan b didapat garis yang membagi antar dua kelas. Selanjutnya memasukan uji untuk mengetahui data tersebut termasuk kedalam kelas -1 atau 1. Jika hasil bernilai positif, maka data tersebut termasuk kedalam kelas 1. Sedangkan jika hasil bernilai negative, maka data tersebut termasuk kedalam kelas -1.

Artificial Neural Network

Artificial Neural Network merupakan sebuah algoritma ML untuk melakukan klasifikasi dan prediksi dengan prinsip disusun dengan struktur dan fungsi otak manusia sebagai model yang ditiru [10]. Pertama lakukan normalisasi data yang bertujuan untuk mempermudah proses pada jaringan dengan diharapkan hasil keluaran

pada pelatihan sesuai dengan fungsi aktivitas yang digunakan dengan cara merubah skala fitur kedalam 0 s.d 1. Persamaan Normalisasi data adalah

$$x_{latih_baru}^{(i)} = \frac{x_{latih_lama}^{(i)} - x_{latih_min}^{(i)}}{x_{latih_max}^{(i)} - x_{latih_min}^{(i)}} \tag{8}$$

Jika terdapat banyak parameter ($x_1, x_2, x_3, \dots, x_n$), maka untuk melakukan inisiasi bobot tergantung banyaknya parameter sehingga bobot adalah ($w_1, w_2, w_3, \dots, w_n$). Karena nilai bobot tidak ditentukan, maka asumsikan terlebih dahulu bobot bernilai $0 < w_i < 1$. Kemudian menghitung Respon unit keluar bertujuan untuk mengeluarkan hasil antara bobot dengan parameter yang telah dinormalisasi dengan persamaan.

$$y_{latih_baru}^{(j)} = \sum_{i=1}^n (x_{latih_baru}^{(i)} w_{lama}^{(i)}) \tag{9}$$

Selanjutnya evaluasi parameter. Evaluasi ini bertujuan apakah keluaran dari data yang telah dilatih sesuai dengan target. Jika keluaran sama dengan target, maka bobot tidak berubah. Jika berubah atau tidak sesuai target, maka nilai bobot akan berubah dengan persamaan

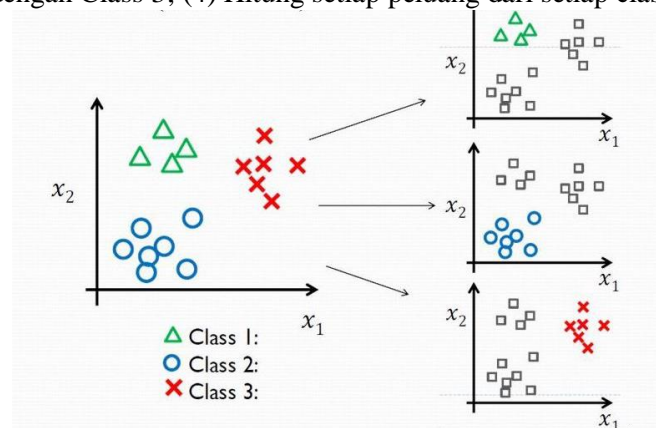
$$w_{(i)baru} = w_{(i)lama} + \alpha \cdot x_{(i)baru} \cdot y_i$$

$$b_{baru} = b_{lama} + \alpha \cdot y_i \tag{10}$$

Ketika bobot (w) sudah ditentukan, maka untuk melakukan pengujian data dilakukan dengan persamaan:

$$y_{test}^{(j)} = \sum_{i=1}^n w_{baru}^{(i)} x_{uji}^{(i)} \tag{11}$$

Algoritma ML pada dasarnya digunakan untuk menyelesaikan masalah dua buah class atau biner. Gambar 1 menampilkan ilustrasi dari Permasalahan tiga class. Apabila Permasalahan yang terjadi menghasilkan tiga buah kelas, maka langkah penyelesaian sebagai berikut : (1) Kelompokkan Class 2 dan 3 menjadi sebuah Class baru, kemudian pisahkan dengan Class 1; (2) Kelompokkan Class 3 dan 1 menjadi sebuah Class baru, kemudian pisahkan dengan Class 2; (3) Kelompokkan Class 1 dan 2 menjadi sebuah Class baru, kemudian pisahkan dengan Class 3; (4) Hitung setiap peluang dari setiap class baru



Gambar 1. Ilustrasi Permasalahan Multiclass

C. Hasil dan Pembahasan

Menyusun Algoritma

Import Library, hal yang dilakukan pertama sebelum melakukan pengolahan data dengan python yaitu melakukan import beberapa library, seperti pandas, numpy dan sklearn.

Import Data Hasil Pertandingan sejak 1872-2022

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
0	1872-11-30	Scotland	England	0	0	Friendly	Glasgow	Scotland	False
1	1873-03-08	England	Scotland	4	2	Friendly	London	England	False
2	1874-03-07	Scotland	England	2	1	Friendly	Glasgow	Scotland	False
3	1875-03-06	England	Scotland	2	2	Friendly	London	England	False
4	1876-03-04	Scotland	England	3	0	Friendly	Glasgow	Scotland	False

Gambar 2. Data Hasil Pertandingan Sejak 1872 – 2022

Menambahkan Kolom Pemenang (“winning_team”) Kolom “winning_team” merupakan kolom yang berisi pemenang diantara kolom “home_team” dan “away_team” dimana dalam menentukan pemenang dilihat dari kolom “home_score” dan “away_score”.

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral	winning_team
0	1872-11-30	Scotland	England	0	0	Friendly	Glasgow	Scotland	False	Draw
1	1873-03-08	England	Scotland	4	2	Friendly	London	England	False	England
2	1874-03-07	Scotland	England	2	1	Friendly	Glasgow	Scotland	False	Scotland
3	1875-03-06	England	Scotland	2	2	Friendly	London	England	False	Draw
4	1876-03-04	Scotland	England	3	0	Friendly	Glasgow	Scotland	False	Scotland

Gambar 3. Penambahan Kolom winning_team

Pembagian Data Latih dan Data Uji, data latih (Train Data) merupakan kumpulan data yang dikumpulkan untuk membangun sebuah algoritma. Sedangkan data uji (Test Data) merupakan kumpulan data yang dikumpulkan untuk menguji suatu algoritma apakah algoritma yang dibangun bekerja dengan baik atau tidak. Data Latih dan Data Uji perlu dipisah karena algoritma dapat mengingat data yang digunakan sehingga prediksi akan selalu bernilai benar untuk data yang telah digunakan sebelumnya. Data yang akan diuji adalah hasil pertandingan FIFA World Cup edisi 2006-2018. Sehingga pembagian data hasil pertandingan sebelumnya dari FIFA World Cup 1930 s.d. FIFA World Cup 2002

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral	winning_team	match_year
24982	2002-06-22	Senegal	Turkey	0.0	1.0	FIFA World Cup	Osaka	Japan	True	Turkey	2002
24984	2002-06-25	Korea Republic	Germany	0.0	1.0	FIFA World Cup	Seoul	Korea Republic	False	Germany	2002
24986	2002-06-26	Brazil	Turkey	1.0	0.0	FIFA World Cup	Saitama	Japan	True	Brazil	2002
24987	2002-06-29	Korea Republic	Turkey	2.0	3.0	FIFA World Cup	Daegu	Korea Republic	False	Turkey	2002
24989	2002-06-30	Germany	Brazil	0.0	2.0	FIFA World Cup	Yokohama	Japan	True	Brazil	2002

Gambar 4. Pemilihan Data Latih

Proses Encoding, algoritma machine learning mengharuskan variable berupa numerik, sehingga perlu perubahan data dari string menjadi numeric. Proses tersebut disebut Encoding[4]. Pada kolom “winning_team”, 2 untuk “home_team”, 1 untuk draw dan 0 untuk “away_team”

	home_team	away_team	winning_team
639	Senegal	Turkey	0
640	Korea Republic	Germany	0
641	Brazil	Turkey	2
642	Korea Republic	Turkey	0
643	Germany	Brazil	0

Gambar 5. Proses Encoding

Memasukan Parameter Terhadap Data

	winning_team	att_home	mid_home	def_home	ovr_home	att_away	mid_away	def_away	ovr_away	home_team_Algeria ...	away_team_Spain	away_team_
0	0	87	87	83	86	77	75	71	74	0 ...	0	
1	2	90	85	84	86	81	78	73	77	0 ...	0	
2	0	88	85	87	87	81	84	82	83	0 ...	0	
3	0	74	75	70	73	70	71	72	71	0 ...	0	
4	2	91	82	81	85	90	85	84	86	0 ...	0	

Gambar 6. Data Terbaru

Melakukan Pelatihan dan Pengujian Data, pengujian data latih dilakukan menggunakan beberapa algoritma ML. Pada penelitian ini semua persamaan tersebut menggunakan library sklearn sehingga output yang keluar adalah peluang “home_team” menang, draw atau “away_team” menang. Beberapa library tersebut adalah Logistic Regression(), KNeighborsClassifier(), MultinomialNB(), LinearSVC(), MLPClassifier(), RandomForestClassifier().

Tabel 1. Hasil Pelatihan dan Pengujian Data

	Logistic Regression	KNN	Naïve-Bayes	SVM	Neural Network	Random Forest
FIFA World Cup 2006	64,06%	64,06%	67,18%	64,06%	67,18%	57,81%
FIFA World Cup 2010	46,87%	48,43%	45,31%	50%	48,43%	40,62%
FIFA World Cup 2014	51,56%	59,37%	56,25%	65,62%	59,37%	46,87%
FIFA World Cup 2018	48,43%	60,93%	57,81%	57,81%	60,93%	50%
Rata-Rata	52,73%	58,19%	56,63%	59,37%	58,97%	48,82%

Evaluasi, berdasarkan table 1 diatas, maka didapatkan beberapa evaluasi diantaranya : (1) Pada FIFA World Cup 2006, algoritma Naïve-Bayes dan Neural Network memiliki tingkat keakurasian tertinggi, yaitu 67,18%; (2) Pada FIFA World Cup 2010, algoritma SVM memiliki tingkat keakurasian tertinggi, yaitu 50%; (3) Pada FIFA World Cup 2014, algoritma SVM memiliki tingkat keakurasian tertinggi, yaitu 65,62%; (4) Pada FIFA World Cup 2018, algoritma KNN dan Neural Network memiliki tingkat keakurasian tertinggi, yaitu 60,93%; (5) Rata-rata algoritma SVM memiliki tingkat keakurasian tertinggi yaitu 59,37%; (6) Algoritma Neural Network memiliki tingkat keakurasian tertinggi pada 2006 dan 2018. Artinya apabila kedua tim yang sama bertemu pada 2006, jika bertemu kembali 12 tahun kemudian akan memiliki peluang kemenangan yang sama; (7) Algoritma SVM memiliki tingkat keakurasian tertinggi pada 2010 dan 2014. Artinya apabila kedua

tim yang sama bertemu pada 2010, jika bertemu kembali 4 tahun kemudian akan memiliki peluang kemenangan yang sama.

Uji Validitas Terhadap UEFA Euro dan Copa America, uji Validitas bertujuan apakah kesimpulan yang didapat saat menguji algoritma pada kasus sebelumnya akan sama dengan kasus yang lain. Dalam hal ini, Validitas dilakukan pada UEFA Euro dan Copa America.

Tabel 2. Uji Validitas Terhadap UEFA Euro

	Logistic Regression	KNN	Naïve-Bayes	SVM	Neural Network	Random Forest
UEFA Euro 2008	51,6%	54,8%	58%	61,2%	54,8%	77,4%
UEFA Euro 2012	40%	60%	53,3%	53,3%	53%	40%
UEFA Euro 2016	44%	46%	44%	42%	48%	46%
UEFA Euro 2021	47,9%	60%	42%	54%	54%	50%
Rata-Rata	45,8%	55,2%	49,3%	52,6%	52,45%	53,3%

Tabel 3. Uji Validitas Terhadap Copa America

	Logistic Regression	KNN	Naïve-Bayes	SVM	Neural Network	Random Forest
Copa America 2015	53,8%	53,8%	50%	53,8%	50%	53,8%
Copa America 2016	45,1%	54,8%	48,3%	54,8%	51,6%	41,9%
Copa America 2019	44%	52%	56%	48%	56%	44%
Copa America 2021	51,8%	48,1%	51,8%	48,1%	44,4%	55,5%
Rata-Rata	48,6%	52,1%	51,5%	51,1%	50,5%	48,8%

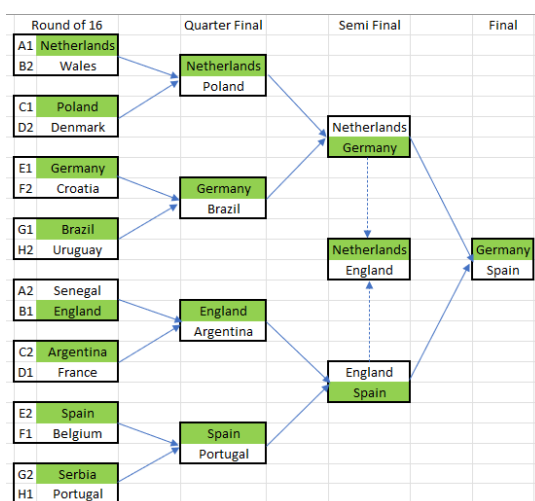
Berdasarkan uji validitas diatas, bahwa pada FIFA World Cup, UEFA Euro maupun Copa America memiliki akurasi tertinggi yang berbeda. Hal ini dikarenakan faktor kekuatan tim dari semua lini berbeda. Atas hal tersebut, apabila akan melakukan sebuah prediksi dengan suatu algoritma maka dilakukan evaluasi terhadap edisi sebelum dari kompetisi tersebut.

Prediksi FIFA World Cup Qatar 2022

Setelah melakukan pelatihan, pengujian, evaluasi dan Validitas algoritma terhadap 4 edisi FIFA World Cup dan UEFA Euro serta Copa America, selanjutnya melakukan prediksi FIFA World Cup Qatar 2022 menggunakan algoritma SVM dan Neural Network dengan python. Apabila kedua tim yang sama bertemu pada 2022, jika bertemu 4 tahun yang lalu atau diluar 12 tahun, menggunakan algoritma SVM. Jika bertemu 12 tahun yang lalu menggunakan alogritma Neural Network . langkah untuk melakukan prediksi sama seperti diatas

Tabel 4. Pemetaan Hasil Prediksi Pada Fase Grup

Grup A					Grup B					Grup C							
No	Tim	W	D	L	P	No	Tim	W	D	L	P	No	Tim	W	D	L	P
1	Netherlands	3	0	0	9	1	England	3	0	0	9	1	Poland	3	0	0	9
2	Senegal	2	0	1	6	2	Wales	2	0	1	6	2	Argentina	2	0	1	6
3	Qatar	1	0	2	3	3	USA	1	0	2	3	3	Mexico	1	0	2	3
4	Ecuador	0	0	3	0	4	Iran	0	0	3	0	4	Saudi Arab	0	0	3	0
Grup D					Grup E					Grup F							
No	Tim	W	D	L	P	No	Tim	W	D	L	P	No	Tim	W	D	L	P
1	France	3	0	0	9	1	Germany	3	0	0	9	1	Belgium	3	0	0	9
2	Denmark	2	0	1	6	2	Spain	2	0	1	6	2	Croatia	2	0	1	6
3	Tunisia	1	0	2	3	3	Costa Rica	1	0	2	3	3	Morocco	1	0	2	3
4	Australia	0	0	1	0	4	Japan	0	0	3	0	4	Canada	0	0	3	0
Grup G					Grup H												
No	Tim	W	D	L	P	No	Tim	W	D	L	P						
1	Brazil	3	0	0	9	1	Portugal	3	0	0	9						
2	Serbia	2	0	1	6	2	Uruguay	2	0	1	6						
3	Switzerland	1	0	2	3	3	S. Korea	1	0	2	3						
4	Cameroon	0	0	3	0	4	Ghana	0	0	3	0						



Gambar 7. Pemetaan Hasil Prediksi pada Knock-Out System

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:
 Pada FIFA World Cup 2006, algoritma Naïve-Bayes dan Neural Network memiliki tingkat keakurasian tertinggi, yaitu 67,18%

Pada FIFA World Cup 2010, algoritma SVM memiliki tingkat keakurasian tertinggi, yaitu 50%

Pada FIFA World Cup 2014, algoritma SVM memiliki tingkat keakurasian tertinggi, yaitu 65,62%

Pada FIFA World Cup 2018, algoritma KNN dan Neural Network memiliki tingkat keakurasian tertinggi, yaitu 60,93%

Rata-rata algoritma SVM memiliki tingkat keakurasian tertinggi yaitu 59,37%

Algoritma SVM memiliki tingkat keakurasian tertinggi pada 2010 dan 2014. Artinya apabila kedua tim yang sama bertemu pada 2010, jika bertemu kembali 4 tahun kemudian akan memiliki peluang kemenangan yang sama.

Algoritma Neural Network memiliki tingkat keakurasian tertinggi pada 2006 dan 2018. Artinya apabila kedua tim yang sama bertemu pada 2006, jika bertemu kembali 12 tahun kemudian akan memiliki peluang kemenangan yang sama

Algoritma SVM dan Neural Network digunakan untuk memprediksi hasil FIFA World Cup Qatar 2022 berdasarkan histori pertandingan kedua tim yang bertemu. Jika 12 tahun yang lalu kedua tim bertemu, algoritma yang digunakan adalah Neural Network. Sedangkan Jika 4 tahun yang lalu atau belum pernah sama sekali bertemu, algoritma yang digunakan adalah SVM.

Hasil prediksi FIFA World Cup Qatar 2022 yaitu Germany berpeluang memenangkan FIFA World Cup Qatar 2022.

Daftar Pustaka

- [1] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [2] J. Brownlee, "Master Machine Learning Algorithms: Discover how they work and implement them from scratch.," *Mach. Learn. Mastery*, pp. 1–163, 2016, [Online]. Available: <http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>
- [3] R. Primartha, *Algoritma Machine Learning*. Bandung: Informatika, 2021.
- [4] Fahrizal, F. O. Reynaldi, and N. Hikmah, "Implementasi Machine Learning pada Sistem PETS Identification Menggunakan Python Berbasis UBuntu," *J. Inf. Syst. Informatics Comput.*, vol. 4, no. 1, pp. 86–91, 2020.
- [5] A. U. Zailani and N. L. Hanun, "Penerapan Algoritma Klasifikasi Random Forest Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera," *Infotech J. Technol. Inf.*, vol. 6, no. 1, pp. 7–14, 2020, doi: 10.37365/jti.v6i1.61.
- [6] N. Jayanti, S. Puspitodjati, and T. Elida, "Teknik Klasifikasi Pohon Keputusan Untuk Memprediksi Kebangkrutan Bank Berdasarkan Rasio Keuangan," *Univ. Stuttgart*, no. Kommit, pp. 101–107, 2008.
- [7] Purnama, *Pengantar Machine Learning*. Bandung: Informatika, 2019.
- [8] P. A. Octaviani, Yuciana Wilandari, and D. Ispriyanti, "Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang," *J. Gaussian*, vol. 3, no. 8, pp. 811–820, 2014, [Online]. Available: [http://download.portalgaruda.org/article.php?article=286497&val=4706&title=PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE \(SVM\) PADA DATA AKREDITASI SEKOLAH DASAR \(SD\) DI KABUPATEN MAGELANG](http://download.portalgaruda.org/article.php?article=286497&val=4706&title=PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG)
- [9] T. Wahyono, *Fundamental of Python for Machine Learning*. Yogyakarta: Gava Media, 2021.
- [10] H. Mukhtar, R. Muhammad, T. Reny Medikawati, and Yoze Rizki, "Peramalan Kedatangan Wisatawan Mancanegara Ke Indonesia Menurut Kebangsaan Perbulannya Menggunakan Metode Multilayer Perceptron," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 2, no. 2, pp. 113–119, 2021, doi: 10.37859/coscitech.v2i2.3324.