

Seleksi Fitur Dan Preferensi Penyerang Terbaik Liga Inggris Berbasis Fisher's Discriminant Ratio, K-Means Clustering Dan Topsis

Alif Sefty W, Lukman Junaedi, Tresna Maulana Fahrudin

Universitas Narotama Surabaya, Jl Arief Rahman Hakim 51, Surabaya, 60117, Indonesia

Universitas Narotama Surabaya, Jl Arief Rahman Hakim 51, Surabaya, 60117, Indonesia

Universitas Narotama Surabaya, Jl Arief Rahman Hakim 51, Surabaya, 60117, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 16 Agustus 2019

Revisi Akhir: 10 September 2019

Diterbitkan Online: 30 September 2019

KATA KUNCI

K-Means

TOPSIS

English Premier League

Football Player

Fisher's Discriminant Ratio

KORESPONDENSI

No HP: 085745587260

E-mail: iluminezetoro@gmail.com

ABSTRACT

Football is a sport that is carried out by several people in a team, each player in a team has their respective roles in a match. One of the important positions in a team is as an attacker whose job is to score goals for the team. This research aims to determine the best attackers in the English Premier League from all teams competing in the 2018-2019 season based on data from the Fifa 19 game. There are several stages in this research, the first step is clustering with K-Means Clustering method, the second step is feature selection using Fisher Discriminant Ratio method (FDR) and then weighted and calculated using TOPSIS (The Technique for Order of Preference by Similarity to Ideal Solution) method. The calculation of the preference value for each player is based on the weight value of the features that have passed the selection stage. The player who has the highest preference value will be chosen as the best player.

1. PENDAHULUAN

Liga Inggris merupakan salah satu liga terbaik yang ada saat ini, pada Liga Inggris terdapat banyak pemain kelas dunia yang bermain di berbagai posisi pada tiap tim yang ada. Pemain – pemain tersebut berposisi sebagai penjaga gawang, pemain belakang, pemain tengah ataupun pemain depan. Dalam sebuah kompetisi tentunya setiap tim yang berpartisipasi memiliki pemain – pemain yang menjadi andalan dan tumpuan untuk memenangkan sebuah laga salah satunya adalah pemain depan. Pada pertandingan sepak bola pemain depan memiliki peran untuk mencetak gol ke gawang lawan, membuat ruang untuk penyerangan serta memberikan umpan untuk pemain lainnya[1]. Pada Liga Inggris setiap tim memiliki beberapa pemain depan dengan kemampuan individu yang mumpuni, umumnya para pemain tersebut memiliki kemampuan individu yang lebih baik dari pemain lainnya sehingga mempermudah mereka dalam mencetak gol.

Persaingan yang sangat kompetitif di setiap musim pada liga Inggris para penyerang harus memiliki *skill* yang baik guna membantu tim dalam pertandingan. Untuk mengetahui penyerang

terbaik yang ada di liga primer inggris kami akan melakukan klasterisasi pada semua penyerang yang ada berdasarkan setiap skill individu. Setiap satu pemain memiliki beberapa fitur yang kemudian kami proses dengan salah satu metode pada data mining untuk mendapatkan informasi penting yaitu dengan metode *K-Means Clustering*[2]. Setelah melakukan tahapan clustering akan dilakukan proses seleksi fitur dengan metode Fisher Discriminant Ratio (FDR) untuk tahapan terakhir penulis melakukan pemilihan penyerang terbaik dengan sistem penunjang keputusan yang menggunakan metode *Topsis*. *Topsis* (The Technique for Order of Preference by Similarity to Ideal Solution) merupakan salah satu metode untuk penyelesaian masalah dengan *Multiple Criteria Decision Making* (MCDM), metode ini kami pilih karena cocok dan baik dalam melakukan seleksi pada beberapa kriteria yang telah ditentukan.

2. TINJAUAN PUSTAKA

2.1 Liga Inggris

Liga Inggris menjadi salah satu liga yang sangat menarik perhatian penonton sepak bola. Pada setiap musimnya selalu

tercipta banyak gol di setiap pertandingannya, setiap klub mempunyai strategi tersendiri untuk mengoptimalkan performa dari setiap penyerang yang mereka miliki. Dalam beberapa tahun terakhir terdapat beberapa pemain yang menjadi kandidat penyerang terbaik pada liga Inggris diantaranya, Sadio Mane, Sergio Aguero, Harry Kane, dan Eden Hazard. Mereka layak disebut sebagai penyerang terbaik karena kontribusi gol mereka untuk klub masing-masing. Rata-rata penyerang terbaik yang dimiliki oleh setiap klub papan atas dapat mencetak 15 gol atau lebih setiap musimnya. Penyerang yang baik memiliki tingkat rasio gol disetiap pertandingan yang tinggi dan juga tingkat konversi peluang yang baik, selain itu penyerang yang baik juga dapat berperan besar membantu terciptanya gol bagi pemain yang lain.

2.2 Data Mining

Data Mining adalah tahapan yang menggunakan ilmu matematika, ilmu statistika serta kecerdasan buatan atau *machine learning* untuk melakukan ekstraksi data serta melakukan pengidentifikasian terhadap informasi yang terkandung dalam data yang diolah. Dalam data mining terdapat beberapa istilah yang biasa digunakan yaitu KDD (*Knowledge discovery in data base*), analisa pola data atau (*Data Analysis*), ekstraksi pengetahuan (*Knowledge Extraction*), Kecerdasan bisnis (*Business Intellegence*), *data archeology*, *data dregging*[3]. Data mining memiliki berbagai fungsi yang dapat meningkatkan dan membantu dalam pengolahan data guna mendapatkan informasi baru untuk pengguna

2.3 K-Means Clustering

K-Means merupakan salah satu algoritma atau metode dalam data mining yang sangat populer untuk digunakan pada dataset dengan jumlah yang besar[4]. Secara garis besar K-Means digunakan untuk membagi atau mempartisi data dalam sebuah cluster[5]. Pengelompokan dilakukan berdasarkan karakteristik data yang sama lalu untuk data yang memiliki karakteristik yang berbeda akan dikelompokkan pada tempat yang lain. Metode K-Means bertugas mengelompokkan data (N) pada klaster (K) dengan nilai rata-rata terdekat[6], selain itu *K-Means* juga melakukan pemodelan tanpa supervisi (*unsupervised*). Sehingga tujuan utama metode ini adalah memperkecil variasi data yang ada pada suatu setiap cluster dan memaksimalkan variasi pada cluster yang lain.

2.4 TOPSIS

Topsis (*Technique for Order of Preference by Similarity to Ideal Solution*) adalah salah satu metode dalam pengambilan keputusan yang diperkenalkan oleh Yoon dan Hwang pada tahun 1981. Metode pengambilan keputusan berfungsi sebagai alat bantu bagi pengambil keputusan untuk memperluas kemampuan pengambil keputusan tanpa menggantikan penilaian dari individu yang mengambil keputusan[7]. Penggunaan metode ini banyak diterapkan pada pengambilan keputusan yang memiliki kriteria banyak atau multikriteria. Metode ini memiliki konsep dimana penentuan alternatif terbaik dilakukan dengan perhitungan dari jarak terpendek solusi ideal positif dan jarak terjauh untuk solusi ideal negatif, pada metode TOPSIS pembobotan pada setiap indikator merupakan tahapan yang krusial[8], banyaknya faktor yang menjadi pertimbangan membuat proses pengambilan keputusan menjadi agak sulit.

Jenis permasalahan tersebut biasanya dinamakan MCDM (*Multiple Criteria Decision Making*), Topsis banyak digunakan karena metode ini sangat sederhana dan mudah untuk dilakukan, dengan konsep yang sederhana serta proses dan langkah – langkah yang mudah dipahami oleh pengguna.

2.5 FDR

FDR (*Fisher's discriminant ratio*) adalah salah satu metode analisa statistik untuk menentukan variabel yang terkait dan variabel tidak terkait berupa variabel kontinu atau binary[9]. Metode ini pertama kali dicetuskan oleh Sir Ronald Fisher pada tahun 1936. FDR pada umumnya digunakan untuk menentukan besarnya perbedaan pada setiap fitur pada dua kelas berbeda. Hasil dari proses FDR adalah untuk fitur yang mempunyai perbedaan yang besar pada rata – rata dari kelas dan varian kecil dari tiap kelas maka nilai FDR yang didapatkan akan tinggi, jika diantara fitur terdapat perbedaan absolut rata – rata yang sama maka fitur dengan varian terkecil akan mendapatkan nilai FDR yang lebih tinggi[10]. Pada penelitian ini metode FDR digunakan untuk menyeleksi fitur yang paling ideal untuk digunakan pada proses selanjutnya.

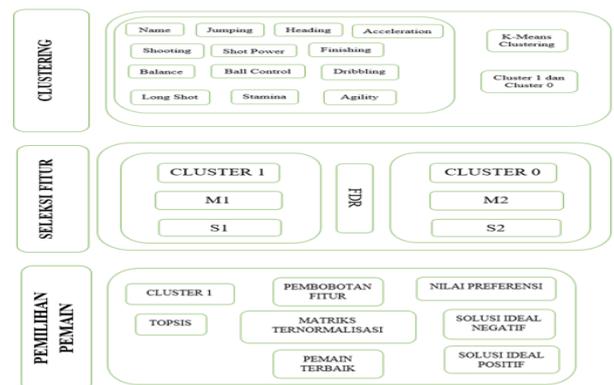
3. METODOLOGI

3.1 Pengumpulan Data

Data yang digunakan berasal dari sebuah game terbitan EA Sports yaitu FIFA 2019. Dalam penelitian ini data yang digunakan adalah data pemain depan dan pemain tengah dari setiap tim yang ada pada Liga Primer Inggris. Untuk jumlah data yang digunakan sebanyak 275 *record* dengan 13 atribut pada setiap pemain

3.2 Analisis Data

Pada bagian ini akan dijelaskan tahapan kerja yang akan dilakukan oleh penulis pada penelitian ini meliputi *clustering*, seleksi fitur dan pemilihan pemain terbaik.

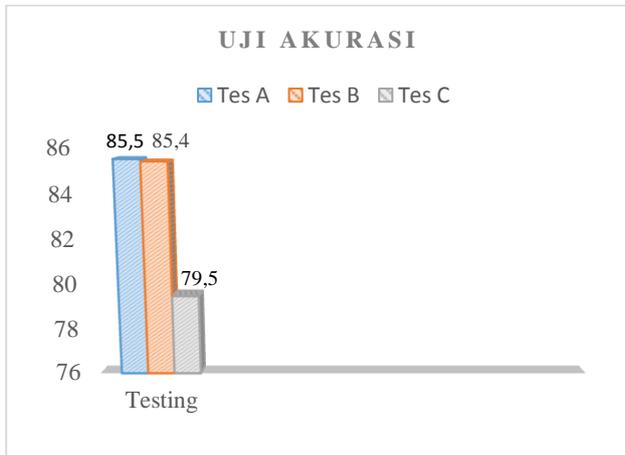


Gambar 1 Desain Sistem

1. Clustering
Berdasarkan atribut yang dimiliki oleh setiap pemain, semua data yang ada akan dilakukan pengelompokan dengan metode *K-Means Clustering*. Semua data tersebut akan dibagi menjadi 2 cluster yaitu cluster 1 dan 0
2. Seleksi Fitur

20% data test 80% data training dan yang terakhir Tes C 30% data tes dan 70% data training.

Dari hasil tiga kali proses pengujian dengan metode split didapati bahwa dengan 10% data testing dan 90 % data training memiliki tingkat akurasi sebesar 85.5%, lalu dengan 20% data test dan 80 % data training memiliki tingkat akurasi 85,4 % dan terakhir 30 % data testing dan 70 % data training memiliki 79.5 % tingkat akurasi.



Grafik 1 Hasil Uji Akurasi

4.2. Seleksi Fitur

Berdasarkan tingkat akurasi pada proses sebelumnya pada proses ini akan melakukan pengujian terhadap setiap fitur – fitur yang ada, untuk kemudian dipilih fitur – fitur paling berpengaruh yang dimiliki oleh setiap pemain pada setiap klaster.

Untuk tahapan pengujian nilai fitur dilakukan dengan empat tahapan antara lain, pertama memisahkan data menjadi dua kelompok, kedua perhitungan nilai rata – rata dari setiap fitur, ketiga perhitungan nilai variance total dari setiap fitur, yang terakhir yaitu perhitungan nilai FDR.

Tabel 3 Mean Cluster 1

Nama Pemain	Acceleration	Stamina
Eden Hazard	93	83
Sergio Aguero	89	77
Harry Kane	68	90
Paul Pogba	71	91
Mohamed Salah	95	85
Alexis Sanchez	87	84
..
Chidiebere Nwakali	77	76
Total	12649	12092
M1	78.57	75.11

Tabel 4 Variance Cluster 1

Nama Pemain	Acceleration	Stamina
Eden Hazard	93	83
Sergio Aguero	89	77
Harry Kane	68	90
Paul Pogba	71	91
Mohamed Salah	95	85
Alexis Sanchez	87	84

..
Chidiebere Nwakali	77	76
Total Variance	12197.565	10859.205
S1	75.761	67.448

Tabel 5 Mean Cluster 0

Nama Pemain	Acceleration	Stamina
Ilkay Gundogan	67	69
Francesc Fabregas	56	60
Olivier Giroud	43	62
Daniel Sturridge	72	56
Granit Xhaka	47	77
Marouane Fellaini	41	75
..
Andrew Eleftheriou	65	54
Total	7305	6937
M2	64.08	60.85

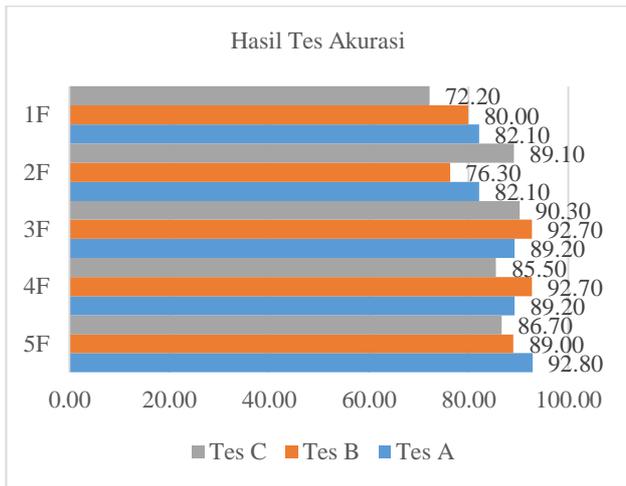
Tabel 6. Variance Cluster 0

Nama Pemain	Acceleration	Stamina
Ilkay Gundogan	67	69
Francesc Fabregas	56	60
Olivier Giroud	43	62
Daniel Sturridge	72	56
Granit Xhaka	47	77
Marouane Fellaini	41	75
..
Andrew Eleftheriou	65	54
Total Variance	12290.29	9866.46
S2	107.81	86.55

Tabel 7 Nilai FDR

No	Fitur	Nilai
1	Stamina	1,3915
2	Acceleration	1,1432
3	Dribbling	1,0862
4	Agility	0,8696
5	Sprint Speed	0,8687
6	Ball Control	0,5449
7	Balance	0,4355
8	Long Shots	0,3284
9	Shooting	0,2710
10	Shot Power	0,2181
11	Finishing	0,1989
12	Jumping	0,0337
13	Heading	0,0067

Setelah mendapatkan nilai FDR dari seluruh fitur yang ada kemudian akan dilakukan pengujian tingkat akurasi dengan menghapus satu persatu fitur dari urutan yang terendah hingga didapatkan hasil akurasi yang baik dengan fitur yang paling berpengaruh. Pengujian dilakukan menggunakan metode yang sama seperti sebelumnya yaitu split dengan 3 skema yaitu Tes A 10% data test 90% data training, Tes B 20% data test 80% data training dan yang terakhir Tes C 30% data tes dan 70% data training.



Grafik 2 Nilai FDR

Dari hasil pengujian mengambil lima fitur dengan nilai terbesar yaitu *stamina, acceleration, dribbling, agility, sprint speed*. Hasil dari seleksi fitur ini nantinya akan menjadi fitur utama dalam proses TOPSIS untuk menentukan pemain terbaik yang ada pada cluster 1

4.3. TOPSIS

Pada bagian ini melakukan beberapa tahapan perhitungan untuk menentukan siapa pemain terbaik yang ada di Liga Inggris dari anggota klaster 1. Untuk tahapannya dibagi menjadi beberapa tahapan diantaranya menentukan ranking dari setiap nilai fitur yang ada pada setiap anggota kelompok, menghitung matriks ternormalisasi, menghitung matriks normalisasi yang sudah terbobot, menentukan solusi ideal positif dan solusi ideal negatif, menghitung jarak antara solusi ideal positif dan solusi ideal negatif, menghitung nilai preferensi setiap pemain berdasarkan fitur – fitur yang ada.

1. Tahap Perangkingan Nilai Fitur

Tabel 5 Kriteria Fitur

No	Kriteria	Nilai
1	Tinggi	1
2	Sedang	0.5
3	Kecil	0

Tabel 6 Ranking Fitur

No	Nama Pemain	F1	F2	F3	F4	F5
1	Eden Hazard	0.5	1	1	1	1
2	Sergio Aguero	0.5	1	1	1	0.5
3	Harry Kane	1	0	0.5	0.5	0.5
4	Paul Pogba	1	0.5	1	0.5	0.5
5	Mohamed Salah	1	1	1	1	1
6	Alexis Sanchez	1	0.5	1	1	0.5
...
161	Ryan Giles	0	0.5	0	0.5	1
Total		83.5	80.5	82	85	80

Untuk F1 = Stamina, F2 = Acceleration, F3 = Dribbling, F4 = Agility, F5 = Sprint Speed

2. Tahap Perhitungan Matriks Ternormalisasi

Tabel 7 Matriks Ternormalisasi

NO	Nama Pemain	F1	F2	F3	F4	F5
1	Eden Hazard	0.006	0.012	0.012	0.012	0.013
2	Sergio Aguero	0.006	0.012	0.012	0.012	0.006
3	Harry Kane	0.012	0.000	0.006	0.006	0.006
4	Paul Pogba	0.012	0.006	0.012	0.006	0.006
5	Mohamed Salah	0.012	0.012	0.012	0.012	0.013
6	Alexis Sanchez	0.012	0.006	0.012	0.012	0.006
...
161	Ryan Giles	0	0.006	0.000	0.006	0.013

3. Tahap Perhitungan Matriks Ternormalisasi Terbobot

Tabel 8 Ranking Fitur

No	Fitur	Nilai
1	Stamina	1
2	Acceleration	0.75
3	Dribbling	0.5
4	Agility	0.25
5	Sprint Speed	0.25

Tabel 9 Matriks Ternormalisasi Terbobot

NO	Nama Pemain	F1	F2	F3	F4	F5
1	Eden Hazard	0.5	0.75	0.5	0.25	0.25
2	Sergio Aguero	0.5	0.75	0.5	0.25	0.125
3	Harry Kane	1	0	0.25	0.125	0.125
4	Paul Pogba	1	0.375	0.5	0.125	0.125
5	Mohamed Salah	1	0.75	0.5	0.25	0.25
...
161	Ryan Giles	0	0.375	0.000	0.125	0.25

4. Tahap Penentuan Solusi Ideal Positif dan Negatif

Tabel 10 Nilai Min dan Max Fitur

	F1	F2	F3	F4	F5
Min	0	0	0	0	0
Max	1	0.75	0.5	0.25	0.25

Tabel 11 Hasil Perhitungan Solusi Ideal Positif dan Negatif

NO	Nama Pemain	F1	F2	F3	F4	F5
1	Eden Hazard	0.500	0.750	0.500	0.250	0.250
2	Sergio Aguero	0.500	0.750	0.500	0.250	0.125
3	Harry Kane	1.000	0.000	0.250	0.125	0.125
4	Paul Pogba	1.000	0.375	0.500	0.125	0.125
5	Mohamed Salah	1.000	0.750	0.500	0.250	0.250
...
161	Ryan Giles	0.000	0.375	0.000	0.125	0.250

5. Tahap Perhitungan Jarak Solusi Ideal Positif dan Negatif

Tabel 12 Hasil Perhitungan Jarak

NO	Nama Pemain	D+	D-
1	Eden Hazard	0.500	1.090
2	Sergio Aguerro	0.515	1.068
3	Harry Kane	0.810	1.046
4	Paul Pogba	0.415	1.192
5	Mohamed Salah	0.000	1.392
...
161	Ryan Giles	1.186	0.468

6. Tahap Perhitungan Nilai Preferensi Setiap Pemain

Tabel 13 Hasil Perhitungan Preferensi

NO	Nama Pemain	V
1	Mohamed Salah	1.000
2	Sadio Mane	1.000
3	Gabriel Fernando de Jesus	0.917
4	Willian Borges da Silva	0.917
5	Heung Min Son	0.885
6	Jamie Vardy	0.824
7	Alexis Sanchez	0.754
8	Paul Pogba	0.742
9	Roberto Firmino	0.742
...
161	Ryan Giles	0.246

Dari hasil perhitungan preferensi setiap pemain kemudian akan diurutkan berdasarkan nilai terbesar hingga terkecil, pemain dengan nilai preferensi terbesar diantara pemain lainnya kemudian akan dipilih sebagai pemain terbaik yang ada di Liga Inggris musim 2018 – 2019. Dari data yang telah dilakukan proses perhitungan dan seleksi maka dapat ditentukan bahwa pemain terbaik Liga Inggris 2018 – 2019 adalah pemain yang memiliki nilai preferensi 1. Dari 161 pemain terdapat 2 pemain yang memiliki nilai preferensi paling tinggi diantara lainnya.

5. KESIMPULAN DAN SARAN

Dari hasil percobaan yang dilakukan dihasilkan dua orang pemain dengan perhitungan nilai preferensi tertinggi dari 275 pemain lainnya yaitu Mohamed Salah dan Sadio Mane, kedua pemain tersebut berasal dari klub Liverpool. Dari seluruh fitur yang diuji yaitu Stamina, Acceleration, Dribling, Agility dan Sprint Speed kedua pemain ini merupakan dua pemain terbaik pada Liga Inggris pada game Fifa 19 ini. Keduanya memiliki hasil perhitungan nilai preferensi yang sama yaitu 1 yang didapatkan dari hasil uji dan perhitungan jarak solusi ideal positif dan solusi ideal negatif.

DAFTAR PUSTAKA

[1] R. A. Siregar, "Seleksi Penyerang Utama Menggunakan K-Means Clustering Dan Sistem Pendukung Keputusan Metode Topsis," vol. 2, no. 1, pp. 35–46, 2017.

[2] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana, and F. Khusnu, "Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means," *Clust. K-Means*, pp. 1–7, 2012.

[3] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data*. 2014.

[4] M. Capó, A. Pérez, and J. A. Lozano, "An efficient

approximation to the K-means clustering for massive data," *Knowledge-Based Syst.*, vol. 117, pp. 56–69, 2017.

[5] S. K. Majhi and S. Biswal, "Optimal cluster analysis using hybrid K-Means and Ant Lion Optimizer," *Karbala Int. J. Mod. Sci.*, vol. 4, no. 4, pp. 347–360, 2018.

[6] Z. Kakushadze and W. Yu, "K-means and cluster models for cancer signatures," *Biomol. Detect. Quantif.*, vol. 13, no. July, pp. 7–31, 2017.

[7] E. Turban, J. E. Aronson, and T.-P. Liang, "Decision Support Systems and Intelligent Systems (7th Ed)," *Fenxi Huaxue*, vol. 32, no. 10, 2005.

[8] Z. Li et al., "An improved approach for water quality evaluation: TOPSIS-based informative weighting and ranking (TIWR) approach," *Ecol. Indic.*, vol. 89, no. February, pp. 356–364, 2018.

[9] T. M. Fahrudin, I. Syarif, and A. R. Barakbah, "Data Mining Approach for Breast Cancer Patient Recovery," *Emit. Int. J. Eng. Technol.*, vol. 5, no. 1, pp. 36–71, 2017.

[10] E. Prasetyo, "Analisis Fitur Tekstur Daun Mangga Dengan Fisher ' s Discriminant Ratio Untuk Pencapaian Fitur Yang Informatif Mangga merupakan salah 2015). Ini menunjukkan," no. January, 2015.

BIODATA PENULIS

Alif Sefty Wiyantoro



Mahasiswa Jurusan Teknik Informatika Fakultas Ilmu Komputer dari Universitas Narotama Surabaya.

Lukman Junaedi



Dosen Prodi Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Narotama, Surabaya, Indonesia. Memperoleh gelar M.Kom. dari Institut Teknologi Sepuluh Nopember (ITS), Surabaya Indonesia pada 2012.

Tresna Maulana Fahrudin



Menyelesaikan jenjang D4 dan S2 di Politeknik Elektronika Negeri Surabaya di Program Studi Teknik Informatika. Saat ini berkarir sebagai Dosen Tetap Fakultas Ilmu Komputer, Universitas Narotama Surabaya.