

KLASIFIKASI TIPE EMOSI AROUSAL PADA SINYAL EEG DENGAN METODE SUPPORT VECTOR MACHINE

AROUSAL TYPE OF EMOTION CLASSIFICATION IN EEG SIGNAL WITH SUPPORT VECTOR MACHINE METHOD

Nur Arviah Sofyan¹, Rita Purnamasari², Sugondo Hadiyoso³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom, Bandung
¹nurarviahs@gmail.com, ²ritapurnamasari@telkomuniversity.ac.id, ³sugondo@telkomuniversity.ac.id

Abstrak

Perkembangan teknologi Brain Computer Interface saat ini telah menyebar dalam kasus mengklasifikasikan emosi berdasarkan sinyal otak (EEG) pada manusia, yang dalam penelitian ini menggunakan data sekunder dari DEAP. Salah satu parameter emosi yang difokuskan di sini adalah tipe emosi *arousal* dengan jangkauan dari rendah (*uninterested*) ke tinggi (*excited*). Penelitian ini menerapkan *Principal Component Analysis* sebagai ekstraksi fitur. Tidak hanya itu, ekstraksi fitur juga dilakukan secara statistik. Sedangkan untuk klasifikasi fitur menggunakan *Support Vector Machine* dengan akurasi maksimum yang hanya mampu mencapai 60% yang menandakan sistem masih membutuhkan perbaikan untuk penelitian selanjutnya.

Kata Kunci: EEG, PCA, SVM, deteksi emosi

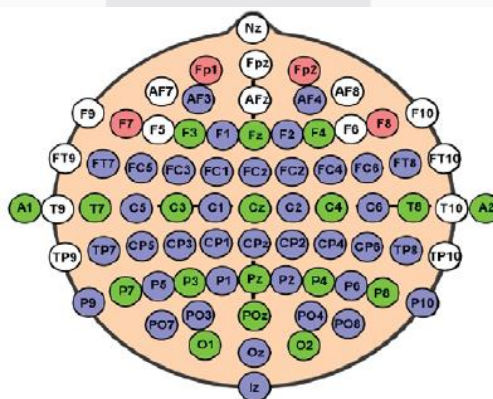
Abstract

The development of Brain Computer Interface technology nowadays has spread out in a case of classifying emotions based on brain signal (EEG) in human, which in this work using a set of secondary data from DEAP. One of the emotion parameters being focused on here is arousal with the range from low (*uninterested*) to high (*excited*). This study is applying *Principal Component Analysis* as the feature extraction. Not only that, feature extraction also being done statistically. As for feature classification is using *Support Vector Machine* with the maximum accuracy that only able to reach 60% which still needs improvements in the system for future works.

Keyword : EEG, PCA, SVM, *emotion detection*

1. Pendahuluan

Emosi manusia berfluktuasi tergantung pada situasi dan / atau kondisi yang saat ini terjadi pada manusia. Dengan kata lain, keadaan fisiologi manusia memiliki dampak besar pada fluktuasi itu sendiri. Dalam penelitian ini, 32 peserta diselidiki berdasarkan DEAP: *A Dataset for Emotion Analysis using EEG, Physiological and Video Signals*. Berdasarkan dataset, untuk pemilihan saluran EEG dari 40 saluran hanya 32 yang dipilih untuk digunakan. Posisi saluran ditunjukkan pada Gambar 1 dimana Coan et al. menunjukkan bahwa aktivitas otak frontal kiri berhubungan dengan emosi positif secara relatif, sedangkan aktivitas otak frontal kanan berhubungan dengan emosi negatif.



Gambar 1 Pemodelan posisi elektroda di kepala manusia [1].

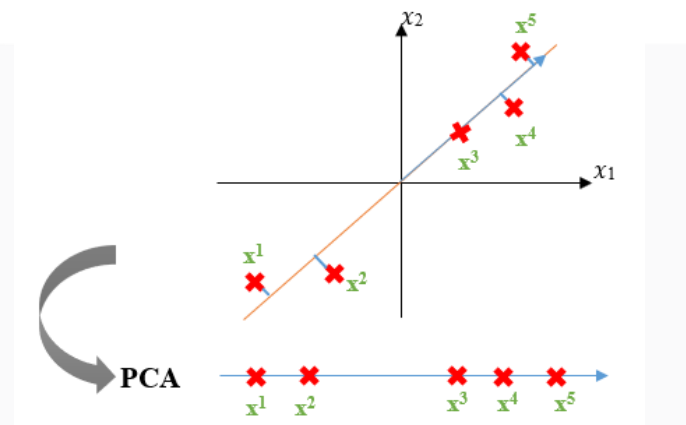
Dataset yang digunakan merupakan data sekunder dari DEAP yang telah dilakukan *down-sampling* menjadi 128 Hz dari *sampling rate* awal yaitu 512 Hz. Tidak hanya itu, dataset ini juga telah diproses sebelumnya dimana artefak dari EOG telah dihilangkan dan dilakukan filter jenis *band pass* 4.0 - 45.0 Hz [2]. Dengan demikian, dalam penelitian ini, data yang sudah diproses dan digunakan berisi 32 file peserta dengan 40 percobaan dan 32 kanal masing-masing peserta.

Untuk metode yang digunakan, *Principal Component Analysis* (PCA) dipilih sebagai ekstraksi fitur dengan mengambil nilai *eigen diagonal* dan juga ciri statistik seperti *entropy* dan *kurtosis* dalam prosesnya. Selain itu juga digunakan metode klasifikasi fitur *Support Vector Machine* (SVM) untuk mengetahui efektivitasnya dalam klasifikasi emosi berdasarkan sinyal EEG.

2. Ekstraksi Fitur

Principal Component Analysis (PCA) merupakan suatu metode ekstraksi ciri yang mengungkap teknik linier dengan tujuan merepresentasikan data dengan vektor yang berdimensi tinggi menjadi data dengan vektor yang berdimensi lebih rendah [3]. Dengan kata lain, metode PCA merupakan suatu penyederhanaan data dengan melakukan transformasi secara linier yang akan membentuk sistem koordinat baru dengan varians maksimum tanpa menghilangkan karakteristik dari data tersebut [4]. Konsep dari sistem PCA yaitu observasi data yang mungkin memiliki korelasi satu sama lain dan dikonversikan menjadi suatu data yang pasti berelasi antara satu dengan yang lainnya [5].

Algoritma PCA mulanya melakukan kalkulasi fitur matriks dari data asli yang kemudian akan menghasilkan nilai *mean* dari tiap matriks lalu dikurangi dari tiap nilai pada dimensi tertentu. Selanjutnya PCA mengkalkulasi matriks kovarians yang nantinya akan menghasilkan *eigen values* dan *eigen vectors*. *Eigenvector* dengan *eigen values* tertinggi merupakan *principal component* dari data tersebut [6]. Hasil akhir yang didapatkan terilustrasikan pada Gambar 2, PCA akan menemukan koordinat baru untuk suatu data dimana ruang sub dimensinya lebih rendah dan jumlah kuadratnya terminimalisasi [7].



Gambar 2 Konsep dasar PCA dalam mentransformasi data ke sumbu *principal component* secara linier.

Selanjutnya, fitur statistik ditambahkan sebagai pelengkap *output* fitur dari PCA disebabkan PCA tidak bisa berdiri sendiri. Fitur statistik ini diekstraksi untuk mengetahui karakteristik aktivitas EEG yang berbeda dari semua subset. Pada penelitian ini dipilih dua fitur statistik yaitu *entropy* dan *kurtosis* sebagai pelengkap fitur PCA sebelumnya. Penerapan fitur statistik ini dilakukan secara keseluruhan pada data yang digunakan.

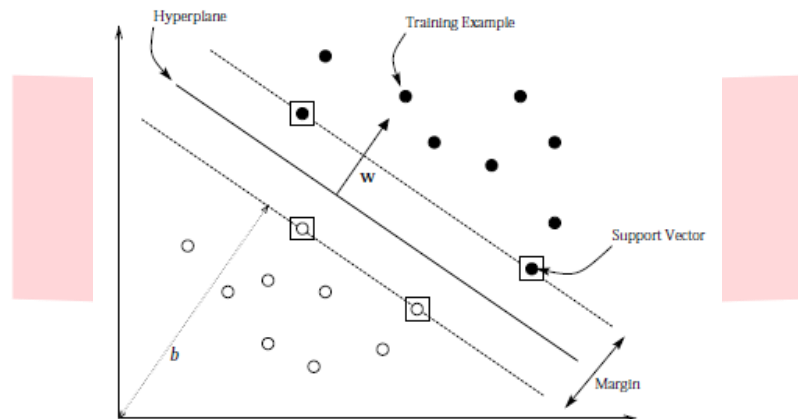
Tabel 1 Persamaan matematis *entropy* dan *kurtosis*.

Jenis Fitur Statistik	Persamaan
Entropy	$-\sum_{i=0}^{N-1} p_i \log_2 p_i$
Kurtosis	$\frac{\sum_{i=1}^N \frac{X_t - \bar{X}}{N}}{s^4}$

Berdasarkan Tabel 1, tertera variabel p_i pada persamaan *entropy* untuk mendefinisikan probabilitas karakter kemunculan nilai i dalam sistem. Kemudian pada persamaan *kurtosis*, variabel \bar{X} mendefinisikan nilai rata-rata, N sebagai ukuran sampel, dan s sebagai standar deviasi sistem. Keluaran dari tahap ekstraksi fitur merupakan 3 ciri (*diagonal eigen value*, *entropy*, *kurtosis*) yang kemudian dilanjutkan ke tahap klasifikasi.

3. Support Vector Machine (SVM)

Support Vector Machine merupakan suatu algoritma berfungsi untuk memisahkan data latih ke dalam dua kelas pada input space yang mengungkap ide pokok menentukan *hyperplane* dengan jarak batas (*margin*) yang maksimal berdasarkan fungsi kernel yang digunakan. Batas maksimal *margin* tersebut dilihat dari jumlah jarak *hyperplane* ke titik terdekat dari tiap data di kedua kelas tersebut [8]. Pada metode algoritma ini dilakukan *plotting* data sebagai titik-titik pada bidang n-dimensi dimana n merupakan jumlah fitur yang dimiliki. Kemudian dilakukan klasifikasi dengan menemukan *hyperplane* yang memisahkan dua kategori data.



Gambar 3 Konsep dasar *hyperplane* pada SVM linear function .

Berdasarkan Gambar 3 [8], dimana w didefinisikan sebagai *hyperplane* vektor normal sedangkan b sebagai bias, dan *support vector*nya sendiri terletak pada bidang *margin*.

SVM memiliki teknik pengklasifikasian yang disebut fungsi kernel. Fungsi kernel sangat berguna pada data berupa *text* karena tergolong *non-linear*. Teknik fungsi kernel membantu mentransformasikan bidang masukan yang sebelumnya berdimensi rendah ke dimensi yang lebih tinggi. Menurut Hsu, beberapa fungsi Kernel yang umumnya digunakan tertera di Tabel 2 [9].

Tabel 2 Persamaan matematis fungsi kernel SVM.

Jenis Kernel	Persamaan
Linier	$x_i \cdot z$
Polynomial	$(\gamma x \cdot z + c)^{degree}$
Radial Basis Function	$\exp(-\gamma x - z ^2)$

Berdasarkan Tabel 2, variabel γ , c , dan *degree* merupakan parameter yang digunakan untuk mendefinisikan tiap kernel tertentu.

Pada tahap ini, dataset yang telah melewati tahap ekstraksi ciri dengan keluaran 3 fitur digabungkan dengan data *rating* partisipan yang telah terklasifikasi sebagai *low arousal (uninterested)* dan *high arousal (excited)* menjadikan data sebesar 1280x4 untuk dilakukan klasifikasi dengan *Support Vector Machine (SVM)*. Masukan untuk klasifikasi terbagi menjadi beberapa *file* Excel berisi data latih dan data uji yang telah dikelompokkan sesuai dengan skenario pengujian. Dalam prosesnya, metode pengklasifikasian dengan SVM dilengkapi dengan tipe validasi *cross fold validation* dengan *kfold* sebanyak 5 kali.

4. Pelatihan Sistem

Tahap ini merupakan suatu tahap dimana data utuh dikelompokkan menjadi data latih dan uji. Pengelompokkan ini kemudian dilakukan berdasarkan kelas *low* dan *high* lalu ada pembagian data sebesar 50% dari data keseluruhan. Secara spesifik, pelatihan dibagi menjadi beberapa skenario sebagai berikut:

- Skenario 1**, pemilihan fungsi kernel terbaik kemudian dibandingkan untuk pemilihan fungsi *trial* terbaik untuk data masukan proses klasifikasi.
- Skenario 2**, dengan menggunakan percobaan terbaik dan fungsi kernel terbaik dilakukan investigasi terhadap ciri yang paling dominan dalam sistem klasifikasi dengan mengkombinasikan 3 ciri sedemikian rupa dalam prosesnya.

5. Akuisisi Data

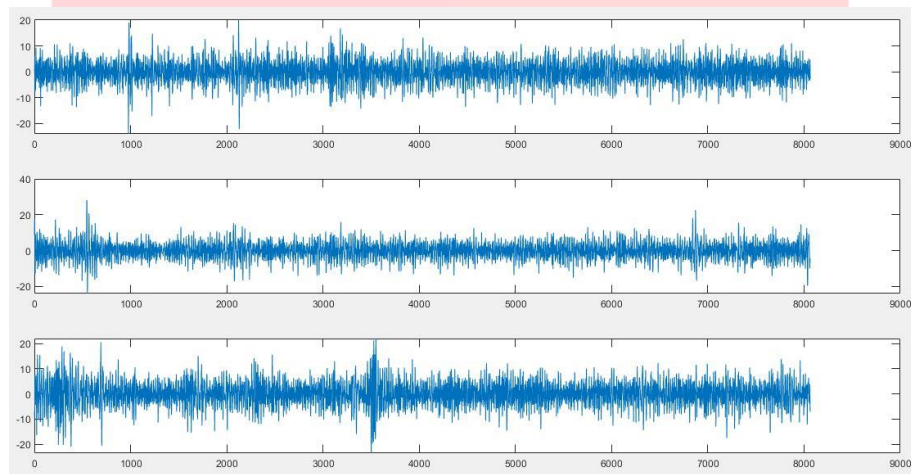
Data preprocessed yang diunduh dari DEAP berisi 32 *file* dengan format (.mat) yang mewakili data untuk 32 partisipan. Tiap *file* partisipan memiliki 2 array yaitu array data dan array label dengan konten masing-masing array dijelaskan pada Tabel 4.

Tabel 3 Konten array tiap *file* partisipan.

Nama Array	Ukuran Array	Konten Array
data	40 x 40 x 8064	<i>video/trial x channel x data</i>
label	40 x 4	<i>video/trial x label (valence, arousal, dominance, liking)</i>

Dari data preprocessed tersebut, dilakukan pembacaan data per 32 subjek, 40 trial, 32 kanal, dan 8064 ukuran data yang menghasilkan penguraian tiap file partisipan menjadi 1 subjek, 1 kanal, dan 1 trial atau disebut juga sebagai s1_t1_k1 sampai dengan s1_t40_k32 untuk partisipan pertama dan seterusnya.

Raw preprocessed sinyal EEG dari dataset DEAP bisa *diplotting* seperti pada Gambar 4 dengan pengambilan pada subjek 1 untuk 3 *trial* utama dengan pilihan kanal secara acak. Pada subjek 1, hasil *trial* 1 tergolong *Uninterested*, sedangkan hasil *trial* 2 dan 3 tergolong *Excited*.



Gambar 4 Plot sinyal EEG subjek 1 dengan 3 *trial* pada kanal 18 (AF4).

Pada tahap ekstraksi ciri, keluaran yang didapatkan terkonversi dalam bentuk file Excel (.xlsx). Ekstraksi ciri yang dilakukan yaitu terhadap 32 partisipan dengan 40 *trial* dan 32 kanal, menjadikan data keluaran sebesar 40960x3. Dari 40960 data tersebut, sebelum masuk ke tahap klasifikasi dilakukan pencarian nilai rata-rata dari 32 kanal tiap *trial* disatukan untuk menjadi 1 kanal tiap *trial*. Maka, hasil akhir keluaran data untuk dilanjutkan ke tahap klasifikasi yaitu sebesar 1280x4 dalam bentuk Excel (.xlsx). Untuk cuplikan hasil ekstraksi ciri oleh partisipan 1 dengan 3 *trial* utama dapat dilihat pada Tabel 5.

Tabel 4 Contoh hasil ekstraksi ciri dengan klasifikasi berdasarkan rating.

CR1	CR2	CR3	CLASS
16.1298	2.215207	3.987568	Uninterested
1218.44	1.596091	18.43995	Excited
222.63	1.788631	36.03375	Excited

6. Hasil Pengujian Skenario 1

Berdasarkan data awal 1280x4 memiliki 40 *trial* per partisipan. Pada Skenario 1 ini, dilakukan pemilihan *trial* terbaik dari tiap partisipan dengan melakukan penelitian pada seluruh *trial* dan juga mengaplikasikan tiap fungsi kernel dalam prosesnya.

Tabel 5 Hasil *tracing* data tiap *trial*.

TRIAL	LINEAR	GAUSSIAN	POLYNOME
1	59.4	59.4	43.8
2	56.3	43.8	46.9
3	75	59.4	65.6
4	53.1	59.4	34.4
5	75	65.6	50
6	40.6	59.4	46.9
7	62.5	65.5	68.8
8	62.5	65.6	62.5
9	53.1	62.5	62.5
10	53.1	56.3	43.8
11	34.4	40.6	43.8
12	56.3	59.4	62.5
13	40.6	59.4	59.4
14	53.1	65.6	56.3
15	59.4	59.4	50
16	56.3	56.3	56.3
17	62.5	68.8	56.3
18	68.8	68.8	78.1
19	78.1	78.1	78.1
20	71.9	71.9	59.4
21	69.7	71.9	71.9
22	50	65.6	65.6
23	65.6	68.8	62.5
24	46.9	59.4	40.6
25	65.6	78.1	78.1
26	62.5	65.6	65.6
27	65.6	71.9	62.5
28	53.1	62.5	53.1
29	53.1	65.6	46.9
30	43.8	59.4	37.5
31	46.9	40.6	59.4
32	59.4	50	68.8
33	53.1	43.8	50
34	71.9	75	62.5
35	87.5	87.5	68.8
36	56.3	71.9	68.8
37	53.1	62.5	50
38	50	59.4	43.8
39	46.9	68.8	65.6
40	71.9	71.9	46.9

Proses seleksi trial dari keseluruhan 40 trial data, didapatkan 13 trial dengan akurasi yang paling menonjol dan mampu membedakan antara dua kelas. Untuk inputan data dari proses seleksi awal ini yaitu data dikelompokkan berdasarkan nomor *trial*nya kemudian dilakukan klasifikasi dengan fungsi kernel masing-masing. Hasil seleksi dari 40 *trial* tersebut ditunjukkan pada Tabel 5. Kemudian dari 13 *trial* tersebut, seleksi dilakukan lagi dengan mengelompokkan data berdasarkan kelas *arousal* untuk mendapatkan *trial* terbaik berdasarkan akurasi terbesar.

Tabel 6 Hasil seleksi *trial* terbaik berdasarkan kelas *arousal*.

Trial	Linear	Polynome	Gaussian
3	56.25%	56.25%	56.25%
8	60%	60%	60%
13	56.25%	56.25%	56.25%
14	43.75%	43.75%	43.75%
21	47.05%	47.05%	47.05%
23	47.05%	47.05%	47.05%
25	50.00%	50.00%	50.00%
27	46.67%	46.67%	46.67%
34	56.25%	56.25%	56.25%
36	43.75%	43.75%	43.75%
39	25.00%	25.00%	25.00%

Kemudian pada Tabel 6, ditunjukkan hasil akurasi dari percobaan 13 *trial* dengan mengelompokkan data berdasarkan kelas arousal terlebih dahulu.. Dalam proses percobaan tiap fungsi kernel, akurasi yang didapatkan tiap *trial* bernilai sama diantara ketiga fungsi kernel tersebut. Untuk *trial* yang paling baik berdasarkan akurasi terbesar yaitu *trial* 8 dengan akurasi 60% pada 3 fungsi kernel baik Linear, Polynomial, maupun Gaussian.

7. Hasil Pengujian Skenario 2

Dari total data sebesar 32x4, kali ini data dikelompokkan sesuai kelas *high arousal (Excited)* dan *low arousal (Uninterested)*. Untuk total data dengan kelas *Excited* sebanyak 20 data dan untuk kelas *Uninterested* sebanyak 12 data, dengan pembagian data uji dan data latih pada Tabel 7.

Tabel 7 Pengelompokkan data uji dan latih berdasarkan tingkat *arousal*.

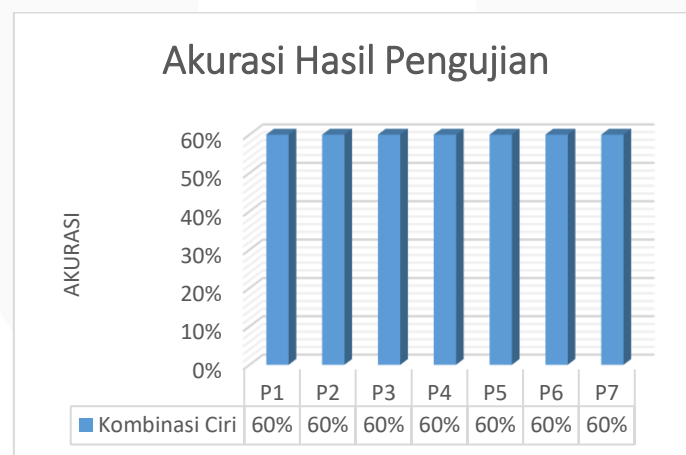
Data Latih		Data Uji	
<i>Excited</i>	<i>Uninterested</i>	<i>Excited</i>	<i>Uninterested</i>
10	7	10	5

Berdasarkan pengelompokkan data latih tersebut, dataset langsung diklasifikasi dengan menggunakan fungsi kernel yaitu Polynomial. Pada Skenario 2, data *input* untuk dijadikan data latih dan data uji dikelompokkan berdasarkan kombinasi ciri yang telah ditentukan pada Tabel 8.

Tabel 8 Variabel kombinasi ciri untuk pengujian sistem.

Variabel	Kombinasi Ciri
P1	PCA
P2	Entropy
P3	Kurtosis
P4	PCA, Entropy
P5	PCA, Kurtosis
P6	Entropy, Kurtosis
P7	PCA, Entropy, Kurtosis

Tahap ini bertujuan untuk mengetahui jenis ciri yang paling mendominasi proses klasifikasi dari keseluruhan ciri yang digunakan. Dari hasil *tracing* data dalam percobaan kombinasi ciri, didapatkan hasil seperti Gambar 5.



Gambar 5 Dominasi ciri berdasarkan hasil akurasi terbesar.

Berdasarkan Gambar 5, ketiga ciri yang telah dikombinasikan sedemikian rupa menghasilkan tingkat akurasi yang sama. Hal ini menunjukkan bahwa ketiga ciri yang digunakan mempengaruhi satu sama lain dalam proses ekstraksi ciri yang sebelumnya telah dilakukan.

8. Analisis

Pada sinyal EEG terdapat suatu aspek yang disebut variansi normal dalam bentuk beberapa gelombang psikologis yang dimiliki tiap orang. Bentuk gelombang psikologis tiap orang yang bervariasi menyebabkan

variansi normal dari sinyal EEG berbeda juga untuk tiap orangnya. Maka dari itu, dilakukan seleksi terbaik pada trial tiap orang karena tiap partisipan mendapatkan stimulus video yang sama dengan durasi yang sama, jadi proses klasifikasi dilakukan per trial dari tiap partisipan. Faktor lainnya juga bisa dijadikan alasan seperti tidak memungkinkannya untuk melakukan generalisir tingkat arousal antar satu dengan yang lainnya dalam satu waktu karena tingkat arousal tiap orang bervariasi meskipun diberikan stimulus video yang sama, respon yang dihasilkan akan berbeda tiap orangnya.

Berdasarkan hasil pengujian, dapat ditarik kesimpulan bahwa dapat tercapai akurasi sistem sebesar 60% pada trial 8 dengan memanfaatkan fungsi kernel Polynomial. Kemudian pada proses penelitian dominasi ciri, ternyata ketiga ciri yang digunakan saling mendominasi satu sama lain dan menghasilkan tingkat akurasi yang sama saat dilakukan tracing berdasarkan kombinasi ciri yang telah ditentukan. Untuk tabel ketepatan klasifikasi sistem ditunjukkan pada Gambar 6.

		<i>Excited</i>	<i>Uninterested</i>
<i>Excited</i>		12	8
<i>Uninterested</i>		3	9
		<u><i>Predicted Class</i></u>	
		<u><i>True Class</i></u>	

Gambar 6 Tabel ketepatan klasifikasi sistem.

9. Kesimpulan

Berdasarkan hasil skenario pelatihan dan pengujian data yang dilakukan dapat ditarik kesimpulan, sebagai berikut:

1. Sistem mampu mengenali dan klasifikasi tiap kelas arousal berdasarkan tabel ketepatan klasifikasi dengan hanya 2 data yang tidak bisa diklasifikasikan dengan baik.
2. Percobaan terbaik terletak pada *trial 8* dengan memanfaatkan fungsi kernel *polynomial* dengan akurasi maksimum 60% dan pada penggunaan kombinasi ciri dominan yaitu ketiga ciri (PCA, entropy, dan kurtosis) saling mendominasi satu sama lain dan menghasilkan akurasi yang sama. Dilihat dari keseragaman hasil dominasi ciri dan akurasi penggunaan fungsi kernel, untuk saat ini sistem masih terbilang mampu dalam melakukan proses klasifikasi dengan baik dengan misklasifikasi sebanyak 11 dari keseluruhan 32 data.

10. Daftar Pustaka

- [1] M. M. Sorkhabi, "Emotion Detection from EEG signals with Continuous Wavelet Analyzing," *Am. J. Comput. Res. Repos.*, vol. 2, no. 4, pp. 66–70, 2014.
- [2] Y. Pan, P. Shen, and L. Shen, "Speech emotion recognition using support vector machine," *Int. J. Smart Home*, 2012.
- [3] N. Thammasan, K. Moriyama, K. Fukui, and M. Numao, "Familiarity effects in EEG-based emotion recognition," *Brain Informatics*, vol. 4, no. 1, pp. 39–50, 2017.
- [4] A. Hilmi, I. Wijayanto, and S. Hadiyoso, "Analisis Perbandingan Pola Sinyal Alfa Dan Beta Eeg Untuk Klasifikasi Kondisi Rileks Pada Perokok Aktif Dengan Menggunakan K-Nearest Neighbor Pattern Comparison Analysis Between Alpha and Beta Eeg Signal for Relaxed Condition Classification on Active Smok," vol. 4, no. 3, pp. 3395–3402, 2017.
- [5] M. A. Lestari, A. Rizal, and U. Suryana, "Deteksi Apnea Tidur Melalui Sinyal Elektrokardiogram Menggunakan Metode Discrete Wavelet Transform , Principal Component Analysis , dan Linear

Discriminat Analysis,” 2015.

- [6] A. P. Gosavi and S. R. Khot, “Emotion Recognition using Principal Component Analysis with Singular Value Decomposition,” *2014 Int. Conf. Electron. Commun. Syst. ICECS 2014*, 2014.
- [7] A. Muntasa and M. K. Sophan, “Ekstraksi Fitur Berbasis 2D-Discrete Cosine Transform dan Principal Component Analysis Untuk Pengenalan Citra Wajah,” vol. 2009, no. Snati, pp. 2–7, 2009.
- [8] P. Michel, “Support Vector Machines in Automated Emotion Classification,” *Human-Computer Interact.*
- [9] I. A. Muis, M. Affandes, I. A. Muis, P. Studi, T. Informatika, and F. Sains, “Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet,” *J. Sains, Teknol. dan Ind.*, vol. 12, no. 2, pp. 189–197, 2015.
- [10] Z. Yan and Y. Yang, “Application of ecoc svms in remote sensing image classification,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 40, no. 2, pp. 191–196, 2014.