MATHunesa
Jurnal Ilmiah Matematika *Volume 09 No 01*
e-ISSN : 2716-506X | p-ISSN : 2301-9115 *Tahun 2021*

KLASIFIKASI VIDEO PEMBELAJARAN DARING YANG MEMBINGUNGKAN SISWA DENGAN ALGORITMA *K-STAR NEAREST NEIGHBOR*

Hanief Jamielatuththooah
Jurusan Matematika, FMIPA, Universitas Negeri Surabaya
e-mail : hanief.17030214011@mhs.unesa.ac.id

Abstrak

Massive Open Online Course (MOOC) adalah salah satu pembelajaran yang dilakukan secara daring melalui rekaman video yang dapat ditonton berulang kali sehingga memudahkan siswa dalam memahami pembelajaran. Namun pembelajaran ini dapat membingungkan siswa karena tidak ada umpan balik secara langsung dari guru. Untuk mengetahui tingkat kebingungan siswa terhadap video yang ditonton dapat menggunakan klasifikasi sinyal EEG. Pada penelitian ini digunakan algoritma *K-Star Nearest Neighbor* sebagai metode untuk mengklasifikasikan video MOOC yang membingungkan siswa melalui sinyal EEG. Penelitian ini bertujuan untuk mengetahui apakah algoritma *K-Star Nearest Neighbor* dapat digunakan untuk klasifikasi sinyal EEG ditinjau dari nilai akurasi. Dataset yang digunakan yaitu dataset *Confused student* EEG *brainwave* data bersumber dari Kaggle yang memiliki 15 atribut serta dua kelas yaitu kelas *not confused* dan *confused*. Data diambil dari 10 subyek dan masing-masing menonton 10 video MOOC berbeda. Sebelum proses klasifikasi, terlebih dahulu dilakukan normalisasi dan *split validation*. Akurasi terbaik untuk label *pre-defined* yaitu label pada video berdasarkan tingkat kebingungan yang diberikan oleh guru mencapai 63.33% dengan *time split* 0.42 detik, sedangkan untuk label *user-defined* yaitu label video berdasarkan tingkat kebingungan yang dialami oleh siswa dengan akurasi terbaik mencapai 73.33% dengan *time split* 0.13 detik.
Kata Kunci: Video MOOC, Sinyal EEG, K-Star Nearest Neighbor

Abstract

Massive Open Online Course (MOOC) is one of the online learning based through video recordings that can be watched repeatedly, making it easier for students to understand learning materials. However, this learning can confuse students because there is no direct feedback from the teacher. EEG signal classification can be used to determine the level of student confusion based on the video. In this study, the K-Star Nearest Neighbor algorithm is used as a method for classifying MOOC videos that confuse students through EEG signals. This study aims to determine whether the K-Star Nearest Neighbor algorithm can be used for EEG signal classification in terms of its accuracy value. The dataset is Confused student EEG brainwave dataset from Kaggle which has 15 attributes and two classes, namely the not confused and confused classes. Data were taken from 10 subjects and each watched 10 MOOC videos. Before the classification process, normalization and split validation are first carried out. The best accuracy for pre-defined labels, namely labels on videos based on the level of confusion given by the teacher, reached 63.33% with the time split is 0.42 seconds, while for user-defined labels, namely video labels based on the level of confusion that came by students with the best accuracy reached 73.33% with a time split of 0.13 seconds.
Keywords: *MOOC Video, EEG Signal, K-Star Nearest Neighbor*

PENDAHULUAN

Saat ini sudah banyak kursus berbasis MOOC (Massive Open Online Course) dengan berbagai pilihan topik pelajaran. Berbeda dari pembelajaran konvensional, MOOC memudahkan siswa untuk mengakses ulang video pembelajaran jika mereka tidak paham. Beberapa peneliti, guru, perguruan tinggi, dan universitas telah mencoba untuk mengkolaborasikan pembelajaran berbasis MOOC dan pembelajaran konvensional (Brame, 2016).

Siswa pada umumnya mengalami kebingungan ketika sulit memahami materi pelajaran yang dijelaskan oleh guru maupun saat belajar mandiri. Pada pembelajaran konvensional guru dapat dengan mudah memberikan tanggapan terhadap kebingungan yang dialami oleh siswa. Namun hal tersebut tidak berlaku untuk pembelajaran berbasis MOOC dimana guru dan murid tidak bertemu secara langsung. Tanpa adanya interaksi tatap muka secara langsung di lingkungan digital, emosi seperti kebingungan sulit untuk dideteksi. Oleh karena itu, pemberian umpan balik harus diberikan untuk

membantu kemajuan siswa sehingga mereka tidak terjebak dalam kebingungan (Lodge et al., 2018).

Lingkungan pembelajaran digital yang sebagian besar memberikan kemudahan siswa untuk belajar secara mandiri berpotensi untuk diciptakan suatu pendeteksi untuk merespon kesulitan siswa, namun potensi tersebut belum terwujud (Lodge et al., 2018). Sinyal EEG dapat digunakan sebagai alternatif untuk meningkatkan pembelajaran berbasis MOOC (Wang et al., 2013). Sinyal EEG direkam untuk menginformasikan aktivitas listrik dalam otak. Beberapa penelitian berfokus pada klasifikasi Sinyal EEG untuk mendeteksi emosi (Li & Jung, 2020), tingkat konsentrasi (Karmila et al., 2016), dan tingkat kebingungan (Kumar et al., 2019). Pada penelitian sebelumnya, klasifikasi sinyal EEG dilakukan untuk mendeteksi kebingungan dalam memahami konten video dalam pembelajaran berbasis MOOC dengan algoritma Gaussian Naive Bayes (Wang et al., 2013) dan beberapa algoritma supervised learning (Kumar et al., 2019).

Pada penelitian ini, peneliti menggunakan data sinyal EEG siswa selama menonton video MOOC untuk diklasifikasikan berdasarkan video yang ditonton dengan algoritma lazy K-Star Nearest Neighbor untuk mendeteksi kebingungan dalam memahami video MOOC. Algoritma K-Star Nearest Neighbor dipilih karena belum ada penelitian yang mengunakan algoritma ini. Penelitian ini bertujuan untuk mengetahui apakah algoritma K-Star Nearest Neighbor dapat digunakan untuk klasifikasi sinyal EEG ditinjau dari nilai akurasi.

KAJIAN TEORI

VIDEO MOOC

Massive Open Online Course (MOOC) adalah kursus online untuk semua orang yang memiliki akses internet dan memiliki motivasi diri dalam belajar dimana saja dan kapan saja (Jordan, 2014). Pengguna MOOC juga lebih mudah untuk belajar secara mandiri, tanpa adanya pengawasan secara langsung oleh guru, tidak harus mengantri saat mendaftar, dan biaya yang dikeluarkan untuk mengikuti kursus berbasis MOOC juga lebih terjangkau (Kloft et al., 2014).

SINYAL EEG

Sinyal EEG merupakan sinyal bioelektrik yang berasal dari permukaan kulit manusia, umumnya sinyal ini bersifat kompleks dan dapat digunakan sebagai sumber informasi fungsi otak (Karmila et al., 2016). EEG direkam dengan aktif AgCl elektroda yang ditempelkan pada bagian-bagian tertentu di kulit kepala. EEG dibagi menjadi empat yaitu gelombang delta (0,5 – 3 Hz) yang muncul ketika seseorang sedang tertidur nyenyak, theta (4 – 7 Hz) ketika seseorang mengantuk atau stress, alpha (8 – 13 Hz) saat keadaan rileks, beta (14 – 30 Hz) muncul ketika seseorang sedang berpikir, dan gamma (30 – 50 Hz) ketika seseorang dalam kesadaran penuh (Karmila et al., 2016).

K-STAR NEAREST NEIGHBOR

K-Star Nearest Neighbor adalah salah satu algoritma *lazy* dan merupakan modifikaasi dari algoritma K-Nearest Neighbor. Nama lain dari K-Star Nearest Neighbor adalah *weighted* K-Nearest Neighbor (Anava & Levy, 2017). K-Star Nearest Neighbor didefinisikan sebagai metode analisis cluster yang bertujuan untuk mempartisi pengamatan 'n' menjadi cluster 'k' dimana setiap pengamatan termasuk dalam cluster dengan mean terdekat (Vijayarani et al., 2013).

Data sample baru, katakan x, ditetapkan ke kelas yang paling sering terjadi di antara k-tetangga terdekat. Jarak pada algoritma K-Star diukur menggunakan entropi yang dihitung dari rata-rata kompleksitas transformasi suatu sampel data ke sample data lainnya, sehingga probabilitas transformasi dihitung secara acak (Tejera Hernández, 2015). K-Star dapat mengatasi masalah pada atribut bernilai real, atribut simbolis, dan nilai yang hilang dengan melakukan pendekatan secara konsisten (Sharma & Jain, 2013).

Fungsi K-Star dapat dihitung menggunakan persamaan (1).

$$K^*(y|x) = -\log_2 P^*(y|x) \quad (1)$$

dimana P^* adalah sumasi dari probabilitas sample data pada kelas y dengan menjumlahkan probabilitas dari x ke setiap sample anggota kelas y (Martínez-López et al., 2016).

Pada algoritma K-Nearest Neighbor hanya menggunakan jarak metrik terurut untuk menentukan k-tetangga terdekat. Namun, penggunaan jarak metrik dapat mempengaruhi kinerja algoritma secara signifikan misalnya pada analisis text, *computer vision*, analisis program, dan lain-lain. Pada algoritma K-Star Nearest Neighbor

jarak metrik yang telah diurutkan tidak langsung digunakan untuk menentukan k-tetangga terdekat namun dilakukan pembobotan terlebih dahulu untuk memperoleh nilai k yang optimal untuk setiap sample data. Kelemahan algoritma K-Star Nearest Neighbor yaitu terbatasnya perspektif geometris karena hanya mempertimbangkan jarak antara *decision point* dan sample data $\{d(x_0, x_i)\}_{i=1}^{n}$, serta mengabaikan relasi geometrik antar sample data $\{d(x_i, x_j)\}_{i,j=1}^{n}$ (Anava & Levy, 2017).

METODE

Pada penelitian ini terdiri dari tiga tahap yaitu pra-pemrosesan data, proses klasifikasi, dan proses evaluasi hasil klasifikasi.

DATASET

Penelitian ini menggunakan *Confused student EEG brainwave data* yang bersumber dari Kaggle (*https://www.kaggle.com/wanghaohan/confused-eeg)*. Data tersebut berasal dari sinyal EEG 10 subyek pelajar yang direkam saat menonton video MOOC. Masing-masing pelajar menonton sebanyak 10 video. Dari 15 atribut hanya digunakan sebanyak 10 atribut yang terdiri dari sinyal delta, theta, gamma1, gamma2, alfa1, alfa2, beta1, beta2, pre-defined label, dan user-defined label. Lima atribut lainnya tidak digunakan karena bukan merupakan sinyal EEG siswa.

Pre-defined label dan user-defined label masing-masing merupakan atribut kelas dengan dua kelas yaitu kelas *not confused* dan *confused*. Label *pre-defined* adalah label pada video berdasarkan tingkat kebingungan yang diberikan oleh guru. Label *user-defined* adalah label video berdasarkan tingkat kebingungan yang dialami oleh siswa. Pemberian label dari guru dan siswa bertujuan untuk mempermudah guru dalam memberikan respon terhadap kebingungan yang dialami siswa dan selanjutnya guru dapat memberikan video MOOC yang lebih mudah untuk dipahami siswa.

PRA-PEMROSESAN DATA

Pra-pemrosesan data dilakukan untuk memperoleh data input yang sesuai dengan algoritma yang digunakan agar mendapatkan akurasi yang baik. Data yang digunakan terdiri dari 1282 baris dan 10 kolom atribut. Kemudian data dikelompokkan berdasarkan SubjectID dan VideoID. Satu data point pada data merupakan sinyal EEG satu subject dengan satu video yang ditonton. Sehingga dari data diperoleh 100 data point. Untuk mempermudah tahap klasifikasi, setiap data point masing-masing dihitung rata-ratanya sehingga diperoleh rata-rata dari sinyal EEG masing-masing siswa berdasarkan VideoID yang ditonton.

Mean data kemudian di normalisasi dengan menghitung nilai Z-score. Z-score dapat dihitung menggunakan persamaan (2).

$$Z_{score} = \frac{x_i - \overline{x}}{\sigma}, untuk\ i = 1,2,\dots,n \quad (2)$$

dimana $x_i \epsilon R$ adalah sample data ke-i, $\overline{x}$ adalah rata-rata data, dan σ adalah standar deviasi. Normalisasi dilakukan untuk memperoleh data dengan standar deviasi yang sama.

Setelah melakukan normalisasi selanjutnya adalah melakukan *split validation* yaitu membagi data menjadi *training data* dan *test data*. Dalam penelitian ini, digunakan *percentage split* dengan perbandingan 70:30. Pembagian data dilakukan dengan melakukan *shuffling data* secara acak kemudian mengambil 70% data dimulai dari indeks pertama untuk *training data* dan sisanya yaitu 30% untuk *test data.*

PROSES KLASIFIKASI

Proses klasifikasi pada peneltian ini diawali dengan menghitung jarak antara *training data* dan *test data* menggunakan *Euclidean Distance* dengan formula seperti pada persamaan (3).

$$D(a,b) = \sqrt{\sum_{i,j}(a_i - b_j)^2} \quad (3)$$

dimana a_i adalah *test data* ke i untuk $i \in \{1,2,\dots,n\}$ dengan n adalah banyak *test data* dan b_j adalah *training data* ke j untuk $j \in \{1,2,\dots,m\}$ dengan m adalah banyak *traing data.* Kemudian jarak diurutkan dari yang terdekat sebagai input pada Algoritma K-Star Nearest Neighbor

ALGORITMA K-STAR NEAREST NEIGHBOR

Algoritma K-Star Nearest Neighbor atau *weighted* K-Nearest Neighbor adalah sebagai berikut :

Misalkan $\beta \in \mathbb{R}^n$ adalah vektor jarak terurut dan $y \in \{0,1\}$ adalah kelas aktual dimana $y = 0$ merupakan data dengan kelas *not confused* dan $y = 1$ untuk data dengan kelas *confused.*

Inisialisasi $\lambda_0 = \beta_1 + 1, k = 0$. Ketika $\lambda_k > \beta_{k+1}$ dan $k < n - 1$ lakukan *update* pada nilai k sedemikian hingga $k = k + 1$.

Kemudian hitung λ_k menggunakan formula pada persamaan (4)

$$\lambda_k = \frac{1}{k}\left(\sum_{i=1}^{k}\beta_i + \sqrt{k + \left(\sum_{i=1}^{k}\beta_i\right)^2 - k\left(\sum_{i=1}^{k}{\beta_i}^2\right)}\right)$$

Kemudian hitung

$$\hat{y} = \sum_i \alpha_i y_i \quad (5)$$

dengan

$$\alpha_i = \frac{max(0,(\lambda_k - \beta_i))}{\sum_{i=1}^{n} max(0,(\lambda_k - \beta_i))} \quad (6)$$

dimana $\hat{y}$ adalah kelas hasil prediksi, dan α adalah vektor bobot (Anava & Levy, 2017).

PROSES EVALUASI HASIL KLASIFIKASI

Hasil klasifikasi dapat dievaluasi menggunakan *confusion matrix* yaitu tabel yang mendeskripsikan performa suatu model klasifikasi dengan membandingkan jumlah data pada kelas hasil prediksi dan kelas aktual.

Tabel 1. Confusion Matrix

		Kelas Aktual	
		True	*False*
Kelas Prediksi	*True*	TP	FP
	False	FN	TN

dimana TP (*True Positive)* adalah jumlah data pada kelas *not confused* yang berhasil terklasifikasi pada kelas *not confused*, TN (*True Negative*) adalah jumlah data pada kelas *confused* yang terklasifikasi sebagai *confused*, FP (*False Positive)* adalah jumlah data pada kelas *not confused* yang terklasifikasi sebagai *confused*, FN (*False Negative)* adalah jumlah data pada kelas *confused* yang terklasifikasi sebagai data pada kelas *not confused*.

Nilai akurasi yaitu tingkat keberhasilan algoritma dalam mengklasifikasi data sample sesuai dengan kelasnya. *Precision* yaitu tingkat keakuratan ekspektasi terhadap prediksi yang dihasilkan dari sistem. *Recall* yaitu tingkat kesuksesan suatu sistem dalam mengklasifikasi data sesuai dengan kelas aktual. Nilai akurasi, *precision*, dan *recall* dapat dihitung menggunakan formula berikut ini.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\% \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan uji coba dataset dengan nilai *entropicAutoBlend* yang berbeda untuk setiap predefined label dan user-defined label. Jika *entropicAutoBlend* bernilai *False* maka atribut kelas dinyatakan sebagai numerik. Jika *entropicAutoBlend* bernilai *True* maka atribut dinyatakan sebagai simbol. Class 0 menyatakan kelas *not confused* dan class 1 menyatakan kelas *confused*.

Tabel 2. Tabel Hasil Klasifikasi (entropicAutoBlend=False)

Label	**Akurasi (%)**	**Test time (s)**	**Class**	***Precision***	***Recall***
Pre-defined	60	0.11	0	0.667	0.588
			1	0.533	0.615
			Avg.	0.609	0.600
User-defined	73.33	0.13	0	0.800	0.706
			1	0.667	0.769
			Avg.	0.742	0.733

Tabel 3. Tabel Hasil Klasifikasi (entropicAutoBlend=True)

Label	**Akurasi (%)**	**Test time (s)**	**Class**	***Precision***	***Recall***
Pre-defined	63.33	0.42	0	0.750	0.529
			1	0.556	0.769
			Avg.	0.666	0.633
User-defined	60	0.45	0	0.629	0.529
			1	0.529	0.692
			Avg.	0.622	0.600

Tabel 2 menunjukkan hasil performa klasifikasi dengan parameter *entropicAutoblend* bernilai *False* dan Tabel 3 menunjukkan hasil performa klasifikasi dengan parameter *entropicAutoBlend* bernilai *True*. Dari kedua tabel tersebut nilai akurasi terbaik untuk label *pre-defined* label mencapai 63.33% untuk atribut simbolik dan akurasi terbaik pada untuk label *user-defined* mencapai 73.33% untuk atribut numerik. Label *pre-defined* mecapai akurasi terbaik dengan atribut simbolik dikarenakan pada proses klasifikasi untuk data yang digunakan adalah data EEG siswa sehingga tidak mewakili label *pre-defined* secara umum. Karena label *pre-defined* merupakan label

yang diberikan sebelum siswa menonton video yang berarti label video ini sebagai acuan secara simbolik saja. Sedangkan untuk label *user-defined* mewakili data EEG secara umum karena label diberikan oleh siswa untuk setiap video yang ditonton.

Nilai tertinggi pada *precision* dan *recall* diperoleh pada label *user-defined* dengan *entropicAutoBlend* bernilai *False*. Berdasarkan kedua tabel diatas, time split yang diperlukan untuk membagi data menjadi *training data* dan *test data* pada data dengan atribut numerik (*entropicAutoBlend* bernilai *False*) lebih cepat dibandingkan pembagian data pada data dengan atribut simbolik (*entropicAutoBlend* bernilai *True*). Dari hasil uji coba diatas akurasi terbaik mencapai 73.33% dan hasil tersebut lebih baik dibandingkan hasil uji coba dari penelitian sebelumnya oleh Vijayarani et al. dengan dataset yang berbeda yaitu klasifikasi text menggunakan K-Star Nearest Neighbor diperoleh akurasi mencapai 68.47% (Vijayarani et al., 2013).

PENUTUP

SIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat diambil kesimpulan bahwa algoritma K-Star Nearest Neighbor dapat mencapai nilai akurasi tertinggi ketika entropicAutoBlend bernilai False untuk data dengan atribut numerik. Sedangkan untuk data dengan atribut simbolik akurasi tertinggi dapat diperoleh menggunakan entropicAutoBlend bernilai True. Time split yang diperlukan untuk atribut numerik lebih cepat dibandingkan atribut simbolik.

DAFTAR PUSTAKA

Anava, O., & Levy, K. Y. (2017). k*-Nearest Neighbors: From Global to Local. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, *NIPS '16*, 9. http://arxiv.org/abs/1701.07266

Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE Life Sciences Education*, *15*(4), es6. https://doi.org/10.1187/cbe.16-03-0125

Jordan, K. (2014). Initial Trends in Enrolment and Completion of Massive Open Online Courses. *International Review of Research in Open and Distance Learning*, *15*(1), 133–160.

Karmila, R., Djamal, E. C., & Nursantika, D. (2016). Identifikasi Tingkat Konsentrasi Dari Sinyal EEG Dengan Wavelet dan Adaptive Backpropagation. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, *0*(0), 2016. https://journal.uii.ac.id/Snati/article/view/6250

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC Dropout over Weeks Using Machine Learning Methods. *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, 60–65. https://doi.org/10.3115/v1/W14-4111

Kumar, B., Gupta, D., & Goswami, R. S. (2019). Classification of Student's Confusion Level in E-Learning using Machine Learning. *International Journal of Innovative Technology and Exploring Engineering*, *9*(2S), 346–351. https://doi.org/10.35940/ijitee.B1092.1292S19

Li, G., & Jung, J. J. (2020). Maximum Marginal Approach on EEG Signal Preprocessing for Emotion Detection. *Applied Sciences*, *10*(21), 7677. https://doi.org/10.3390/app10217677

Lodge, J. M., Kennedy, G., Lockyer, L., Arguel, A., & Pachman, M. (2018). Understanding Difficulties and Resulting Confusion in Learning: An Integrative Review. *Frontiers in Education*, *3*(June), 1–10. https://doi.org/10.3389/feduc.2018.00049

Martínez-López, Y., Madera-Quintana, J., & De Varona, I. L. (2016). Study of The Performance of The K* Algorithm in International Databases. *Revista Politécnica*, *12*(23), 51–56.

Sharma, T. C., & Jain, M. (2013). WEKA Approach for Comparative Study of Classification Algorithm. *(IJARCCE) International Journal of Advanced Research in Computer and Communication Engineering*, *2*(4), 1925–1931. www.ijarcce.com

Tejera Hernández, D. C. (2015). An Experimental Study of K* Algorithm. *International Journal of Information Engineering and Electronic Business*, *7*(2), 14–19. https://doi.org/10.5815/ijieeb.2015.02.03

Vijayarani, M. S., Muthulakshmi, M. M., & Professor, A. (2013). Comparative Analysis of Bayes and Lazy Classification Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(8). www.ijarcce.com

Wang, H., Li, Y., Hu, X., Yang, Y., Meng, Z., & Chang, K. M. (2013). Using EEG to Improve Massive Open Online Courses Feedback Interaction. In *AIED Workshop*.