



SYSTEMATIC LITERATUR REVIEW (SLR): METODE, MANFAAT, DAN TANTANGAN LEARNING ANALYTICS DENGAN METODE DATA MINING DI DUNIA PENDIDIKAN TINGGI

Yansen Makleat¹⁾, Fauzia Ramadhan²⁾, Abdul Cholis³⁾, dan Sudin Saepudin⁴⁾

^{1, 2, 3, 4)}Program Studi Sistem Informasi, Universitas Nusa Putra

Jl. Raya Cibatu Cisaat No.21, Cibolang Kaler, Kec. Cisaat, Kab. Sukabumi, Jawa Barat 43155

e-mail: yansen.makleat_si20@nusaputra.ac.id¹⁾, fauzia.ramadhan_si20@nusaputra.ac.id²⁾,

abdul.cholis_si20@nusaputra.ac.id³⁾, sudin.saepudin@nusaputra.ac.id⁴⁾

*Korespondensi: e-mail: fauzia.ramadhan_si20@nusaputra.ac.id

ABSTRAK

Digitalisasi data pendidikan memungkinkan peneliti untuk dengan mudah mengakses data penting, terperinci, dan informatif untuk pengambilan keputusan yang diperlukan. Data mining dalam pendidikan berkaitan dengan aplikasi teknik penemuan pengetahuan, pembelajaran mesin, dan statistik berbasis informasi dari data pendidikan. Untuk mengukur tingkat kompetensi mahasiswa dalam menggunakan media *Learning Analytics* lalu sebagai tujuan untuk meningkatkan kualitas pendidikan tinggi saat ini kerap kali mengimplementasikan Educational Data Mining (EDM) maka dalam metodologi penelitian peneliti akan menyoroti pemanfaatan data mining (DM) di bidang pendidikan dan juga pendekatan yang dapat digunakan secara memadai untuk memprediksi perilaku dan kinerja mahasiswa di dunia pendidikan melalui *systematic literature review* dan juga pendekatan dengan menggunakan klasifikasi Smooth Support Vector Machine (SSVM) dan teknik algoritma kernel K-Means Clustering yang dapat digunakan secara memadai untuk memprediksi atau meramalkan perilaku dan kinerja mahasiswa. Hasil yang peneliti harapkan dapat sesuai dengan tujuan penelitian ini yaitu dapat mengetahui efektifitas penggunaan *data mining* di perguruan tinggi maka penulis menggunakan K-Means Clustering dan SVM untuk mengetahui setiap kategori perilaku dan kinerja dari mahasiswa dengan mengambil 10 sampel mahasiswa yang telah menggunakan *Learning Analytics* untuk sistem belajarnya yakni LMS (*Learning Management System*). Dan tujuannya untuk menjawab *literatur review* peneliti terkait penelitian-penelitian sebelumnya yakni efektivitas dari penggunaan *Education Data Mining* dalam melihat efektifitas dan tantangan dari *Learning Analytics* di perguruan tinggi.

Kata Kunci: *Learning Analytics*, Pendidikan Tinggi, *K-Means Clustering*, SSVM

ABSTRACT

Digitization of data education enables researchers to easily access important, detailed and informative data to make the necessary decisions. Data mining in education is related to the application of knowledge discovery techniques, machine learning, and information statistics from data education. To measure the level of competence of students in using Learning Analytics media, then as a goal to improve the quality of higher education at this time, Educational Data Mining (EDM) is often implemented, so in research methodology researchers will apply the use of data mining (DM) in the field of education and also approaches that can be used adequately to predict student behavior and performance in education through a systematic literature review and also an approach using the Smooth Support Vector Machine (SSVM) classification and the kernel K-Means Clustering algorithm technique which can be used adequately to predict or predict student behavior and performance. The results that the researchers hope are in accordance with the objectives of this study, namely being able to find out the effectiveness of using data mining in tertiary institutions, so the authors use K-Means Clustering and SVM to find out each category of student behavior and performance by taking 10 samples of students who have used Learning Analytics for the system. learning namely LMS (Learning Management System). And the aim is to answer research literature studies related to previous studies, namely the effectiveness of using Education Data Mining in seeing the effectiveness and challenges of Learning Analytics in higher education



Keywords: *Learning Analytics, Higher Education, K-Means Clustering, SSVM*

I. PENDAHULUAN

Saat ini, perkembangan pada teknologi informasi terus berlanjut merambah ke seluruh lini masa salah satunya ke dunia pendidikan. Pendidikan tidak hanya sekedar berperan sebagai media *transfer of knowledge*, namun pendidikan juga berfungsi sebagai sarana berkembangnya kemampuan para siswa, baik akademik maupun non-akademik. Penerapan *data mining* sendiri adalah sebagai solusi bagi dunia pendidikan untuk menemukan beberapa pengetahuan baru dan pola tersembunyi yang nantinya membantu pengambilan keputusan dalam lingkup manajemen pendidikan yang tujuannya untuk meningkatkan sistem pendidikan karena sistem pendidikan berbasis web saat ini terus dikembangkan karena dalam beberapa fiturnya dapat menangkap perilaku peserta didik secara signifikan[1]. Perkembangan EDM merupakan proses transformasi yang telah banyak dihadapi oleh berbagai peneliti dalam berbagai bidang aplikasi termasuk pemodelan domain pendidikan, analisis komponen pembelajaran, profil pengguna, pengetahuan pengguna, pemodelan perilaku, pengalaman pemodelan, analisis pembelajaran, dan analisis tren[2].

Saat ini memprediksi kinerja belajar siswa merupakan sebuah tantangan dan menjadi tugas penting dalam dunia pendidikan. Prediksi kinerja akademik melalui *Learning Analytics (LA)*. *Learning analytics* diartikan sebagai pengukuran, pengumpulan, analisis, dan pelaporan data terkait peserta didik, LA menjadi salah satu dari bidang teknologi pendidikan yang menjadi perhatian para peaktisi pendidikan untuk diteliti lebih lanjut[3]. LA tidak hanya membantu siswa dalam mengendalikan pembelajaran mereka sendiri dan menjadi siswa yang mandiri tetapi juga memungkinan para pendidik untuk mengidentifikasi resiko siswa dalam kemungkinan kegagalan[4].

Salah satu dari Algoritma pada data mining yang digunakan dalam perhitungan manual dan menggunakan tools SPSS dalam penelitian ini dengan K-Means Clustering. K-Means dapat membagi data, kemudian mengelompokkannya ke beberapa kelompok yang memiliki karakteristik yang sama. Kemudian, berdasarkan perbedaan yang ada di antara masing-masing cluster, setiap cluster dapat dipisahkan. Ini adalah hasil cluster data yang dapat dipertimbangkan. Saat membuat fungsi keputusan dapat dibedakan, metode Smooth Support Vector Machine (SVM) adalah algoritma SVM konvensional yang memperkenalkan batasan kehalusan pada margin. [5].

Berdasarkan latar belakang tersebut peneliti menggabungkan metode SVM dengan K-Means Clustering dalam Systematic Literature Review yang peneliti lakukan, tujuannya adalah untuk mengklasifikasikan dan mengetahui perkembangan EDM dan Learning Analytics dalam dunia pendidikan saat ini dengan mengelompokkan literatur-literatur penelitian dan menemukan jawaban terkait manfaat dan tantangan EDM dan Learning Analytics dalam Pendidikan Tinggi.

II. METODE PENELITIAN

A. Systematic Literature Review

Pengumpulan data dalam penelitian SLR[6] termasuk pertanyaan penelitian, proses pencarian, standar batasan dan masukan, penilaian kualitas, pengumpulan data, analisis data, dan penyimpangan laporan. Berikut ini adalah penjelasannya:

1. Pertanyaan Penelitian (RQ)

Pada penelitian ini, pertanyaan peneliti (RQ1) adalah sebagai berikut:

RQ1. Apa saja hasil belajar yang dipengaruhi oleh pendekatan pembelajaran pada *Educational Data Mining*?

RQ2. Bagaimanakah strategi pembelajaran *Education Data Mining* yang diimplementasikan di perguruan tinggi?

RQ3. Bagaimana hasil penelitian mengenai strategi pembelajaran *Education Data Mining*?

2. Proses Pencarian

Search engine (Google Chrome) digunakan untuk melakukan pencarian dengan menggunakan sumber yang relevan dan sederhana diakses untuk menjawab pertanyaan penelitian dengan alamat situs Science Direct, Google Scholar, Scopus, IEEE sebagai data sekunder.

3. Kriteria Batasan dan masukan

Pada tahap ini, data yang ditemukan digunakan untuk penelitian SLR. Jika studi memenuhi kriteria



berikut, maka studi tersebut dianggap layak untuk digunakan dalam penelitian.

- a. Rentang waktu data yang digunakan 2018-2023.
- b. Data tersebut diperoleh dari sumber Science Direct, Google Scholar, Scopus, IEEE.
- c. Data yang digunakan hanya berhubungan dengan penggunaan *Education Data Mining*.

4. Kualitas Penelitaian

Selanjutnya, pertanyaan kriteria kualitas penilaian akan digunakan untuk mengevaluasi hasil penelitian SLR. Kriteria ini termasuk sebagai berikut:

QA1. Apakah jurnal diterbitkan pada tahun 2018-2023?

QA2. Apakah pada jurnal memberikan informasi yang digunakan dalam penggunaan strategi pembelajaran *Education Data Mining*?

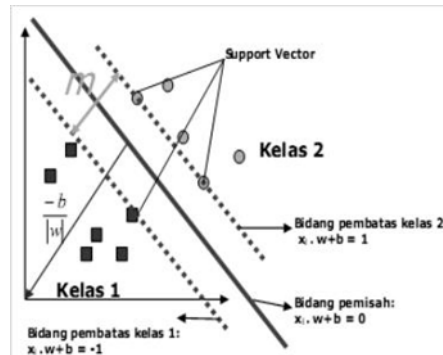
QA3. Apakah hasil dan diskusi tentang penerapan strategi pembelajaran ditulis dalam artikel jurnal *Education Data Mining* di perguruan tinggi?

B. Preprocessing Data

1. *Data Cleaning*: Tahap ini melibatkan identifikasi dan penanganan terhadap data yang hilang, tidak lengkap, atau salah. Hal ini meliputi penghapusan data duplikat, penanganan missing values, dan penanganan outliers yang dapat mempengaruhi analisis [7].
2. *Integrasi Data*: Jika data yang digunakan berasal dari berbagai sumber, maka perlu dilakukan integrasi data untuk menggabungkan data tersebut menjadi satu dataset yang konsisten [8]. Hal ini melibatkan pemadanan skema data, penyatuan data yang saling terkait, dan pemilihan atribut yang relevan.
3. *Transformasi Data*: Pada tahap ini, data dapat diubah atau ditransformasikan agar sesuai dengan kebutuhan analisis. Contohnya, normalisasi data untuk menghilangkan perbedaan skala [9], atau melakukan transformasi logaritmik untuk mengatasi distribusi yang tidak normal.
4. *Reduksi Dimensi*: Jika dataset memiliki jumlah atribut yang besar, teknik reduksi dimensi dapat digunakan untuk mengurangi kompleksitas data. Hal ini dapat dilakukan melalui pemilihan atribut yang paling informatif atau melalui teknik seperti *Principal Component Analysis (PCA)* untuk menghasilkan representasi data yang lebih ringkas namun masih menjaga informasi penting.
5. *Discretization*: *Discretization* dilakukan jika variabel yang kontinu perlu diubah menjadi variabel diskrit. Hal ini dilakukan untuk mempermudah analisis dan penggunaan algoritma yang membutuhkan variabel diskrit.
6. *Sampling*: Jika dataset memiliki ukuran yang sangat besar, teknik sampling dapat digunakan untuk mengambil *subset* data yang lebih kecil [10]. Ini dapat mempercepat waktu komputasi dan mengurangi beban memori yang diperlukan dalam analisis.

C. Smooth Support Vector Machine (SVM)

Dalam hal klasifikasi dan regresi, support vector machines (SVM) merupakan teknik prediksi. Hyperplane, atau dalam ruang yang besar fungsi pemisah akan dicari untuk memaksimalkan jarak (margin) antara dua kelas data. Proses klasifikasi menggunakan SVM secara sederhana dapat digambarkan sebagai upaya untuk menemukan bahwa garis adalah fungsi pemisah terbaik, yang memiliki kemampuan untuk membedakan dua kelas data dari satu sama lain dalam ruang input dengan nilai +1, -1 [11].



Gambar 1. Konsep Fungsi Pemisah pada SVM

Gambar 1 menunjukkan beberapa data kelas -1 dan +1. Data kelas -1 disimbolkan kotak, dan data kelas +1 disimbolkan bulat. Fungsi terbaik untuk memisahkan data adalah yang dapat memisahkan kedua kelas data dengan nilai margin (m) terbesar, yang menunjukkan jarak antara data dari masing-masing kelas dan fungsi pemisah[12]. Karena kedua kelas berada di bidang pembatas dan dekat dengan fungsi pemisah terbaik, garis tebal yang memisahkan mereka menunjukkan fungsi pemisah terbaik. Hanya data support vector yang digunakan selama proses yang menghasilkan fungsi pemisah terbaik[13].

D. K-Means Clustering

Algoritma K-Means[14] menggunakan prosedur berulang untuk mengumpulkan data kelompok karena didasarkan pada menentukan jumlah awal kelompok dengan mengetahui nilai centroid awalnya. Setelah iterasi K-Means berhenti, algoritma K-Means mengelompokkan semua item data dalam dataset ke dalam suatu cluster. Jumlah cluster dihitung dengan melacak cluster yang paling dekat satu sama lain di seluruh objek[15].

III. HASIL DAN PEMBAHASAN

Hasil Penelitian

Tabel 1. *Literature Review*

Nama, Tahun	Permasalahan	Data	Fokus Penelitian	Metode Algoritma	Variabel	Kontribusi
(Abadi et al., 2018)[16]	Karena input siswa berasal dari berbagai daerah, departemen pemasaran harus menggunakan strategi khusus untuk mempromosikan calon siswa.	100 record data siswa	Strategi promosi	K-Means	Alamat rumah, sekolah asal, transportasi, dan alasan memilih sekolah	Metode K-Means yang digunakan untuk mendukung strategi promosi dalam menjaring calon mahasiswa baru.
(Purba et al., 2018)[17]	Siswa yang putus sekolah dapat terjadi karena berbagai alasan, termasuk orang tua yang tidak mendukung,	36 record data akademik mahasiswa	Prediksi mahasiswa DO	K-Means	Data penilaian kualitas dan IPK	Universitas dapat menggunakan model ini untuk mengetahui situasi mahasiswa



	siswa yang malu-malu, semangat belajar yang rendah, dan kurangnya waktu untuk berperilaku.					dan membantu dalam proses pengambilan keputusan preventif untuk mencegah mahasiswa keluar.
(Negara & Doni, 2020)[18]		Data pengolahan kas	Perbandingan performa algoritma data mining pada dataset tabungan mahasiswa	decision tree C4.5 dan metode K-Means	Haji savings program Magnitude of Savings Total Savings Id_member	Menentukan strategi mencari santri yang berpotensi menabung tabungan haji

Untuk penerapan algoritma yang digunakan untuk riset ini yakni SVM dan K-Means maka kami akan mengambil 10 data mahasiswa secara acak untuk kemudian diteliti lebih lanjut. Terlebih dahulu, sebelum pengolahan metode K-Means Clustering dapat dilakukan, Kota dan jurusan serta data nominal lainnya harus dalam bentuk angka dimana nantinya akan berbentuk frekuensi di tiap kategori kinerja dan perilaku mahasiswa.

Setelah data siswa diubah menjadi angka, algoritma pengelompokan K-Means dapat digunakan untuk mengelompokkan data. Beberapa langkah harus dilakukan sebelum algoritma dapat mengelompokkan data menjadi beberapa cluster:

1. Tentukan berapa banyak cluster yang ingin Anda miliki. Data yang ada saat ini akan dikelompokkan menjadi tiga kelompok dalam penelitian ini.
2. Menentukan titik pusat awal dari setiap kelompok. Titik pusat awal dipilih secara acak dalam penelitian ini, dan titik pusat awal dari setiap kelompok ditemukan.
3. Tempatkan setiap data pada cluster. Metode hard K-Means digunakan dalam penelitian ini untuk membagi semua data ke dalam cluster, yang berarti bahwa setiap cluster akan berada paling dekat dengan titik pusatnya. Untuk mengetahui cluster mana yang paling dekat dengan data, kita harus menghitung jarak dari siswa pertama ke pusat cluster pertama.
 - a. Menetapkan (X, C_k) ($1 \leq i \leq N$, $1 \leq k \leq K$) dengan nilai awal, membentuk K cluster awal C_1, C_2, \dots, C_K .
 - b. Untuk setiap cluster C_k , hitung C_k dan $g(C_k)$.

$$|C_k| = \sum_{i=1}^N \delta(x_i, C_k)$$

$$g(C_k) = \frac{1}{|C_k|^2} \sum_{i=1}^N \sum_{j=1}^N \delta(u_j, C_k) \delta(u_i, C_k) H(x_i, x_j)$$

- c. Untuk setiap sampel pelatihan x_i dan klaster C_k , hitung $f(x_i, C_k)$ Dan kemudian tetapkan x_i ke cluster terdekat.

$$f(x_i, C_k) = \frac{2}{|C_k|} \sum_{j=1}^N \delta(u_j, C_k) H(x_i, x_j)$$

$$\delta(x_i, C_k) = \begin{cases} 1, & f(x_i, C_k) + g(C_k) < f(x_i, C_j) + g(C_j) \\ & \text{for all } j \neq k \\ 0, & \text{otherwise} \end{cases}$$



- d. Ulangi langkah 2 dan 3 sampai konvergen.
- e. Pada tiap cluster C_k pilih sampel yang paling dekat dengan bagian pusat sebagai perwakilan dari C_k , $m_k = \text{Arg min } D(\Phi(X_i), Z_k)$.

Eksperimen dilakukan dalam dua tahap yaitu pelatihan dan tes atau pengujian. Pelatihan dilakukan untuk masing-masing kelas 80 dan 90 persen dari jumlah total kumpulan data. Pengujian dilakukan untuk setiap kelas dengan menggunakan semua dataset. Banyaknya nilai berlabel 1 diartikan sebagai prediksi nilai akhir mahasiswa. Contohnya dibawah ini:

```
Result test for Matrix data training :
4.0000 4.0000 4.0000 4.0000 0.9750 = 1 True
4.0000 4.0000 4.0000 4.0000 0.9750 = 1 True
4.0000 4.0000 4.0000 4.0000 0.9750 = 1 True
4.0000 4.0000 4.0000 4.0000 0.9750 = 1 True
4.0000 4.0000 4.0000 4.0000 0.9750 = 1 True
4.0000 4.0000 4.0000 4.0000 0.9750 = 1 True
4.0000 3.0000 2.0000 1.0000 -0.9421 = -1 True
3.0000 4.0000 2.0000 4.0000 -0.9909 = -1 True
2.0000 1.0000 3.0000 4.0000 -0.9517 = -1 True
3.0000 2.0000 4.0000 2.0000 -0.9437 = -1 True
1.0000 2.0000 1.0000 3.0000 -0.9423 = -1 True
2.0000 3.0000 2.0000 3.0000 -0.9427 = -1 True
1.0000 2.0000 3.0000 3.0000 -0.9427 = -1 True
2.0000 3.0000 4.0000 2.0000 -0.9814 = -1 True
3.0000 4.0000 2.0000 4.0000 -0.9909 = -1 True
2.0000 1.0000 2.0000 4.0000 -0.9515 = -1 True
2.0000 3.0000 4.0000 2.0000 -0.9814 = -1 True
2.0000 3.0000 4.0000 4.0000 -0.9421 = -1 True
2.0000 1.0000 3.0000 2.0000 -0.9425 = -1 True
3.0000 4.0000 2.0000 4.0000 -0.9909 = -1 True

Prediction = 100.00 percentage
Success for training...!!!
```

Gambar 2. Contoh Data Program pada SVM

IV. PEMBAHASAN

Literatur review mencakup berbagai komponen analisis sehingga dapat mengetahui proses pendidikan untuk meningkatkan pembelajaran dan mengungkapkan manfaat penting dari penggunaan informasi dalam pendidikan tinggi. Penggunaan analisis akademik juga merupakan keuntungan tambahan, karena implementasi algoritma untuk titik data yang berbeda untuk meningkatkan pembelajaran menghasilkan perubahan. Dengan melakukan analisis menyeluruh dari data yang besar, peneliti dapat menemukan informasi bermanfaat yang dapat membantu institusi pendidikan, siswa, guru, dan peneliti.

Penelitian ini diharapkan menghasilkan temuan yang linier yang konsisten dengan temuan penelitian sebelumnya. Tinjauan literatur menunjukkan masalah LA dengan pelacakan dan pengumpulan data, analisis data, hubungannya dengan ilmu pendidikan, peningkatan lingkungan belajar, teknologi baru, dan masalah etika yang berkaitan dengan hukum dan privasi. Penelitian ini telah menjelaskan berbagai cara learning analytics (LA) digunakan untuk menunjukkan cara-cara di mana penggunaan big data dapat membantu pendidikan, serta masalah yang dihadapi oleh pihak yang terlibat dalam proses pendidikan.

V. KESIMPULAN

Untuk mengetahui efektifitas penggunaan *data mining* di perguruan tinggi maka penulis menggunakan K-Means Clustering dan SVM untuk mengetahui setiap kategori perilaku dan kinerja dari mahasiswa dengan mengambil 10 sampel mahasiswa yang telah menggunakan *Learning Analytics* untuk sistem belajarnya yakni LMS (*Learning Management System*). Dan tujuannya untuk menjawab *literatur review* peneliti terkait penelitian-penelitian sebelumnya yakni efektivitas dari penggunaan *Education Data Mining* dalam melihat efektifitas dan tantangan dari *Learning Analytics* di perguruan tinggi. Selain itu, penelitian ini ingin menunjukkan bahwa analisis pembelajaran menekankan pentingnya memahami cara menggunakan data untuk meningkatkan proses pendidikan di semua tingkatan dan mengoptimalkan hasilnya.



REFERENCES

- [1] Bagno, R.B., Freitas, J.S., 2022. Setting the three-stage r&d shared portfolio methodology: an innovative approach to industry–university collaboration. *Rev. Gest. Ahead-Print*. <https://doi.org/10.1108/REGE-07-2021-0108>
- [2] Beuter Júnior, N., Faccin, K., Volkmer Martins, B., Balestrin, A., 2019. Knowledge-based dynamic capabilities for sustainable innovation: the case of the green plastic project. *Sustainability* 11, 2392. <https://doi.org/10.3390/su11082392>
- [3] González-Pérez, L.I., Ramírez-Montoya, M.S., 2022. Components of education 4.0 in 21st century skills frameworks: systematic review. *Sustainability* 14, 1493. <https://doi.org/10.3390/su14031493>
- [4] Keinz, P., Marhold, K., 2021. Technological competence leveraging projects via intermediaries: viable means to outbound open innovation and mediated capability building? *Int. J. Proj. Manag.* 39, 196–208. <https://doi.org/10.1016/j.ijproman.2020.10.006>
- [5] Li, L., Wang, Y., Xu, Y., Lin, K.Y., 2022. Meta-learning based industrial intelligence of feature nearest algorithm selection framework for classification problems. *J. Manuf. Syst.* 62, 767–776. <https://doi.org/10.1016/j.jmsy.2021.03.007>
- [6] E. Triandini, S. Jayanatha, A. Indrawan, G. Werla Putra, and B. Iswara, “Metode Systematic Literature Review untuk Identifikasi Platform dan Metode Pengembangan Sistem Informasi di Indonesia”, *Indonesian J. of Inf. Syst.*, vol. 1, no. 2, pp. 63–77, Feb. 2019.
- [7] Ramírez-Montoya, M.S., Castillo-Martínez, I.M., Sanabria-Z, J., Miranda, J., 2022. Complex thinking in the framework of Education 4.0 and Open Innovation-a systematic literature review. *J. Open Innov. Technol. Mark. Complex* 8, 4. <https://doi.org/10.3390/joitmc8010004>
- [8] Yang, H.W., Fang, S.C., 2018. Open innovation: Powered by people, enabled by IT. In *22nd Pacific Asia Conference on Information Systems-Opportunities and Challenges for the Digitized Society: Are We Ready?*, PACIS 2018. Association for Information Systems.
- [9] Zhang, X., Chu, Z., Ren, L., Xing, J., 2023. Open innovation and sustainable competitive advantage: the role of organizational learning. *Technol. Forecast. Soc. Change* 186, 122114. <https://doi.org/10.1016/j.techfore.2022.122114>
- [10] Zobel, A.K., Hagedoorn, J., 2020. Implications of open innovation for organizational boundaries and the governance of contractual relations. *Acad. Manag. Perspect.* 34, 400–423. <https://doi.org/10.5465/amp.2016.0175>
- [11] Maqsood, A., (2013) "Role of data mining in education sector." *International Journal of Computer Science and Mobile Computing* 2.4:374-383.
- [12] Parneet, K., Manpreet, S., Gurpreet, S., (2015) Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector, *Procedia Computer Science*, Volume 57.
- [13] Dipak Kumar Jana, Prajna Bhunia, Sirsendu Das Adhikary, Anjan Mishra, Analyzing of salient features and classification of wine type based on quality through various neural network and support vector machine classifiers, *Results in Control and Optimization*, Volume 11, 2023, 100219, ISSN 2666-7207, <https://doi.org/10.1016/j.rico.2023.100219>.
- [14] Madhulatha, T. S., 2012, An Overview On Clustering Methods, *IOSR Journal of Engineering*, II(4), pp.719-725.
- [15] Bangoria, B., Mankad, N. & Pambhar, V., 2013. A survey on efficient enhanced K-Means clustering algorithm. *International journal for scientific research & development*, I(9), pp.1698-700.



- [16] S. Abadi et al., “Application model of k-means clustering: Insights into promotion strategy of vocational high school,” *Int. J. Eng. Technol.*, vol. 7, no. 2.27 Special Issue 27, pp. 182–187, 2018, doi: 10.14419/ijet.v7i2.11491.
- [17] Windania Purba et al 2018 The effect of mining data k-means clustering toward students profile model drop out potential *J. Phys.: Conf. Ser.* 1007 012049
- [18] Y D P Negara and A F Doni 2020. Comparison of Data Mining Algorithm Performance on Student Savings Dataset *J. Phys.: Conf. Ser.* 1569 022081