

## ***Learning Analytics dan Educational Data Mining pada Data Pendidikan***

Selly Anastassia Amellia Kharis<sup>1, a)</sup>, Arman Haqqi Anna Zili<sup>2, b)</sup>

<sup>1</sup>Universitas Terbuka

<sup>2</sup>Universitas Indonesia

Email: <sup>a)</sup>selly@ecampus.ut.ac.id, <sup>b)</sup>armanhaz@sci.ui.ac.id

### **Abstrak**

Salah satu kelebihan dari pemanfaatan teknologi pada masa pandemi di dunia pendidikan adalah meningkatnya *digital footprint*. *Digital footprint* dapat berupa data mengenai diskusi antara pendidik dan peserta didik, nilai diskusi, nilai tugas, nilai ujian, kehadiran, frekuensi mengakses materi, dan sebagainya. Dengan bertambahnya *digital footprint*, semakin banyak hal yang dapat diketahui dan digali. Penggalian data pendidikan (*educational data mining*) telah dilakukan beberapa negara untuk menganalisis dan menyelesaikan isu-isu dalam bidang pendidikan. *Educational Data Mining* telah secara luas dipergunakan untuk berbagai keperluan seperti analisa gaya belajar siswa, kajian efektivitas bahan ajar, prediksi tren kinerja peserta didik, simulasi pengambilan keputusan, dan lain-lain. Penelitian ini bertujuan untuk menjelaskan pengertian, penggunaan, dan dampak *Learning Analytics* dan *Educational Data Mining* pada data pendidikan. Penelitian ini menggunakan metode studi literatur dengan pendekatan kualitatif. Sumber-sumber pada penulisan artikel ini berasal dari jurnal dan buku. Teknik analisis yang digunakan pada penelitian ini adalah teknik analisis isi. *Learning Analytics* dan *Educational Data Mining* menggabungkan beberapa ilmu seperti statistika, *data mining*, *machine learning*. *Learning Analytics* dan *Educational Data Mining* memiliki potensi untuk dikembangkan pada data pendidikan di Indonesia. *Learning Analytics* dan *Educational Data Mining* dapat memberikan informasi kepada institusi, pendidik, dan peserta didik sehingga mendukung analisis prediksi dan pada akhirnya dapat meningkatkan motivasi, kinerja dan hasil dari suatu proses pembelajaran. Dengan melacak, menggabungkan, dan menganalisis *digital footprint* peserta didik, jalur baru untuk kebijakan pendidikan hadir secara lebih terbuka sesuai dengan pertumbuhan *database* peserta didik.

**Kata kunci:** *educational data mining, learning analytics, machine learning*

### **PENDAHULUAN**

Sejak masa pandemi COVID-19, seluruh kegiatan masyarakat menjadi dibatasi oleh pemerintah. Kegiatan belajar mengajar yang biasanya dilakukan dengan tatap muka dibatasi dan diubah menjadi daring (dalam jaringan). Perubahan tersebut tercantum pada Surat Edaran Nomor 4 Tahun 2020 Kementerian Pendidikan dan Kebudayaan mengenai pelaksanaan pendidikan dalam masa darurat COVID-19. Kementerian menganjurkan agar kegiatan belajar mengajar sekolah dilakukan secara daring (Kemdikbud, 2020). Sekolah hingga perguruan tinggi akhirnya menerapkan pembelajaran daring. Dalam masa pembelajaran daring, seluruh guru, dosen, siswa, dan mahasiswa diminta agar memanfaatkan teknologi dalam proses belajar mengajar. Berbagai perangkat lunak pembelajaran *e-learning* berkembang dan mulai digunakan oleh pendidik dan peserta didik.

Interaksi antara pendidik dan peserta didik di *e-learning* tercatat dan meninggalkan *digital footprint*. *Digital footprint* tersebut dapat berupa data mengenai diskusi antara pendidik dan peserta didik, nilai diskusi, nilai tugas, nilai ujian, kehadiran, frekuensi mengakses materi, dan sebagainya. Apabila *digital footprint* ini dihimpun akan menjadi sekumpulan data pendidikan yang akan sangat bermanfaat dalam

analisis pembelajaran (*learning analytics*). *Learning analytics* diartikan sebagai pengukuran, pengumpulan, analisis, dan pelaporan data terkait peserta didik dan konteksnya dengan tujuan untuk memahami dan mengoptimalkan pembelajaran dan lingkungan dimana hal itu terjadi (Long & Siemens, 2011). *Learning analytics* telah menjadi salah satu bidang teknologi pendidikan yang semakin mendapat perhatian dari para peneliti dan praktisi pendidikan di satu dekade terakhir (Hui & Kwok, 2019). Para peneliti dan praktisi pendidikan tertarik untuk mempelajari *learning analytics* dari dua buah aspek, yaitu aspek pedagogis (seperti kurikulum pembelajaran mandiri, proses pembelajaran) dan aspek teknologi (seperti *data mining*, teknik visualisasi data, sistem manajemen pembelajaran) (Lee dkk, 2020). Pemerintah, universitas, organisasi pengujian, dan penyelenggara *massive open online course* (MOOC) mengumpulkan data mengenai peserta didik dan cara belajarnya.

Dunia pendidikan memiliki data yang berlimpah dan berkesinambungan mengenai siswa dan alumni yang dihasilkannya. Sebagian besar data yang tersedia sayangnya belum dalam format yang rapi dan terorganisir dengan baik. Oleh karena itu, dibutuhkan keterampilan untuk mengakses dan mengubah data sehingga dapat dihasilkan informasi yang bermanfaat dalam membuat kebijakan atau keputusan dan berdasarkan hal ilmiah. *Learning analytics* akan melakukan penambangan data (*data mining*) pada data pendidikan yang disebut dengan *Educational Data Mining* (EDM) untuk mendapatkan informasi tersirat dari pembelajaran yang telah dilakukan. EDM fokus pada pengembangan dan penerapan berbagai metode untuk mengekstrak informasi penting pada data pendidikan (Avella, J.T, dkk, 2016). LA dan EDM akan membentuk ekosistem untuk mengumpulkan, memproses, melaporkan, dan mengerjakan data digital secara terus menerus untuk meningkatkan proses pendidikan (Aldowah dkk, 2019).

Penerapan *data mining* di dunia pendidikan masih dalam tahap awal dan perlu mendapat perhatian lebih. Proses penggunaan EDM dan LA untuk meningkatkan kualitas pengambilan keputusan telah menjadi tantangan bagi institusi pendidikan tinggi pada hari ini (Sacin dkk, 2009). Penelitian-penelitian mengenai *learning analytics* dan EDM di Indonesia masih sedikit. Padahal data pendidikan di Indonesia terus bertambah seiring bertambahnya kegiatan akademik dan non akademik khususnya pada masa pandemi ini. Berlimpahnya data pendidikan di Indonesia berpotensi untuk digali sehingga menghasilkan pola-pola baru dari data yang bermanfaat untuk efektivitas proses pembelajaran. Hal ini membuka peluang diterapkannya *learning analytics* dan EDM untuk pengelolaan pendidikan yang lebih baik dan efektif (Firdaus, 2017). Sudah saatnya perkembangan teknologi yang terjadi digunakan untuk menganalisis data pendidikan sehingga keputusan dan kebijakan yang diambil oleh institusi lebih terbuka dan berdasarkan data yang ada. Penggunaan *learning analytics* dan *educational data mining* yang sudah ada perlu dipaparkan sehingga dapat memberikan gambaran penggunaan data pendidikan untuk pengambilan keputusan dan pencegahan kegagalan peserta didik pada proses pembelajaran. Oleh karena itu, penelitian ini bertujuan untuk menjelaskan pengertian, penggunaan, dan dampak *Learning Analytics* dan *Educational Data Mining* pada data pendidikan.

## METODE

Metode penelitian ini menggunakan pendekatan kualitatif dan menerapkan metode penelitian studi kepustakaan. Penelitian ini bertujuan untuk memahami fenomena yang ada secara deskriptif dalam bentuk kata-kata dan bahasa tanpa ada pengujian hipotesis dan hubungan. Menurut Bogdan dan Taylor, penelitian kualitatif merupakan suatu penelitian yang digunakan untuk menghasilkan data deskriptif berupa kata-kata tertulis atau lisan dari orang-orang dan perilaku yang dapat diamati (Moleong, 2010). Sedangkan penelitian studi kepustakaan menurut Sugiyono dalam Milya Sari adalah kajian teoritis, referensi, serta literatur ilmiah lainnya yang berkaitan dengan budaya, nilai, dan norma yang berkembang pada situasi sosial yang diteliti (Sari, 2020). Literatur yang digunakan dalam kajian ini bersumber dari artikel, buku, dan jurnal *online* yang berkaitan dengan judul kajian. Adapun langkah-langkah yang dilakukan dalam kajian literatur ini di antaranya mengidentifikasi sumber-sumber yang relevan, mereview literatur, menulis literatur dan mengaplikasikan literatur pada kajian yang akan dilakukan (Rahayu dkk, 2019). Teknik analisis yang digunakan dalam penelitian ini adalah teknik analisis isi (*content analysis*). Teknik ini menggunakan pengolahan data deskriptif, yaitu dengan menjelaskan suatu permasalahan, gejala, kejadian sebagaimana adanya berdasarkan data yang ada dan

bukan untuk menguji kebenaran hipotesis.

## HASIL DAN PEMBAHASAN

### *Data Mining*

Kumpulan informasi dan teknologi pendidikan yang kompleks telah mendorong perguruan tinggi melakukan reformasi pendidikan, salah satunya berupa datafikasi pembelajaran (Mayer-Schonberger & Cukier, 2014). Penggabungan dan penganalisisan setiap *bit* dan *byte* data akan memunculkan potensi untuk mengungkapkan wawasan baru pada proses dan hasil pembelajaran siswa. Setiap siswa merupakan sumber data potensial. Menurut Kementerian Pendidikan dan Kebudayaan dalam Statistik Pendidikan 2021, ada 45,21 juta siswa di Indonesia pada tahun ajaran 2021/2022 yang terdiri dari 24,84 juta siswa sekolah dasar (SD); 10,09 juta siswa sekolah menengah pertama (SMP); 5,01 juta siswa sekolah menengah atas (SMA), dan 5,25 juta siswa sekolah menengah kejuruan (SMK) (Badan Pusat Statistik, 2021). Banyaknya peserta didik berpotensi pada berlimpahnya data yang terpendam untuk proyek *Big Data* dan *data mining*.

*Data mining* adalah proses yang menggunakan teknik statistik, matematika, *artificial intelligent*, dan *machine learning* untuk mencari pola atau informasi pada data terpilih (Mardi, 2017). *Data mining* merupakan salah satu langkah dalam *knowledge discovery in database* (KDD). Proses KDD secara garis besar terdiri dari pembersihan data (*data cleaning*), integrasi data (*data integration*), pemilihan data (*data selection*), transformasi data (*data transformation*), penggalian data (*data mining*), evaluasi pola (*pattern evaluation*), dan penyajian pengetahuan (*knowledge presentation*) (Ayub, 2007). Pembersihan data (*data cleaning*) menghilangkan *noise* dan data yang tidak relevan atau konsisten. Pada beberapa data pendidikan terdapat isian yang tidak sempurna seperti data yang hilang, tidak valid, atau salah ketik. Untuk mencegah terjadinya bias pada hasil KDD, maka data tidak sempurna tersebut dibuang. Langkah selanjutnya adalah integrasi data (*data integration*). Integrasi data adalah penggabungan berbagai data dari berbagai sumber ke dalam satu *database* baru. Penggabungan data dapat dilakukan pada fitur yang sama atau menambahkan fitur sehingga didapat informasi yang lebih besar. Selanjutnya adalah seleksi data (*data selection*). Dari tahap penggabungan data, tidak semua data digunakan. Seleksi data akan menyeleksi data yang sesuai untuk dianalisis sebelum tahap penggalian informasi dimulai. Selanjutnya dilakukan transformasi data (*data transformation*) untuk menyamakan format sesuai dengan metode penggalian data yang digunakan. Selanjutnya adalah penggalian data (*data mining*). Metode penggalian data sangat beragam bergantung pada tujuan dan proses yang ingin dilakukan dalam KDD secara keseluruhan. Setelah penggalian data dilakukan, pola atau informasi hasil ekstraksi dievaluasi (*pattern evaluation*). Hal ini bertujuan untuk menilai apakah pola atau informasi tersebut sudah sesuai dengan hipotesa atau belum. Berbagai perbaikan dapat dilakukan kembali pada proses *data mining* jika hasil yang diperoleh tidak sesuai dengan hipotesa. Terakhir adalah penyajian pengetahuan (*knowledge presentation*). Pola atau informasi hasil ekstraksi dari proses *data mining* perlu disajikan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

### *Educational Data Mining*

Proses analisis dari data-data yang diambil dari institusi pendidikan menggunakan teknik *data mining* untuk menghasilkan informasi yang berguna baik dalam penelitian maupun praktik-praktik dalam bidang pendidikan dikenal sebagai *educational data mining* (Blanco dkk, 2019). Pesatnya penggunaan teknologi dalam pendidikan membuat data pendidikan tersedia begitu banyak sehingga *Educational Data Mining* (EDM) mampu menyediakan sejumlah besar informasi yang relevan dan gambaran yang lebih jelas tentang peserta didik dan proses belajarnya. EDM menggunakan berbagai macam data mulai dari *intelligence computer tutor*, *classic computer-based educational systems*, kelas

online, data-data akademik yang ada di institusi pendidikan, dan data ujian yang terstandarkan (Romero dkk, 2010). Data tersebut dianalisis menggunakan teknik *Data Mining* (DM) untuk memecahkan masalah pendidikan. Sama halnya dengan teknik ekstraksi DM lainnya, proses pada EDM juga mengekstraksi hal-hal yang menarik, dapat ditafsirkan, dan bermanfaat dari data pendidikan. Hasil ekstraksi ini kemudian digunakan untuk menambah pengetahuan tentang fenomena pendidikan, siswa, dan sejumlah pengaturan agar proses belajar dan mengajar dapat berlangsung secara optimal. Dengan mengembangkan pendekatan komputasi yang menggabungkan data dan teori diyakini dapat membantu meningkatkan kualitas proses belajar mengajar. Dari sudut pandang praktis, EDM memungkinkan pengguna untuk mengekstrak pengetahuan dari data siswa. Pengetahuan ini dapat digunakan dalam cara yang berbeda seperti untuk memvalidasi dan mengevaluasi sistem pendidikan, meningkatkan kualitas proses pembelajaran, dan meletakkan landasan atau kurikulum bagi proses pembelajaran yang lebih efektif menyesuaikan dengan kebutuhan dan kemampuan siswa.

Ide serupa telah diterapkan dengan sukses terutama dalam bisnis data pada bidang yang berbeda, seperti pada sistem *e-commerce* untuk meningkatkan keuntungan penjualan karena menyesuaikan dengan kebutuhan dan kemampuan konsumen atau target pasarnya. Dengan demikian, keberhasilan penerapan teknik DM mendorong penerapannya di berbagai domain pengetahuan lainnya. Contoh lebih khususnya lagi adalah DM telah diterapkan ke data pendidikan untuk tujuan penelitian seperti meningkatkan proses belajar dan membimbing siswa untuk belajar atau memperoleh pemahaman yang lebih mendalam sesuai dengan pola belajar dan kemampuan siswa tersebut. Walaupun EDM membuat kemajuan yang relatif lebih sedikit jika dibandingkan penerapan DM pada bidang lain namun hal ini tetap dirasa berguna dalam dunia pendidikan.

Banyak tugas atau masalah dalam lingkungan pendidikan yang telah dikelola atau diselesaikan melalui EDM. Baker menyarankan empat bidang utama aplikasi EDM, yaitu meningkatkan model pembelajaran siswa, meningkatkan model domain atau bahan ajar, mempelajari dukungan pedagogis yang disediakan oleh perangkat lunak pembelajaran, dan melakukan penelitian ilmiah tentang pembelajaran dan peserta didik (Baker, 2009). Lima pendekatan/metode yang tersedia yaitu: prediksi, pengelompokan, melihat hubungan atau korelasi, pengolahan data untuk menilai kualitas siswanya, dan penemuan metode belajar baru dari model yang dibentuk. Sementara Castro mengelompokkan tugas EDM menjadi empat bidang yang berbeda, yaitu aplikasi yang berhubungan dengan penilaian kinerja belajar siswa, adaptasi kursus atau subjek dan rekomendasi pembelajaran untuk menyesuaikan pembelajaran siswa berdasarkan perilaku individu siswa, mengembangkan metode untuk mengevaluasi materi dalam kursus online, pendekatan yang menggunakan umpan balik dari siswa dan guru dalam kursus e-learning, dan model deteksi untuk mengungkap perilaku belajar siswa (Castro, 2007).

### *Learning Analytics*

Berbagai industri seperti kesehatan, perbankan, asuransi, penerbangan, hiburan dan telekomunikasi telah lama melihat keuntungan dalam memanfaatkan wawasan atau data historis yang dibawa oleh analisis data dalam skala besar (Kiron dkk, 2012). Dari jalur penerbangan yang dioptimalkan hingga asuransi kesehatan yang menggunakan model prediktif, penggunaan data skala besar (*big data*) telah menguntungkan industri dan dapat melihat perilaku konsumen. Sebaliknya, sektor pendidikan memiliki progress yang lambat, atau bisa dibilang yang paling berhati-hati dalam hal memanfaatkan beragam data yang telah dihasilkan dan dikumpulkan selama pembelajaran siswa. Kurangnya pemanfaatan data cukup mengejutkan, mengingat bahwa teknologi pendidikan seperti *learning system management* (LMS) sebenarnya sudah cukup mapan dan sebagian besar dianggap sebagai sumber daya inti untuk praktik pengajaran kontemporer.

Organisasi pendidikan baru-baru ini mulai terjun ke dunia analisis data dan pembelajaran mesin yang sangat dalam untuk menggunakan wawasan terdahulu atau data historis menjadi informasi baru untuk perbaikan kualitas pengajaran dan pengalaman belajar siswa. Pada awal 2011, sekelompok kecil

peneliti pendidikan menjadi tuan rumah Konferensi Pertama Analisis Pembelajaran Internasional (LAK'11) di Banff, Kanada. Salah satu tujuan utama dari pertemuan ini adalah untuk mendefinisikan dan membatasi ruang lingkup fokus penelitian yang muncul tentang memahami pembelajaran siswa melalui penggunaan pembelajaran mesin, *data mining* dan metode visualisasi data. Hasil dari konferensi ini termasuk pembentukan *Society for Learning Analytics Research* (Solar) dan mendefinisikan *learning analytics* atau analisis pembelajaran sebagai "pengukuran, pengumpulan, analisis, dan pelaporan data tentang peserta didik dan konteksnya bertujuan untuk memahami dan mengoptimalkan pembelajaran dan lingkungan di mana hal itu terjadi" (Siemens dkk, 2011). Dari pertemuan kecil tersebut di lapangan telah terjadi hal yang besar dan dramatis dalam minat terhadap *learning analytics* atau analisis pembelajaran. Hal tersebut dapat dilihat melalui pendanaan penelitian, publikasi dan komersialisasi terkait teknologi yang diterapkan pada data pendidikan. Ketertarikan dalam mempelajari *learning analytics* berasal dari bidang yang terkoneksi ke penggunaan teknologi dalam pendidikan karena mereka juga merasakan manfaat dalam mengatasi tantangan yang sering dikaitkan dengan praktek mengajar kontemporer. Misalnya, analisis pembelajaran dapat membantu dalam menyediakan penilaian dan umpan balik yang dipersonalisasi dan tepat waktu dalam skala kecil hingga ukuran kelas yang besar.

Lembaga pendidikan modern dituntut untuk menyeimbangkan perannya selain untuk memenuhi kepentingan publik juga harus mampu memenuhi kebutuhan untuk tetap layak secara finansial. Jika lembaga pendidikan tidak menguntungkan tentu akan berpengaruh terhadap keberlangsungan lembaga tersebut. Pekerjaan awal dalam analisis pembelajaran terlihat memberikan solusi untuk menyeimbangkan hal tersebut. Meningkatnya keberagaman dan permintaan siswa dalam pembelajaran sementara pendanaan dari pemerintah malah mengalami pengurangan membutuhkan model pendidikan yang lebih hemat biaya (efisien dan efektif) namun tetap mampu mempertahankan kualitas pengajaran yang tinggi. Secara sederhana situasi ini mengakibatkan peningkatan ukuran kelas dan adopsi teknologi pada pembelajaran yang makin meluas untuk mempromosikan akses yang lebih fleksibel ke dunia pendidikan.

*Learning analytics* menggunakan data pembelajaran siswa yang tersedia untuk menetapkan indikator awal siswa dan mengelompokkan siswa berdasarkan pola belajar dan prestasi akademiknya. Walaupun analisis pembelajaran dianggap sebagai bidang penelitian baru, konsep menganalisis data tentang peserta didik dan pola belajar mereka bukanlah hal baru bagi dunia pendidikan. Pemanfaatan berbagai bentuk teknologi pembelajaran dan data pembelajaran sudah diketahui dan dilakukan pada penelitian pendidikan jauh sebelum munculnya istilah *learning analytics*. Banyak teknik dan metode penambangan data yang sekarang umum digunakan dalam pembelajaran penelitian analitik, seperti jaringan sosial atau analisis wacana yang telah lama ada pada sejarah di luar bidang pendidikan (Baker & Yacef, 2009). Namun, pembentukan *learning analytics* sebagai bidang diskrit telah berfungsi sebagai katalis untuk penggabungan beberapa domain penelitian, metode dan teori pembelajaran untuk memberikan informasi atau pandangan yang baru, kesempatan penelitian untuk memahami proses pembelajaran. Sebagai contoh, *learning analytics* berpotensi untuk meningkatkan pengalaman belajar siswa melalui umpan balik yang sesuai dan dukungan pada fase yang tepat dari pengajar.

### ***Artificial Intelligence***

*Artificial intelligence* (AI) di bidang pendidikan seringkali digunakan untuk *Educational Data Mining* (EDM) dan *Learning analytics* (LA) (Baek & Doleck, 2020). Istilah kecerdasan buatan atau *artificial intelligence* (AI) pertama kali dicetuskan oleh John McCarthy pada tahun 1956 ketika ia mengadakan konferensi akademik pertama tentang topik tersebut. Tetapi perjalanan untuk memahami apakah mesin benar-benar dapat berpikir telah dimulai jauh sebelum itu. Di dalam tulisan karya Vannevar Bush "As We May Think" ia mengusulkan sebuah sistem yang memperkuat pengetahuan dan pemahaman orang lain (Bush, 1945). Lima tahun kemudian Alan Turing menulis makalah tentang

gagasan bahwa mesin mampu mensimulasikan manusia dan kemampuan untuk melakukan hal-hal cerdas, seperti bermain catur (Turing, 1950)

Tidak ada yang bisa menyangkal kemampuan komputer untuk memproses logika. Tetapi banyak orang masih meragukan apakah sebuah mesin benar-benar dapat berpikir. Sebagai contoh ada hal yang disebut dengan argumen 'ruang Cina'. Bayangkan beberapa orang terkunci di sebuah ruangan dimana mereka diberikan catatan dalam bahasa Cina. Mereka boleh menggunakan seluruh buku di perpustakaan, aturan, dan tabel pencarian sehingga mereka akan dapat menghasilkan tanggapan yang valid dari catatan dalam bahasa Cina tersebut. Tapi apakah mereka benar-benar 'mengerti' bahasanya? Argumen tersebut dapat juga diterapkan pada perdebatan apakah komputer dapat berpikir atau tidak karena komputer akan selalu menerapkan pencarian fakta dan hafalan, mereka tidak akan pernah bisa 'mengerti' subjek. Argumen ini telah dibantah dengan berbagai cara oleh para peneliti, tetapi hal itu juga dapat merusak kepercayaan masyarakat terhadap mesin. Sebagian masyarakat takut jika suatu saat mesin akan menguasai manusia jika memang benar mesin dapat berpikir.

Kecerdasan buatan atau *artificial intelligence* (AI) telah dipelajari selama beberapa dekade dan masih merupakan salah satu subjek yang paling sulit dipahami di bidang komputer sains. Hal ini karena betapa besar dan samar subjek yang dibahas dan dipelajarinya. Subjek AI dapat membahas tentang mesin yang mampu melakukan hal sederhana hingga yang benar-benar mampu berpikir untuk mencari algoritma yang dapat digunakan untuk memainkan *board games*. Kemampuan mesin ini dapat diaplikasikan di hampir setiap cara atau bidang dimana kita menggunakan alat bantu komputer.

Penelitian dan penggunaan AI di bidang pendidikan terhitung masih cukup baru. Salah satu contoh penggunaan AI di bidang pendidikan adalah seperti yang sudah dilakukan oleh Da Yang Tan dan Chin Wei Cheah pada tahun 2021 di Singapura (Tan & Cheah, 2021). Mereka membuat aplikasi pembelajaran *online* menggunakan AI untuk subjek fisika. Aplikasi tersebut disajikan dengan tampilan menarik menggunakan animasi seperti sebuah *game*. Selain berperan sebagai tutor pribadi untuk siswa, aplikasi tersebut juga dapat menentukan level kompetensi siswa melalui serangkaian evaluasi. Gambaran level kompetensi tadi diharapkan dapat memberikan informasi lebih untuk guru yang akan mengajar fisika di kelas tersebut. Guru diharapkan dapat membuat rencana pengajaran menjadi lebih optimal atau dapat juga merancang kurikulum yang menyesuaikan kebutuhan dan kemampuan peserta didiknya.

Baik EDM maupun LA sama-sama menggunakan *machine learning* untuk kepentingan penelitiannya. Banyak ulasan yang membahas perkembangan EDM dan LA yang cukup pesat ini seiring dengan perkembangan kekuatan komputer dan keterbukaan data yang terjadi pada dekade terakhir ini. (Lemay, Baek, & Doleck, 2021). Komputer pada masa kini sudah mampu untuk memproses data yang sedemikian banyak dan juga didukung dengan hadirnya era keterbukaan informasi membuat penelitian terkait EDM dan LA meningkat pesat. Masih menurut Lemay, dkk., dalam lima tahun terakhir ini, penelitian terkait EDM dan LA mengerucut pada penelitian tentang perilaku dan kompetensi siswa atau peserta didik. EDM yang dilakukan dapat digunakan untuk mengumpulkan data mengenai frekuensi belajar (akses materi) dan nilai ujian siswa. Sementara LA kemudian dapat digunakan untuk melihat korelasi atau implikasi apakah benar siswa yang frekuensi belajarnya lebih banyak akan memperoleh nilai yang lebih baik. Konsep *machine learning* yang sering digunakan untuk EDM dan LA diatas adalah klasifikasi atau regresi.

### ***Implementasi Machine Learning pada data pendidikan***

*Machine learning* adalah bidang ilmu yang mengembangkan algoritma atau model yang dapat belajar dari data untuk menggali pengetahuan yang ada pada data tersebut seperti proses belajar pada manusia. *Machine learning* dapat membantu manusia untuk mengelola berbagai data khususnya data yang besar, kompleks, dan butuh respon cepat. Berdasarkan input yang diberikan pada data pelatihan, *machine learning* dikelompokkan menjadi dua jenis yaitu *supervised* dan *unsupervised* (Engelbrecht,

2007). Pada *supervised learning*, data latih disertai dengan nilai target, sehingga masing-masing data latih berupa pasangan data, misal  $\{x_i, t_i\}, i = 1, 2, \dots, N$  dengan  $x$  adalah data latih dan  $t$  adalah nilai target. Tujuan *supervised learning* adalah membangun suatu algoritma atau model yang dapat menghasilkan output yang paling sesuai dengan nilai target ( $t_i$ ) untuk semua data latih ( $x_i$ ). *Supervised learning* biasanya digunakan dalam klasifikasi, aproksimasi, kontrol, pemodelan dan identifikasi, pemrosesan sinyal dan optimisasi. Beberapa metode supervised learning di antaranya linear regression, ridge regression, polynomial regression, perceptron, logistic regression (Du & Swamy, 2014).

*Supervised learning* pada bidang pendidikan dapat diwujudkan salah satunya sebagai peramalan kesuksesan siswa setelah menjalani sebagian proses pembelajaran. Misalnya pada suatu subjek pembelajaran tertentu sudah diperoleh data mengenai frekuensi akses materi, nilai kuis, nilai pre test, nilai post test, nilai tugas dan nilai ujian selama beberapa tahun terakhir. Hasil akhir lulus atau tidaknya siswa dapat dijadikan sebagai nilai target sehingga jika mesin atau komputer dilatih dengan input data-data tersebut maka mesin akan dapat meramalkan siswa dengan profil tertentu dapat lulus atau tidak pada subjek tersebut. Hal ini tentunya akan sangat bermanfaat untuk guru yang ingin merubah strategi atau pola pengajaran dan sebagainya setelah mengetahui hasil peramalan mesin terhadap performa peserta didiknya.

Sementara pada *unsupervised learning*, data latih tidak disertai nilai target, misal  $\{x_i\}, i = 1, 2, \dots, N$  dengan  $x$  adalah data latih. Tujuan *unsupervised learning* adalah membangun algoritma atau model yang dapat mendeskripsikan struktur tersembunyi pada data. *Unsupervised learning* dikenal juga sebagai model variabel tersembunyi (*latent variable models*). *Unsupervised learning* biasanya digunakan untuk *clustering*, *vector quantization*, *feature extraction*, *signal coding*, dan *data analysis*. Beberapa metode *unsupervised learning* di antaranya k-means, fuzzy c-means, *principal component analysis*, dan sebagainya (Du & Swamy, 2014).

*Unsupervised learning* pada bidang pendidikan dapat diwujudkan salah satunya sebagai pengelompokan siswa ke dalam kelompok-kelompok yang memiliki profil mirip. Salah satu skenario yang dapat dilakukan adalah setelah menjalani setengah proses pembelajaran pada suatu subjek tertentu, guru menginput data mengenai frekuensi akses materi, nilai kuis, nilai pre test, nilai post test, nilai tugas dan nilai ujian untuk dipelajari oleh mesin. Mesin kemudian akan dapat membentuk kelompok-kelompok siswa yang memiliki kompetensi atau perilaku yang hampir seragam. Guru kemudian dapat menentukan jenis *treatment* berbeda yang mungkin bisa diterapkan pada kelompok tersebut. Kelompok yang dinilai cukup *advance* kompetensinya dapat diberikan target penelitian lanjutan sementara kelompok yang dianggap masih kurang dapat diberikan tambahan jam pelajaran untuk membahas materi-materi yang tertinggal dengan lebih intens dan menyesuaikan kecakapan peserta didik.

### ***Manfaat EDM dan Learning Analytics pada Pendidikan***

Kemampuan untuk melakukan prediksi kinerja peserta didik dan identifikasi kesulitan yang dihadapi merupakan kegiatan yang cukup signifikan dan bermanfaat baik bagi peserta didik, pendidik, maupun institusi pendidikan. Berbekal kemampuan tersebut, selanjutnya dapat dilakukan langkah-langkah perbaikan, seperti misalnya melakukan intervensi melalui bimbingan tambahan atau dapat dijadikan sebagai indikator peringatan dini [Campbell dkk, 2007]. Jika dimanfaatkan secara efektif, hal tersebut juga dapat membantu institusi pendidikan untuk meningkatkan proses pembelajaran dan mengembangkan strategi yang tepat sehingga dapat menurunkan tingkat *dropout* (B. Daniel, 2015). Namun, pengukuran kinerja akademis seorang murid juga dihadapkan pada tantangan bahwa hal tersebut sangat mungkin dipengaruhi oleh faktor yang berbeda-beda. Keterkaitan antara variabel dan faktor-faktor untuk melakukan prediksi kinerja murid bersifat nonlinier. Konsep machine learning mungkin tidak sepenuhnya dapat diaplikasikan secara langsung pada jenis data dan permasalahan tersebut [B. Guo dkk, 2015].

Sejauh ini walaupun penelitian terkait EDM dan LA sudah meningkat cukup pesat jumlahnya namun penerapan nyata dan langkah-langkah perbaikan berdasarkan ramalan atau saran dari mesin (komputer) masih terlihat jarang dilakukan. Menurut Kaufmann para guru atau pengajar cenderung

lebih percaya saran dan masukan dari atasan atau pengawasnya daripada saran yang diberikan oleh mesin atau komputer. Dibutuhkan Kerjasama dan kesepahaman dari peneliti data pendidikan dan instansi pendidikan yang diteliti untuk mulai bersama menggunakan hasil penelitian selama hasilnya digunakan untuk kemajuan dan kebaikan semua pihak.

## PENUTUP

### Kesimpulan

Salah satu dampak dari kemajuan teknologi adalah semakin melimpahnya *digital footprint* pada dunia pendidikan. Dengan melimpahnya *digital footprint*, penggalian data semakin mungkin dilakukan. *Educational data mining* (EDM) adalah implementasi teknik penggalian data yang digunakan untuk menganalisis dan menyelesaikan isu-isu dalam dunia pendidikan. EDM dipergunakan untuk berbagai keperluan seperti deteksi kegagalan atau keberhasilan seorang peserta didik, analisa gaya belajar siswa, kajian efektivitas bahan ajar, prediksi tren kinerja peserta didik, simulasi pengambilan keputusan, identifikasi peserta didik yang diprediksi akan mengalami *dropout* dan lain-lain. Dari informasi yang didapat EDM dilakukan *Learning Analytics* untuk menganalisis langkah perbaikan dan intervensi yang dapat dilakukan. Jika langkah perbaikan dan intervensi yang dilakukan efektif, hal tersebut dapat membantu institusi pendidikan untuk meningkatkan kualitas pembelajaran yang dilakukannya dan mencegah kegagalan pada peserta didik sehingga secara tidak langsung menurunkan angka *dropout* peserta didik.

## REFERENSI

- A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, dan B. Navarro Colorado (2019) "A Systematic Review of Deep Learning Approaches to Educational Data Mining," *Complexity*, Vol. 1, hal. 1-22
- Aldowah, H., Al-Samarraie, H. & Fauzy, W.M. (2019). *Educational data mining and learning analytics for 21<sup>st</sup> century higher education: A review and synthesis. Telematics and Informatics*, 37, 13-49
- Avella, J.T., Kebritchi, M., Nunn, S. & Kanai, T. (2016). *Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. Online Learning*, 20 (2), 13-29
- Ayub, Mewati. (2007). *Proses data mining dalam pembelajaran berbantuan komputer. Jurnal Sistem Informasi Universitas Kristen Maranatha* Vol. 2(1)
- B. Daniel. (2015). "Big Data and Analytics in Higher Education: Opportunities and Challenges," *British J. of Educ. Technol.*, Vol. 46, No. 5, hal. 904- 920
- B. Guo, R. Zhang, G. Xu, C. Shi, dan L. Yang. (2015) "Predicting Students Performance in Educational Data Mining," 2015 Int. Symp. on Educ. Technol. (ISET), hal. 125-128
- Badan Pusat Statistik. (2021). *Statistik Pendidikan 2021*. Jakarta: Badan Pusat Statistik, 13
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17
- Bush, Vannevar. 1945. *As We May Think*. The Atlantic Monthly. July 1945
- C. Romero, S. Ventura, M. Pechenizkiy, & R.S. Baker (2010). *Handbook of Educational Data Mining*, Boca Raton, USA: CRC Press.
- Du, K.-L., & Swamy, M. (2014). *Neural Networks and Statistical Learning*. Verlag: Springer
- Engelbrecht, A. P. (2007). *Introduction to Computational Intelligence 2nd Edition*. Chichester: John Wiley & Sons Ltd.
- F. Castro, A. Vellido, A. Nebot, and F. Mugica (2007), "Applying data mining techniques to e-learning problems," in *Evolution of teaching and learning paradigms in intelligent environment*, pp. 183–221, Springer.
- Firdaus, D. (2017) "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer". *Jurnal Format* Volume 6 Nomor 2 ISSN 2089-5615
- Hui, Y.K., & Kwok, L.F (2019). A review on learning analytics. *International Journal of Innovation and Learning* 25(2), hal. 197-222.

- J.P. Campbell, P.B. DeBlois, & D.G. Oblinger (2007), "Academic Analytics: A New Tool for a New Era," *EDUCAUSE Review*, Vol. 42, No. 4, hal. 40-57
- Kementerian Pendidikan dan Kebudayaan. (2020, Maret 24). Mendikbud Terbitkan SE tentang Pelaksanaan Pendidikan dalam Masa Darurat Covid-19. Retrieved from [Kemdikbud.go.id: https://www.kemdikbud.go.id/main/blog/2020/03/mendikbud-terbitkan-se-tentang-pelaksanaan-pendidikan-dalam-masa-darurat-covid19](https://www.kemdikbud.go.id/main/blog/2020/03/mendikbud-terbitkan-se-tentang-pelaksanaan-pendidikan-dalam-masa-darurat-covid19)
- Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M. (2012). Analytics: The widening divide. *MIT Sloan Management Review*, 53(2), 1.
- Lamay, D.J, Baek, C., Doleck, T. (2021). "Comparison of Learning Analytics and Educational Data Mining: A Topic of Modeling Approach". *Computers and Education: Artificial Intelligence 2* 100016
- Lee, LK., Cheung, S.K.S. & Kwok, LF (2020). *Learning analytics: current trends and innovative practices. J. Comput. Educ.* 7, 1–6.. <https://doi.org/10.1007/s40692-020-00155-8>
- Lexy J. Moleong (2010). Metode Penelitian Kualitatif. Bandung: Remaja Rosdakarya, hal. 6
- Long, P. & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 31-40.
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4. 5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, 2(2), 213-219.
- Mayer-Schonberger, V., & Cukier, K. (2014). *Learning with big data: The future of education*. New York: Houghton Mifflin Harcourt.
- Kementerian Pendidikan dan Kebudayaan. (2020, Maret 24). Mendikbud Terbitkan SE tentang Pelaksanaan Pendidikan dalam Masa Darurat Covid-19. Retrieved from [Kemdikbud.go.id: https://www.kemdikbud.go.id/main/blog/2020/03/mendikbud-terbitkan-se-tentang-pelaksanaan-pendidikan-dalam-masa-darurat-covid19](https://www.kemdikbud.go.id/main/blog/2020/03/mendikbud-terbitkan-se-tentang-pelaksanaan-pendidikan-dalam-masa-darurat-covid19)
- Rahayu, T., Syafril, S., Wekke, I.S., Erlinda, R. (2019). "Teknik Menulis Review Literatur dalam Sebuah Artikel Ilmiah". <https://doi.org/10.31227/osf.io/z6m2y>.
- R. S. Baker & K. Yacef (2009) "The state of educational data mining in 2009: A review and future visions," *JEDM-Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17.
- Sacin, C.V., Agapito, J.B, Shafti, L., Ortigosa, A. (2009). *Recommendation in higher education using data mining techniques*. In: Paper presented at the Educational Data Mining.
- Sari, M., Asmendri. (2020). Penelitian Kepustakaan (*Library Research*) dalam Penelitian Pendidikan IPA. *Natural Science: Jurnal Penelitian Bidang IPA dan Pendidikan IPA* 6(1), 41-53.
- Siemens, G., Long, P., Gašević, D., & Conole, G. (2011). Call for Papers, 1st International Conference Learning Analytics & Knowledge (LAK 2011). Retrieved from <https://tekri.athabascau.ca/analytics/call-papers>
- Tan, D.Y & Cheah, C.W. Developing a gamified AI-enabled online learning application to improve students' perception of university physics. *Computers and Education: Artificial Intelligence 2* (2021) 10032
- Turing, Alan. 1950. Computing Machinery and Intelligence. *Mind* 49, 433 – 460