

Kajian Penggunaan Data Log Mahasiswa untuk Berbagai Permasalahan Analisis Pembelajaran

(A Review: Usage of Student Log Data for Several Learning Analytics Problems)

Sri Suning Kusumawardani^{1*}, Syukron Abu Ishaq Alfarozi²

Abstract—An online learning system is a very crucial thing nowadays to prevent the spread of COVID-19 virus. However, this system is very difficult to maintain student motivation and engagement because there is no direct interaction between teacher and student. This study reviewed the use of student log data for the needs of learning analytics to predict student performance or drop-out trends from a course by looking at the student interaction log data with the system and student demographic data using open data, namely the Open University Learning Analytics Dataset (OULAD). From reviews of several research articles that refer to these data, we can see: 1) the common problems, i.e., prediction of drop-out student, prediction of student performance and engagement; 2) the features used during modeling, i.e., demographics and interactions, either summarized daily or weekly with various feature representations; 3) learning analysis methods that use machine learning algorithm, i.e., Decision Tree (DT), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM). This paper also discusses the risk mitigation process of students, planning and designing data systems that support learning analytics, and problems that are often encountered during the modeling process.

Intisari—Pada saat ini, penyelenggaraan sistem pembelajaran daring menjadi hal yang penting di tengah pandemi untuk menekan persebaran virus COVID-19. Namun, sistem ini sangat sulit menjaga motivasi dan tingkat keterlibatan mahasiswa karena tidak ada interaksi langsung antara pengajar dengan mahasiswa. Makalah ini meninjau penggunaan data log mahasiswa untuk kebutuhan analisis pembelajaran guna memprediksi kinerja atau kecenderungan *drop-out* mahasiswa dari suatu mata kuliah dengan melihat pada data log interaksi mahasiswa dengan sistem dan data demografis mahasiswa menggunakan suatu data terbuka, yaitu *Open University Learning Analytics Dataset* (OULAD). Dari tinjauan beberapa artikel penelitian yang merujuk pada *dataset* tersebut, ada beberapa hal yang perlu ditinjau: 1) permasalahan yang sering diangkat, yaitu prediksi kecenderungan gagal dari mata kuliah tertentu, prediksi kinerja, dan prediksi keterlibatan mahasiswa; 2) fitur yang digunakan pada saat pemodelan, yaitu fitur demografis dan interaksi, baik yang diringkas secara harian atau mingguan dengan berbagai representasi fitur; 3) metode analisis pembelajaran yang secara khusus menggunakan metode pembelajaran mesin yang sering digunakan, yaitu *Decision Tree* (DT), *Support Vector Machine* (SVM), *Multi-Layer Perceptron*

(MLP), dan *Long Short-Term Memory* (LSTM). Makalah ini juga mendiskusikan proses mitigasi dari mahasiswa yang berisiko, perancangan sistem data yang mendukung analisis pembelajaran, dan permasalahan yang sering ditemui pada saat proses pemodelan.

Kata Kunci— *Learning Analytics*, MOOCs, Pembelajaran Daring, Data Log, Prediksi.

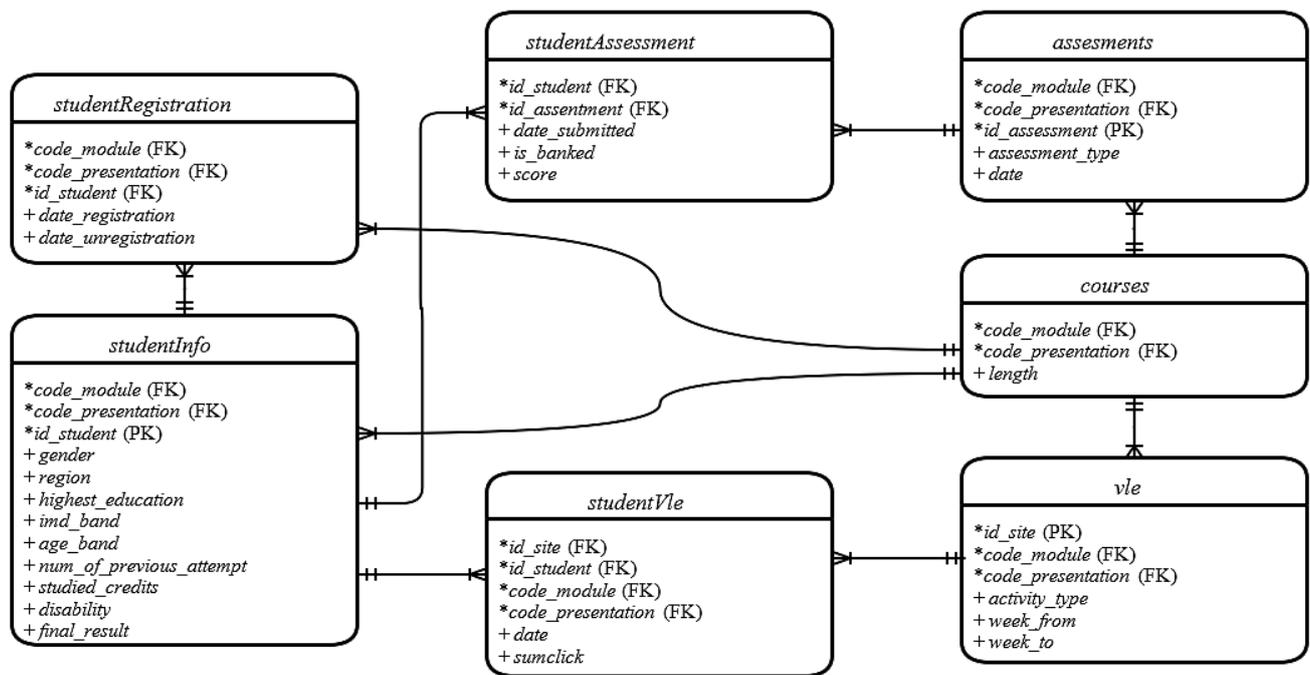
I. PENDAHULUAN

Perkembangan teknologi informasi memberikan dampak yang sangat baik terhadap dunia pendidikan dalam upaya membangun sistem pendidikan berkelanjutan. Dengan penggunaan teknologi, sistem pendidikan dapat diakses tanpa batasan geografis dan waktu untuk semua orang, salah satunya dengan suatu sistem pembelajaran masif, secara daring, dan terbuka, atau yang dikenal dengan *Massive Open Online Courses* (MOOCs). Saat ini, di tengah pandemi COVID-19, hampir semua institusi, terutama institusi pendidikan, menjalankan sistem pembelajaran daring.

Surat Edaran Kementerian Pendidikan dan Kebudayaan Indonesia Nomor 15 Tahun 2020 tentang pedoman penyelenggaraan belajar dari rumah dalam masa darurat penyebaran *coronavirus disease* (COVID-19) mengharuskan para mahasiswa untuk melakukan belajar secara daring dalam upaya menekan angka persebaran COVID-19 ini. Namun, tidak adanya tatap muka secara langsung menyebabkan sulitnya memonitor kinerja mahasiswa, menjaga motivasi belajar, dan interaksi antara pengajar dan mahasiswa. Dalam hal ini, mahasiswa lebih banyak berhubungan dengan sistem pengelolaan pembelajaran (*e-Learning*) daripada dengan pengajar.

Analisis pembelajaran, yang menjadi bagian sistem *e-Learning*, merupakan suatu metode untuk mencari faktor-faktor yang berhubungan dengan kinerja belajar mahasiswa, dalam hal ini bisa berupa suatu alat prediksi atau alat analisis faktor. Pentingnya analisis pembelajaran untuk meningkatkan kinerja belajar atau memprediksi kecenderungan gagal (*drop-out*) merupakan hal yang krusial dalam kasus pembelajaran daring. Dalam hal *e-Learning*, faktor-faktor tersebut hanya bisa didapatkan dari data personal dan aktivitas log mahasiswa dengan sistem. Data log merupakan suatu data yang diperoleh dari sistem yang merekam segala aktivitas yang dilakukan oleh *user*. Pada kasus analisis pembelajaran, data interaksi mahasiswa dengan sistem mempunyai potensi dimanfaatkan untuk mengukur tingkat keterlibatan atau motivasi belajar dari mahasiswa.

^{1*,2} Departemen Teknik Elektro dan Teknologi Informasi Fakultas Teknik Universitas Gadjah Mada, Jl. Grafika No.2, Yogyakarta, INDONESIA 55281 (email: ^{1*}suning@ugm.ac.id, ²syukron@outlook.com)



Gbr. 1 Entity relationship diagram pada OULAD [1] *) keys, +) atribut.

Pada makalah ini, ditinjau beberapa artikel yang berhubungan dengan analisis pembelajaran yang secara spesifik mengolah *Open University Learning Analytics Dataset* (OULAD) [1]. Berbeda dengan penelitian yang membahas MOOCs secara umum [2], makalah ini fokus pada perkembangan riset pada OULAD sehingga diharapkan dapat menjadi rujukan penelitian selanjutnya tentang analisis pembelajaran untuk mengangkat permasalahan yang belum pernah diangkat. OULAD merupakan suatu data terbuka yang diberikan oleh *Open University* (OU) dan memiliki kelengkapan yang cukup komprehensif dengan fitur demografis dan fitur interaksi harian dari 22 modul mata kuliah dari Ilmu Sosial, Ilmu Alam, Matematika, dan Teknik, dengan 32.593 mahasiswa dan jutaan interaksi (10.644.280 data). Tinjauan ini dimaksudkan untuk menganalisis penggunaan data log mahasiswa untuk kebutuhan analisis pembelajaran, baik dari segi permasalahan, fitur-fitur yang digunakan, maupun metode yang digunakan. Metode-metode yang ditinjau dan dianalisis menggunakan pembelajaran mesin yang sampai saat ini merupakan metode yang sangat sering digunakan, dilihat dari kinerja dan tingkat fleksibilitasnya.

Pada makalah ini diajukan beberapa pertanyaan penelitian, antara lain sebagai berikut.

- P1:** Apa saja permasalahan dalam konteks analisis pembelajaran yang diangkat dalam pengolahan OULAD?
P2: Apa saja fitur-fitur yang digunakan pada OULAD untuk analisis pembelajaran?
P3: Apa saja metode atau model pembelajaran mesin yang digunakan untuk OULAD?

Dari pertanyaan tersebut, juga didiskusikan panduan untuk merancang suatu sistem data yang mendukung metode analisis pembelajaran dengan mempertimbangkan hal-hal yang telah ada pada OULAD, cara mitigasi pembelajaran untuk

meningkatkan kinerja yang seharusnya dilakukan, dan beberapa permasalahan terhadap model dan data yang sering dijumpai pada tahap pemodelan.

II. DATA TERBUKA DAN OULAD

Pada bagian ini dibahas tentang data terbuka (*open data*) untuk keperluan analisis pembelajaran, berbagai contoh data terbuka, dan alasan *OULA dataset* (OULAD) dipilih sebagai bahan penelitian pada makalah ini.

A. Data Terbuka untuk Analisis Pembelajaran

Data terbuka merupakan suatu basis data yang tersedia untuk publik sedemikian hingga dapat diolah dan diekstrak informasi dan pengetahuannya. Berbagai data terbuka untuk analisis pembelajaran yang tersedia untuk umum antara lain OULAD, HarvardX, MITx, dan Coursera Forum. Dari berbagai data terbuka tersebut, OULAD memiliki kelengkapan fitur yang lebih baik dari yang lainnya, yaitu data demografis, data tugas, dan data interaksi harian dengan lingkungan pembelajaran virtual (*Virtual Learning Environment, VLE*).

Dalam hal analisis pembelajaran, data terbuka sangatlah penting untuk mengetahui perkembangan analisis pembelajaran, terutama dalam bidang pembelajaran mesin. Selain dapat diperuntukkan sebagai standar acuan kinerja (*benchmarking*) dari metode yang digunakan, juga sebagai panduan desain basis data karena sifatnya yang terbuka (semua dapat mengakses dan mengolahnya). Lebih lanjut lagi, struktur data tersebut dapat diterapkan pada data institusi yang tertutup dengan menggunakan teknologi analisis pembelajaran yang sama, seperti halnya yang telah diterapkan pada data terbuka.

OULAD didesain untuk kebutuhan analisis pembelajaran, baik dari kelengkapan fiturnya maupun desain basis datanya, seperti yang diperlihatkan pada Gbr. 1. *Dataset* lain yang juga

diperuntukkan bagi analisis pembelajaran yaitu *dataset KDD Cup 2010* dan *KDD Cup 2015*. Namun, meskipun sama-sama menyajikan data interaksi dengan VLE, berbeda dengan OULAD tersebut, *dataset* yang lain tidak dilengkapi dengan fitur demografis dan fitur historis mahasiswa terhadap mata kuliah yang pernah diambil sebelumnya.

Pada makalah ini digunakan OULAD untuk kebutuhan kajian permasalahan di bidang analisis pembelajaran yang telah dilakukan di banyak penelitian untuk memberikan suatu gambaran umum hal-hal yang dapat dilakukan dengan sebuah data analisis pembelajaran. OULAD dipilih berdasarkan struktur data dan kelengkapan fiturnya dibandingkan dengan *dataset* yang lain. Selain itu, OULAD diambil dari suatu mata kuliah yang sesungguhnya yang diselenggarakan oleh OU sehingga dapat memberikan gambaran yang nyata dari suatu data di suatu lembaga pendidikan. Selain itu, setiap data yang ada di OULAD memiliki empat label hasil akhir, yaitu *withdrawn*, *pass*, *fail*, dan *distinction*. Seperti yang disampaikan pada sebuah penelitian, faktor-faktor yang memengaruhi capaian pembelajaran (*learning outcome*) antara lain yaitu tingkat keterlibatan (*engagement*), pencapaian (*achievement*), dan tingkat undur diri (*attrition*) [3]. Dengan demikian, label hasil akhir tersebut bisa menentukan faktor pencapaian dan tingkat undur diri yang dapat digunakan untuk menentukan capaian pembelajaran tercapai atau tidak, sementara untuk tingkat keterlibatan dapat dilihat dari data interaksi mahasiswa, baik dengan sistem (VLE) ataupun tugas. Secara umum, capaian pembelajaran menurut model Biggs [4] terdiri atas capaian kuantitatif (keahlian dan fakta-fakta pembelajaran), kualitatif (struktur dan proses transfer pembelajaran), dan afektif (keterlibatan atau penerapan keahlian yang didapatkan). Di sini yang dibahas hanya faktor-faktor penunjang capaian pembelajaran tersebut, dengan menggunakan analisis pembelajaran karena untuk capaian pembelajaran itu sendiri tentunya sangatlah berbeda untuk setiap modul mata kuliah.

B. OULA Dataset (OULAD)

OULAD merupakan suatu basis data terbuka dari OU yang menyimpan data log dari mahasiswa secara lengkap, yaitu data aktivitas interaksi dan juga data demografis [1]. Data tersebut didapat dari interaksi mahasiswa dengan sistem pembelajaran jarak jauh dari tahun 2013 sampai 2014. Setiap mata kuliah di OU disebut dengan modul yang dapat dipresentasikan di beberapa waktu yang berbeda. Periode presentasi dikodekan dengan tahun dan bulan mulainya. Sebagai contoh, 2013B yang berarti dimulai pada tahun 2013 pada bulan Februari. Sementara kode modul menentukan mata kuliah yang akan diberikan dengan lingkup/domain Ilmu Sosial, MIPA, dan Teknik (STEM), seperti yang ditunjukkan pada Tabel I. Dalam Tabel I tersebut, setiap modul dipresentasikan beberapa kali (dua sampai empat kali pada tahun 2013-2014). Dalam hal ini, presentasi yang diberikan di dalam OULAD hanya B (Februari) dan J (Oktober) dengan tujuh mata kuliah (kode: AAA-GGG).

Secara singkat, OULAD¹ terdiri atas tujuh tabel data, seperti yang ditunjukkan pada Gbr. 1, yang disimpan dalam format *Comma Separated Values* (CSV). Tabel data tersebut dapat

TABEL I
KODE MODUL BERDASARKAN DOMAIN

Kode	Domain	Presentasi	Mahasiswa
AAA	Ilmu Sosial	2	748
BBB	Ilmu Sosial	4	7.909
CCC	MIPA dan Teknik (STEM)	2	4.434
DDD	MIPA dan Teknik (STEM)	4	6.272
EEE	MIPA dan Teknik (STEM)	3	2.934
FFF	MIPA dan Teknik (STEM)	4	7.762
GGG	Ilmu Sosial	3	2.534

dikelompokkan menjadi tiga, yaitu data demografis, data modul mata kuliah, dan data aktivitas interaksi.

1) *Data Demografis*: Setiap mahasiswa yang terdaftar pada suatu modul dicatat pada tabel data *studentInfo* yang merupakan tabel atribut demografis dari mahasiswa. Tabel ini menyimpan data rentang umur (*age band*), wilayah (*region*), pendidikan terakhir (*highest education*), dan lainnya, seperti yang tertera pada Gbr. 1 (tabel *studentInfo*), dengan atribut tambahan hasil akhir (*final result*) yang dikategorikan sebagai *withdrawn*, *pass*, *fail*, dan *distinction*.

2) *Data Modul Mata Kuliah*: Data modul terdiri atas tiga tabel data, yaitu *courses*, *assessment*, dan VLE. Setiap modul, *courses*, dengan durasi (*length*, dalam hari) tertentu memiliki sejumlah tugas pada tabel *assessments* dengan tipe tugas tertentu (*assessment type*) dan batas akhir pengumpulan tugas (*date*). Setiap modul juga memiliki beberapa VLE yang menunjang materi dan segala sumber daya pembelajaran seperti misalnya, forum diskusi, materi pembelajaran, dan lain sebagainya (20 tipe sumber daya/aktivitas yang berbeda).

3) *Data Aktivitas Interaksi*: Data aktivitas merupakan data log interaksi mahasiswa dengan sistem OU, ditunjukkan dengan hubungan *one-to-many* dari tabel *studentInfo* yang terdiri atas tiga tabel data, yaitu *studentRegistration* yang menyimpan tanggal mahasiswa melakukan registrasi dan pembatalan registrasi (dalam hal ini *final result* menjadi *withdrawn*), *studentAssessment* yang menyimpan nilai dan waktu mahasiswa melakukan *submission* terhadap suatu tugas tertentu, dan *studentVle* yang menyimpan aktivitas mahasiswa dengan VLE berupa tanggal dan jumlah klik terhadap suatu sumber daya pembelajaran dari suatu modul tertentu.

III. METODOLOGI

Dalam pengumpulan referensi digunakan metode pencarian sederhana untuk menunjang penelitian dalam makalah ini. Proses pengumpulan referensi untuk tinjauan penggunaan OULAD dalam analisis pembelajaran menggunakan mesin pencari Google Scholar dengan kriteria inklusi dan eksklusi seperti yang ditunjukkan pada Tabel II.

Pertama, dilakukan pencarian dengan judul artikel OULAD, yaitu "*open university learning analytics dataset*" [1], kemudian artikel-artikel yang merujuk pada artikel OULAD tersebut diambil berdasarkan kriteria inklusi.

Kedua, dilakukan penyaringan berdasarkan kriteria eksklusi terhadap artikel-artikel yang didapatkan pada langkah pertama, dengan mengambil beberapa artikel yang membahas tentang

¹ Deskripsi yang lebih rinci pada https://analyse.kmi.open.ac.uk/open_dataset.

TABEL II
KRITERIA INKLUSI DAN EKSKLUSI

Kriteria Inklusi
1. Artikel yang merujuk pada “ <i>open university learning analytics dataset</i> ” [1].
2. Artikel yang membahas tentang pengolahan OULAD dengan metode pembelajaran mesin.
Kriteria Eksklusi
1. Terpublikasi sebelum tahun 2018.
2. Tidak membahas pengolahan data dalam analisis pembelajaran.
3. Hanya membahas tentang MOOC.
4. Hanya dirujuk sebagai metode pengumpulan data.

pengolahan OULAD dalam konteks analisis pembelajaran. Selain konteks tersebut, artikel yang didapat tidak menjadi bagian analisis pada makalah ini.

Ketiga, dilakukan analisis terhadap artikel-artikel yang telah didapatkan sehingga dapat dilihat permasalahan, fitur, dan metode yang digunakan dalam analisis pembelajaran, khususnya untuk OULAD, yaitu untuk menjawab pertanyaan penelitian yang telah diajukan.

IV. HASIL DAN PEMBAHASAN

Pada bagian ini dibahas setiap pertanyaan penelitian yang telah dideskripsikan pada bagian sebelumnya. Setiap pertanyaan penelitian dibahas pada subbagian masing-masing. Secara ringkas, Gbr. 2 menjelaskan pengelompokan artikel-artikel yang mengolah OULAD yang dibagi berdasarkan permasalahannya, fitur yang digunakannya, dan metode atau model pembelajaran mesin yang digunakannya.

A. Permasalahan yang Diangkat pada OULAD (P1)

Dari beberapa artikel yang ditinjau, sebagian besar artikel membahas permasalahan klasifikasi biner, yaitu undur diri (*withdrawn*) atau tidak. Sebagai contoh, diprediksi kemungkinan mahasiswa akan keluar/undur diri dari mata kuliah tertentu, yaitu modul CCC2013B, berdasarkan aktivitas mahasiswa sampai 40% dari total tugas terkumpul [5]. Referensi [6] juga mengangkat permasalahan tentang *drop out* atau tidak, berdasarkan relasi durasi pemberian tugas dan aktivitas mingguan mahasiswa. Demikian juga dengan beberapa referensi yang lainnya, yang mengangkat permasalahan yang sama, yaitu mahasiswa akan undur diri (*withdrawn*) atau tidak, baik pada satu modul tertentu atau keseluruhan modul [7]–[10].

Di sisi lain, penelitian lain mengangkat permasalahan klasifikasi kinerja mahasiswa berdasarkan tiga label di *final result*, yaitu *fail*, *pass*, dan *distinction* [11]. Referensi [12] memprediksi tingkat keterlibatan mahasiswa dalam suatu modul dan melihat variabel yang sangat berpengaruh berdasarkan data demografis mahasiswa. Referensi [13] memprediksi kategori akhir mahasiswa dengan beberapa klasifikasi biner, yaitu *pass-fail*, *distinction-fail*, *distinction-pass*, dan *withdrawn-pass*, dengan mendesain beberapa model yang berbeda untuk setiap permasalahan. Sementara itu, selain

memprediksi enam kategori berdasarkan rentang nilai akhir, yaitu “100-90”, “90-80”, “80-70”, “70-60”, “60-50”, “<50”, sebuah penelitian juga mengangkat permasalahan regresi terhadap nilai akhir tersebut, dengan mengombinasikan fungsi biaya keduanya dengan suatu pembobotan tertentu [14].

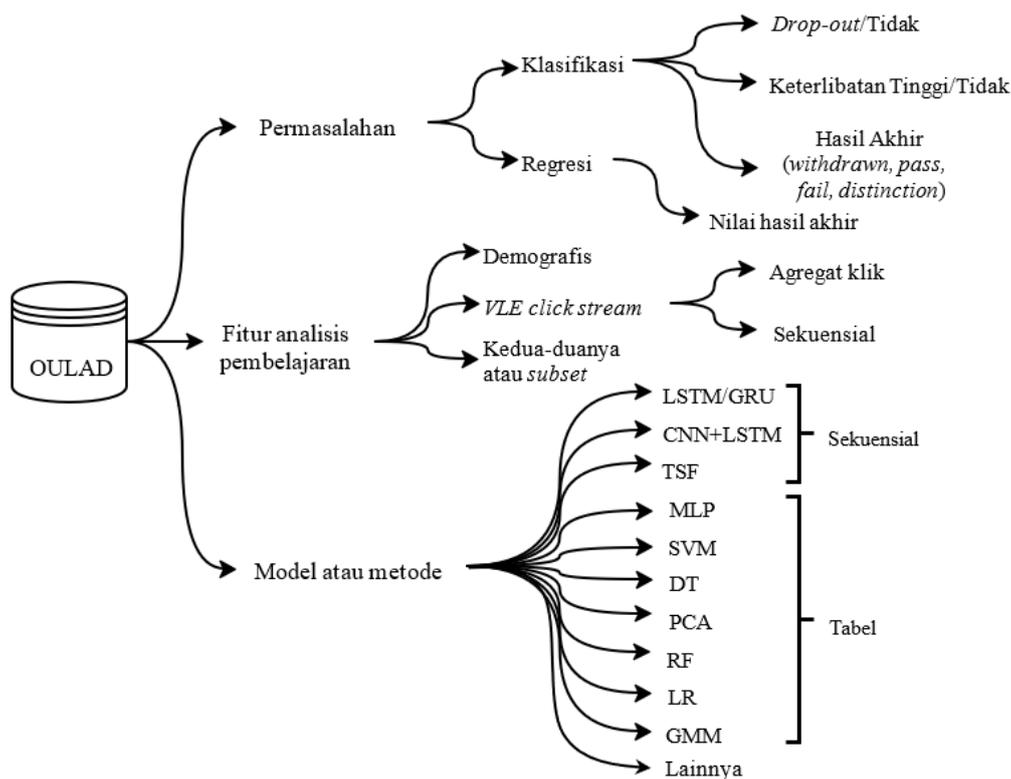
Dari uraian beberapa artikel penelitian tentang beragam permasalahan yang diangkat dalam pengolahan OULAD tersebut, dapat ditarik kesimpulan bahwa mayoritas permasalahan yang diangkat adalah tentang undur diri (*withdrawn*) atau tidak, yaitu delapan dari tiga belas artikel yang ditinjau (61%, – lihat Tabel III di bagian akhir makalah), dengan beberapa variasi permasalahan, yaitu menggunakan keseluruhan modul atau hanya sebagian modul saja. Secara umum, permasalahan-permasalahan yang diangkat berdasar pada kategori hasil akhir, yaitu *withdrawn*, *pass*, *fail*, *distinction*, dan kombinasinya.

Di lain sisi, berdasarkan penelusuran, masih belum ada penelitian yang membahas pemanfaatan OULAD untuk memprediksi nilai setiap tugas yang diberikan untuk setiap modul mata kuliah berdasarkan *click stream* interaksi dengan sistem. Tugas merupakan suatu komponen penting dalam menentukan hasil akhir dari mahasiswa sehingga dengan memprediksi setiap tugas yang diberikan, didapatkan faktor-faktor atau kecenderungan mahasiswa, yaitu memiliki kinerja yang baik atau tidak. Dari fitur *click stream* yang didapat dari prediksi nilai tugas tentunya bisa juga digunakan lebih lanjut untuk prediksi kinerja mahasiswa itu sendiri, yang merupakan salah satu faktor penunjang capaian pembelajaran sehingga dapat meningkatkan capaian pembelajaran yang diharapkan.

B. Fitur yang Digunakan pada OULAD (P2)

Berbagai macam cara para peneliti merepresentasikan fitur-fitur untuk memecahkan berbagai permasalahan telah diuraikan pada subbagian sebelumnya. Pada subbagian ini diuraikan beberapa rekayasa fitur pada OULAD sehingga bisa digunakan untuk analisis pembelajaran. Secara garis besar, seperti yang ditunjukkan pada Gbr. 2, rekayasa fitur dari OULAD dibagi menjadi tiga bagian, yaitu fitur demografis, fitur aktivitas mahasiswa dan sistem pembelajaran (VLE), dan kombinasi atau *subset* dari keduanya.

Fitur demografis tertera pada tabel *studentInfo* yang menjelaskan data pribadi terbatas untuk setiap mahasiswa. Referensi [11] hanya menggunakan fitur demografis untuk memprediksi kategori *pass*, *fail*, dan *distinction*. Dengan demikian, dapat dilihat pengaruh data pribadi terbatas tersebut terhadap kinerja mahasiswa. Dalam hal ini, dijelaskan juga pengaruh terbesar dalam konteks fitur demografis, yaitu berdasarkan wilayah, tingkat kemiskinan yang dilihat dari atribut *imd band*, dan tingkat pendidikan [11]. Fitur demografis tidak tergantung pada aktivitas interaksi mahasiswa sehingga pemodelan prediksi bisa dilakukan pada saat mahasiswa telah melakukan registrasi, baik sebelum ataupun sesudah modul dimulai. Namun, dengan hanya menggunakan fitur demografis, model akan cenderung bias dengan mengesampingkan tingkat keterlibatan atau partisipasi mahasiswa dengan sistem. Dalam hal ini, tidak semua mahasiswa dari suatu daerah, dengan level pendidikan atau kemiskinan tertentu, memiliki potensi kinerja



Gbr. 2 Hierarki pengolahan OULAD secara umum.

yang sama. Dengan demikian, fitur interaksi perlu juga dilihat agar lebih objektif.

Selanjutnya, mayoritas peneliti menggunakan fitur interaksi dari mahasiswa untuk kebutuhan analisis pembelajaran [5]–[9], [13], [14]. Data aktivitas mahasiswa secara harian ini dapat diproses melalui dua cara, yaitu dijadikan tabel (tabular: 9 dari 13 artikel, 69%) atau dijadikan data sekuensial (deret waktu: 31%). Terdapat beberapa cara penyajian data dalam bentuk tabular untuk penggunaan fitur interaksi. Referensi [9] menyajikan tabel untuk suatu modul tertentu dengan menjadikannya tabel fitur klik dengan jumlah kolom sebanyak total hari pada modul tersebut, sedemikian hingga atribut kolom pertama merupakan ringkasan klik pada hari pertama, kolom kedua pada hari kedua, dan seterusnya. Beberapa metode ringkasan klik harian dengan mengesampingkan jenis sumber daya diperkenalkan, antara lain *binary* (pada hari tertentu ada aktivitas atau tidak), *normalize* (total klik dibagi maksimal total klik yang ada), dan *counts* (total klik). Tidak hanya itu, berbagai kombinasi fitur juga dilakukan, yaitu demografis saja, aktivitas harian saja, atau keduanya dengan hasil cukup signifikan (aktivitas harian *binary* F1 = 90%, demografis dan aktivitas harian *binary* F1 = 91%). Ada beberapa hal yang perlu diperhatikan, yaitu dengan menggunakan fitur tabular dengan panjang tetap, dalam hal ini sampai akhir modul, akan menyebabkan keterlambatan dalam memprediksi mahasiswa yang gagal karena sudah di akhir presentasi modul, sementara sebagian mahasiswa sudah melakukan undur diri di awal atau tengah presentasi modul dan menunjukkan tidak ada aktivitas sampai akhir modul. Namun, hal ini dapat dimanfaatkan untuk menginvestigasi fitur-fitur

yang memiliki pengaruh signifikan terhadap kinerja mahasiswa.

Beberapa peneliti mencoba menggunakan panjang fitur yang berbeda untuk dapat melakukan mitigasi terhadap mahasiswa yang berisiko lebih awal, seperti membagi menjadi beberapa rentang panjang fitur, yaitu sampai kuartil 1 (25% dari lama modul dalam hari), kuartil 2 (50%), kuartil 3 (75%), dan kuartil 4 (100%), dengan 54 fitur ringkasan berdasarkan jenis sumber daya yang dijumlah sampai kuartil tertentu [13]. Dengan demikian, mitigasi kegagalan dalam pembelajaran dapat dimitigasi tertentu. Selain itu, ada peneliti membagi rentang waktu dari fitur aktivitas berdasarkan batas akhir setiap tugas, dalam hal ini tugas sangat berpengaruh pada hasil akhir mahasiswa, sehingga kecenderungan mahasiswa untuk undur diri diprediksi pada setiap submisi tugas berakhir [5], [6]. Selain itu, fitur interaksi juga dapat disajikan dalam bentuk sekuensial, seperti halnya deret waktu, sehingga prediksi kinerja atau kecenderungan mahasiswa melakukan *drop out* dapat dilakukan secara harian atau mingguan. Fitur sekuensial yang digunakan dalam beberapa literatur sangat bervariasi, yaitu jumlah klik harian dari semua sumber daya VLE [14], jumlah klik pada satu sumber daya VLE tertentu (*Forum* atau *OUcontent* atau *resource* VLE) [8], dan ringkasan dari semua sumber daya VLE (setiap sumber daya VLE memiliki atribut tersendiri) dalam mingguan [7]. Di samping itu, untuk meningkatkan kinerja, fitur sekuensial dengan fitur tabel demografis bisa digabungkan [14].

Dari uraian tersebut, dapat disimpulkan bahwa penggunaan fitur yang berbeda menunjukkan fokus permasalahan yang berbeda. Berdasarkan pemanfaatannya, fitur dapat dibagi

menjadi dua, yaitu sebagai analisis faktor risiko dari sistem pembelajaran atau sistem pendeteksian dini pada mahasiswa yang memiliki kecenderungan gagal dalam suatu mata kuliah.

1) *Analisis Faktor*: Fitur yang digunakan biasanya merupakan fitur dari keseluruhan rentang waktu, yaitu dari suatu mata kuliah dimulai sampai berakhir, baik yang digunakan untuk permasalahan klasifikasi maupun regresi. Fokus utama dari penggunaan keseluruhan rentang waktu dari fitur bukan terletak pada kinerja yang diperoleh karena tentunya kinerjanya cenderung lebih baik daripada menggunakan rentang waktu yang sebagian, tetapi yang seharusnya lebih ditekankan adalah faktor yang sangat memengaruhi mahasiswa menjadi gagal atau berhasil dalam suatu mata kuliah. Dengan demikian, dari faktor-faktor tersebut, bisa dilakukan intervensi lebih terhadap faktor-faktor yang signifikan untuk meningkatkan tingkat kesuksesan mahasiswa. Dengan kata lain, dengan menggunakan analisis faktor, perbaikan sistem *e-Learning* bisa dilakukan dengan mempertimbangkan faktor-faktor penting yang telah dianalisis.

2) *Sistem Pendeteksian Dini*: Penggunaan fitur dengan rentang waktu yang lebih pendek akan sangat berguna untuk proses mitigasi terhadap mahasiswa yang dianggap berisiko dalam pengajaran. Berbeda dengan analisis faktor, semakin kecil rentang waktu yang digunakan, semakin besar pemanfaatannya dari sisi mitigasi mahasiswa yang berisiko. Namun, semakin kecil rentang waktu dari fitur yang digunakan akan menyebabkan kinerja dari sistem ini menurun karena tidak adanya cukup informasi yang diberikan. Sebagai contoh, prediksi mahasiswa akan gagal atau tidak dilakukan di awal perkuliahan dimulai. Tentunya hal ini akan sangat sulit dilakukan untuk mendapatkan tingkat kepercayaan yang tinggi sehingga pada sistem ini kinerja dari suatu model menjadi hal yang sangat penting untuk menunjukkan tingkat kepercayaan dari sistem tersebut. Pemilihan rentang waktu bisa dilakukan dengan berbagai cara, seperti sampai pertengahan mata kuliah atau dibagi menjadi empat rentang waktu yang berbeda [13], yaitu kuartil 1 sampai 4 (tentunya kuartil 4 akan memiliki kinerja yang lebih tinggi, tetapi pemanfaatannya cenderung tidak besar, karena pendeteksian dilakukan di akhir).

C. Model Pembelajaran Mesin atau Metode yang Digunakan pada OULAD (P3)

Metode pembelajaran mesin yang digunakan untuk melakukan analisis pembelajaran dalam pengolahan OULAD sangat bervariasi, antara lain *Decision Tree* (DT) [5], [9]–[12], [15], [16]; *Multilayer Perceptron* (MLP)/*Artificial Neural Network* (ANN) [17], seperti pada [10], [13]; *Support Vector Machine* (SVM) [18], seperti pada [5], [9], [10]; *Logistic Regression* (LR) [15], seperti pada [5], [9]; *Long Short-Term Memory* (LSTM) [19], seperti pada [7], [14]; *Gated Recurrent Unit* (GRU) [20]; *Random Forest* (RF) [5], [9]; *Adaptive RF* (ARF) [21]; *Distributed RF* (DRF), *Gradient Boosting Machine* (GBM), dan *Generalized Linear Model* (GLM) [22]; *Gaussian Mixture Models* (GMM) [23], seperti pada [6]; dan *Time Series Forest* (TSF) [24], seperti pada [8]. Selain itu, ada juga yang menggabungkan dengan metode *Principal*

Component Analysis (PCA) untuk mereduksi fitur sebelum diproses dengan algoritme pembelajaran mesin [5] dan menggabungkan LSTM dengan *Convolutional Neural Network* (CNN) [25], seperti pada [14].

Selanjutnya, ditinjau dari jenis fiturnya, algoritme pembelajaran tersebut dapat dibagi menjadi dua kelompok, yaitu algoritme dengan fitur sekuensial atau tabular. Fitur sekuensial erat hubungannya dengan deret waktu (*time series*), baik secara harian ataupun mingguan. Kelompok algoritme ini terdiri atas LSTM atau RNN dan TSF. Keuntungannya adalah penggunaan fitur sekuensial prediksi kinerja dapat dilakukan sesuai waktu yang tersedia (kapan saja). Namun, perlu dicatat bahwa semakin awal prediksi kinerja dilakukan, semakin menurun tingkat keakuratan algoritme, karena tidak cukup tren data yang bisa mewakili interaksi mahasiswa pada suatu modul tertentu. Sementara itu, algoritme yang lain masuk ke kategori fitur tabel, yang jauh lebih mudah dan sederhana pengolahan/pemrosesan algoritmenya dibandingkan fitur sekuensial.

Tabel III (di bagian akhir makalah) meringkas artikel-artikel yang menjadi tinjauan utama makalah ini berdasarkan metode dan fitur yang digunakan, keluaran tugas, dan lingkup modul OULAD yang digunakan. Pada tabel tersebut dapat dilihat penggunaan OULAD untuk berbagai kebutuhan dengan tingkat kinerja yang beragam berdasarkan metrik akurasi, *F1-score*, dan *Area Under the Curve* (AUC). Mayoritas permasalahan yang diangkat adalah tentang undur diri (*withdrawn*) atau tidak. Di sisi lain, fitur dapat berupa tabular (data demografis, interaksi/aktivitas, tugas) atau sekuensial (data interaksi/aktivitas dalam bentuk deret waktu).

V. DISKUSI

Bagian ini mendiskusikan cara untuk memanfaatkan data-data log pada sebuah *e-Learning* dengan mempertimbangkan hal-hal yang telah dilakukan oleh OU. Diskusi mencakup penyajian arsitektur data sistem pembelajaran daring, mitigasi pembelajaran, permasalahan terhadap model, dan data.

A. Penyajian Arsitektur Data Sistem Pembelajaran Daring

Desain arsitektur data dari sistem pembelajaran daring untuk memenuhi kebutuhan analisis pembelajaran seharusnya sudah dipersiapkan pada tahap perencanaan pembuatan sistem sehingga arsitektur data dapat mendukung analisis pembelajaran. Dalam hal ini, berdasarkan OULAD, data log mahasiswa sangatlah penting untuk mengetahui kinerja mahasiswa. Sistem *e-Learning* yang ditemui sering kali hanya menyimpan data personal atau demografis dari mahasiswa, sedangkan data aktivitas sangat sedikit, hanya terbatas pada data interaksi dengan tugas tertentu. Dalam hal ini, perlu diukur tingkat keterlibatan mahasiswa dengan menggunakan data log harian sehingga bisa dilakukan mitigasi kemungkinan *drop out* atau bisa meningkatkan kinerja belajar dengan perlakuan atau penyesuaian lingkungan yang berbeda. Pada dasarnya perancangan data yang terstruktur menjadi kebutuhan mendasar pada analisis pembelajaran. Selain itu, semakin banyak data yang disimpan akan lebih baik untuk kebutuhan analisis di waktu yang akan datang, meskipun pada saat ini

belum diketahui kebutuhan yang akan menggunakannya. Namun, kekurangan dari penyimpanan data yang sangat banyak adalah dibutuhkan kapasitas penyimpanan data yang lebih besar.

Dari uraian tersebut, dengan mempertimbangkan hal yang telah dilakukan di OULAD, entitas basis data untuk kebutuhan sistem daring penuh setidaknya memerlukan tiga hal, yaitu data demografis, data modul/mata kuliah, dan data interaksi dengan sistem. Dengan tiga entitas tersebut, sudah bisa dilakukan analisis pembelajaran, baik untuk analisis faktor ataupun untuk sistem pendeteksian dini. Untuk sistem pendeteksian dini, entitas basis data diharapkan memiliki fitur rentang waktu, terutama untuk entitas data interaksi dengan sistem atau tugas dari modul tersebut. Selain itu, bisa juga ditambahkan entitas data untuk menyimpan histori pembelajaran dari mahasiswa sehingga dapat digunakan untuk fitur mahasiswa memiliki pengetahuan sebelumnya atau tidak, sebelum teregistrasi pada modul mata kuliah tertentu. Kelengkapan fitur dari desain basis data yang dibuat akan memudahkan proses analisis pembelajaran dan diharapkan dapat meningkatkan kualitas dan capaian pembelajaran. Sebagai contoh, entitas interaksi dapat disiapkan dengan membuat suatu tabel interaksi, seperti halnya *studentVLE* dan *studentAssessment* yang menyimpan interaksi mahasiswa dengan modul *e-learning* dan tugas/ujian. Untuk entitas demografis, pada umumnya semua *e-learning* menyimpan entitas ini pada awal pendaftaran, seperti halnya tabel *studentInfo*. Selanjutnya, entitas historis menyimpan data-data mata kuliah yang pernah diambil oleh setiap mahasiswa, misalnya dengan membuat suatu tabel *studentCourses* yang terdiri atas *studentId* dan *courseId*.

Pembahasan desain basis data tersebut berdasar pada sudut pandang mahasiswa. Untuk melengkapi fitur yang komprehensif, dibutuhkan tidak hanya data mahasiswa, tetapi juga data pengajar jika kelas tersebut menggunakan pendekatan kelas berbasis tutorial. Dalam hal ini, faktor yang juga memengaruhi kinerja dan kelulusan suatu proses pembelajaran, selain sistem dan mahasiswa, adalah sisi pengajar itu sendiri. Dengan demikian, kebutuhan akan data untuk analisis pembelajaran, khususnya dengan metode pembelajaran mesin, dapat diolah dengan mudah, akurat, dan komprehensif.

B. Mitigasi Pembelajaran

Semakin awal dilakukannya prediksi kinerja atau kemungkinan untuk *drop-out* akan semakin baik dan bisa dilakukan mitigasi atau penyesuaian terhadap kelompok mahasiswa yang berisiko mundur atau gagal pada suatu mata kuliah. Mempertimbangkan hal ini, suatu model pembelajaran mesin yang akan digunakan untuk memprediksi kelompok berisiko haruslah bisa digunakan sebelum mata kuliah berakhir atau sebelum mahasiswa undur diri dari kelas. Beberapa penelitian melakukan prediksi dengan menggunakan semua fitur aktivitas sampai akhir modul yang tidak terlalu bermanfaat dalam mitigasi mahasiswa. Akan tetapi, prediksi di akhir modul dapat digunakan sebagai analisis faktor-faktor yang berpengaruh terhadap kecenderungan mahasiswa gagal atau mundur dari suatu mata kuliah tertentu. Hal ini sangat berhubungan erat dengan hal yang telah dibahas pada bagian IV-B-1 dan IV-B-2. Mitigasi pembelajaran dapat dilakukan

menggunakan pendekatan analisis faktor dengan meningkatkan modul pembelajaran (intervensi pada modul pembelajaran). Sebagai contoh, suatu modul/subbab materi tertentu memiliki pengaruh positif yang signifikan terhadap kecenderungan kesuksesan mahasiswa dalam suatu mata kuliah tertentu. Dengan demikian, diperlukan peningkatan pembelajaran pada modul tersebut atau penambahan materi penunjang yang dapat dijadikan penjelasan tambahan. Selain itu, bisa juga dilakukan sistem pendeteksian dini (intervensi pada mahasiswa). Lebih lanjut lagi, sistem pendeteksian dini dapat digabungkan dengan metodologi *Self-Regulated Learning* (SRL) [26], tergantung pada tingkat kategori risiko dari mahasiswa, yaitu dengan memberikan suatu tindakan akademis terhadap mahasiswa-mahasiswa yang berisiko tinggi, baik dengan komunikasi langsung meningkatkan keterlibatan dalam suatu kelas tertentu atau dengan cara yang lainnya sesuai risiko dan budaya pada suatu institusi tertentu.

C. Permasalahan terhadap Model dan Data

Kuantitas data sangat berpengaruh terhadap kinerja model, sedangkan secara umum kualitas data dapat ditinjau dari dua hal, yaitu kelengkapan dan cakupan. Kelengkapan data dapat dilihat dari atribut yang ada. Data yang berkualitas tidak kehilangan atribut yang penting dalam pemodelan sehingga kinerja algoritme pembelajaran mesin lebih baik (tingkat akurasi lebih tinggi). Sementara itu, cakupan data adalah data keberagaman dari contoh (dalam hal ini mahasiswa) sehingga mencakup semua tipe contoh. Dengan demikian, model dapat memprediksi berbagai tipe mahasiswa. Solusi dari hal ini adalah dengan mengumpulkan sebanyak-banyaknya data sehingga kinerja model dapat lebih dipertanggungjawabkan. Dengan data yang sedikit, suatu model cenderung *overfitting* karena cakupan data yang kecil sehingga model tidak mampu melakukan generalisasi dan tentunya tingkat akurasi menjadi dipertanyakan. Selain permasalahan yang telah disebutkan, permasalahan yang lain adalah data kategori yang tidakimbang (*imbalanced data*), yang menyebabkan model menjadi bias pada kategori yang paling besar (mayoritas) [10]. Hal ini dapat diselesaikan dengan menyeimbangkan kategori tersebut dengan menggunakan berbagai macam metode *over-sampling* atau *under-sampling*.

VI. KESIMPULAN

Pengolahan OULAD didasarkan pada dua jenis fitur, yaitu fitur demografis dan fitur aktivitas interaksi mahasiswa dengan sistem. Dari penjelasan pada bagian-bagian sebelumnya, dapat ditarik kesimpulan sebagai berikut.

Mayoritas permasalahan yang diangkat yaitu 61% tentang prediksi mahasiswa berisiko *drop out/withdrawn* atau tidak, baik menggunakan fitur interaksi, demografis, atau keduanya. Mayoritas peneliti, yaitu sebanyak 92%, menggunakan fitur interaksi, dalam hal ini data log mahasiswa, untuk prediksi kinerja atau kecenderungan mahasiswa yang berisiko *drop out* (tidak melanjutkan pembelajaran/mundur. Selain itu, ada juga peneliti yang menggunakan hanya fitur demografis atau keduanya untuk analisis pembelajaran, baik digunakan untuk analisis faktor maupun pemodelan prediksi.

Berdasarkan penggunaan rentang waktu fiturnya, model yang dibangun dapat dikategorikan menjadi dua, yaitu analisis faktor untuk mencari faktor yang sangat memengaruhi kinerja mahasiswa dan sistem pendeteksian dini untuk mendeteksi mahasiswa yang berisiko. Berdasarkan pengolahan fitur untuk pemrosesan pembelajaran mesin, secara umum model dibagi menjadi dua, yaitu fitur tabular yang umumnya digunakan pada mayoritas algoritme pembelajaran dan sekuensial yang berupa deret waktu dengan model TSF, LSTM, atau GRU.

Beberapa saran telah direkomendasikan, yaitu berkaitan dengan penyajian data untuk sistem pembelajaran daring, mitigasi pembelajaran, dan permasalahan yang sering muncul pada pemodelan yang dikaitkan dengan kuantitas dan kualitas data, yaitu sebagai berikut. Pembangunan sistem data setidaknya terdiri atas tiga entitas, yaitu data demografis, data modul/mata kuliah, dan data interaksi dengan sistem. Dapat juga ditambahkan data histori atau data pengajaran untuk mendukung analisis pembelajaran yang lebih komprehensif. Mitigasi pembelajaran dapat dilakukan dengan dua hal, yaitu menggunakan pendekatan analisis faktor dengan meningkatkan modul pembelajaran (intervensi pada modul pembelajaran) atau dengan sistem pendeteksian dini (intervensi pada mahasiswa seperti halnya SRL). Permasalahan yang sering dijumpai pada saat pemodelan antara lain kurangnya kelengkapan dan cakupan data, *overfitting*, dan jumlah kelas data yang tidak berimbang (*imbalanced dataset*).

Makalah ini dapat digunakan sebagai panduan mengolah fitur interaksi dan demografis serta cara pemodelan yang telah atau banyak digunakan oleh berbagai peneliti sehingga menginspirasi munculnya pemodelan baru atau sistem penyediaan data untuk keperluan analisis pembelajaran yang baru dan efisien, dengan mempertimbangkan kelengkapan dan cakupan data, seperti halnya yang ada pada OULAD dan beberapa metode serta pengolahan fitur yang telah dijelaskan.

REFERENSI

- [1] J. Kuzilek, M. Hlosta, dan Z. Zdrahal, "Open University Learning Analytics Dataset," *Sci. Data*, Vol. 4, hal. 1-8, Nov. 2017.
- [2] P.M. Moreno-Marcos, C. Alario-Hoyos, P.J. Muñoz-Merino, dan C.D. Kloos, "Prediction in MOOCs: A Review and Future Research Directions," *IEEE Trans. Learn. Technol.*, Vol. 12, No. 3, hal. 384-401, Jul. 2019.
- [3] O. Pilli dan W. Admiraal, "Students' Learning Outcomes in Massive Open Online Courses (MOOCs): Some Suggestions for Course Design," *Yükseköğretim Dergisi*, Vol. 7, No. 1, hal. 46-71, 2017.
- [4] J. Biggs dan C. Tang, *Teaching For Quality Learning At University*, 4th ed., Maidenhead, UK: Open University Press, 2011.
- [5] E.G. Poitras, R.F. Behnagh, dan F. Bouchet, "A Dimensionality Reduction Method for Time Series Analysis of Student Behavior to Predict Dropout in Massive Open Online Courses," dalam *Adoption of Data Analytics in Higher Education Learning and Teaching*, D. Ifenthaler and D. Gibson, Eds., Cham, Switzerland: Springer International Publishing, 2020, hal. 391-406.
- [6] R. Alshabandar, A. Hussain, R. Keight, A. Laws, dan T. Baker, "The Application of Gaussian Mixture Models for the Identification of At-Risk Learners in Massive Open Online Courses," *2018 IEEE Congress on Evolutionary Computation (CEC)*, Jul. 2018, hal. 1-8.
- [7] S.-U. Hassan, H. Waheed, N.R. Aljohani, M. Ali, S. Ventura, dan F. Herrera, "Virtual Learning Environment to Predict Withdrawal by Leveraging Deep Learning," *Int. J. Intell. Syst.*, Vol. 34, No. 8, hal. 1935-1952, 2019.
- [8] L. Haiyang, Z. Wang, P. Benachour, dan P. Tubman, "A Time Series Classification Method for Behaviour-Based Dropout Prediction," *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, 2018, hal. 191-195.
- [9] H. Heuer dan A. Breiter, "Student Success Prediction and the Trade-Off between Big Data and Data Minimization," dalam *DeLFI 2018 - Die 16. E-Learning Fachtagung Informatik*, D. Krömker dan U. Schroeder, Eds., Bonn, Germany: Gesellschaft für Informatik e.V., 2018, hal. 219-230.
- [10] F. Hlioui, N. Aloui, dan F. Gargouri, "Withdrawal Prediction Framework in Virtual Learning Environment," *Int. J. Serv. Sci. Manag. Eng. Technol.*, Vol. 11, No. 3, hal. 47-64, Jul. 2020.
- [11] S. Rizvi, B. Rienties, dan S.A. Khoja, "The Role of Demographics in Online Learning: A Decision Tree Based Approach," *Comput. Educ.*, Vol. 137, hal. 32-47, Agu. 2019.
- [12] M. Hussain, W. Zhu, W. Zhang, dan R. Abidi, "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores," *Comput. Intell. Neurosci.*, Vol. 2018, hal. 1-21, Okt. 2018.
- [13] H. Waheed, S.-U. Hassan, N.R. Aljohani, J. Hardman, S. Alelyani, dan R. Nawaz, "Predicting Academic Performance of Students from VLE Big Data Using Deep Learning Models," *Comput. Hum. Behav.*, Vol. 104, hal. 1-13, Mar. 2020.
- [14] X. Song, J. Li, S. Sun, H. Yin, P. Dawson, dan R. Doss, "SEPN: A Sequential Engagement Based Academic Performance Prediction Model," *IEEE Intell. Syst.*, hal. 1-9, 2020.
- [15] T. Hastie, R. Tibshirani, dan J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, 2nd ed., New York, USA: Springer, 2016.
- [16] S.L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, Vol. 16, No. 3, hal. 235-240, Sep. 1994.
- [17] D.E. Rumelhart, G.E. Hinton, dan R.J. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, Vol. 323, hal. 533-536, Okt. 1986.
- [18] C. Cortes dan V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, Vol. 20, No. 3, hal. 273-297, Sep. 1995.
- [19] S. Hochreiter dan J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, Vol. 9, No. 8, hal. 1735-1780, Nov. 1997.
- [20] Y. He, R. Chen, X. Li, C. Hao, S. Liu, G. Zhang, dan B. Jiang, "Online At-Risk Student Identification using RNN-GRU Joint Neural Networks," *Information*, Vol. 11, No. 10, hal. 1-12, Okt. 2020.
- [21] G. Casalino, G. Castellano, A. Mannavola, dan G. Vessio, "Educational Stream Data Analysis: A Case Study," *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)*, 2020, hal. 232-237.
- [22] N. Jha, I. Ghergulescu, dan A.-N. Moldovan, "OULAD MOOC Dropout and Result Prediction using Ensemble, Deep Learning and Regression Techniques," *11th International Conference on Computer Supported Education*, 2020, hal. 154-164.
- [23] D.A. Reynolds, T.F. Quatieri, dan R.B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digit. Signal Process.*, Vol. 10, No. 1, hal. 19-41, Jan. 2000.
- [24] H. Deng, G. Runger, E. Tuv, dan M. Vladimirov, "A Time Series Forest for Classification and Feature Extraction," *Inf. Sci.*, Vol. 239, hal. 142-153, Agu. 2013.
- [25] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, dan L.D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Comput.*, Vol. 1, No. 4, hal. 541-551, Des. 1989.
- [26] M.E. Alonso-Mencía, C. Alario-Hoyos, J. Maldonado-Mahauad, I. Estévez-Ayres, M. Pérez-Sanagustín, dan C.D. Kloos, "Self-regulated Learning in MOOCs: Lessons Learned from a Literature Review," *Educ. Rev.*, Vol. 72, No. 3, hal. 319-345, Mei 2020.

TABEL III
RINGKASAN TINJAUAN ARTIKEL-ARTIKEL YANG MENGOLAH OULAD

Artikel	Fitur	Metode	Modul	Keluaran Tugas	Kinerja Terbaik
<i>A Dimensionality Reduction Method for Time Series Analysis of Student Behaviour to Predict Dropout in MOOCs</i> [5]	Agregat <i>clickstream</i> per interval tugas tabular	DT, RF, LR, SVM, PCA	CCC 2013B	<i>Withdrawn/ Tidak</i> Sampai 40% total tugas	DT: Akurasi 0,78
<i>The Application of Gaussian Mixture Models for the Identification of At-Risk Learners in MOOCs</i> [6]	Agregat <i>clickstream</i> per interval tugas tabular + demografi	GMM, KNN, LR	BBB2013J	<i>Withdrawn/ Tidak</i> Sampai 40% total tugas	GMM: <i>F1-Score</i> 0,835
<i>VLE to Predict Withdrawal by Leveraging Deep Learning</i> [7]	Agregat <i>clickstream</i> Mingguan deret waktu	LSTM	Semua modul	<i>Withdrawn/ Tidak</i> Per minggu	Akurasi 0,972
<i>The Role of Demographics in Online Learning: A Decision Tree Based Approach</i> [11]	Demografi tabular	DT	AAA 2013 BBB 2014 DDD 2013 EEE 2013	<i>Distinction/Fail/Pass</i>	Akurasi 0,831
<i>A Time Series Classification Method for Behaviour-Based Dropout Prediction</i> [8]	Agregat <i>clickstream</i> harian deret waktu	TFS	Semua modul	<i>Withdrawn/Tidak</i> Per modul	Akurasi 0,940
<i>Student Success Prediction and the Trade-Off between Big Data and Data Minimization</i> [9]	Agregat <i>clickstream</i> harian tabular	SVM, DT, LR, dan RF	Semua modul	<i>Withdrawn/Tidak</i>	SVM: Akurasi 0,880
<i>Student Engagement Predictions in An e-Learning System and Their Impact on Student Course Assessment Scores</i> [12]	Agregat <i>clickstream</i> per interaksi tabular	DT dan variannya	AAA 2013J	<i>High Engagement/Tidak</i>	DT: Akurasi 0,885
<i>SEPN: A Sequential Engagement Based Academic Performance Prediction Model</i> [14]	Agregat <i>clickstream</i> harian deret waktu + demografi	CNN+LSTM	Semua modul	Klasifikasi/Regresi nilai hasil akhir	Akurasi 0,610
<i>Predicting Academic Performance of Students from VLE Big Data using Deep Learning Models</i> [13]	Agregat <i>clickstream</i> per interaksi tabular + demografi	ANN, SVM, +LR	Semua modul	<i>Pass-Fail</i> <i>Distinction-Pass</i> <i>Distinction-Fail</i> <i>Withdrawn-Pass</i>	ANN: <i>Withdrawn-Pass</i> : Akurasi 0,932
<i>Withdrawal Prediction Framework in VLE</i> [10]	Agregat <i>clickstream</i> per interaksi tabular + demografi	RF+SMOTE +seleksi fitur	DDD2013B	<i>Withdrawn/Tidak</i>	Akurasi 0,874

TABEL III (LANJUTAN)
RINGKASAN TINJAUAN ARTIKEL-ARTIKEL YANG MENGOLAH OULAD

Artikel	Fitur	Metode	Modul	Keluaran Tugas	Kinerja Terbaik
<i>Online At-Risk Student Identification Using RNN-GRU Joint Neural Networks</i> [20]	Agregat <i>clickstream</i> Mingguan deret waktu + demografi + tugas	RNN/GRU	Semua modul	<i>Pass & Distinction - Fail</i>	Akurasi 0,90
<i>OULAD MOOC Dropout and Result Prediction using Ensemble, Deep Learning and Regression Techniques</i> [22]	Agregat <i>clickstream</i> per interaksi per hari tabular + demografi + tugas	DRF, GBM, ANN, GLM	Semua modul	<i>Withdrawn/Tidak Fail/Tidak</i>	GBM: AUC 0,90
<i>Educational Stream Data Analysis: A Case Study</i> [21]	Agregat <i>clickstream</i> per interaksi per hari tabular + demografi + tugas	RF/ARF +seleksi fitur	Semua modul	<i>Fail-Pass</i>	Akurasi 0,905

Catatan: Interaksi pada OULAD terdiri atas dua puluh aktivitas