Journal of SAINTEKS
Vol 4, No 1, November 2026, page 146-151
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

Klasifikasi Multilabel Artikel Academic Burnout Berbasis Study Demands-Resources Menggunakan TF–IDF dan Linear SVM

Wendi Amsuri Nasution*, Dewi Shinta Wulandari Lubis

Program Studi Manajemen, Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia
Email: [1,*]mrs.wendhi@gmail.com, [2]dewishintawulandari83@gmail.com
Email Penulis Korespondensi: mrs.wendhi@gmail.com

Abstrak-Literatur *academic burnout* mahasiswa membahas tuntutan akademik, dukungan lingkungan belajar, dan kapasitas psikologis individu yang sering muncul secara bersamaan dalam satu penelitian. Kerangka *Study Demands-Resources* (SD-R) digunakan karena memberikan pembedaan konseptual antara tuntutan yang berpotensi menguras energi, sumber daya eksternal yang mendukung proses belajar, dan sumber daya personal yang menunjang kemampuan adaptasi mahasiswa. Pembedaan tersebut memungkinkan pengelompokan artikel berdasarkan peran konstruk yang dianalisis, bukan sekadar kategori faktor internal dan eksternal. Penelitian ini bertujuan mengevaluasi penerapan *Term Frequency-Inverse Document Frequency* (*TF–IDF*) dan *Linear Support Vector Machine* (*Linear SVM*) untuk mengklasifikasikan artikel *academic burnout* secara *multilabel* ke dalam *Study Demands* (D), *Study Resources* (R), dan *Personal Resources* (P). Eksperimen menggunakan 113 artikel berlabel, terdiri atas 90 data latih dan 23 data uji. Hasil evaluasi menunjukkan *macro F1* sebesar 73,56% dan *micro F1* sebesar 75,29%, dengan F1 masing-masing sebesar 78,57% pada D, 63,16% pada R, dan 78,95% pada P. Meskipun P memperoleh skor F1 tertinggi, model belum mampu mengidentifikasi kasus negatif pada kategori tersebut. Temuan ini menunjukkan bahwa pendekatan *multilabel* dapat mendukung pengelompokan literatur berdasarkan konstruk SD-R, tetapi kinerja antarkategori belum merata. Penelitian ini berkontribusi pada pengembangan pendekatan komputasional untuk penelusuran literatur *academic burnout*, dengan keterbatasan pada ukuran data uji, pemilihan sampel secara terarah, dan konsistensi anotasi.

Kata Kunci: Academic Burnout; Klasifikasi Multilabel; Study Demands-Resources; TF–IDF; Linear SVM

1. PENDAHULUAN

Academic burnout merupakan persoalan yang dikaji dalam berbagai konteks pendidikan tinggi, termasuk pendidikan kedokteran, kedokteran gigi, dan program studi lainnya. Penelitian Rahmatpour et al., (2019) , Bolatov et al., (2022), Javed et al., (2024), serta Z. Liu et al., (2023) menunjukkan luasnya perhatian terhadap kondisi tersebut pada mahasiswa. Kajian lain menempatkan kelelahan akademik dalam hubungan dengan keterlibatan belajar, efikasi akademik, dan kecenderungan meninggalkan studi (Jeon et al., 2022; Marôco et al., 2020). Keragaman fokus ini memperkaya pemahaman mengenai *academic burnout*, tetapi juga membuat penelusuran literatur memerlukan pembedaan yang jelas antara faktor yang diteliti, konteks penelitian, dan keluaran yang diamati.

Dalam literatur, tuntutan akademik dapat dikaitkan dengan tekanan belajar, beban tugas, dan kondisi pembelajaran. Olson et al., (2025) membahas stres, *burnout*, dan keterlibatan belajar pada mahasiswa dari bidang akademik berbeda, sedangkan Xia et al., (2025) menelaah tekanan akademik dan kepuasan terhadap kehidupan sekolah. Song et al., (2022a) menempatkan pembelajaran digital dalam pembahasan *academic burnout*. Di sisi lain, dukungan sosial dan lingkungan belajar menjadi perhatian Ye et al., (2021) serta Olson et al., (2023) . Sementara itu, penelitian Gu et al., (2023) , Y. Liu et al., (2023), dan Önder et al., (2025) memperlihatkan bahwa faktor psikologis maupun perilaku individu juga sering muncul dalam kajian yang sama. Ragam konstruk tersebut tidak dapat diperlakukan sebagai satu kelompok faktor yang memiliki fungsi setara.

Kerangka *Study Demands-Resources* (SD-R) menyediakan dasar konseptual untuk membedakan peran konstruk tersebut. Kerangka ini berkaitan dengan pengembangan pemikiran *Job Demands-Resources* yang memisahkan tuntutan dan sumber daya berdasarkan fungsinya (Bakker & Demerouti, 2017; Demerouti et al., 2001). Dalam konteks pendidikan tinggi, Lesener et al. (2020) memperkenalkan kerangka SD-R secara empiris, sedangkan Salmela-Aro et al., (2022) mengkaji tuntutan, sumber daya, dan kebutuhan psikologis dalam kaitannya dengan *burnout* serta keterlibatan studi. Berdasarkan pembedaan ini, *Study Demands* (D) merujuk pada tuntutan akademik, *Study Resources* (R) pada dukungan dari lingkungan studi, dan *Personal Resources* (P) pada kapasitas psikologis positif individu. Kategori tersebut tidak sekadar membagi faktor menjadi internal dan eksternal, melainkan memperhatikan peran konstruk dalam proses belajar. Pembedaan konseptual tersebut penting ketika unit analisisnya berupa artikel ilmiah. Satu artikel dapat mengkaji tuntutan akademik sekaligus dukungan sosial dan kapasitas adaptif mahasiswa. Penelitian mengenai dukungan sosial dan *burnout* (Ye et al., 2021), misalnya, menunjukkan relevansi sumber daya lingkungan, sedangkan kajian tentang *grit* dan kelincahan belajar (Jeon et al., 2022) mengarahkan perhatian pada kapasitas individu. Kedua kelompok konstruk dapat muncul bersama tuntutan studi. Karena itu, pengelompokan artikel ke dalam satu kategori tunggal berisiko menyederhanakan isi penelitian. Klasifikasi multilabel memungkinkan satu dokumen memperoleh lebih dari satu label tanpa menganggap ketiga konstruk saling eksklusif.

Penelusuran manual tetap diperlukan untuk menilai kedudukan variabel dalam setiap studi, tetapi penerapannya pada kumpulan artikel menuntut konsistensi interpretasi. Istilah *stress*, *support*, dan *coping* dapat muncul sebagai variabel penelitian, bagian dari latar belakang, atau sekadar konteks pembahasan. Kemunculan istilah saja belum cukup untuk menentukan label. Pada penelitian ini, dua penilai independen menghasilkan perbedaan penetapan sedikitnya satu label D, R, atau P pada 68 dari 113 artikel. Perbedaan tersebut merupakan variasi interpretasi terhadap konstruk, bukan bukti

Copyright © 2026 the author, Page 146
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 146-151
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

bahwa salah satu penilai melakukan kesalahan. Keadaan ini memperlihatkan pentingnya definisi operasional yang membatasi keputusan pelabelan berdasarkan peran konstruk, bukan semata kemunculan kata.

Konsekuensi klasifikasi yang kurang tepat juga perlu diperhatikan. Prediksi negatif yang keliru dapat menyebabkan artikel relevan terlewat dari penelusuran, sedangkan prediksi positif yang keliru dapat memasukkan artikel yang tidak memenuhi definisi kategori. Kedua kesalahan itu memiliki dampak berbeda ketika peneliti menyusun bahan bacaan mengenai tuntutan studi, dukungan institusional, atau sumber daya personal. Oleh sebab itu, evaluasi klasifikasi tidak cukup bertumpu pada satu ukuran kinerja keseluruhan. Hasil pada masing-masing label dan jenis kesalahan prediksi perlu diperiksa agar kemampuan model tidak dinilai lebih luas daripada yang ditunjukkan data.

Untuk merepresentasikan teks, penelitian ini menggunakan *Term Frequency-Inverse Document Frequency* (TF–IDF), yaitu pembobotan istilah yang memiliki landasan dalam penelitian temu kembali informasi (Salton & Buckley, 1988). Klasifikasi dilakukan menggunakan *Linear Support Vector Machine* (Linear SVM), yang bertumpu pada pendekatan *support-vector networks* (Cortes & Vapnik, 1995). Kombinasi tersebut diterapkan pada judul dan abstrak artikel untuk memprediksi label D, R, dan P. Pemilihan metode ini diarahkan pada evaluasi klasifikasi dokumen berdasarkan fitur tekstual; penelitian tidak membandingkannya dengan algoritma lain dan tidak mengukur kondisi *burnout* pada mahasiswa secara langsung.

Penelitian terdahulu mengenai *academic burnout* telah menggunakan beragam tujuan analisis. Chen & Chen, (2025) menerapkan pendekatan fsQCA untuk menelaah konfigurasi kondisi yang berkaitan dengan *academic burnout*, sedangkan Sanjari et al., (2023) melakukan pengelompokan mahasiswa berdasarkan *burnout* dan kepuasan terhadap bidang studi. Kajian pengukuran, seperti López-Gómez et al., (2025) , juga menempatkan instrumen penilaian *burnout* sebagai fokus penelitian. Studi tersebut memberikan bukti substantif dan pendekatan analitis yang berbeda, tetapi unit analisisnya bukan klasifikasi multilabel judul dan abstrak artikel menurut konstruk SD-R. Dalam cakupan literatur yang ditelaah, evaluasi TF–IDF dan Linear SVM untuk tujuan khusus tersebut belum menjadi fokus pembahasan. Pernyataan ini menunjukkan ruang kajian yang dituju, bukan klaim bahwa belum pernah ada penelitian klasifikasi teks terkait *burnout*.

Dalam perspektif *Human Resource Development* (HRD), pembedaan D, R, dan P berguna untuk menata penelusuran bukti mengenai tuntutan pembelajaran, dukungan kelembagaan, dan kapasitas adaptif mahasiswa. Literatur tentang pengalaman belajar dapat menjadi bahan pertimbangan dalam kajian pengembangan kompetensi dan kesejahteraan mahasiswa sebelum memasuki dunia kerja. Namun, hubungan tersebut bersifat konseptual dan informasional. Hasil klasifikasi artikel tidak membuktikan efektivitas program pendampingan, tidak mengukur kualitas sumber daya manusia, dan tidak menggantikan penelaahan hasil penelitian primer sebelum suatu keputusan pengembangan diambil.

Berdasarkan uraian tersebut, penelitian ini bertujuan mengevaluasi penerapan TF–IDF dan Linear SVM untuk klasifikasi multilabel artikel *academic burnout* berdasarkan konstruk D, R, dan P. Kinerja dinilai melalui *precision*, *recall*, dan *F1-score* per kategori, *macro F1*, *micro F1*, serta matriks kebingungan. Kontribusi penelitian diarahkan pada evaluasi kemampuan dan batas klasifikasi dokumen berdasarkan kerangka SD-R, termasuk implikasi variasi interpretasi pelabelan terhadap penggunaan hasil untuk penelusuran literatur.

2. METODE PENELITIAN

2.1 Desain Penelitian dan Sumber Data

Penelitian ini menggunakan pendekatan kuantitatif komputasional untuk mengevaluasi klasifikasi *multilabel* artikel ilmiah mengenai *academic burnout* berdasarkan kerangka *Study Demands-Resources* (SD-R). Unit analisis adalah artikel ilmiah yang direpresentasikan melalui judul dan abstrak. Kategori klasifikasi meliputi *Study Demands* (D), *Study Resources* (R), dan *Personal Resources* (P).

Data bibliografis diperoleh dari OpenAlex sebanyak 955 rekaman, Scopus 291 rekaman, dan PubMed 174 rekaman, sehingga terkumpul 1.420 rekaman awal. Setelah penghapusan duplikasi, diperoleh 1.110 artikel unik. Dari jumlah tersebut, 846 artikel memiliki abstrak dan 264 artikel tidak memiliki abstrak.

Sebanyak 113 artikel dipilih secara terarah (*purposive sampling*) berdasarkan penilaian *reviewer* terhadap relevansi substansi artikel dengan *academic burnout* dan konstruk SD-R. Kriteria inklusi mencakup artikel yang membahas *academic burnout* mahasiswa, memiliki judul dan abstrak yang dapat dianalisis, serta dinilai relevan dengan tujuan klasifikasi. Rekaman duplikat, artikel tanpa abstrak, dan artikel yang tidak memenuhi kriteria relevansi tidak digunakan dalam *dataset* eksperimen.

Pemilihan 113 artikel didasarkan pada penilaian relevansi, bukan pengambilan sampel acak. Oleh karena itu, (Page et al., 2021)*dataset* eksperimen tidak dimaksudkan untuk merepresentasikan keseluruhan publikasi *academic burnout*. Jumlah artikel pada setiap tahap pengelolaan data dilaporkan secara deskriptif dan tidak diklaim sebagai alur seleksi tinjauan sistematis yang sepenuhnya memenuhi PRISMA 2020.

2.2 Definisi Operasional dan Pelabelan

Pelabelan dilakukan berdasarkan tiga konstruk SD-R. *Study Demands* (D) mencakup tuntutan akademik yang dianalisis dalam artikel, seperti beban tugas, tekanan evaluasi, dan tuntutan waktu. *Study Resources* (R) mencakup sumber daya eksternal yang mendukung kegiatan belajar, seperti dukungan dosen, teman sebaya, dan institusi. *Personal Resources*

Copyright © 2026 the author, Page 147
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 146-151
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

(P) mencakup kapasitas psikologis positif individu, seperti efikasi diri, resiliensi, dan strategi *coping* adaptif. Setiap kategori dinyatakan dalam label biner, yaitu 1 apabila konstruk terkait diidentifikasi dalam artikel dan 0 apabila tidak diidentifikasi. Ketiga kategori bersifat tidak saling eksklusif sehingga satu artikel dapat memperoleh lebih dari satu label positif. Penilaian independen oleh *reviewer* menunjukkan perbedaan penetapan sedikitnya satu label pada 68 dari 113 artikel. Perbedaan tersebut dipahami sebagai variasi interpretasi terhadap konstruk yang dibahas, bukan sebagai bukti kesalahan salah satu *reviewer*. Kondisi ini mencerminkan ambiguitas konseptual yang dapat memengaruhi konsistensi anotasi. Label dalam *dataset* eksperimen diperlakukan sebagai acuan operasional untuk pelatihan dan evaluasi model, bukan sebagai kebenaran konseptual yang mutlak.

2.3 Representasi Teks dan Klasifikasi

Judul dan abstrak setiap artikel digabungkan sebagai masukan klasifikasi. Teks direpresentasikan menggunakan *Term Frequency-Inverse Document Frequency* (*TF–IDF*) dengan fitur *unigram* dan *bigram*. Pendekatan ini menghasilkan representasi numerik berdasarkan bobot istilah yang muncul dalam dokumen. *Dataset* berlabel dibagi menjadi 90 artikel latih dan 23 artikel uji, dengan proporsi sekitar 80:20 dan *random seed* 42. Klasifikasi menggunakan *Linear Support Vector Machine* (*Linear SVM*) untuk memprediksi tiga label secara independen. Prediksi setiap kategori kemudian digabungkan menjadi keluaran *multilabel* sehingga satu artikel dapat diklasifikasikan ke dalam beberapa konstruk SD-R. Penelitian ini mengevaluasi penerapan *TF–IDF* dan *Linear SVM* pada *dataset* yang tersedia, tanpa melakukan perbandingan dengan algoritma klasifikasi lain maupun pemetaan manual.

2.4 Evaluasi Model

Kinerja model dievaluasi menggunakan *precision*, *recall*, dan *F1-score* pada masing-masing kategori. *Precision* menunjukkan proporsi prediksi positif yang tepat, sedangkan *recall* menunjukkan proporsi kasus positif aktual yang berhasil dikenali. *F1-score* digunakan untuk menilai keseimbangan kedua metrik tersebut. Evaluasi keseluruhan menggunakan *macro F1* dan *micro F1*. *Macro F1* dihitung sebagai rata-rata F1 dari ketiga kategori, sedangkan *micro F1* dihitung berdasarkan gabungan hasil klasifikasi biner seluruh kategori. Analisis kesalahan dilakukan melalui matriks kebingungan setiap kategori yang terdiri atas *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Hasil evaluasi digunakan untuk menggambarkan kinerja klasifikasi pada 23 artikel uji, dengan mempertimbangkan keterbatasan ukuran sampel dan variasi interpretasi dalam pelabelan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi Klasifikasi *Multilabel*

Eksperimen dilakukan terhadap 113 artikel berlabel, terdiri atas 90 artikel latih dan 23 artikel uji. Evaluasi menggunakan tiga kategori *Study Demands-Resources* (SD-R), yaitu *Study Demands* (D), *Study Resources* (R), dan *Personal Resources* (P). Hasil klasifikasi menunjukkan perbedaan kinerja antarkategori sebagaimana disajikan pada Tabel 1.

Tabel 1. Kinerja Klasifikasi per Kategori

Kategori	Precision (%)	Recall (%)	F1-score (%)	Support
Study Demands (D)	73,33	84,62	78,57	13
Study Resources (R)	60,00	66,67	63,16	9
Personal Resources (P)	75,00	83,33	78,95	18

Kategori P memperoleh F1 tertinggi sebesar 78,95%, diikuti D sebesar 78,57%. Selisih keduanya hanya 0,38 poin persentase sehingga tidak cukup untuk menyatakan bahwa P memiliki keunggulan kinerja yang berarti. Sebaliknya, R mencatat F1 terendah sebesar 63,16%, atau 15,79 poin persentase di bawah P. Perbedaan ini menunjukkan bahwa kemampuan model mengidentifikasi konstruk SD-R belum merata.

Tabel 2. Ringkasan Evaluasi *Multilabel*

Ukuran	Precision (%)	Recall (%)	F1-score (%)
Micro-average	71,11	80,00	75,29
Macro-average	69,44	78,21	73,56

Nilai *micro F1* sebesar 75,29% dan *macro F1* sebesar 73,56% memperlihatkan kinerja agregat yang relatif berdekatan. Namun, kedua ukuran tersebut belum menggambarkan seluruh karakteristik kesalahan model. *Macro F1* memberikan bobot yang sama kepada setiap kategori, sedangkan *micro F1* menggabungkan keputusan klasifikasi seluruh label. Oleh karena itu, hasil per kategori dan matriks kebingungan tetap diperlukan untuk menilai keterbatasan model.

Tabel 3. Matriks Kebingungan per Label

Label	TN	FP	FN	TP
D	6	4	2	11

Copyright © 2026 the author, Page 148
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 146-151
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

Label	TN	FP	FN	TP
R	10	4	3	6
P	0	5	3	15

Matriks kebingungan memperlihatkan perbedaan penting yang tidak terlihat dari peringkat F1. Pada kategori D, model mengenali 11 dari 13 artikel positif, tetapi masih menghasilkan empat prediksi positif yang keliru. Pada kategori R, tiga dari sembilan artikel positif tidak berhasil dikenali. Kondisi ini menunjukkan risiko terlewatnya artikel yang membahas sumber daya lingkungan belajar apabila prediksi model digunakan sebagai satu-satunya dasar penyaringan.

Kategori P memperlihatkan persoalan berbeda. Model berhasil mengenali 15 dari 18 artikel positif, tetapi seluruh lima artikel negatif diprediksi sebagai positif. Dengan demikian, F1 P yang tertinggi tidak disertai kemampuan mengenali kasus negatif pada data uji. Temuan ini menunjukkan bahwa skor agregat maupun F1 per kategori perlu dibaca bersama jenis kesalahan prediksinya.

3.2 Pembahasan Kinerja dan Karakteristik Konstruk SD-R

Perbedaan kinerja antarkategori menunjukkan bahwa representasi *TF–IDF* dan *Linear SVM* belum memiliki kemampuan yang sama dalam mengenali ketiga konstruk SD-R. Pada kategori D, istilah seperti *academic workload*, *academic stress*, dan tekanan evaluasi dapat menjadi penanda tekstual yang berkaitan dengan tuntutan studi. Namun, kemunculan istilah tersebut tidak selalu menunjukkan bahwa tuntutan akademik menjadi variabel yang dianalisis. Artikel dapat menyebut stres sebagai latar belakang masalah tanpa menjadikannya konstruk utama penelitian.

Kategori R memiliki (Song et al., 2022b; Ye et al., 2021) *precision* dan *recall* paling rendah. Sumber daya studi dapat dinyatakan melalui beragam istilah, seperti *social support*, *peer support*, dukungan dosen, atau dukungan institusional. Selain variasi istilah, suatu bentuk dukungan dapat ditempatkan sebagai variabel utama, variabel pendamping, maupun konteks penelitian. Keragaman tersebut berpotensi menyulitkan pengenalan kategori berdasarkan bobot kata. Akan tetapi, penelitian ini belum melakukan analisis kesalahan pada tingkat artikel untuk membuktikan bahwa variasi terminologi merupakan penyebab rendahnya F1 kategori R.

Pada kategori P, persoalan utamanya bukan kemampuan menemukan artikel positif, melainkan ketidakmampuan model mengenali artikel negatif pada data uji. Hasil ini penting karena konstruk psikologis individual tidak seluruhnya merupakan sumber daya personal. Efikasi diri, resiliensi, dan (Jeon et al., 2022; Önder et al., 2025) *coping* adaptif dapat memenuhi definisi P, sedangkan kecemasan, kesepian, dan kondisi psikologis negatif lainnya tidak otomatis termasuk kategori tersebut. Apabila pembedaan ini tidak diperhatikan, keluaran klasifikasi berisiko memperluas cakupan P melampaui definisi konseptualnya.

Dengan demikian, temuan eksperimen tidak cukup diringkas sebagai keberhasilan klasifikasi *multilabel*. Kinerja D dan P yang berdekatan, kelemahan identifikasi R, serta kegagalan mengenali kasus negatif P memperlihatkan bahwa setiap kategori memerlukan interpretasi tersendiri. Model dapat membantu mengidentifikasi artikel yang berpotensi relevan, tetapi hasil prediksinya belum layak menggantikan pembacaan substantif.

3.3 Ambiguitas Pelabelan dan Batas Operasional *Codebook*

Persoalan penting lainnya terletak pada penetapan label acuan. Penilaian independen oleh dua *reviewer* menunjukkan perbedaan sedikitnya satu label pada 68 dari 113 artikel. Angka tersebut menggambarkan frekuensi perbedaan penilaian dalam *dataset* eksperimen, bukan tingkat kesalahan *reviewer* atau estimasi ketidaksepakatan pada seluruh literatur *academic burnout*. Perbedaan penilaian dapat muncul karena satu artikel membahas beberapa konstruk yang berdekatan atau karena peran suatu variabel tidak dinyatakan secara tegas dalam judul dan abstrak.

Dalam konteks ini, *codebook* perlu dipahami sebagai perangkat operasional untuk membatasi makna kategori, bukan sekadar daftar istilah. Penetapan label D bergantung pada apakah tuntutan akademik benar-benar menjadi konstruk yang dianalisis, bukan hanya disebutkan sebagai latar belakang. Label R mensyaratkan adanya sumber daya yang berasal dari lingkungan studi, sedangkan label P dibatasi pada kapasitas psikologis positif individu. Ketiga kategori tersebut juga tidak saling eksklusif sehingga artikel yang menganalisis tuntutan akademik dan dukungan lingkungan dapat memperoleh label D dan R secara bersamaan.

Batas kategori menjadi lebih sulit diterapkan ketika judul dan abstrak tidak menjelaskan kedudukan suatu variabel. Sebagai contoh, istilah *coping* dapat merujuk pada kapasitas adaptif individu, tetapi tidak setiap penyebutan *coping* memberikan informasi yang cukup untuk menetapkan P=1. Demikian pula, istilah *support* belum tentu menunjukkan sumber daya lingkungan studi apabila sumber dukungan dan perannya tidak dijelaskan. Contoh-contoh tersebut menunjukkan titik keputusan yang perlu diperhatikan dalam penerapan *codebook*, bukan bukti bahwa kasus tertentu merupakan penyebab perbedaan penilaian pada 68 artikel.

Perbedaan interpretasi *reviewer* memiliki implikasi langsung terhadap evaluasi klasifikasi. Model mempelajari hubungan antara teks dan label yang tersedia dalam *dataset*. Apabila batas konseptual suatu label memungkinkan penafsiran berbeda, hasil prediksi yang tidak sesuai dengan label acuan belum tentu mencerminkan ketidakmampuan model mengenali seluruh makna artikel. Sebaliknya, prediksi yang sesuai dengan label acuan juga tidak membuktikan bahwa model telah memahami kedudukan konstruk secara konseptual.

Oleh karena itu, skor evaluasi dalam penelitian ini menunjukkan kesesuaian prediksi terhadap label acuan operasional, bukan validitas mutlak penafsiran konstruk SD-R. Penelitian belum menyediakan analisis perbedaan penilaian menurut kategori maupun rincian penyelesaian perbedaan label. Keterbatasan tersebut membatasi kesimpulan

Copyright © 2026 the author, Page 149
This Journal is licensed under a Creative Commons Attribution 4.0 International License

mengenai konsistensi *codebook*, tetapi tidak menghilangkan nilai hasil eksperimen sebagai evaluasi awal klasifikasi *multilabel* pada *dataset* yang digunakan.

3.4 Implikasi bagi Penelusuran Literatur dan Pengembangan SDM

Hasil penelitian menunjukkan bahwa klasifikasi *multilabel* dapat digunakan untuk membantu pengorganisasian artikel *academic burnout* berdasarkan tuntutan studi, sumber daya lingkungan, dan sumber daya personal. Pemisahan ini memungkinkan peneliti menelusuri konstruk yang berbeda tanpa memaksakan satu kategori tunggal pada artikel yang membahas beberapa faktor sekaligus.

Meskipun demikian, penggunaan model perlu disesuaikan dengan karakteristik kesalahan setiap kategori. Pada R, adanya artikel positif yang tidak teridentifikasi menunjukkan bahwa hasil prediksi negatif tidak dapat dijadikan dasar tunggal untuk mengecualikan artikel dari telaah literatur. Pada P, kegagalan mengenali kasus negatif menunjukkan perlunya pemeriksaan kembali artikel yang diprediksi positif. Dengan demikian, keluaran model lebih tepat digunakan sebagai alat bantu pengelompokan dan penentuan prioritas pembacaan.

Dalam perspektif manajemen dan pengembangan sumber daya manusia, pengelompokan D, R, dan P dapat membantu penelusuran penelitian mengenai tuntutan pembelajaran, dukungan kelembagaan, dan kapasitas adaptif mahasiswa. Ketiga tema tersebut relevan dengan proses pengembangan kompetensi di pendidikan tinggi. Namun, manfaat penelitian ini terbatas pada pengorganisasian bukti ilmiah. Hasil klasifikasi tidak membuktikan efektivitas program pendampingan, pengembangan kompetensi, maupun intervensi untuk mengurangi (Javed et al., 2024; Y. Liu et al., 2023) *academic burnout*.

Interpretasi hasil juga dibatasi oleh pemilihan 113 artikel secara terarah dan penggunaan 23 artikel sebagai data uji. Distribusi label pada *dataset* tersebut tidak dapat digunakan untuk menyimpulkan seberapa sering konstruk D, R, dan P muncul dalam keseluruhan publikasi *academic burnout*. Selain itu, perbedaan penilaian *reviewer* menunjukkan bahwa konsistensi definisi operasional tetap menjadi aspek penting dalam penggunaan klasifikasi berbasis teks.

Dengan mempertimbangkan batasan tersebut, kontribusi utama penelitian ini terletak pada identifikasi karakteristik kinerja dan kesalahan klasifikasi untuk masing-masing konstruk SD-R. Evaluasi tidak hanya menunjukkan kemampuan model menghasilkan prediksi *multilabel*, tetapi juga memperlihatkan batas penggunaannya ketika kategori memiliki cakupan konseptual yang berdekatan.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa klasifikasi *multilabel* menggunakan *TF–IDF* dan *Linear SVM* dapat membantu mengelompokkan artikel *academic burnout* berdasarkan konstruk *Study Demands*, *Study Resources*, dan *Personal Resources*. Namun, perbedaan kinerja antarkategori memperlihatkan bahwa keberhasilan klasifikasi tidak cukup dinilai melalui skor agregat. Kesulitan mengenali artikel mengenai sumber daya studi dan kegagalan mengidentifikasi kasus negatif pada kategori sumber daya personal menegaskan pentingnya membaca karakteristik kesalahan sebelum menggunakan hasil model untuk penelusuran literatur. Kontribusi konseptual penelitian terletak pada penerapan kerangka *Study Demands-Resources* sebagai dasar klasifikasi dokumen yang memungkinkan beberapa konstruk dikenali dalam satu artikel. Pendekatan ini mempertahankan perbedaan antara tuntutan akademik, dukungan lingkungan belajar, dan kapasitas psikologis positif mahasiswa. Perbedaan interpretasi antar *reviewer* juga memperlihatkan bahwa batas operasional konstruk tetap menentukan kualitas pelabelan. Dengan demikian, klasifikasi otomatis tidak menghilangkan kebutuhan akan penilaian konseptual, tetapi menyediakan cara terstruktur untuk menerapkan kategori yang telah ditetapkan. Bagi bidang *Human Resource Development* (HRD), hasil penelitian dapat dimanfaatkan untuk membantu penelusuran bukti ilmiah mengenai tuntutan pembelajaran, dukungan kelembagaan, dan pengembangan kapasitas adaptif mahasiswa. Pengelompokan tersebut berguna sebagai tahap awal dalam menyusun kajian mengenai pengembangan kompetensi dan kesejahteraan mahasiswa di pendidikan tinggi. Akan tetapi, keputusan mengenai program pendampingan atau pengembangan sumber daya manusia tetap memerlukan penelaahan terhadap metode dan temuan studi primer. Prediksi kategori artikel tidak dapat diperlakukan sebagai bukti efektivitas suatu intervensi. Pemanfaatan hasil penelitian dibatasi oleh pemilihan artikel secara terarah, jumlah data uji yang terbatas, dan variasi interpretasi dalam pelabelan. Oleh karena itu, model lebih tepat digunakan sebagai alat bantu pengorganisasian literatur daripada sebagai dasar tunggal untuk menentukan relevansi atau menarik kesimpulan substantif. Nilai tambah penelitian ini bukan hanya tersedianya hasil klasifikasi, melainkan juga teridentifikasinya batas kemampuan model dalam membedakan konstruk SD-R yang berdekatan secara konseptual.

REFERENCES

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, *22*(3), 273–285. https://doi.org/10.1037/ocp0000056

Bolatov, A. K., Seisembekov, T. Z., Smailova, D. S., & Hosseini, H. (2022). Burnout syndrome among medical students in Kazakhstan. *BMC Psychol*, *10*(1), 193. https://doi.org/10.1186/s40359-022-00901-w

Chen, J., & Chen, G. (2025). Academic burnout among Chinese college students: A study based on FSQCA method. *Acta Psychologica*, *253*. https://doi.org/10.1016/j.actpsy.2025.104701

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. https://doi.org/10.1007/BF00994018

Cuevas-Caravaca, E., Sánchez-Romero, E. I., & Antón-Ruiz, J. A. (2024). Academic Burnout, Personality, and Academic

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Variables
in University Students. *Eur J Investig Health Psychol Educ*, *14*(6), 1561–1571. https://doi.org/10.3390/ejihpe14060103 Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, *86*(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499

Gu, J., Wu, P., Luo, Y., He, X., Fu, L., Liu, H., Lin, F., Xu, Q., & Wu, X. (2023). Internet addiction, loneliness, and academic burnout among Chinese college students: a mediation model. *Front Psychiatry*, *14*, 1176596. https://doi.org/10.3389/fpsyt.2023.1176596

Javed, M. Q., Ahmad, Z., Muhammad, M., Binrayes, A., Niazi, I., Nawabi, S., Abulhamael, A. M., & Habib, S. R. (2024). Burnout level evaluation of undergraduate dental college students at middle eastern university. *BMC Med Educ*, *24*(1), 1155. https://doi.org/10.1186/s12909-024-06149-9

Jeon, M. K., Lee, I., & Lee, M. Y. (2022). The multiple mediating effects of grit and learning agility on academic burnout and learning engagement among Korean university students: a cross-sectional study. *Ann Med*, *54*(1), 2710–2724. https://doi.org/10.1080/07853890.2022.2122551

Jeon, M., Lee, I., & Lee, M. (2022). The multiple mediating effects of grit and learning agility on academic burnout and learning engagement among Korean university students: a cross-sectional study. *Annals of Medicine*. https://doi.org/10.1080/07853890.2022.2122551

Lesener, T., Pleiss, L. S., Gusy, B., & Wolter, C. (2020). The Study Demands-Resources Framework: An Empirical Introduction. *Int J Environ Res Public Health*, *17*(14), 5183. https://doi.org/10.3390/ijerph17145183

Liu, Y., Zhang, Y., Peng, C., Li, Y., & Tan, Q. (2023a). Cumulative Ecological Risk and Academic Burnout in Chinese College Students: A Moderated Mediation Model. *International Journal of Environmental Research and Public Health*, *20*(3). https://doi.org/10.3390/ijerph20031712

Liu, Y., Zhang, Y., Peng, C., Li, Y., & Tan, Q. (2023b). Cumulative Ecological Risk and Academic Burnout in Chinese College Students: A Moderated Mediation Model. *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph20031712

Marôco, J., Assunção, H., Harju-Luukkainen, H., Lin, S. W., Sit, P. S., Cheung, K. C., Maloa, B., Ilic, I. S., Smith, T. J., & Campos,
J. (2020). Predictors of academic efficacy and dropout intention in university students: Can engagement suppress burnout?
PLoS One. https://doi.org/10.1371/journal.pone.0239816

Olson, N., Oberhoffer-Fritz, R., Reiner, B., & Schulz, T. (2023). Study related factors associated with study engagement and student burnout among German university students. *Front Public Health*, *11*, 1168264. https://doi.org/10.3389/fpubh.2023.1168264

Önder, İ., Önder, A. N., & Güven Yıldırım, E. (2025). Morningness and academic burnout among university students: the mediating role of procrastination. *BMC Psychol*, *13*(1), 630. https://doi.org/10.1186/s40359-025-02969-6

Rahmatpour, P., Chehrzad, M., Ghanbari, A., & Sadat-Ebrahimi, S. R. (2019). Academic burnout as an educational complication and promotion barrier among undergraduate students: A cross-sectional study. *J Educ Health Promot*, *8*, 201. https://doi.org/10.4103/jehp.jehp_165_19

Salmela-Aro, K., Upadyaya, K., Ronkainen, I., & Hietajärvi, L. (2022). Study Burnout and Engagement During COVID-19 Among University Students: The Role of Demands, Resources, and Psychological Needs. *J Happiness Stud*, *23*(6), 2685–2702. https://doi.org/10.1007/s10902-022-00518-1

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Sanjari, E., Majidian Dehkordi, F., & Raeisi Shahraki, H. (2023). Clustering Undergraduate Students Based on Academic Burnout and Satisfaction from the Field Using Partitioning around Medoid. *Comput Math Methods Med*, *2023*, 8898939. https://doi.org/10.1155/2023/8898939

Song, W., Wang, Z., & Zhang, R. (2022). Classroom Digital Teaching and College Students' Academic Burnout in the Post COVID-19 Era: A Cross-Sectional Study. *International Journal of Environmental Research and Public Health*, *19*(20). https://doi.org/10.3390/ijerph192013403

Ye, Y., Huang, X., & Liu, Y. (2021). Social support and academic burnout among university students: A moderated mediation model. *Psychology Research and Behavior Management*, *14*, 335–344. https://doi.org/10.2147/PRBM.S300797

Copyright © 2026 the author, Page 151
This Journal is licensed under a Creative Commons Attribution 4.0 International License