

Klasifikasi Penghasilan Alumni Universitas Telkom Pada Data *Tracer Study* Menggunakan Algoritma *Support Vector Machine* dan *Logistic Regression*

Eric Deo Alamsyah¹, Indwiarti², Hilda Fahlana³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹erickdeoalamsyah@students.telkomuniversity.ac.id, ²indwiarti@telkomuniversity.ac.id,

³hildafahlana@telkomuniversity.ac.id,

Abstrak

Dalam menghadapi perubahan dinamika pada era digital, salah satu tujuan institusi pendidikan menghasilkan lulusan yang terampil dan mampu bersaing dalam dunia kerja. *Tracer Study* bertujuan mengukur umpan balik dari setiap lulusan. Universitas Telkom melakukan *Tracer Study* untuk mendapatkan informasi mengenai kinerja alumni, salah satunya dilihat dari tingkat penghasilan alumni. Pada penelitian ini dilakukan klasifikasi tingkat penghasilan alumni berdasarkan 7 aspek kompetensi lulusan menggunakan metode analisis *Support Vector Machine*, *Logistic Regression*, dan *Logistic Regression* dioptimasi *AdaBoost*. Proses pemodelan menggunakan 7 aspek kompetensi dan 1 aspek tingkat penghasilan alumni Universitas Telkom. Hasil klasifikasi menunjukkan bahwa pada model *Logistic Regression* menghasilkan *accuracy* sebesar 0.30, *precision* 0.30, *recall* 0.30, dan *F1 Score* 0.28. Pada model *Logistic Regression* dioptimasi *AdaBoost*, performa justru mengalami penurunan dengan *accuracy* 0.27, *precision* berkisar antara 0.22 hingga 0.42, *recall* 0.26 hingga 0.28, dan *F1-Score* antara 0.17 hingga 0.39. Model *Support Vector Machine* menunjukkan sedikit peningkatan performa dengan *accuracy* 0.32 hingga 0.36, *precision* 0.43 hingga 0.49, *recall* 0.32 hingga 0.36, dan *F1-Score* 0.37 hingga 0.41. Meskipun model *Support Vector Machine* menunjukkan hasil yang lebih baik, terutama dalam *precision* dan *F1-Score*, ketiga model masih kesulitan dalam membedakan setiap kelas dengan benar. Manfaat dari penelitian ini dapat memberikan rekomendasi kepada institusi, kompetensi apa saja yang berpengaruh terhadap tingkat penghasilan lulusan.

Kata kunci : *Tracer Study*, *Support Vector Machine*, *Logistic Regression*, *AdaBoost*

Abstract

In facing the changing dynamics of the digital era, one of the objectives of educational institutions is to produce graduates who are competent and competitive in the job market. *Tracer Study* aims to measure feedback from each graduate. Telkom University conducted a *Tracer Study* to obtain information about alumni performance, one of which is seen from the level of alumni income. In this study, the classification of alumni income levels based on 7 aspects of graduate competence using the *Support Vector Machine* analysis method, *Logistic Regression*, and *Logistic Regression* optimised by *AdaBoost*. The modelling process uses 7 aspects of competence and 1 aspect of Telkom University alumni income level. The classification results show that the *Logistic Regression* model produces *accuracy* of 0.30, *precision* 0.30, *recall* 0.30, and *F1 Score* 0.28. In the *AdaBoost* optimised *Logistic Regression* model, performance actually decreased with *accuracy* 0.27, *precision* ranging from 0.22 to 0.42, *recall* 0.26 to 0.28, and *F1-Score* between 0.17 to 0.39. The *Support Vector Machine* model showed a slight increase in performance with *accuracy* 0.32 to 0.36, *precision* 0.43 to 0.49, *recall* 0.32 to 0.36, and *F1-Score* 0.37 to 0.41. Although the *Support Vector Machine* model showed better results, especially in *precision* and *F1-Score*, all three models still struggled to distinguish each class correctly. The benefits of this research can provide.

Keywords: *Tracer Study*, *Support Vector Machine*, *Logistic Regression*, *AdaBoost*

1. Pendahuluan

Dalam era digital saat ini, data menjadi peran krusial dalam pengambilan keputusan. Salah satu metode pengumpulan data adalah *Tracer Study*. *Tracer Study* membantu perguruan tinggi melacak informasi para alumni untuk mendapatkan informasi dalam membuat kebijakan yang bermanfaat bagi pengembangan perguruan tinggi [1]. Hal ini bertujuan untuk mengukur dan memantau kinerja serta umpan balik dari para alumni dengan memperoleh indikator yang dibutuhkan dalam dunia kerja.

Penelitian *Tracer Study* penting untuk mengetahui seberapa efektif perguruan tinggi mempersiapkan lulusan untuk bersaing dalam dunia kerja dan menentukan strategi dalam meningkatkan kualitas pendidikan. Penelitian ini menggunakan teknik analisis *Support Vector Machine*, *Logistic Regression*, dan *Logistic Regression* dioptimasi *AdaBoost* pada dataset *Tracer Study*. Temuan dari *Tracer Study* ini dapat memberikan peluang bagi universitas untuk mengambil tindakan yang dapat mengatasi masalah yang dihadapi oleh lulusan, dengan tujuan meningkatkan pengalaman belajar dan proses transisi individu [2].

Pemilihan variabel aspek kompetensi sebagai prediksi klasifikasi kelas didasarkan pada pentingnya relevansi kurikulum yang diajarkan oleh universitas terhadap keberhasilan karir alumni. Kompetensi yang dimiliki oleh alumni mencerminkan kualitas pendidikan dan kesiapan dalam menghadapi dunia kerja. Oleh karena itu variabel kompetensi, seperti kompetensi di bidang ilmu pengetahuan, teknologi informasi, komunikasi, manajemen waktu, kerja sama tim, pengembangan diri, dan kompetensi etika, dipilih untuk mengetahui sejauh mana keterkaitannya dengan kesuksesan karir alumni. Kemudian variabel pendapatan dipilih sebagai variabel target karena dianggap sebagai indikator yang paling jelas dalam mengukur keberhasilan karir alumni.

Logistic Regression merupakan sebuah model klasifikasi probabilistik diskriminatif yang bekerja dengan input *vector* yang bernilai nyata. *Vector* input yang diklasifikasikan dalam model ini disebut fitur, dan tidak ada batasan pada fitur yang berkorelasi [3]. Model ini juga memperoleh fungsi klasifikasi dan informasi yang menjelaskan fungsi tersebut disimpan. Fungsi tersebut mempresentasikan batas antara kelas yang aman dan kelas yang tidak aman [4]. *Support Vector Machine (SVM)* merupakan sebuah proses pembelajaran yang melibatkan hipotesis yang mencakup fungsi linear dalam ruang fitur dengan dimensi tinggi. Proses pelatihan algoritma pembelajaran ini merupakan sebuah optimasi, di mana bias pembelajaran diterapkan. [5]. Selain itu, optimasi *AdaBoost* digunakan sebagai alat pra-pemrosesan yang secara otomatis memilih fitur-fitur paling penting dari data berdimensi tinggi [6].

Penelitian ini mengeksplorasi efektifitas algoritma *Support Vector Machine*, *Logistic Regression*, dan *Logistic Regression* dioptimasi *AdaBoost* dalam mengklasifikasikan variabel pendapatan dalam dataset *Tracer Study* alumni Universitas Telkom. Masalah pada penelitian ini melibatkan pertanyaan bagaimana mengimplementasikan kedua algoritma tersebut dan metode mana yang lebih unggul dalam memprediksi hasil *Tracer Study*. Oleh karena, tujuan penelitian ini mengimplementasikan algoritma *Support Vector Machine*, *Logistic Regression* dan *Logistic Regression* dioptimasi *AdaBoost* untuk mengklasifikasikan variabel pendapatan pada dataset *Tracer Study* dan menganalisis metode mana yang paling efektif dan efisien dengan mempertimbangkan parameter *accuracy*, *precision*, *recall*, dan *F1-Score*. Dengan demikian, Penelitian ini diharapkan mampu memberikan sebuah wawasan baru yang lebih mendalam mengenai efektifitas kurikulum dan program studi dalam mempersiapkan lulusan untuk sukses di pasar kerja.

2. Metodologi

A. Alur Pemodelan

Berikut merupakan alur pembuatan model meliputi tahapan yang dilakukan sebagai berikut:

1. Pengumpulan Data

Penelitian ini menggunakan data *Tracer Study* alumni Universitas Telkom. *Tracer Study* merupakan kegiatan universitas untuk mengetahui informasi para alumni, seperti profil pribadi, data pendidikan, pekerjaan, dan lainnya. Hal ini dilakukan setiap tahun. Memastikan data mencakup fitur (variabel independen) dan label (variabel dependen atau kelas yang akan diprediksi).

2. Pra-Pemrosesan Data

Melakukan pemrosesan data untuk mengatasi *Missing Values*, *Outlier*, atau kesalahan data lainnya. Jika diperlukan, melakukan transformasi data seperti *encoding* variabel kategorikal menjadi numerik.

3. Pembagian Data

Pembagian data adalah pengelompokan data ke dalam kelompok pelatihan, atau validasi sebagai metode evaluasi model. Pendekatan ini memberikan kesempatan untuk menemukan *hyperparameter* model dan juga memperkirakan kinerja generalisasi. Pemilihan algoritma pemisahan data dapat berdampak positif pada perkiraan kinerja secara umum. Setiap algoritma membagi data menjadi dua *subset*, yaitu data latih yang digunakan untuk pelatihan model dan data uji untuk evaluasi.

4. Data Latih

Data latih digunakan untuk melatih model pembelajaran mesin. Selama tahap ini, model menggunakan input dari data latih dan menyesuaikan parameter internalnya untuk melakukan prediksi yang sesuai dengan data yang diberikan.

5. Data Uji

Data uji digunakan untuk menguji kinerja model yang telah menjalani proses pelatihan. Setelah model dilatih menggunakan data latih, data uji digunakan dalam mengevaluasi sejauh mana model dapat menghasilkan sebuah prediksi yang akurat pada data yang belum pernah