

# Klasifikasi dan Prediksi Kemampuan Literasi Siswa dengan Machine Learning

Oleh: Heri Retnawati, Mathilda Susanti, Limala Ratni S. Kharismawati, Purwoko Haryadi Santoso, Kartianom

## ABSTRAK

Literasi merupakan kemampuan dan keterampilan yang penting untuk dikuasai oleh peserta didik di abad ke-21 karena sangat terkait dengan kemampuan pemecahan masalah di berbagai konteks dan dalam kehidupan sehari-hari. Prediksi literasi dengan menggunakan berbagai variabel dan klasifikasinya perlu dilakukan untuk memberikan masukan pada kebijakan pendidikan, namun berdasarkan riset yang ada, hal ini belum banyak dilakukan, termasuk menggunakan *machine learning*. Terkait dengan hal tersebut, penelitian ini bertujuan untuk melakukan analisis prediksi dan klasifikasi literasi berbahasa siswa di Indonesia dengan menggunakan *machine learning*.

Penelitian ini merupakan penelitian deskriptif eksploratif dengan pendekatan kuantitatif. Data penelitian dikumpulkan dengan menggunakan data dokumentasi dari Lembaga Bahasa SEAQIL Indonesia. Variabel yang terlibat dari penelitian ini yakni capaian literasi, sikap pribadi, penggunaan Bahasa, dukungan orang tua, fasilitas sekolah, pemanfaatan perpustakaan, dan kebiasaan membaca siswa yang dikumpulkan dari berbagai instrumen. Validitas instrumen dianalisis dengan menggunakan analisis faktor konfirmatori, dan reliabilitas diestimasi dengan koefisien reliabilitas bersyarat dari IRT. Analisis data dilakukan dengan menggunakan *machine learning* dengan dibantu *software R*, dengan memilah data menjadi dua bagian, yaitu 80% data *training* dan 20% data validasi.

Berdasarkan hasil penelitian, dapat disimpulkan bahwa pada prediksi menggunakan model LR, terdapat lima variabel yang berkaitan dengan faktor siswa, tiga variabel berkaitan dengan faktor sekolah, dan dua variabel berkaitan dengan faktor provinsi. Terlihat bahwa model LR didominasi oleh variabel-variabel yang berkaitan dengan faktor siswa dalam menjelaskan capaian literasi siswa, namun tingkat kepentingan atau daya prediksinya tidak setinggi variabel dari faktor sekolah dan provinsi. Di sisi lain, model SVM, KNN, dan RF justru mengidentifikasi tiga variabel yang berkaitan dengan faktor siswa, lima variabel berkaitan dengan faktor sekolah, dan dua variabel berkaitan dengan faktor provinsi. Ketiga model ini menunjukkan bahwa variabel-variabel yang berkaitan sekolah dan provinsi merupakan faktor krusial dalam menjelaskan capaian literasi siswa. Pada klasifikasi, hasil penelitian menunjukkan bahwa variabel non-kognitif seperti kualitas sekolah dan asal daerah siswa digabung dengan nilai siswa pada mata pelajaran Bahasa merupakan fitur terpenting dalam memprediksi kemampuan literasi siswa Indonesia. Variabel non-kognitif mampu menghasilkan kinerja prediksi yang paling baik dibanding jika kita menggombinasikannya dengan hasil penilaian kognitif di dalam model. Implikasi dalam tataran teoritis dan praktik dari hasil temuan didiskusikan dalam artikel ini.

Kata Kunci: *Literasi, klasifikasi dan prediksi, machine learning*