

KLASIFIKASI KEMAMPUAN LULUSAN SMK DI INDUSTRI MENGUNAKAN EXTREME GRADIENT BOOSTING (XGBOOST), RANDOM FOREST, DAN LOGISTIC REGRESSION

Oleh:

Afikah Agustiningsih,

Yulian Findawati

Informatika

Universitas Muhammadiyah Sidoarjo

April, 2023



Pendahuluan

Sumber Daya Manusia (SDM) merupakan aset yang paling penting dalam kemajuan pembangunan suatu Negara, khususnya pada suatu organisasi/lembaga pendidikan, sehingga sangat penting agar sumber daya manusia yang ada dapat direncanakan sebaik-baiknya, demi terwujudnya tujuan organisasi atau lembaga. Sumber Daya Manusia disadari sepenuhnya memiliki dampak yang sangat besar dan dianggap sebagai kunci utama dalam meningkatkan mutu pendidikan [1]. Salah satu penyebab baik tidaknya SDM suatu Negara adalah dengan melihat kualitas pendidikan.

Pendidikan memegang peranan penting dalam kemajuan suatu Negara, dimana hal ini menjadi suatu kepentingan bagi negara yang ingin berkembang, maju serta mampu bersaing secara global [2]. Oleh karena itu diperlukan adanya sistem yang dapat membantu pihak sekolah untuk melakukan klasifikasi lulusan peserta didik sebagai bahan untuk evaluasi lulusan di tahun berikutnya.

Pertanyaan Penelitian (Rumusan Masalah)

- Bagaimana cara membangun sistem untuk klasifikasi kemampuan lulusan SMK bersaing di Industri ?
- Bagaimana cara Randomized Search Cross Validation bekerja dalam mencari parameter terbaik ?
- Bagaimana cara melakukan klasifikasi dengan algoritma XGBoost, Random Forest dan Logistic Regression ?

Metode

Penelitian ini menggunakan beberapa metode diantaranya yaitu:

- Tahap preprocessing menggunakan encoding categorical data, handling *imbalanced* data, normalisasi data dengan minmax scaler
- Randomized Search Cross Validation untuk melakukan tuning parameter
- XGBoost, Random Forest dan Logistic Regression untuk melakukan klasifikasi lulusan SMK

Hasil

Pembagian data pada penelitian ini terbagi menjadi 2 yaitu data train 80% dan data test 20%. Pengujian dilakukan menggunakan confusion matrix. Berikut merupakan table perbandingan yang dihasilkan oleh algoritma XGBoost, Random Forest dan logistic Regression.

Algoritma	Accuracy	precision	recall	F1 score
XGBoost	67%	67%	67%	67%
Random Forest	67%	67%	67%	67%
Logistic Regression	50%	50%	50%	50%

Pembahasan

Pada tabel diatas dapat diketahui hasil dari pengukuran tiga metode machine learning dalam confusion matrix skor akurasi, precision, recall dan F1-score paling tinggi yaitu 67% yang diperoleh dari algoritma XGBoost dan Random Forest, sedangkan untuk algoritma Logistic Regression mendapatkan skor 50% untuk skor akurasi, precision, recall dan F1-score.

Temuan Penting Penelitian

Data yang digunakan pada penelitian merupakan data credential yang diambil dari SMK Negeri 1 Bangil. Data tersebut terdiri dari 1126 record dengan beberapa indikator diantaranya adalah tahun kelulusan, jurusan, USBN pengetahuan, USBN keterampilan, rata-rata rapor, NUS dan NS. Sedangkan data status dari tracer study merupakan target perbandingan hasil klasifikasi. Penelitian ini menggunakan data peserta didik kelas 12 tahun ajaran 2020 dan 2021.

Manfaat Penelitian

Penelitian ini di harapkan dapat bermanfaat bagi lembaga sekolah untuk menyusun strategi agar menghasilkan kualitas peserta didik yang lebih baik di tahun berikutnya sehingga dapat bersaing di Industri sesuai dengan kompetensi keahlian masing-masing.

Referensi

- [1] Y. Sangsurya, M. Muazza, and R. Rahman, "Perencanaan Sumber Daya Manusia Dalam Peningkatan Mutu Pendidikan Di Sd Islam Mutiara Al Madan Kota Sungai Penuh," *J. Manaj. Pendidik. Dan Ilmu Sos.*, vol. 2, no. 2, pp. 766–778, 2021, doi: 10.38035/jmpis.v2i2.644.
- [2] S. A. Nurfatimah, S. Hasna, and D. Rostika, "Membangun Kualitas Pendidikan di Indonesia dalam Mewujudkan Program Sustainable Development Goals (SDGs)," *J. Basicedu*, vol. 6, no. 4, pp. 6145–6154, 2022, doi: 10.31004/basicedu.v6i4.3183.
- [3] H. Priyono, R. Sari, and T. Mardiana, "Klasifikasi Pemilihan Jurusan Sekolah Menengah Kejuruan Menggunakan Gradient Boosting Classifier," *J. Inform.*, vol. 9, no. 2, pp. 131–139, 2022, doi: 10.31294/inf.v9i2.12654.
- [4] L. W. Kusuma, "Prediksi Kemampuan Lulusan SMK untuk Dapat Bersaing Di Dunia Kerja dengan Menggunakan Naïve Bayes : Studi Kasus SMK Buddhi Tangerang," *Prediksi Kemamp. Lulusan SMK untuk Dapat Bersaing Di Dunia Kerja dengan Menggunakan Naïve Bayes Stud. Kasus SMK Buddhi Tangerang*, vol. 1, pp. 56–63, 2019.
- [5] Y. Septiani and P. F. Ariyani, "Penerapan Algoritma Naive Bayes Menentukan Klasifikasi Tingkat Kelulusan Siswa SMK Media Informatika Jakarta Application of The Naive Bayes Algorithm Determining Classification of Students' Graduation Level of Jakarta Media Informatika Vocational School," no. September, pp. 607–613, 2022.
- [6] E. Purwaningsih and E. Nurelasari, "Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa," *Syntax J. Inform.*, vol. 10, no. 01, pp. 46–56, 2021, doi: 10.35706/syji.v10i01.5173.
- [7] I. Muslim and K. Karo, "Implementasi Metode XGBoost dan Feature Importance untuk Klasifikasi pada Kebakaran Hutan dan Lahan," *J. Softw. Eng. Inf. Commun. Technol.*, vol. 1, no. 1, pp. 10–16, 2020.
- [8] M. Rizky Mubarak, Muliadi, and R. Herteno, "Hyper-Parameter Tuning pada XGBoost Untuk Prediksi Keberlangsungan Hidup Pasien Gagal Jantung," *Kumpul. J. Ilmu Komput.*, vol. 9, no. 2, pp. 391–401, 2022.
- [9] P. R. Sihombing and I. F. Yuliaty, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 417–426, 2021, doi: 10.30812/matrik.v20i2.1174.
- [10] A. K. Santoso, A. Noviriandini, A. Kurniasih, B. D. Wicaksono, and A. Nuryanto, "Klasifikasi Persepsi Pengguna Twitter Terhadap Kasus Covid-19 Menggunakan Metode Logistic Regression," *JIK (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 234–241, 2021.
- [11] A. Tjalla and M. Mahdiyah, "Data Kategorik dalam Penelitian : Review Bibliometrik," vol. 9, no. 1, pp. 796–802, 2023, doi: 10.58258/jime.v9i1.4814/http.
- [12] A. N. Kasanah, M. Muladi, and U. Pujiyanto, "Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 196–201, 2019, doi: 10.29207/resti.v3i2.945.
- [13] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [14] E. Agustin, A. Eviyanti, and N. L. Azizah, "Deteksi Penyakit Epilepsi Melalui Sinyal EEG Menggunakan Metode DWT dan Extreme Gradient Boosting," vol. 7, pp. 117–127, 2023, doi: 10.30865/mib.v7i1.5412.
- [15] M. Noveanto, H. Sastypratiwi, and H. Muhardi, "Uji Akurasi Klasifikasi Emosi Pada Lirik Lagu Bahasa Indonesia Emotion Classification Accuracy Test in Indonesian Song Lyrics," vol. 10, no. 3, pp. 311–318, 2022, doi: 10.26418/justin.v10i3.56804.

