

Comparison of Models for Classification of Learning Achievement of Middle School Students in Indonesia in 2019 using the Support Vector Machine Algorithm, Conditional Inference Trees, and Random Forest

Perbandingan Model untuk Klasifikasi Capaian Belajar Siswa SMP di Indonesia Tahun 2019 menggunakan Algoritma *Support Vector Machine*, *Conditional Inference Trees*, dan *Random Forest*

Alfina Nurpiana^{1*}, Arie Wahyu Wijayanto^{2*}

Abstract

Indonesian JHS students' learning achievement is still low. During 2015-2019, the average national exam score for Indonesian JHS has always decreased. In the last national examination, the average national exam score was 52.82 and was included in the bad category. This certainly needs to be a concern for local governments and the education office. Therefore, it is necessary to form a classification model that can be used to identify cities/districts in Indonesia which are categorized as bad or enough. This study discusses the comparison of models for the classification of learning achievement categories as seen from the average 2019 JHS results in 514 districts/cities in Indonesia using the Support Vector Machine (SVM), Conditional Inference Trees (Ctree), and Random Forest (RF) algorithms. The three algorithms were chosen because of their respective advantages, namely the SVM algorithm is known to be very powerful, Ctree as an improvement from the usual decision tree, and RF to represent ensemble learning. The independent variables used are education budget, classroom conditions, school accreditation, and teacher qualifications. From the results of this study, it has been found that the SVM algorithm produces the highest accuracy (0,80), recall (0,97), kappa statistics (0,38), and F1-score (0,87) compared to the Ctree and RF algorithms, while only precision (0,80) has the same value as the Ctree algorithm. So, the SVM algorithm produces the best model for the classification of district/city learning achievement categories in Indonesia based on education budget, classroom conditions, school accreditation, and teacher qualifications.

Keywords: learning achievements, classification, SVM, conditional inference tree, random forest

* Politeknik Statistika STIS

Email address: ¹211810142@stis.ac.id, ²ariewahyu@stis.ac.id



Abstrak

Capaian belajar siswa SMP Indonesia masih rendah. Pada rentang 2015-2019, rata-rata nilai UN SMP Indonesia selalu mengalami penurunan. Pada UN terakhir, rata-rata nilai UN SMP Indonesia 2019 sebesar 52,82 dan termasuk dalam kategori kurang. Hal ini tentu perlu menjadi perhatian bagi pemerintah daerah dan dinas pendidikan. Oleh karena itu, perlu dibentuk sebuah model klasifikasi yang dapat digunakan untuk mengidentifikasi kota/kabupaten di Indonesia termasuk kategori kurang atau cukup. Penelitian ini membahas perbandingan model untuk klasifikasi kategori capaian belajar yang dilihat dari rata-rata nilai hasil UN SMP tahun 2019 di 514 kabupaten/kota di Indonesia menggunakan algoritma *Support Vector Machine* (SVM), *Conditional Inference Trees* (Ctree), dan *Random Forest* (RF). Ketiga algoritma itu dipilih karena kelebihan masing-masing, yaitu algoritma SVM dikenal sangat *powerful*, Ctree sebagai perbaikan dari *decision tree* biasa, serta RF untuk mewakili *ensemble learning*. Variabel independen yang digunakan yaitu anggaran pendidikan, kondisi ruang kelas, akreditasi sekolah, dan kualifikasi guru. Dari hasil penelitian ini telah didapatkan bahwa algoritma SVM menghasilkan *accuracy* (0,80), *recall* (0,97), *kappa statistics* (0,38), dan *F1-score* (0,87) yang tertinggi dibandingkan algoritma Ctree dan RF, sedangkan hanya *precision* (0,80) yang nilainya sama dengan algoritma Ctree. Maka, algoritma SVM menghasilkan model terbaik untuk klasifikasi kategori capaian belajar kabupaten/kota di Indonesia berdasarkan anggaran pendidikan, kondisi ruang kelas, akreditasi sekolah, dan kualifikasi guru.

Kata kunci: capaian belajar, klasifikasi, SVM, *conditional inference tree*, *random forest*

1. PENDAHULUAN

Pendidikan merupakan amanat yang tercantum dalam dasar negara Indonesia, yaitu Undang-Undang Dasar 1945 pasal 31, dimana pendidikan adalah hak dan kewajiban seluruh warga negara [19]. Dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024, pemerintah berkomitmen untuk meningkatkan kualitas sumber daya manusia yang salah satu upayanya melalui peningkatan kualitas pendidikan [18]. Kualitas pendidikan dapat dilihat dari capaian belajar siswa.

Sejak tahun 2011 hingga 2019, evaluasi sistem pendidikan nasional di Indonesia menggunakan Ujian Nasional (UN). Selama lima tahun terakhir (2015-2019), rata-rata nilai UN Sekolah Menengah Pertama (SMP) Indonesia selalu mengalami penurunan. Pada UN terakhir, rata-rata nilai UN SMP Indonesia 2019 sebesar 52,82. Jika dirinci setiap kabupaten/kota di Indonesia, rata-rata nilai UN SMP 2019 hanya ada kategori cukup (jika nilai UN antara 55 dan 71) atau kurang (jika nilai UN kurang dari sama dengan 55), tidak ada yang termasuk ke dalam kategori baik atau sangat baik [11]. Pencapaian kompetensi siswa yang dilihat melalui rata-rata nilai UN tersebut menunjukkan bahwa masih rendahnya kualitas sistem pendidikan di Indonesia.

Terdapat beberapa faktor yang memengaruhi capaian belajar siswa. Kondisi bangunan sekolah, khususnya ruang kelas yang baik akan meningkatkan capaian belajar siswa. Kerusakan bangunan pada ruang kelas terdiri dari kerusakan ringan, kerusakan sedang, dan kerusakan berat. Jika kondisi ruang kelas rusak, maka nilai siswa juga cenderung kurang baik [5]. Lingkungan belajar kelas juga memengaruhi capaian belajar, sesuai dengan hasil penelitian [14] pada siswa menengah. Selain itu, akreditasi sekolah yang merupakan nilai kelayakan dari sekolah juga memengaruhi capaian belajar siswa. Hal tersebut dikarenakan akreditasi merupakan instrumen penting untuk meningkatkan kualitas pendidikan [16]. Hasil penelitian oleh [12] menunjukkan bahwa kualifikasi guru memengaruhi capaian pendidikan siswa. Selain kondisi ruang kelas, akreditasi sekolah, dan kualifikasi guru, anggaran pendidikan juga berpengaruh terhadap kualitas lulusan siswa. Berdasarkan penelitian [20] ditunjukkan bahwa pengeluaran pendidikan berkaitan

dengan peningkatan capaian belajar siswa, meskipun hanya di antara pengeluaran yang ada di bawah *threshold*.

Berdasarkan permasalahan yang ada, maka penelitian ini bertujuan membentuk model untuk klasifikasi capaian hasil belajar siswa yang dilihat dari nilai UN SMP tiap kabupaten/kota tahun 2019 apakah masuk ke dalam kategori kurang atau cukup berdasarkan faktor-faktor yang telah disebutkan menggunakan metode *data mining* sehingga kebijakan yang dibuat pemerintah dapat tepat sasaran. Variabel dependen yang digunakan adalah rata-rata hasil nilai UN yang dibagi menjadi dua kategori, yaitu cukup dan kurang. Adapun variabel independen yang digunakan adalah variabel terkait anggaran pendidikan, kondisi ruang kelas, akreditasi sekolah, dan kualifikasi guru. Penelitian ini akan menggunakan tiga algoritma untuk komparasi hasil klasifikasi, yaitu *Support Vector Machine* (SVM), *Conditional Inference Tree* (Ctree), dan *Random Forest* (RF). SVM adalah teknik yang mengkombinasikan perluasan *solution vector* pada *support vectors* dengan teknik solusi dari pengoptimalan *hyperplane*, konvolusi *dot-product*, dan *soft margins* untuk menerima kesalahan pada *training set* [4]. Kelebihan dari SVM adalah menghasilkan akurasi yang tinggi dan dapat menggunakan data terbatas dengan dimensi tinggi. Sementara itu, Ctree adalah *decision tree* khusus yang merupakan perbaikan dari *decision tree* biasa. Penelitian oleh [21] menunjukkan bahwa Ctree didasarkan pada teori *conditional inference* yang didefinisikan dengan baik. RF menggunakan metode *ensemble* dengan pengembangan dari teknik *bootstrap and aggregating*, dimana *training set* akan disampel menggunakan metode *random sampling with replacement* [8]. Beberapa kelebihan dari RF adalah tidak terjadinya *overfitting* seiring penambahan jumlah pohon dan proses relatif cepat [3].

Banyak penelitian yang telah menggunakan pendekatan SVM, *decision tree*, dan RF untuk melakukan klasifikasi. Penelitian yang dilakukan oleh [17] menunjukkan hasil klasifikasi menggunakan SVM lebih baik dari *decision tree*, dengan akurasi masing-masing sebesar 0,82 dan 0,73. Selain itu, penelitian terkait perbandingan beberapa algoritma klasifikasi (RF, SVM, dan RF-SVM) dengan melihat beberapa evaluasi antara lain *recall*, presisi, dan *F-measure* oleh [1] menunjukkan bahwa metode terbaik berturut-turut adalah RF-SVM, SVM, dan RF

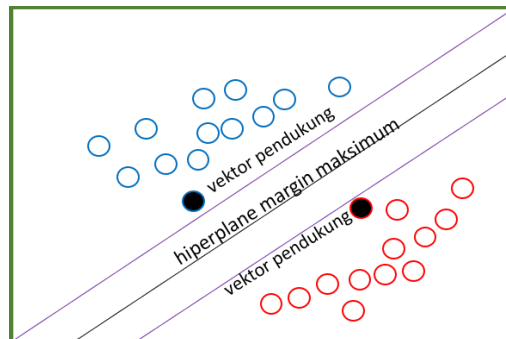
Pada penelitian komparasi perbandingan beberapa metode klasifikasi lain menunjukkan *decision tree* atau RF yang lebih baik. Seperti penelitian oleh [15] yang menunjukkan bahwa akurasi *decision tree* lebih tinggi daripada *k-nearest neighbor*, *naive bayes*, RF, dan SVM. Penelitian lainnya oleh [10] yang mendapati bahwa metode terbaik untuk klasifikasi adalah RF dengan akurasi 95.41% dibandingkan algoritma lainnya, yaitu C4.5 *decision tree*, *k-nearest neighbors*, dan *naive bayes*. Penelitian [7] dan [13] juga menunjukkan bahwa berdasarkan validasi menggunakan *receiver operating characteristics* (ROC) *curve* dan *accuracy*, model RF lebih baik daripada regresi logistik. Sedangkan, pada penelitian [9] menghasilkan bahwa kinerja algoritma SVM dan RF adalah sama baiknya dalam melakukan klasifikasi.

2. METODE PENELITIAN

2.1. Support Vector Machine (SVM)

SVM adalah metode klasifikasi untuk data linear dan nonlinear. SVM mencari hiperplane margin maksimum menggunakan vektor pendukung dan margin [8]. Visualisasi algoritma SVM ditampilkan pada Gambar 2.1. Adapun tahapan kerja algoritma SVM adalah sebagai berikut [2].

1. *Plotting* objek sebagai titik dalam ruang berdimensi n .
2. Dilakukan klasifikasi dengan mendesain hiperplane yang membedakan beberapa target. Dimana dalam ruang dua dimensi, hiperplane adalah garis yang membagi bidang menjadi dua bagian atau target.
3. Hiperplane margin maksimum terbaik akan membagi *dataset* menjadi beberapa target.

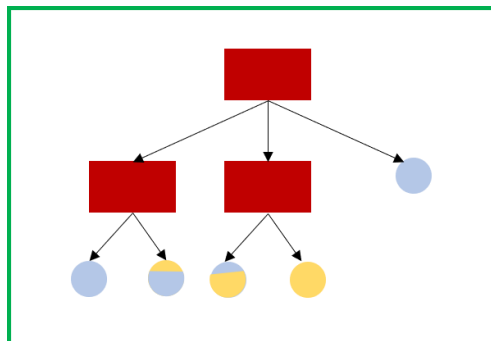


Gambar 2.1. Visualisasi algoritma SVM

2.2. Conditional Inference Trees (Ctree)

Ctree adalah *decision tree* khusus yang menggunakan *recursive partitioning* dari variabel dependen berdasarkan nilai korelasi [6]. Ctree menggunakan uji signifikansi yang merupakan uji permutasi yang memilih kovariat untuk memisahkan dan mengulang variabel [6]. Visualisasi *decision tree* hasil algoritma Ctree ditampilkan pada Gambar 2.2. Adapun tahapan kerja algoritma Ctree adalah sebagai berikut [6].

1. Uji hipotesis nol global antara variabel input dan variabel dependen secara acak kemudian pilih variabel input dengan nilai p tertinggi dengan variabel dependen.
2. Lakukan pemisahan biner pada variabel input yang dipilih.
3. Lakukan langkah 1 dan 2 secara rekursif.

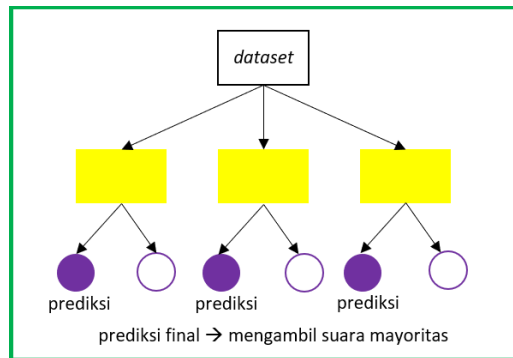


Gambar 2.2. Visualisasi algoritma Ctree

2.3. Random Forest (RF)

RF menggunakan metode *ensemble* dengan pengembangan dari teknik *bagging* (*bootstrap and aggregating*) [8]. *Training set* akan disampel menggunakan metode *random sampling with replacement* sejumlah *tree* yang diinginkan. Penggunaan *bagging* dan *random feature* membuat *uncorrelated forest of decision trees*. Visualisasi algoritma RF ditampilkan pada Gambar 2.3. Adapun tahapan kerja algoritma RF adalah sebagai berikut.

1. Sampel diambil secara *bootstrap* dan didapatkan data ke-1.
2. Dilakukan klasifikasi untuk mendapatkan pohon ke-1 menggunakan data ke-1.
3. Ulangi pengambilan sampel untuk mendapatkan pohon ke-2 hingga pohon ke-k.
4. Setiap pohon menghasilkan prediksi yang berbeda.
5. Dilakukan *aggregating* menggunakan suara mayoritas hasil prediksi dari pohon-pohon tersebut.
6. Target yang mendapatkan suara mayoritas dijadikan prediksi final.



Gambar 2.3. Visualisasi algoritma RF

2.4. Dataset

Dataset terdiri dari data yang diambil dari *website* Neraca Pendidikan Daerah Kemdikbud (<https://npd.kemdikbud.go.id/>). Atribut yang digunakan berjumlah dua belas atribut dan memiliki 514 *record* yang mewakili kabupaten/kota. Atribut kategori rata-rata nilai UN merupakan hasil kategorisasi rata-rata nilai UN tiap kabupaten/kota menjadi dua kategori (kategori kurang dan cukup). Deskripsi dari *dataset* tersebut terdapat dalam Tabel 2.1.

Tabel 2.1. Atribut *Dataset*

No	Nama Atribut	Kode Atribut	Jenis
1	Kode Kabupaten/Kota	kode	Integer
2	Wilayah (Kabupaten/Kota)	wilayah	Karakter
3	Kategori rata-rata nilai UN	kategori	Kategorik 0: kurang 1: cukup
4	Anggaran Pendidikan	anggaran	Numerik
5	Persentase Ruang Kelas dengan Kondisi Rusak Ringan	rusakringan_persen	Numerik
6	Persentase Ruang Kelas dengan Kondisi Rusak Sedang	rusaksedang_persen	Numerik
7	Persentase Ruang Kelas dengan Kondisi Rusak Berat	rusakberat_persen	Numerik
8	Persentase Ruang Kelas dengan Kondisi Rusak Total	rusaktotal_persen	Numerik
9	Persentase Jumlah Sekolah Akreditasi Sekolah "B"	akreb_persen	Numerik
10	Persentase Jumlah Sekolah Akreditasi Sekolah "C"	akrec_persen	Numerik
11	Persentase Jumlah Sekolah yang Tidak Terakreditasi	akrett_persen	Numerik
12	Persentase Guru dengan Pendidikan di bawah D-4/S-1	bukansarjana_persen	Numerik

2.5. Preprocessing

Sebelum membentuk model klasifikasi, dilakukan *preprocessing* yang bertujuan untuk menyiapkan data sehingga dapat siap digunakan untuk analisis. Tahap *preprocessing* meliputi *data cleaning*, *data integration*, *data reduction*, serta *data transformation* dan *data discretization*. Tahap ini bertujuan meningkatkan akurasi model pengklasifikasi.

2.6. Statistik Deskriptif

Statistik deskriptif menyajikan suatu kelompok data untuk menginterpretasi data. Statistik deskriptif dalam penelitian ini digunakan untuk melihat apakah data yang digunakan *imbalance* atau tidak.

2.7. Model Klasifikasi

Model klasifikasi yang diusulkan adalah algoritma SVM, Ctree, dan RF. Data yang sudah lengkap siap diuji menggunakan algoritma SVM, Ctree, dan RF dengan menyeleksi atribut dan

pemilihan label terlebih dahulu. Setelah seleksi data akan diuji validasi dengan menggunakan *data training* dan *data testing*.

2.8. Pengujian dan Evaluasi

Split validation adalah validasi sederhana yaitu dengan membagi *dataset* secara *random* menjadi *data training* dan *data testing* [2]. Data yang telah di-*preprocessing*, selanjutnya dilakukan *training* untuk menghasilkan model klasifikasi. Kemudian, dilakukan *testing* atau pengujian terhadap model klasifikasi yang dihasilkan. *Testing* dilakukan menggunakan *confusion matrix* untuk menguji ketiga model yang dihasilkan. Evaluasi terhadap model tersebut dilihat dari nilai *accuracy*, *precision*, *recall*, dan *kappa statistics*. Selain itu, dilihat pula dari nilai *F1-score* yang dihasilkan. *F1-score* dapat menjadi pertimbangan pada data yang *imbalance*.

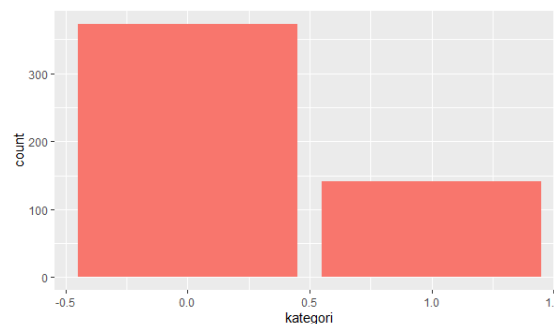
3. HASIL DAN PEMBAHASAN

Pengujian data rata-rata hasil nilai UN SMP ini menggunakan *dataset* sebanyak 514 *record*. Sebelum data digunakan, dilakukan *preprocessing* terlebih dahulu. Dalam *dataset*, tidak terdapat *missing value* sehingga data dapat langsung digunakan. Selain itu, tidak pula dilakukan transformasi karena satuan dari tiap atribut berbentuk persentase serta tidak terpaut terlalu jauh.

Tabel 3.1. *Sample dataset*

kode	wilayah	kategori	anggaran	rusak ringan_persen	rusak sedang_persen	rusak berat_persen	rusak total_persen	akrec_persen	akrett_persen	bukan sarjana_persen
10100	Kab. Kepulauan Seribu	0	0	34,04	0	2,13	0	0	0	2,2
16000	Kota Jakarta Pusat	1	0	50,4	3,15	0,22	0	2,59	0	3,89
16100	Kota Jakarta Utara	1	0	54,72	1,19	1,39	0	2,13	0	3,53
16200	Kota Jakarta Barat	1	0	48,56	2,3	0,69	0	2,14	1,07	4,57
16300	Kota Jakarta Selatan	1	0	45,24	3,04	0,71	0	0,47	0,47	2,71
16400	Kota Jakarta Timur	1	0	47,05	2,38	1,98	0	2,61	1,87	3,19

Atribut kategori merupakan hasil *preprocessing*, dimana atribut kategori (kategori rata-rata nilai hasil UN) dengan *value* kurang menjadi 0 dan *value* cukup menjadi 1. *Dataset* yang akan diuji disajikan pada Tabel 3.1.



Gambar 3.1. Sebaran kategori rata-rata nilai UN SMP pada seluruh data

Pemeriksaan apakah data *imbalance* dilakukan menggunakan aplikasi RStudio dengan fungsi *ggplot*. *Output* dari *running* tersebut digambarkan pada Gambar 3.1. Pada gambar tersebut terlihat bahwa data *imbalance*, yaitu jumlah kategori rata-rata nilai UN SMP kabupaten/kota yang kurang

(kode 0) lebih banyak dua kali lipat dari jumlah kategori rata-rata nilai UN SMP kabupaten/kota yang cukup (kode 1).

Dengan demikian, pada data *imbalance*, perlu untuk melihat evaluasi tertimbangnya, yaitu dalam bentuk *F1-score/Kappa statistics*. Kasus pada penelitian ini mementingkan *false positif* untuk dihindari sehingga selain *accuracy*, *precision* juga perlu diperhatikan saat evaluasi model.

Pengujian menggunakan *confusion matrix* setiap ketiga model ditunjukkan pada tabel 3.2, tabel 3.3, dan tabel 3.4. Hasil klasifikasi model menggunakan tiga teknik *data mining* (SVM, Ctree, dan RF) terhadap target untuk hasil nilai UN SMP kabupaten/kota tahun 2019 ditunjukkan pada tabel 3.5. Adapun validasi menggunakan *accuracy*, *precision*, *recall*, *kappa statistics*, dan *F1-score*. Validasi dilakukan dengan membagi data menjadi dua, yaitu *training data* dan *testing data*. Pembagian data dilakukan secara acak menggunakan fungsi *sample* di Rstudio, dengan pembagian sebesar 60 persen data untuk *training data* dan 40 persen data untuk *testing data*.

Penerapan algoritma SVM pada klasifikasi data rata-rata nilai UN menghasilkan nilai *accuracy* sebesar 0,80 dimana merupakan yang tertinggi dibanding algoritma Ctree dan RF. Nilai *precision* algoritma SVM, Ctree, dan RF berturut-turut adalah 0,80, 0,75, dan 0,80. *Kappa statistics* dan *F1-score* juga dilihat menimbang data yang digunakan untuk pengujian adalah data yang *imbalance*. Hasil pada tabel 3.5 menunjukkan bahwa algoritma SVM memiliki nilai *kappa statistics* dan *F1-score* yang lebih tinggi dibanding algoritma Ctree dan RF.

Tabel 3.2. *Confusion matrix* model SVM

	0 (Kurang)	1 (Cukup)
0 (Kurang)	145	38
1 (Cukup)	4	19

Tabel 3.3. *Confusion matrix* model Ctree

	Kurang	Cukup
0 (Kurang)	137	45
1 (Cukup)	12	12

Tabel 3.4. *Confusion matrix* model RF

	Kurang	Cukup
0 (Kurang)	139	35
1 (Cukup)	10	22

Tabel 3.5. Perbandingan hasil pengujian

	SVM	Ctree	RF
<i>Accuracy</i>	0,80	0,72	0,78
<i>Precision</i>	0,80	0,75	0,80
<i>Recall</i>	0,97	0,92	0,93
<i>Kappa statistics</i>	0,38	0,16	0,36
<i>F1-score</i>	0,87	0,83	0,86

4. KESIMPULAN

Pengujian klasifikasi capaian belajar siswa menggunakan aplikasi Rstudio pada data 514 *record* yang mewakili kabupaten/kota di Indonesia telah dilakukan menggunakan algoritma SVM, Ctree, dan RF. Hasil penelitian menggunakan algoritma SVM menghasilkan *accuracy* sebesar 0,80 dan *precision* sebesar 0,80, algoritma Ctree menghasilkan *accuracy* sebesar 0,72 dan *precision* sebesar 0,75, serta algoritma RF menghasilkan *accuracy* sebesar 0,78 dan *precision* sebesar 0,80. Dengan demikian, diperoleh bahwa berdasarkan evaluasi, algoritma SVM menghasilkan model klasifikasi terbaik.

Dengan mengetahui model terbaik untuk klasifikasi capaian belajar menggunakan hasil nilai UN siswa SMP, ke depannya pemerintah daerah dan dinas pendidikan dapat membuat kebijakan untuk mengatasi masalah kabupaten/kota dengan capaian belajar termasuk kategori kurang dengan melihat beberapa faktor, yaitu anggaran pendidikan, kondisi ruang kelas, akreditasi sekolah, dan kualifikasi guru. Saran untuk penelitian selanjutnya yaitu sebaiknya melakukan komparasi metode klasifikasi lainnya, seperti *neural networks*.

DAFTAR PUSTAKA

- [1] Al Amrani, Y., Lazaar, M. & El Kadiri, K. E. 2018. Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis. *Procedia Computer Science*, Vol. 127, 511-520.
- [2] Aryani, Y. & Wijayanto, A.W. 2021. Klasifikasi Pengembalian Radar dari Ionosfer Menggunakan SVM, Naive Bayes dan Random Forest. *Komputika: Jurnal Sistem Komputer*, Vol. 10, No. 2.
- [3] Breiman, L. & Cutler, A. 2004. RfTools for Predicting and Understanding Data. *Interface Work*, 1-62.
- [4] Cortes, C. & Vapnik, V. 1995. Support-vector networks. *Machine learning*, Vol. 20, No. 3, 273-297.
- [5] Earthman, G. I. 2017. The Relationship Between School Building Condition and Student Achievement: A Critical Examination of the Literature. *Journal of Ethical Educational Leadership*, Vol. 4, No. 3, 1-6.
- [6] GeeksforGeeks. 2020. Conditional Inference Trees in R Programming. GeeksforGeeks. <https://www.geeksforgeeks.org/conditional-inference-trees-in-r-programming/>. [20 Desember 2021]
- [7] Ghosh, A. & Dey, P. 2021. Flood Severity assessment of the coastal tract situated between Muriganga and Saptamukhi estuaries of Sundarban delta of India using Frequency Ratio (FR), Fuzzy Logic (FL), Logistic Regression (LR) and Random Forest (RF) models. *Regional Studies in Marine Science*, Vol. 42, 1-15.
- [8] Han, J., Kamber, M. & Pei, J. 2012. *Data Mining Concepts and Techniques*, Third Edition. Elsevier Inc., San Fransisco.
- [9] Iman, Q. & Wijayanto, A. W. 2021. Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode Random Forest dan Support Vector Machine. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, Vol. 9, No. 2, 178-184.
- [10] Kemala, I. & Wijayanto, A. W. 2021. Perbandingan Kinerja Metode Bagging dan Non-Ensemble Machine Learning pada Klasifikasi Wilayah di Indonesia menurut Indeks Pembangunan Manusia. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, Vol. 9, No. 2, 269-275.
- [11] Kemdikbud. 2019. Hasil UN dan IIUN: 2019. Kemdikbud. <https://npd.kemdikbud.go.id/?appid=hasilun&tahun=2019>. [4 Desember 2021]
- [12] Lee, S. W., & Lee, E. A. 2020. Teacher qualification matters: The association between cumulative teacher qualification and students' educational attainment. *International Journal of Educational Development*, Vol. 77.

- [13] Li, C., Liao, C., Meng, X., Chen, H., Chen, W., Wei, B. & Zhu, P. 2021. Effective Analysis of Inpatient Satisfaction: The Random Forest Algorithm. *Patient Preference and Adherence*, Vol. 15, 691-703.
- [14] Malik, R. H., & Rizvi, A. A. 2018. Effect of Classroom Learning Environment on Students' Academic Achievement in Mathematics at Secondary Level. *Bulletin of Education and Research*, Vol. 40, No. 2, 207-218.
- [15] Maulidah, M., Gata, W., Aulianita, R. & Agustyaningrum, C.I. 2020. Decision Tree Classification Algorithm for Recommended Books by Book Category. *E-Bisnis: Jurnal Ilmiah Ekonomi dan Bisnis*, Vol. 13, No. 2, 89-96.
- [16] Nguyen, C. H. 2016. Accreditation of Schools in Vietnam: Achievements, Challenges, and Lessons Learnt. *Journal of Science of Hnue*, Vol. 61, No. 11, 91-98.
- [17] Nurfaizah, Hariguna, T., & Romadon, Y. I. 2019. The accuracy comparison of vector support machine and decision tree methods in sentiment analysis. *Journal of Physics: Conference Series*, No. 1367.
- [18] Republik Indonesia. Peraturan Presiden Nomor 18 Tahun 2020 tentang Rencana Pembangunan Jangka Menengah Nasional Tahun 2020-2024.
- [19] Republik Indonesia. Undang-Undang Dasar 1945.
- [20] Vegas, E., & Coffin, C. 2015. When Education Expenditure Matters: An Empirical Analysis of Recent International Data. *Comparative Education Review*, Vol. 59, No. 2, 289-304.
- [21] Wang, S., Liu, Y., Cairano-Gilfedder, C. D., Titmus, S., Naim, M. N. & Syntetos, A. A. 2018. Reliability Analysis for Automobile Engines: Conditional Inference Trees. *Procedia CIRP*, Vol. 72, 1392-1397.