

PREDIKSI KEBERADAAN SPESIES AMFIBI DI DEKAT SEBUAH TANDON AIR MENGUNAKAN *DEEP LEARNING*

**Riski Darmawan¹, Hasyir Daffa Ibrahim², Nathanael Victor Darenoh³, Irfan Ardiansyah⁴,
Novanto Yudistira⁵**

^{1, 2, 3, 4, 5} Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹riskidarmawan@student.ub.ac.id, ²melan_gnrx@student.ub.ac.id, ³natvic_dreh@student.ub.ac.id
⁴irfanardiansyah@student.ub.ac.id, ⁵yudistira@ub.ac.id

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak

Spesies amfibi merupakan salah satu indikator kondisi suatu ekosistem karena reaksi natural sebuah hewan amfibi terhadap perubahan alam di sekitarnya. Penelitian ini ditujukan untuk menguji fenomena keberadaan hewan amfibi pada suatu daerah tertentu, seperti tandon air. Data diperoleh melalui GIS atau *Geographic Information System* dan juga Inventarisasi langsung pada tempat dengan membuat laporan analisis dampak lingkungan yang disebut *Environmental Impact Assessment* (EIA). Penggunaan algoritma pembelajaran mesin digunakan untuk memprediksi keberadaan hewan amfibi pada daerah tersebut. Algoritma yang digunakan dalam penelitian mencakup Naive Bayes, Decision Trees dan Support Vector Machine dengan jumlah data sebanyak 189 data serta mempertimbangkan 12 fitur untuk setiap datanya dan mencapai akurasi 46,57% dengan algoritma Naive Bayes, 69,42% dengan algoritma Decision Tree dan 72,85% dengan algoritma Support Vector Machine. Lebih lanjut, algoritma Decision Tree melakukan prediksi dengan nilai presisi paling tinggi pada 56,71%, diikuti dengan algoritma Naive Bayes sebesar 47% dan algoritma Support Vector Machine sebesar 30,71%. Hasil pengujian data menunjukkan bahwa algoritma yang paling cocok untuk digunakan dalam penelitian ini merupakan algoritma Decision Tree karena rerata nilainya yang konsisten dan yang terbaik diantara ketiga algoritma yang dibandingkan. Banyaknya data dan fitur pada dataset tersebut yang akan menentukan baik buruknya hasil dari algoritma klasifikasi yang digunakan. Perbandingan antara jumlah data dan fitur yang tidak seimbang juga akan memiliki hasil klasifikasi yang buruk. Sehingga perlu data yang lebih banyak saat melakukan inventarisasi lapangan sebelum membuat hasil yang akan dijadikan sebagai dataset pengujian sebuah algoritma pembelajaran mesin.

Kata kunci: *Amfibi, Pembelajaran Mesin, Deep Learning, Ekosistem*

PREDICTION OF THE PRESENCE OF AMPHIBIAN SPECIES NEAR WATER RESERVOIR USING DEEP LEARNING

Abstract

Amphibian species indicate the condition of an ecosystem because of its natural reaction of an amphibian towards changes in their natural surroundings. This research examines the phenomenon of an amphibian presence in water reservoirs. The data are obtained through GIS and also a direct inventorize on the spot by making an environmental impact analysis report called Environmental Impact Assessment. The machine learning algorithms are used to predict the presence of amphibians in the area. The Algorithms used includes Naive Bayes, Decision Trees and Support Vector Machine algorithms with a total of 189 instance of data while considering the 12 features for each data reaching an accuracy of 46.57% with the Naive Bayes algorithm, followed by 69.42% with the Decision Trees algorithm and 72.85% with the Support Vector Machine algorithm. Furthermore, the Decision Trees algorithm makes predictions with the highest precision value of 56.71%, followed by the Naive Bayes algorithm at 47% and the Support Vector Machine algorithm at 30.71%. The findings show that the most optimal algorithm is the Decision Trees algorithm based on its consistent average value compared to the others. The amount of data and features of a dataset determines whether the results of a classification algorithm used are good or bad. It is also found that the imbalanced amount of data and features in this dataset has a poor impact on the classification result. Thus, more data is needed when conducting a field inventorize before making a dataset employed as testing data for a machine learning algorithm.

Keywords: *Amphibians, Machine Learning, Deep Learning, Echosystem*

1. PENDAHULUAN

Spesies amfibi merupakan kelas dari hewan bertulang belakang yang dapat hidup di dua alam, yakni di darat dan di air. Sebagian besar amfibi ditemukan di dekat genangan air (Arya et al., 2021). Kemampuan hidup di dua alam itu menyebabkan spesies amfibi memiliki ciri khas ikan dan reptil. Hingga akhir tahun 2018, terdapat hampir 8.000 spesies amfibi yang diketahui, dengan rincian 88% (~7.000) adalah katak (anurans), 9% (~700) adalah salamander, dan sekitar 3% (>200) sisanya adalah caecilian (Wake and Koo, 2018). Daur hidup amfibi dimulai dari telur yang diletakkan di air, kemudian berkembang menjadi berudu yang bernapas menggunakan insang. Saat dewasa, amfibi bernapas menggunakan paru-paru sebagai pernapasan utama dan kulit sebagai pernapasan sekunder.

Amfibi merupakan hewan berdarah dingin seperti reptil. Kulit amfibi cukup unik sehingga membutuhkan kondisi lingkungan tertentu untuk menjadi tempat tinggal serta berkembang biak (Baker et al., 2011). Amfibi disebut juga bio-indikator (indikator alami) lingkungan yang sangat sensitif (Unglaub et al., 2018). Lingkungan yang terlalu banyak terpapar sinar matahari dapat merusak sel hewan amfibi. Di samping itu, terlalu banyaknya hembusan angin dapat menyebabkan dehidrasi sehingga kulit hewan amfibi menjadi kering. Oleh karena itu, amfibi akan menjadi hewan yang pertama kali mati ketika habitatnya terganggu, tidak sesuai, atau terkontaminasi oleh bahan kimia.

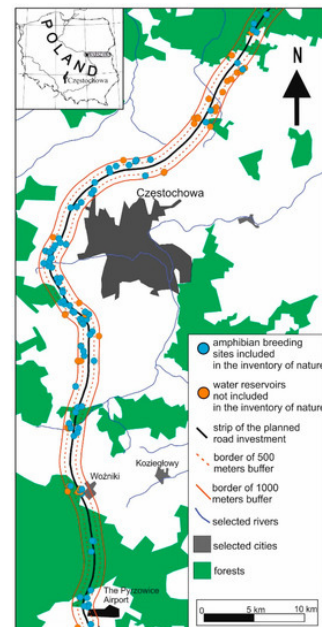
Menurut Huda (2018), berdasarkan perspektif ekologi, beberapa jenis amfibi memiliki sensitivitas yang tinggi terhadap suhu, kelembapan, dan perubahan lingkungan serta berperan sebagai konsumen primer seperti serangga atau hewan invertebrata lainnya. Spesies amfibi akan merespon perubahan kecil yang terjadi di lingkungan sekitar mereka. Respon yang dihasilkan amfibi dapat menunjukkan adanya fragmentasi habitat, *ecosystem stress*, dampak pestisida, dan lain-lain.

Hingga saat ini, cukup banyak penelitian yang menghasilkan solusi mutakhir untuk permasalahan spesies amfibi dan lingkungannya. Namun, kebanyakan studi tersebut mencakup wilayah geografis yang sangat luas (Blachnik, Soltysiak and Dabrowska, 2019). Oleh sebab itu, prediksi keberadaan spesies amfibi tertentu tidak dapat dilakukan secara akurat dan tidak cukup tepat. Isu terkait pemodelan distribusi spesies juga telah dikaji banyak penulis, seperti pembuatan peta kejadian spesies tanpa prediksi hanya dengan melibatkan kerangka kerja GIS (Cogălniceanu et al., 2013; Blank and Blaustein, 2014; Dujsebayaeva and Malakhov, 2017). Selain itu, sudah banyak juga yang mengkaji isu tersebut pada bidang penelitian yang memastikan generalisasi yang dihasilkan tinggi (Munguía et al., 2012; Chen, 2013; Tiago, Pereira and Capinha, 2017; Ali et al., 2018).

Menurut Siahaan, Dewi, and Darmawan, (2019), amfibi memiliki peranan penting dalam menjaga keseimbangan ekosistem yaitu sebagai bagian dari sistem rantai makanan. Selain sebagai konsumen sekunder dalam rantai makanan, amfibi juga membantu manusia dalam memelihara lahan pertanian sebagai pengendali hama serangga. Namun demikian, penelitian tentang keberadaan spesies amfibi dalam menilai keseimbangan alam masih sangat kurang. Mengingat pentingnya peran tersebut, penelitian ini ditujukan untuk memprediksi spesies amfibi apa saja yang ada dalam suatu ekosistem dengan kondisi tertentu, dalam kasus ini ekosistemnya mencakup daerah sekitar sebuah tandon air. Dengan menerapkan beberapa algoritma klasifikasi, diharapkan dapat diketahui algoritma apa yang memiliki performa paling baik untuk memprediksi keberadaan amfibi tersebut.

2. METODE PENELITIAN

Untuk dapat memprediksi keberadaan spesies amfibi, sebuah dataset dikumpulkan oleh inventarisasi lapangan proyek jalan tol A dan B di Polandia. Dataset yang dihasilkan inventarisasi lapangan diperoleh dari GIS (*Geographic Information System*) dan informasi gambar satelit. Selain itu, data juga diperoleh dari informasi yang dikumpulkan oleh inventarisasi alam yang membuat laporan analisis dampak lingkungan atau *Environmental Impact Assessment* (EIA).



Gambar 1. Lokasi bagian jalan raya A1 yang sedang dianalisis

Gambar 1 menunjukkan lokasi bagian jalan raya A1, yaitu lokasi dari proyek jalan A. Proyek jalan A terletak di sepanjang perbatasan utara Provinsi Silesia yang memiliki panjang sekitar 75 km, namun untuk penelitiannya melibatkan hanya

500 m dari area lahan tersebut. Inventarisasi lapangan ini mulai dilakukan pada tahun 2010 dan 2011. Hasil inventarisasi tersebut dilengkapi dengan observasi Marek Sołtysiak yang dilakukan antara tahun 2014 dan 2016. Proyek pertama ini mencakup sebanyak 80 lokasi penangkaran amfibi. Inventarisasi proyek jalan B disiapkan di sekitar dua varian dari ruas jalan raya Beskidy yang direncanakan pada bagian Bielsko-Biala - Wadowice dari jalan raya S52. Panjang jalan ini kurang lebih 60 km. Selama inventarisasi yang dilakukan pada tahun 2010, didapatkan bahwa terdapat 125 lokasi dan potensial munculnya amfibi pada area tersebut.

Metodologi inventarisasi yang digunakan bernama Metode *Herpetology* yang meliputi analisis peta, literatur dan arsip data hasil analisis serta melakukan observasi lapangan. Seperti pada penelitian A, pengamatan keberadaan amfibi dilakukan pada musim semi pada daerah penelitian yang lebarnya adalah 500 m untuk masing - masing lokasi A dan B. Hasil dari Inventarisasi ini akan menjadi dataset untuk memprediksi keberadaan hewan amfibi pada kedua lokasi pengamatan tersebut dan dapat diunduh pada alamat <https://archive.ics.uci.edu/ml/datasets/Amphibians>.

Dataset Amphibians yang digunakan dalam penelitian ini merupakan dataset yang memiliki fitur kategorikal dan fitur numerik yang bersifat kontinu dengan total 189 baris data dan 16 atribut. Terdapat 7 kelas pada dataset ini yang didefinisikan dengan notasi "0" yang menyatakan tidak adanya spesies amfibi di daerah tersebut dan notasi "1" yang menyatakan adanya spesies amfibi di daerah tersebut. Kelas hewan amfibi meliputi Green frogs, Brown frogs, Common toad, Fire-bellied toad, Tree frog, Common newt, dan Great crested newt.

Dataset Amphibians dibuat menggunakan metode herpetologi yang baik, sehingga fitur yang terdapat pada dataset ini sudah mencakup sebagian besar aspek yang mempengaruhi keberadaan amfibi di area tertentu. Fitur-fitur pada dataset ini terdiri dari:

1. Luas permukaan penampungan air
2. Jumlah penampungan air di habitat
3. Jenis penampungan air
4. Keberadaan vegetasi di tempat penampungan air
5. Jenis *land cover* paling dominan (pertama/kedua/ketiga) di sekitar waduk
6. Kegunaan / fungsi penampungan air
7. Keberadaan kegiatan memancing
8. Persentase akses dari tepi penampungan air ke daerah yang belum berkembang
9. Jarak minimum dari penampungan air ke jalan raya
10. Jarak minimum dari penampungan air ke bangunan
11. Pemeliharaan tempat penampungan air (Faktor kebersihan)
12. Jenis tepi penampungan air

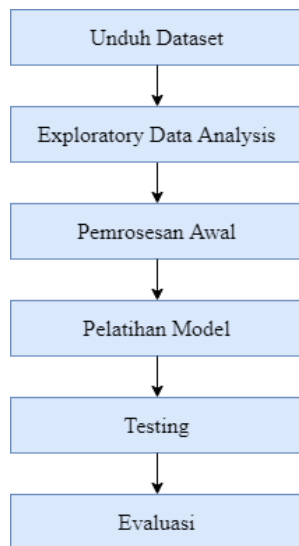
Alur proses pembelajaran mesin dimulai dengan mengunduh dataset yang sudah disediakan pada alamat sebelumnya serta mempelajari atau mendapatkan informasi dataset secara singkat. Informasi dataset yang didapatkan akan dijadikan cerminan proses apa saja yang akan dilakukan pada *preprocessing* atau persiapan data. *Preprocessing* dalam penelitian ini meliputi penghapusan kolom yang tidak digunakan dalam proses prediksi, menangani data penciran atau *outlier*, *encoding*. Sebelum proses *encoding*, didapati bahwa fitur kategorikal pada dataset dinyatakan dalam angka, sehingga fitur tersebut dapat disalahartikan sebagai data numerik. Oleh karena itu, semua nilai data pada fitur kategorikal akan disubstitusikan menjadi nilai yang sebenarnya. Kemudian proses *encoding* dapat dilanjutkan untuk setiap data selain numerik sehingga jumlah fitur yang ada berjumlah 69. Tahap *preprocessing* kemudian dilanjutkan dengan proses normalisasi menggunakan strategi Z-Score untuk data numerik. Setiap tahap *preprocessing* tersebut ditujukan untuk menghasilkan data yang baik, mudah dibaca, dan siap digunakan pada tahap berikutnya.

Setelah normalisasi, selanjutnya dataset akan dibagi menjadi data latih dan data uji untuk kepentingan tahap pelatihan model, *testing* dan evaluasi. Data latih ditujukan untuk proses training model pembelajaran mesin, sedangkan data uji ditujukan sebagai landasan prediksi kelas dan sebagai tolak ukur hasil evaluasi algoritma pembelajaran mesin. Data latih dan data uji dibagi menggunakan rasio perbandingan sebesar 80 : 20. Kemudian fitur dan label setiap data latih dan data uji akan dipisahkan untuk memudahkan proses selanjutnya. Setelah semua tahapan persiapan dataset yang digunakan selesai, maka proses pembelajaran mesin dapat dilanjutkan pada tahap modelling.

Terdapat tiga Algoritma klasifikasi yang digunakan dan akan dibandingkan mana algoritma yang cocok atau memiliki performa yang baik untuk digunakan dalam prediksi dataset ini. Ketiga algoritma tersebut yakni Naive Bayes, Decision Tree dan Support Vector Machine. Setelah model setiap algoritma dibuat, tahap prediksi dapat dimulai dengan memasukan model dari setiap algoritma klasifikasi dan data yang diujikan sebagai parameter prediksi.

Tahap evaluasi menjadi tahapan akhir dari proses pembelajaran mesin pada penelitian ini. Evaluasi dilakukan terhadap ketiga algoritma klasifikasi yang digunakan dengan *classification report*. *Classification report* dibuat dengan membandingkan kelas hasil prediksi dengan kelas sebenarnya sehingga terbentuk sebuah *confusion matrix*. *Classification report* mencakup *accuracy*, *precision*, *recall*, dan *f1-score* untuk setiap kelas. Kemudian hasil *classification report* setiap kelas tersebut akan dihitung rata-ratanya untuk mengetahui performa setiap algoritma klasifikasinya.

Secara keseluruhan, alur proses pembelajaran mesin menggunakan *deep learning* untuk dataset Amphibians diilustrasikan dalam Gambar 2.



Gambar 2. Alur Proses Pembelajaran Mesin

3. EKSPERIMEN DAN HASIL

Percobaan pembelajaran mesin dilakukan sesuai dengan alur proses pembelajaran mesin yang ditunjukkan pada Gambar 2. Namun, kami mendapati kendala pada proses *modelling* yang tidak dapat dieksekusi begitu saja seperti pada dataset lain. Hal ini disebabkan karena terdapat tujuh kelas yang menyatakan kehadiran spesies amfibi pada lingkungan tersebut. Dengan kata lain, dataset amphibians merupakan permasalahan regresi atau permasalahan kuantitas. Di sisi lain, algoritma yang digunakan adalah algoritma klasifikasi yang digunakan untuk permasalahan pengelompokan.

Untuk mengatasi masalah tersebut, tahap *modelling* dilakukan dengan strategi pembuatan model untuk masing-masing kelas pada dataset. Model yang didapatkan kemudian digunakan untuk memprediksi masing-masing kelas yang bersangkutan, yaitu memprediksi notasi “0” dan “1” yang menandakan ada tidaknya spesies amfibi tersebut di daerah dengan kondisi yang telah ditentukan.

Ketiga algoritma klasifikasi yang digunakan yakni Naive Bayes, Decision Tree, dan Support Vector Machine akan diimplementasikan menggunakan bantuan *library* Scikit-Learn. Maka dari itu pada tahap pemodelan dilakukan import *library* Scikit-Learn untuk setiap algoritma klasifikasinya. Setelah tahap pelatihan model dan prediksi selesai, masalah baru muncul pada tahap evaluasi. *Classification report* tidak dapat digunakan sekaligus untuk mengevaluasi ketujuh kelas yang ada pada dataset. Untuk mengatasi masalah tersebut, digunakan strategi yang sama seperti pada tahap *modelling* dan prediksi. *Classification report* dibuat untuk masing-masing kelas spesies amfibi sehingga

setiap algoritma klasifikasi memiliki sebanyak tujuh *classification report* yang mencakup *accuracy*, *precision*, *recall* dan *f1-score*. Ketujuh *classification report* pada masing-masing algoritma ini kemudian dihitung rata-ratanya untuk mewakili dan mengetahui performa total algoritma klasifikasi yang digunakan secara umum.

3.1. Hasil Algoritma Naive Bayes

Algoritma Naive Bayes menjadi algoritma pertama yang digunakan dalam proses pembelajaran mesin. Pada Naive Bayes, nilai posterior didapatkan dari operasi prior, *likelihood* atau *conditional probability* dan *evidence*. *Likelihood* pada Naive Bayes memiliki tiga model yang dapat digunakan sesuai dengan jenis dataset. *Likelihood* model Gaussian digunakan pada percobaan ini karena fitur dataset Amphibians bersifat kontinu.

Tabel 1. *Classification Report* Naive Bayes

Evaluasi	Hasil
Accuracy	0.4657142857142857
Precision	0.4700000000000001
Recall	0.5642857142857143
F1-Score	0.38574826564830056

Berdasarkan data pada Tabel 1, dapat dilihat bahwa nilai *accuracy* algoritma Naive Bayes kurang baik dengan nilai yang bahkan tidak mencapai 50% begitu pula dengan nilai *precision* dan *F1-score* nya. Di samping itu, nilai *recall* relatif memiliki nilai yang cukup baik dimana mencapai 56%. Dengan kata lain, algoritma Naive Bayes pada dataset ini hanya dapat menghasilkan perbandingan antara *True Positive* (TN) dengan semua nilai yang memang positif (Prinsip *Recall*) dengan hasil yang baik saja. Hal ini disebabkan karena jumlah data yang relatif sedikit padahal fitur yang dimiliki banyak dan kompleks sehingga algoritma Naive Bayes belum bisa membuat probabilitas dan statistik yang akurat. Oleh karena itu, algoritma Naive Bayes pada dataset ini dapat dikatakan kurang baik dalam menghasilkan sebuah prediksi.

3.2. Hasil Algoritma Decision Tree

Algoritma Decision Tree menjadi algoritma kedua yang digunakan dalam proses pembelajaran mesin. Pada algoritma ini, terdapat *hyperparameter* berupa algoritma yang digunakan untuk pembentukan cabang. Model pembentukan cabang pada Decision Tree sendiri ada 3, yaitu GINI Index, Information Gain, dan Gain Ratio. Pada percobaan ini, model pembentukan cabang yang digunakan adalah GINI Index.

Tabel 2. *Classification Report* Decision Tree

Evaluasi	Hasil
Accuracy	0.6942857142857143
Precision	0.5671428571428572
Recall	0.6171428571428572
F1-Score	0.5812548130458912

Berdasarkan data pada Tabel 2, dapat dilihat bahwa keempat nilai yang diperoleh menggunakan algoritma Decision Tree baik karena nilai *classification report* nya melebihi 50%. Nilai *accuracy* dan *recall* yang dihasilkan menggunakan algoritma Decision Tree bagus, sedangkan untuk nilai *precision* dan *F1-score*-nya dapat dikatakan secara relatif bagus. Hasil *classification report* yang baik seperti ini berarti algoritma ini tidak menghasilkan prediksi yang benar saja, tetapi juga hasil prediksi tersebut juga baik. Prediksi baik di sini diartikan sebagai prediksi dengan hasil *confusion matrix* yang baik atau dengan kata lain tidak mengakibatkan kesalahan fatal. Hal ini disebabkan karena algoritma Decision Tree yang mampu mengeliminasi data yang tidak diperlukan dan konsepnya yang melakukan eksplorasi data secara keseluruhan sehingga dapat membuat pola klasifikasi yang dapat memprediksi hasil lebih akurat dan tepat. Di samping itu, Decision Tree memang disebut juga sebagai CART yang merupakan akronim dari *Classification and Regression Tree* sehingga selain melakukan klasifikasi, Decision Tree juga dapat melakukan regresi (Zhu et al., 2018). Hal ini menyebabkan pola klasifikasi yang dibuat juga akan memiliki hubungan sebab-akibat yang kuat antara suatu fitur dengan fitur lainnya sehingga sebagai hasilnya kami bisa mendapatkan nilai *accuracy* dan *precision* yang baik. Oleh karena itu, algoritma Decision Tree pada dataset ini dapat dikatakan cocok dengan dataset yang dimiliki sehingga mampu menghasilkan prediksi dengan hasil yang menyerupai keadaan sesungguhnya.

3.3. Hasil Algoritma Support Vector Machine

Algoritma Support Vector Machine menjadi algoritma ketiga dan yang terakhir digunakan dalam proses pembelajaran mesin. Pada algoritma ini, terdapat dua skema pembentukan *hyperplane* atau pemisah antar dua atau lebih kelas, yaitu one-vs-all dan one-vs-one. Percobaan ini menggunakan skema one-vs-one dengan *hyperparameter gamma=auto* dan *hyperparameter* lainnya dikonfigurasi dengan nilai default.

Tabel 3. *Classification Report* Support Vector Machine

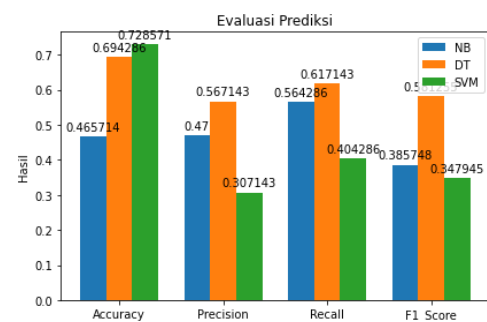
Evaluasi	Hasil
Accuracy	0.7285714285714285
Precision	0.3071428571428571
Recall	0.4042857142857143
F1-Score	0.34794456547468594

Berdasarkan data pada Tabel 3, dapat dilihat bahwa nilai *accuracy* algoritma Support Vector Machine sangat baik bila dibandingkan dengan dua algoritma sebelumnya. Namun dengan nilai *precision* 30%, *recall* 40%, *F1-score* 34%, dapat diterjemahkan bahwa hasil prediksi algoritma ini memiliki banyak hasil benar, namun sangat

berpotensi menghasilkan kesalahan prediksi yang fatal. Sehingga Support Vector Machine dapat dikatakan memiliki hasil yang paling buruk. Hal ini disebabkan karena algoritma Support Vector Machine memiliki perhitungan yang lebih matang dan sistematis yaitu dengan mencari *hyperplane* sehingga menghasilkan akurasi yang tinggi, namun eksplorasi data tidak dilakukan sebaik saat algoritma Decision Tree. Oleh karena itu, algoritma Support Vector Machine pada dataset ini dapat dikatakan menghasilkan prediksi yang akurat namun ketepatan prediksinya sangat tidak konsisten maka dari ketiga algoritma yang digunakan dapat disimpulkan bahwa Support Vector Machine merupakan algoritma yang paling tidak cocok untuk digunakan terhadap dataset ini.

3.4. Performa Ketiga Algoritma Klasifikasi

Untuk dapat melihat perbandingan performa dari ketiga algoritma klasifikasi, maka kami menggunakan sebuah histogram untuk memvisualisasi perbandingan tersebut dengan tujuan kemudahan dalam membaca. Histogram yang dibuat berisikan nilai rata-rata dari *accuracy*, *precision*, *recall* dan *F1-score* untuk setiap hasil prediksi dari *classification report* algoritma klasifikasi. Warna biru merupakan notasi untuk algoritma Naive Bayes, oranye untuk algoritma Decision Tree dan hijau untuk algoritma Support Vector Machine.



Gambar 3. Ringkasan Performa Ketiga Algoritma Klasifikasi

Melalui Gambar 3 dapat dilihat bahwa performa yang paling konsisten dan relatif baik untuk tingkat *accuracy*, *precision*, *recall* dan *F1-score* adalah ketika menggunakan algoritma Decision Tree. Hal ini disebabkan karena untuk menentukan baik atau tidaknya sebuah algoritma tidak hanya bertumpu terhadap satu tipe performa seperti jumlah prediksi benar (*accuracy*) saja. Dapat diamati bahwa untuk Support Vector Machine memiliki performa *accuracy* paling tinggi namun untuk pada tingkat *precision*, *recall* dan *F1-score*-nya paling rendah dibanding ketiga algoritma. Sedangkan untuk algoritma Naive Bayes memiliki nilai yang relatif konsisten namun saat dibandingkan dengan algoritma Decision Tree untuk keempat nilainya masih lebih rendah. Maka dari itu, berdasarkan performa konsisten yang dihasilkan oleh

algoritma Decision Tree, dapat disimpulkan bahwa Decision Tree merupakan algoritma yang paling cocok digunakan untuk memprediksi keberadaan hewan Amfibi pada sebuah tandon air.

4. KESIMPULAN

Baik buruknya hasil prediksi suatu algoritma tidak hanya diukur dari jumlah data yang benar diprediksi saja. Meskipun tingkat akurasi sangat tinggi, namun jika data yang diprediksi berpotensi menyebabkan kesalahan fatal atau tidak konsisten, maka penggunaan algoritma tersebut tetap tidak dapat dikatakan baik. Faktor lain seperti tingkat *precision*, *recall* dan *F1-score* juga memegang peranan penting untuk menentukan kualitas prediksi. Maka dari itu diperlukan sebuah *classification report* untuk dapat mengukur secara lebih detail kemampuan sebuah algoritma dalam menyelesaikan suatu masalah.

Berdasarkan hasil eksperimen yang telah dilakukan, dapat ditarik sebuah kesimpulan bahwa algoritma Decision Tree merupakan algoritma yang paling cocok digunakan untuk memprediksi keberadaan hewan amfibi di sebuah tandon air pada dataset ini. Karena jika dilihat dari fitur dataset yang relatif kompleks serta kemampuan algoritma Decision Tree untuk mengeksplorasi data secara keseluruhan, maka penggunaan algoritmanya sangat cocok untuk dataset ini.

Beberapa hal yang perlu diperhatikan dalam percobaan ini adalah banyaknya data dan fitur yang tersedia dari dataset serta tipe data yang digunakan. Karena ketiga hal tersebut yang menentukan baik buruknya hasil dari algoritma klasifikasi yang digunakan. Seperti halnya pada algoritma Naive Bayes yang digunakan pada percobaan ini memiliki hasil yang buruk karena perbandingan antara jumlah data dan fitur tidak seimbang serta sifat pada setiap fitur yang dapat dikatakan kompleks. Maka dari itu perlu data yang lebih banyak saat melakukan inventarisasi lapangan sebelum membuat hasil yang akan dijadikan sebagai dataset pengujian sebuah algoritma pembelajaran mesin.

DAFTAR PUSTAKA

- ALI, W. ET AL. (2018) 'Diversity and habitat preferences of amphibians and reptiles in Pakistan: a review', *Journal of Asia-Pacific Biodiversity*. Elsevier Ltd, 11(2), pp. 173–187. doi: 10.1016/j.japb.2018.01.009.
- ARYA, F. ET AL. (2021) 'Quagga: Jurnal Pendidikan dan Biologi Inventarisasi Awal Jenis Amfibi di Kawasan Malibu Anai (Provinsi Sumatera Barat) dengan Keterangan Habitatnya Quagga: Jurnal Pendidikan dan Biologi Volume 13, Nomor 1, Januari 2021, pp. 82–87', 13, pp. 82–87. doi: 10.25134/quagga.v13i1.3738.Received.
- BAKER, J. ET AL. (2011) *Amphibian Habitat Management Handbook*, Management. Bournemouth: Amphibian and Reptile Conservation.
- BLACHNIK, M., SOŁTYSIAK, M. AND DABROWSKA, D. (2019) 'Predicting presence of amphibian species using features obtained from GIS and satellite images', *ISPRS International Journal of Geo-Information*, 8(3), pp. 13–22. doi: 10.3390/ijgi8030123.
- BLANK, L. AND BLAUSTEIN, L. (2014) 'A multi-scale analysis of breeding site characteristics of the endangered fire salamander (*Salamandra atra*) at its extreme southern range limit', *Hydrobiologia*, 726(1), pp. 229–244. doi: 10.1007/s10750-013-1770-8.
- CHEN, Y. H. (2013) 'Habitat suitability modeling of amphibian species in southern and central China: Environmental correlates and potential richness mapping', *Science China Life Sciences*, 56(5), pp. 476–484. doi: 10.1007/s11427-013-4475-3.
- COGĂLNICEANU, D. ET AL. (2013) 'Diversity and distribution of amphibians in Romania', *ZooKeys*, 296, pp. 35–57. doi: 10.3897/zookeys.296.4872.
- DUJSEBAYEVA, T. N. AND MALAKHOV, D. V. (2017) 'The model of ranodon sibiricus ecological niche: GIS and remote sensing approach', *Russian Journal of Herpetology*, 24(3), pp. 171–192. doi: 10.30906/1026-2296-2019-24-3-171-192.
- HUDA, N. (2018) 'Inventarisasi Keanekaragaman Amfibi di Kawasan Wisata Air Terjun Bajuin Kabupaten Tanah Laut', *Jurnal Pendidikan Hayati*, 4(2), pp. 85–92.
- MUNGUÍA, M. ET AL. (2012) 'Equilibrium of global amphibian species distributions with climate.', *PloS one*, 7(4), pp. 1–9. doi: 10.1371/journal.pone.0034420.
- SIAHAAN, K., DEWI, B. S. AND DARMAWAN, A. (2019) 'Keanekaragaman Amfibi Ordo Anura di Blok Perlindungan dan Blok Pemanfaatan Hutan Pendidikan Konservasi Terpadu, Taman Hutan Raya Wan Abdul Rachman', *Jurnal Sylva Lestari*, 7(3), pp. 370–378.
- TIAGO, P., PEREIRA, H. M. AND CAPINHA, C. (2017) 'Using citizen science data to estimate climatic niches and species distributions', *Basic and Applied Ecology*. Elsevier GmbH, 20, pp. 75–85. doi: 10.1016/j.baae.2017.04.001.
- UNGLAUB, B. ET AL. (2018) 'The relationships between habitat suitability, population size and body condition in a pond-breeding

- amphibian', Basic and Applied Ecology, 27, pp. 20–29. doi: <https://doi.org/10.1016/j.baae.2018.01.002>.
- WAKE, D. B. AND KOO, M. S. (2018) 'Amphibians', Current Biology, 28(21), pp. R1237–R1241. doi: 10.1016/j.cub.2018.09.028.
- ZHU, F. ET AL. (2018) 'A classification algorithm of cart decision tree based on mapreduce attribute weights', International Journal of Performability Engineering, 14(1), pp. 17–25. doi: 10.23940/ijpe.18.01.p3.1725.