

Pemodelan Prediksi Perdagangan Satwa Liar Menggunakan Algoritma *k-Nearest Neighbor*

Predictive Modeling of Wildlife Trade Using the k-Nearest Neighbor Algorithm

ANGELA PRIMASARI^{1*}, IRMAN HERMADI¹, MEDRIA HARDHIENATA²

Abstrak

Perdagangan satwa liar dengan pengawasan yang tidak optimal dapat merupakan salah satu ancaman yang mampu memberikan dampak signifikan bagi keberlangsungan keanekaragaman hayati. Beberapa spesies hewan seperti mamalia dan reptil saat ini terancam punah dengan adanya perdagangan satwa liar yang tidak terpantau dengan baik. Praktik perdagangan satwa liar kerap ditemukan di berbagai media, termasuk situs internet, karena kemudahan akses yang ditawarkan kepada masyarakat luas. Namun jika data tersebut diolah secara manual maka dibutuhkan waktu, tenaga, dan upaya yang cukup besar. Oleh sebab itu, dibutuhkan sebuah pendekatan berbasis kecerdasan buatan untuk mengidentifikasi wilayah dengan aktivitas perdangan satwa liar yang tinggi dan berpotensi melampaui batas perdagangan satwa liar yang diizinkan. Penelitian ini bertujuan untuk membangun sebuah model yang dapat melakukan prediksi area yang rawan perdagangan satwa liar. Algoritma yang digunakan dalam penelitian ini adalah algoritma *clustering K-Means* untuk mengelompokkan wilayah berdasarkan tingkat kerawanan aktivitas perdagangan satwa liar, serta algoritma *k-Nearest Neighbor* untuk melakukan prediksi tingkat kerawanan wilayah perdagangan satwa liar. Selain itu, metode *Market Basket Analysis* digunakan untuk mengidentifikasi pola asosiasi dalam perdagangan satwa liar antar negara. Data yang digunakan merupakan data perdagangan satwa liar dari berbagai negara pada tahun 2018 hingga 2020. Dengan menggunakan pendekatan *clustering*, dalam penelitian ini diklasifikasikan tiga wilayah dengan potensi kerawanan perdagangan satwa liar, yaitu wilayah dengan resiko rendah, sedang, dan tinggi. Hasil penelitian menunjukkan bahwa model prediksi yang dibangun mampu memprediksi wilayah rawan perdagangan satwa liar dengan tingkat akurasi model *training* sebesar 99% dengan data impor dan 100% dengan data ekspor. Setelah dievaluasi dengan *3-cross fold validation*, akurasi model yang diperoleh adalah sebesar 97% untuk data impor dan 98% untuk data ekspor. Hasil akurasi model *testing* dalam penelitian ini adalah sebesar 100% dengan data impor maupun ekspor. Melalui pendekatan metode *market basket analysis*, penelitian ini menyimpulkan dengan data yang dipertimbangkan belum ditemukan pola asosiasi yang kuat dalam aktivitas perdagangan satwa liar antara satu negara spesifik dengan negara lainnya.

Kata Kunci: *k-Nearest neighbor, machine learning, market basket analysis, perdagangan satwa liar, prediksi*

Abstract

Wildlife trade conducted under suboptimal supervision can pose a significant threat to the long-term sustainability of biodiversity. Several animal species, including mammals and reptiles, are now facing extinction due to the lack of proper monitoring of wildlife trade. Wildlife trade practices are frequently found across various media platforms, including websites, due to the ease of access they offer to the public. However, if the data is processed manually, it requires considerable time, effort, and resources. To address this issue, an artificial intelligence-based approach is needed to help process wildlife trade data so that areas vulnerable to wildlife trafficking can be predicted quickly and accurately. Therefore, an artificial intelligence-based approach is needed to identify regions with high wildlife trade activity that are at risk of exceeding the permitted trade limits. This study aims to develop a model capable of predicting regions at risk of wildlife trade. The algorithms used in this study are the K-Means clustering algorithm to classify regions based on the risk level of wildlife trade areas, and the k-Nearest Neighbor algorithm to predict the risk level of wildlife trade areas. In addition, the Market Basket Analysis method is used to identify association patterns in wildlife trade between countries. The data used consists of wildlife trade data from various countries from 2018 to 2020. Using the clustering approach, this study classifies

¹ Program Studi Ilmu Komputer, Sekolah Sains Data, Matematika dan Informatika, Institut Pertanian Bogor, Bogor 16680

² Program Studi Kecerdasan Buatan, Sekolah Sains Data, Matematika dan Informatika, Institut Pertanian Bogor, Bogor 16680

*Penulis Korespondensi : Tel/Faks: 0821-25243700; Surel: angelaangela@apps.ipb.ac.id

three levels of risk for wildlife trafficking: low, medium, and high. The results of the study show that the predictive model developed is capable of identifying areas vulnerable to wildlife trade, achieving a training accuracy of 99% with import data and 100% with export data. After being evaluated using 3-fold cross-validation, the model achieved an accuracy of 97% for import data and 98% for export data. The testing accuracy obtained in this study was 100% for both import and export data. Through the market basket analysis approach, this study concludes that, based on the data considered, no strong association patterns have been found in wildlife trade activities between specific countries.

Keywords: K-nearest neighbor, machine learning, market basket analysis, prediction, wildlife trade.

PENDAHULUAN

Perdagangan satwa liar meliputi segala perdagangan dan pertukaran sumber daya hewan liar oleh manusia. Perdagangan satwa liar mencakup hewan yang masih hidup atau produk turunannya seperti kulit, bahan obat, bahan makanan, yang memberikan pendapatan baik bagi kelompok masyarakat dan juga berkontribusi terhadap perekonomian nasional. Pada tahun 1998-2007, lebih dari 35 juta hewan diperdagangkan dari negara-negara Asia Tenggara (Nijman, 2009). Hampir 30 juta diambil dari alam liar dan 4.5 juta dari penangkaran (Nijman, 2009). Selain itu sekitar 3.8 juta satwa liar hidup dan 1.4 juta bagian hewan utuh diimpor ke China dari Asia Tenggara dan perdagangan reptil dari Asia Tenggara 79% diimpor dari negara Indonesia, Laos, dan Malaysia (Jiao *et al.* 2021). Pada tahun 2015-2019, sekitar 58% ikan hias dieksport dari Brazil ke Jerman, Jepang, dan China (Melo de Sousa *et al.* 2021). Sementara Maroko, Mesir, Amerika Serikat dinyatakan sebagai negara asal perdangangan satwa liar, dan Spanyol, UK, Belanda, serta Prancis merepresentasikan 79% sebagai negara tujuan perdagangan satwa liar (Hitchens & Blakeslee, 2020).

Faktor pemicu perdagangan satwa liar diantaranya permintaan pasar yang tinggi, keuntungan yang besar, resiko yang relatif rendah, dan konsekuensi hukuman yang relatif kecil (Fukushima *et al.* 2021). Akibat dari perdagangan satwa liar yang belum terpantau secara optimal tersebut, saat ini spesies zebra menjadi spesies satwa liar yang terancam punah karena setiap tahun sekitar 100,000 *specimen* terjual (Melo de Sousa *et al.* 2021). Selain itu, spesies ular *phyton Burma* juga mengalami penurunan spesies dan tergolong satwa liar diduga karena perdagangan *illegal*. Sekitar 61.6% spesies reptil, burung, dan mamalia mengalami penurunan populasi, sekitar 16.4% populasi burung, mamalia, dan reptil punah, dan sekitar 98.6% populasi *Phocoena sinus* menurun pada tahun 2011-2019 (Cardoso *et al.* 2021).

Bentuk perdagangan satwa liar umumnya banyak ditemukan di *website*, karena media ini cenderung lebih mudah diakses oleh masyarakat luas. Perdagangan ini dapat bersifat legal maupun ilegal, tergantung pada jenis spesies yang diperdagangkan, perizinan yang dimiliki, serta regulasi yang berlaku di negara terkait. Pengolahan data perdagangan satwa liar jika dilakukan secara manual dari *website* membutuhkan upaya, *resource*, dan juga waktu yang cukup lama untuk meneliti pola perdagangan satwa liar. Oleh karena itu pendekatan *artificial intelligent* dalam beberapa studi telah diusulkan untuk menyelesaikan masalah ini. Algoritma *artificial intelligent* dapat membantu mengklasifikasikan dan memprediksikan daerah rawan perdagangan satwa liar dengan waktu yang singkat dan *resource* yang lebih sedikit (Salamat *et al.* 2021). Selain itu, pendekatan berbasis kecerdasan buatan juga diperlukan untuk mengidentifikasi wilayah dengan aktivitas perdangan satwa liar yang tinggi dan berpotensi melampaui batas perdagangan satwa liar yang diizinkan.

Beberapa penelitian sebelumnya telah melakukan klasifikasi dan prediksi tingkat kerawanan suatu daerah dengan algoritma *artificial intelligent*. Pada tahun 2020, penelitian Lee dan Roberts (2020) meneliti data perdagangan satwa liar dari situs CITES dalam jangka waktu 14 minggu. Penelitian tersebut menggunakan metode hierarkial *clustering*. Hasil menunjukkan bahwa klaster perdagangan satwa liar dapat dibagi menjadi 8 klaster, dengan China, Jerman, Belanda, UK, dan Polandia yang mendominasi perdagangan. Selain itu, penelitian (Kumar, Puran, & Dhar, 2011) menggunakan algoritma *Probabilistic Neural Network* dan *k-Nearest Neighbor* untuk mengklasifikasi gambar hewan. Penelitian ini menggunakan skala perbandingan data *traning* dan *testing* sebesar 7:3. Hasilnya menunjukkan bahwa prediksi dengan menggunakan algoritma *K-Nearest neighbor* lebih tinggi akurasinya dibandingkan

dengan algoritma yang lain yang dipertimbangkan. Sementara algoritma K-Means dan *market basket analysis* digunakan juga oleh penelitian Roodpishi dan Nashtaei (2015) untuk mengklasifikasi dan menganalisis asosiasi data pelanggan pada bidang industri. Hasil penelitian menunjukkan bahwa data dibagi menjadi 5 klaster, dan pola asosiasi data dapat disimpulkan dengan nilai minimum *support* 10% dan nilai minimum *confidence* 90%.

Pada penelitian perdagangan satwa liar yang sudah dilakukan oleh Lee dan Roberts (2020), jumlah jenis hewan yang diperdagangkan dan jumlah kasus perdagangan satwa liar belum dipertimbangkan. Selain itu, belum banyak penelitian dilakukan untuk memprediksi tingkat kerawanan suatu daerah berdasarkan jumlah kasus dan jumlah jenis hewan yang diperjualbelikan. Berdasarkan hal tersebut, pada penelitian ini dilakukan memprediksi kerawanan wilayah yang berpotensi melakukan perdagangan satwa liar menggunakan algoritma K-Means dan *k-Nearest Neighbor* dengan mempertimbangkan jumlah kasus dan jumlah jenis hewan yang diperdagangkan. Selain itu, metode *Market Basket Analysis* digunakan untuk mengidentifikasi pola asosiasi dalam perdagangan satwa liar antar negara. Pada penelitian ini, data yang digunakan merupakan data perdagangan satwa liar secara legal dari berbagai negara yang terdapat pada situs CITES.

METODE

Data yang digunakan dalam penelitian ini merupakan data kasus perdagangan satwa liar yang diunduh dari situs <https://trade.cites.org/>. Data perdagangan satwa liar yang digunakan dalam penelitian ini merupakan data perdagangan secara legal yang dilakukan oleh berbagai negara di dunia dalam jangka waktu tahun 2018 sampai tahun 2020. Data yang digunakan memiliki fitur dan tipe data seperti pada Tabel 1. Data terdiri dari 16 kolom dan 179.524 baris.

Tabel 1 Fitur dan tipe data perdagangan satwa liar yang digunakan dalam penelitian berdasarkan <https://trade.cites.org/>

No	Fitur	Tipe Data
1	<i>Year</i>	Int
2	<i>Appendix</i>	String
3	<i>Taxon</i>	String
4	<i>Class</i>	String
5	<i>Order</i>	String
6	<i>Family</i>	String
7	<i>Genus</i>	String
8	<i>Term</i>	String
9	<i>Importer Reported Quantity</i>	Int
10	<i>Exporter Reported Quantity</i>	Int
11	<i>Unit</i>	String
12	<i>Importer</i>	String
13	<i>Exporter</i>	String
14	<i>Origin</i>	String
15	<i>Purpose</i>	String
16	<i>Source</i>	String

Praproses Data

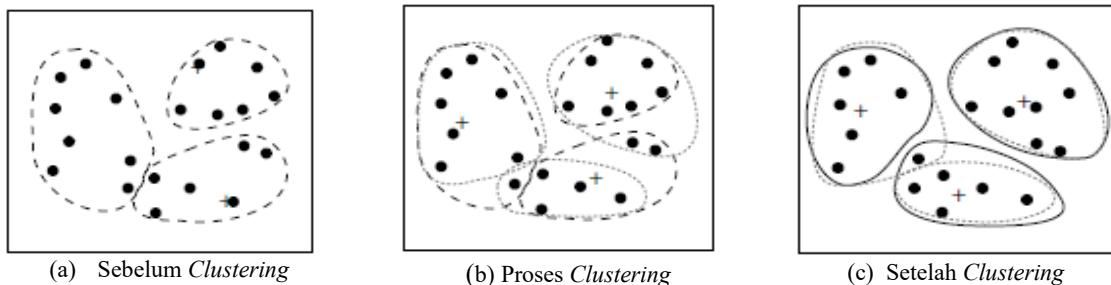
Tahapan praproses merupakan tahapan awal yang dilakukan untuk mengubah data input menjadi data dengan format yang tepat dan siap untuk diproses. Praproses terdiri dari *data integration*, *data cleaning*, *data reduction* dan *data transformation* (Febrian et al. 2022), serta seleksi fitur (Ilbeigipour, Albadvi, dan Noughabi, 2022). Pada tahap *data integration*, tipe data diubah menjadi tipe data yang dapat dibaca oleh sistem. Selanjutnya dilakukan *data cleaning* dimana dilakukan penyaringan data yang tidak lengkap, tidak akurat dan data yang hilang. *Data reduction* merupakan proses untuk mengurangi volume data yang disimpan, sehingga dapat menurunkan waktu pemrosesan dan biaya penyimpanan. *Data transformation* adalah proses mengubah data mentah menjadi format yang sesuai untuk dianalisis atau diproses oleh sistem komputer. Seleksi fitur merupakan bagian dari tahapan praproses yang bertujuan untuk mengurangi volume data dengan memilih subset fitur yang paling relevan, serta mengeliminasi variabel yang tidak berkontribusi signifikan terhadap pemodelan (Ilbeigipour, Albadvi, dan

Noughabi, 2022). Pada tahapan seleksi fitur, dilakukan pengolahan data untuk mendapatkan jumlah kasus per negara dan jumlah jenis satwa yang diperjualbelikan di setiap negara. Kedua fitur ini dipilih karena jumlah spesies yang diperdagangkan dalam data CITES tidak memiliki korelasi dengan volume perdagangan (Slabova *et al.* 2020), sehingga fitur-fitur kuantitatif lain yang ada pada data mentah tidak digunakan untuk *clustering*. Sementara penelitian Lee dan Roberts (2020) yang juga menggunakan data CITES, memberikan hasil bahwa jenis hewan yang diperdagangkan dalam setiap negara merupakan salah satu fitur yang *valid* untuk digunakan dalam penelitian. Berdasarkan dua referensi penelitian tersebut maka pada penelitian ini diputuskan untuk menggunakan fitur jumlah kasus per negara dan jumlah jenis hewan yang diperjualbelikan per negara.

Dalam penelitian ini, dilakukan juga penghapusan baris yang kosong, berisi nilai N dan berisi nilai NaN. Data yang berisi nilai N atau NaN merupakan data yang merepresentasikan nilai yang kosong (*missing value*). Setelah dilakukan penghapusan baris yang kosong, dimensinya berkurang menjadi 25,839 baris dan 16 kolom. Selain itu dilakukan juga seleksi fitur menjadi dua fitur yang paling relevan, yaitu jumlah jenis hewan dan jumlah kasus per lokasi.

Clustering menggunakan K-Means

Data perdagangan satwa liar yang digunakan dalam penelitian ini pada tahap awal belum memiliki label tingkat kerawanan. Oleh karena itu, dalam penelitian ini algoritma K-Means berperan memberikan pelabelan tingkat kerawanan perdagangan satwa liar. *Clustering K-Means* bertujuan untuk mengelompokkan data yang mirip dan menemukan pola berdasarkan kemiripan (Eshimiakhe dan Lawal 2022). Proses pengelompokan data menggunakan algoritma K-Means dapat dilihat pada Gambar 1.



Gambar 1 Proses pengelompokan data menggunakan algoritma K-Means (Kumar, Puran, dan Dhar 2011)

Setelah melalui praproses data, dilakukan perhitungan jumlah jenis hewan yang diperdagangkan dan jumlah kasus pada setiap negara. Data tersebut akan dikelompokkan berdasarkan tingkat kemiripan dengan menggunakan algoritma K-Means. *Clustering* dengan algoritma K-Means diawali dengan mendefinisikan nilai k yang paling optimal untuk mengetahui jumlah *cluster* daerah yang rawan melakukan aktivitas perdagangan satwa liar. Nilai k dalam *clustering* dapat ditentukan dengan metode *elbow* dan metode *silhouette* (Saputra, Saputra, dan D. Oswari, 2020). Metode *elbow* mengukur jarak setiap data dengan *centroid*. Nilai k yang paling baik dari metode *elbow* dapat dilihat dari grafik dengan lekuk yang signifikan. Nilai k yang paling optimal dapat ditentukan berdasarkan nilai kohesi dan nilai separasi. Nilai kohesi digunakan untuk mengukur seberapa erat keterikatan data dalam sebuah klaster, sedangkan nilai separasi mengukur seberapa jauh satu klaster dengan klaster yang lain. Pada metode *elbow*, nilai kohesi dan nilai separasi dapat dihitung menggunakan nilai *Sum Squared Error* (SSE) sebagaimana didefinisikan pada Persamaan 1 (Syakur *et al.* 2018).

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - C_k\|_2^2, \quad (1)$$

dengan K adalah banyaknya cluster yang terbentuk, S_k adalah himpunan data *points* yang dialokasikan ke *cluster* k , C_k titik pusat dari *cluster* ke- k , dan x_i merupakan *data point* di cluster

k. Sementara itu pada metode *silhouette* juga dihitung nilai kohesi dan separasi. Nilai *silhouette* dihitung menggunakan rumus pada Persamaan 2 dan Persamaan 3 (Saputra *et al.* 2020).

$$S = 1 - \frac{\text{cohesion measure}}{\text{separation measure}} \quad (2)$$

(Jika nilai kohesi < nilai separasi)

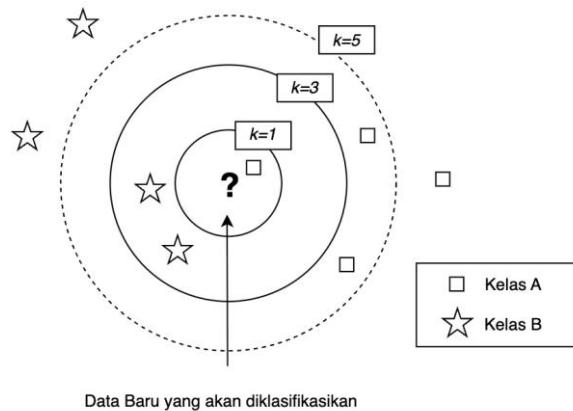
$$S = \frac{\text{separation measure}}{\text{cohesion measure}} - 1 \quad (3)$$

(Jika nilai kohesi > nilai separasi)

Nilai *silhouette* yang tinggi menunjukkan bahwa objek berada dalam klaster yang tepat dan jauh dari klaster lainnya, sehingga mencerminkan kualitas clustering yang baik.

Pemodelan Prediksi Menggunakan K-Nearest Neighbor

Setelah mengklasifikasikan data perdagangan satwa liar menjadi beberapa klaster berdasarkan tingkat kerawanan dengan algoritma K-Means, selanjutnya dibangun model prediksi dengan algoritma *k-Nearest Neighbor* (*k*-NN) untuk memprediksi tingkat kerawanan area perdagangan satwa liar dengan data baru. Data baru dikelompokkan berdasarkan jumlah jenis hewan dan jumlah kasus perdagangan hewan pada setiap negara. Ilustrasi klasifikasi dengan algoritma *k*-NN dapat dilihat pada Gambar 2.



Gambar 1 Ilustrasi klasifikasi data baru pada algoritma *k*-NN

Algoritma *k*-NN diawali dengan mendefinisikan nilai *k* yang paling ideal untuk menentukan berapa jumlah data tetangga terdekat yang optimal untuk memprediksi data baru. Nilai *k* memiliki dampak yang signifikan terhadap tingkat akurasi klasifikasi saat menggunakan algoritma *k*-NN (Arnita *et al.* 2022). Nilai *k* yang umum dipilih adalah bilangan ganjil untuk menghindari hasil voting yang imbang, dan dipilih berdasarkan nilai akurasi tertinggi pada data validasi.

Ketika ada data baru yang akan diprediksi tingkat kerawannya, maka algoritma *k*-NN akan menghitung jarak data baru tersebut dengan jarak *k* tetangga terdekatnya. Jarak antara data baru dengan data tetangga terdekat dihitung dengan menggunakan *euclidean distance* dengan rumus seperti pada Persamaan 4 (Anak Bundak *et al.* 2021) sebagai berikut:

$$d = \sqrt{(m_{x_1} - m_{x_2})^2 + (m_{y_1} - m_{y_2})^2}, \quad (4)$$

dengan d adalah jarak antara data baru dengan data tetangga terdekat, m_{x_1} adalah jumlah jenis hewan pada data baru m_{x_2} adalah jumlah jenis hewan pada data tetangga terdekat, m_{y_1} merupakan jumlah kasus pada data baru, dan m_{y_2} merupakan jumlah kasus pada data tetangga terdekat. Jarak dihitung dengan menarik garis lurus antara dua titik. Dalam hal ini adalah data tetangga terdekat dan data baru yang akan diprediksi. Algoritma k -NN digunakan dalam penelitian ini untuk memprediksi tingkat kerawanan perdagangan satwa liar suatu wilayah pada data baru berdasarkan data tetangga terdekat.

Analisis Asosiasi Antar Negara Menggunakan *Market Basket Analysis*

Selain prediksi tingkat kerawanan pada suatu daerah, penelitian ini juga memprediksi apakah ada keterkaitan antar negara dalam melakukan perdagangan satwa liar, maka pada tahapan yang terakhir dilakukan *market basket analysis* (Tatiana & Mikhail, 2018). *Market basket analysis* merupakan salah satu algoritma asosiasi yang berfokus menemukan aturan-aturan asosiasi pada kombinasi antar item dalam *dataset*. Analisis asosiasi menggunakan algoritma *market basket analysis* ini dilakukan untuk menentukan apakah ada kecenderungan suatu negara selalu melakukan ekspor atau impor ke negara tertentu, sehingga negara-negara tersebut dapat lebih diperhatikan. Dalam algoritma *market basket analysis* terdapat beberapa nilai yaitu nilai *support* yang menunjukkan seberapa sering kejadian terjadi, kemudian nilai *confidence* yang menunjukkan kemungkinan kejadian *item* yang ada di kanan aturan dapat mempengaruhi kejadian *item* yang berada di kiri, dan nilai *lift* yang menunjukkan kekuatan koneksi antar *item*.

Nilai *support* pada algoritma *market basket analysis* dihitung dengan menggunakan Persamaan 5 (Tatiana dan Mikhail 2018).

$$\text{Support } (X) = \frac{\text{Count}(X)}{N} \quad (5)$$

Selain itu nilai *confidence* dihitung dengan menggunakan Persamaan 6 (Tatiana dan Mikhail 2018).

$$\text{Confidence } (X \rightarrow Y) = \frac{S(X \cup Y)}{S(X)} \quad (6)$$

Sementara nilai *lift* dihitung dengan menggunakan Persamaan 7 (Tatiana dan Mikhail 2018).

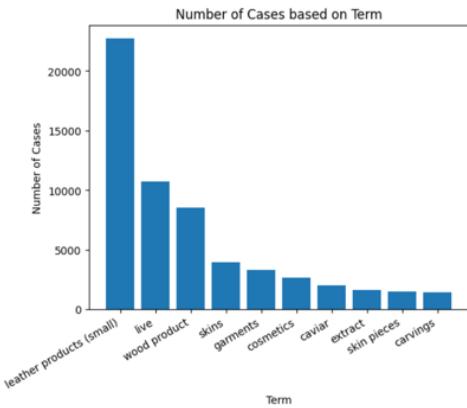
$$\text{Lift } (X \rightarrow Y) = \frac{S(X \cup Y)}{S(X)S(Y)}, \quad (7)$$

dengan (X) merupakan jumlah kejadian perdagangan satwa liar dalam negara X , (Y) jumlah kejadian perdagangan satwa liar dalam negara Y , $\text{Count } (X)$ jumlah seluruh kejadian perdagangan satwa liar dalam negara X , N jumlah seluruh kejadian perdagangan satwa liar pada seluruh negara, $S(X \cup Y)$ adalah nilai *support* dari negara X maupun Y yang melakukan perdagangan satwa liar, $S(X)$ adalah nilai *support* dari negara X , $S(Y)$ adalah nilai *support* dari negara Y . Nilai *support* menunjukkan seberapa sering perdagangan satwa liar antar negara terjadi. Jika nilai *support* besar, maka sering terjadi perdagangan satwa liar antar kedua negara tersebut. Nilai *confidence* menunjukkan kejadian perdagangan antar dua negara yang saling mempengaruhi. Jika nilai *confidence* besar, maka perdagangan satwa liar di negara yang satu mempengaruhi perdagangan satwa liar di negara yang lain. Sementara nilai *lift* menunjukkan kekuatan koneksi antar negara. Sehingga jika nilai *lift* besar, maka kedua negara tersebut saling berkaitan erat dalam melakukan perdagangan satwa liar.

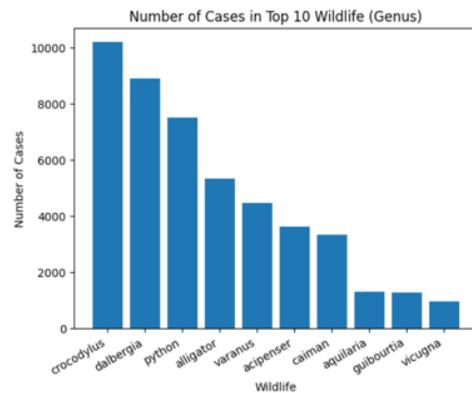
HASIL DAN PEMBAHASAN

Praproses Data

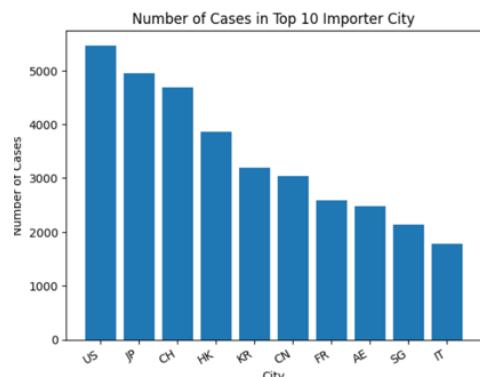
Pada tahap praproses, selain dilakukan pembersihan terhadap data yang redundan, tidak konsisten, dan kosong, juga dilakukan pengelompokan jumlah kasus untuk melihat data terbanyak dalam setiap kategori. Visualisasi untuk lebih memahami data dalam setiap kategori yang digunakan dalam penelitian ini dapat dilihat pada Gambar 3 hingga Gambar 8.



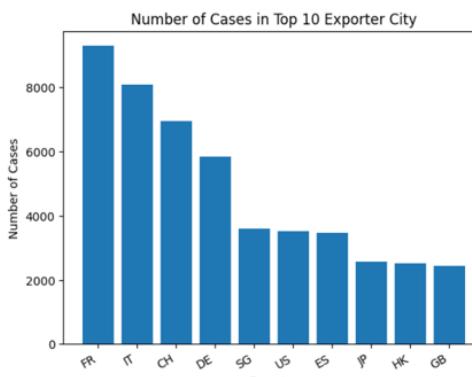
Gambar 2 Jumlah kasus berdasarkan cara jenis produk yang diperdagangkan



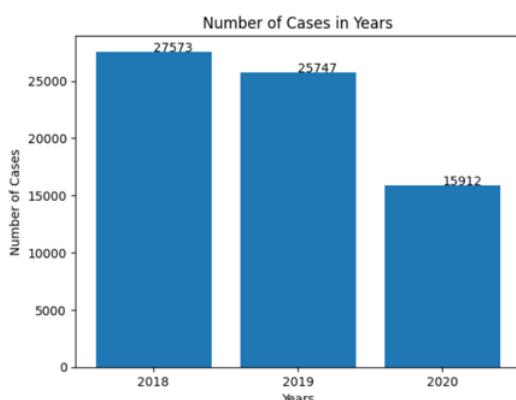
Gambar 3 Jumlah kasus berdasarkan genus



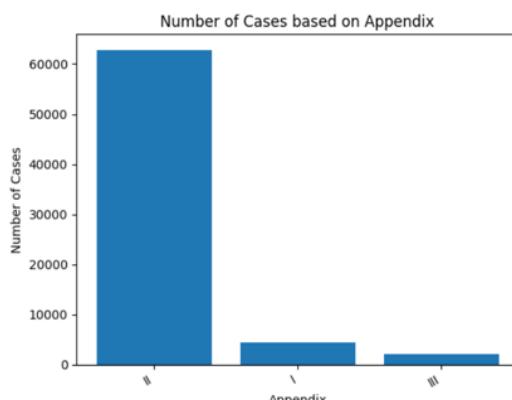
Gambar 5 Jumlah kasus 10 besar negara importir



Gambar 6 Jumlah kasus 10 besar negara eksportir



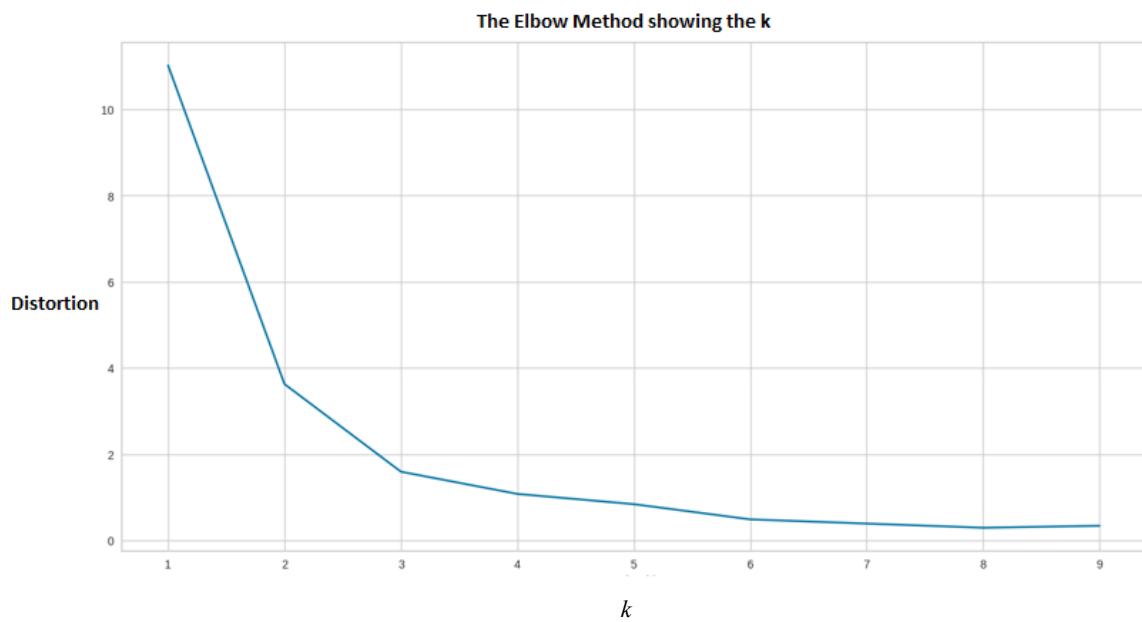
Gambar 7 Jumlah kasus per tahun



Gambar 8 Jumlah kasus berdasarkan Appendix

Clustering

Pada tahap *clustering*, nilai k pada *clustering* ditentukan dengan metode *elbow*. Sumbu-X pada Gambar 9 menunjukkan nilai k , sedangkan sumbu-Y menunjukkan nilai SSE (*Sum Squared Error*).

Gambar 9 Metode *elbow*

Selanjutnya, untuk melakukan validasi hasil metode *elbow*, dilakukan juga metode *silhouette* terhadap nilai $k = 2$, $k = 3$, dan $k = 4$. Nilai *silhouette* untuk $k=2$ adalah 0.892, untuk $k=3$ adalah 0.850, dan untuk $k=4$ adalah 0.772. Nilai *silhouette* yang paling mendekati 1 menunjukkan nilai k yang paling optimal. Nilai silhoutte untuk $k=2$ dan $k=3$ pada Gambar 9 keduanya menunjukkan nilai yang cukup baik (di atas 0,8). Dalam penelitian ini diputuskan nilai k yang digunakan dalam penelitian ini bukan nilai $k=2$ melainkan nilai $k = 3$ agar klaster yang dihasilkan dapat merepresentasikan kelas dengan tingkat kerawanan perdagangan satwa liar dengan tingkat yang lebih beragam (3 klaster). Data perdagangan satwa liar dibagi menjadi 3 klaster yaitu daerah dengan tingkat kerawanan tinggi, sedang, dan rendah. Hasil K-Means *clustering* dalam penelitian ini dapat di lihat pada Tabel 2 dengan data negara importir dan pada Tabel 3 dengan data negara eksportir.

Tabel 2 Hasil *clustering* negara importir+

Tingkat Kerawanan	Jumlah Kasus		Jenis Hewan		Contoh beberapa negara dalam <i>cluster</i>
	Min	Max	Min	Max	
Rendah	1	503	1	111	Jerman, Brazil, Taiwan, Australia, Kanada
Sedang	621	1372	72	144	United States, Cina, Singapur, Prancis
Tinggi	1824	2778	158	290	Jepang, Hongkong, Korea

Tabel 3 Hasil *clustering* negara eksportir

Tingkat Kerawanan	Jumlah Kasus		Jenis Hewan		Contoh beberapa negara dalam <i>cluster</i>
	Min	Max	Min	Max	
Rendah	1	346	1	63	Thailand, Norwegia, Vietnam, Kamboja, Brazil
Sedang	454	1006	29	108	Kanada, Indonesia, Spanyol, Belgia
Tinggi	1650	3667	44	128	Italia, Prancis, Switzerland

Tahap Pemodelan dengan Algoritma k -NN

Pada pemodelan dengan algoritma k -NN, data dibagi menjadi data *training* dan data *testing*. Data yang digunakan untuk melakukan pemodelan dengan k -NN merupakan data yang telah dilabeli pada tahap *clustering*. Dalam penelitian ini, data dibagi dengan komposisi 70% data *training* dan 30% data *testing*. Rasio data *training* yang melebihi 90% tidak disarankan untuk digunakan karena akan menimbulkan hafalan atau *memorization* (Uçar *et al.* 2020).

Berdasarkan Tabel 5 dan Tabel 6, diperoleh hasil akurasi model K-Nearest Neighbors (KNN) terhadap data negara importir dan eksportir. Secara umum, model menunjukkan performa yang sangat baik, dengan akurasi pelatihan (*training*) yang konsisten tinggi di semua nilai k , yaitu berkisar antara 97% hingga 100%. Pada data negara importir, nilai $k=3$ dan $k=5$ menghasilkan akurasi testing tertinggi yaitu 100%, disertai dengan validasi silang (*3-cross fold validation*) yang cukup stabil di angka 97% dan 98%. Namun, ketika nilai k meningkat menjadi 7 dan 9, terjadi sedikit penurunan akurasi model *testing* menjadi masing-masing 98% dan 96%.

Hal serupa juga terlihat pada data negara eksportir (lihat Tabel 6). Nilai $k=3$ dan $k=5$ memberikan akurasi testing terbaik sebesar 100%, dengan hasil validasi silang sebesar 98%. Akan tetapi, pada nilai $k=7$ dan $k=9$ akurasi testing menurun menjadi 95% dan 93%. Penurunan akurasi seiring bertambahnya nilai k mengindikasikan bahwa model menjadi terlalu umum dan kurang sensitif terhadap pola lokal, yang berdampak pada penurunan kemampuan prediksi terhadap data uji. Oleh karena itu, pada tahap akhir, dipilih $k=3$ karena memberikan kombinasi terbaik antara akurasi model *training*, validasi silang, dan *testing*. Selain itu, nilai k yang lebih kecil cenderung lebih responsif terhadap pola lokal dalam data yang kompleks seperti perdagangan satwa liar. Tabel 7 menunjukkan beberapa contoh keluaran dari data pemodelan k -NN dengan $k=3$.

Tabel 4 Akurasi KNN dengan Data Negara Importir

Nilai k	Akurasi <i>Training</i>	<i>3-Cross Fold Validation</i>	Akurasi <i>Testing</i>
3	99%	97%	100%
5	99%	98%	100%
7	99%	97%	98%
9	98%	97%	96%

Tabel 5 Akurasi KNN dengan Data Negara Eksportir

Nilai k	Akurasi <i>Training</i>	<i>3-Cross Fold Validation</i>	Akurasi <i>Testing</i>
3	100%	98%	100%
5	99%	98%	100%
7	97%	98%	95%
9	97%	98%	93%

Tabel 7 Data pemodelan K-NN

Jumlah Kasus	Jumlah Jenis Hewan	Tingkat Kerawanan
1325	99	sedang
1103	76	sedang
2382	164	tinggi
242	51	rendah
2457	80	tinggi
710	63	sedang
3390	133	tinggi

Tahap Prediksi Kecenderungan Perdagangan Hewan

Dalam penelitian ini, untuk melihat adanya asosiasi perdagangan satwa liar antar negara digunakan algoritma *Market Basket Analysis*. Dengan nilai *support* 0.005 dan nilai *confidence* 0.1, *min lift* = 2 dan *min length* = 2 diperoleh kecenderungan bahwa negara Kanada cenderung melakukan ekspor ke US, sedangkan Belanda dan Vietnam cenderung melakukan ekspor ke Korea. Namun kemungkinan kejadian ini sangat kecil karena nilai *support*-nya hanya 0.5% dan asosiasi kejadian antara kedua negara pun hanya 10%. Sehingga dapat disimpulkan bahwa hampir tidak teridentifikasi pola keterikatan perdagangan satwa liar yang kuat antar negara tersebut.

SIMPULAN

Pada penelitian ini dilakukan klasifikasi tingkat kerawanan daerah perdagangan satwa liar dengan metode *clustering K-Means*. Dengan metode tersebut data perdagangan satwa liar diklasifikasikan menjadi tiga daerah dengan tingkat kerawanan (resiko) rendah, sedang, dan tinggi. Selanjutnya setelah pelabelan data dilakukan, digunakan metode *k-NN* untuk memprediksi tingkat kerawanan daerah perdagangan satwa liar. Untuk pemodelan prediksi tingkat kerawanan daerah dengan metode *k-Nearest Neighbor (k-NN)* ditentukan berdasarkan 3 data tetangga terdekat. Pada penelitian ini, metode *clustering K-Means* dan juga metode *K-NN* telah berhasil digunakan untuk melakukan klastering dan prediksi dengan akurasi yang tinggi, yakni akurasi *training* sebesar 99% untuk data impor dan 100% untuk data ekspor. Setelah dievaluasi dengan 3 *cross fold validation* akurasi pada penelitian ini menjadi 97% untuk data impor dan 98% dengan data ekspor. Hasil akurasi model *testing* adalah sebesar 100% untuk data impor maupun ekspor. Hasil penelitian dalam penelitian ini juga menunjukkan bahwa praktik perdagangan satwa liar memiliki pola yang abstrak antar negara dan belum ditemukan ada kecenderungan asosiasi yang kuat antara satu negara dengan negara lain. Hal ini disimpulkan dari hasil keluaran dari algoritma *market basket analysis* yang memiliki nilai *support* dan nilai *confidence* yang cenderung kecil yaitu 0.005 dan 0.1.

Salah satu keterbatasan dalam penelitian ini adalah data yang digunakan tidak memiliki parameter waktu yang spesifik, sehingga belum dapat diprediksi apakah ada keterkaitan antara maraknya perdagangan satwa liar dengan musim atau momen tertentu. Penambahan parameter waktu yang lebih detail, seperti bulan atau musim kejadian, dapat dilakukan pada penelitian berikutnya untuk memahami bagaimana faktor temporal berperan dalam meningkatkan akurasi prediksi tingkat kerawanan perdagangan satwa liar. Selain itu, perluasan jumlah dan cakupan data, baik dari sisi wilayah geografis maupun periode waktu yang lebih panjang, juga disarankan untuk memperkaya variasi pola dan meningkatkan generalisasi model.

DAFTAR PUSTAKA

- Anak BCE, Rahman MA, Karim AMK, Osman NH. 2021. Fuzzy rank cluster top k Euclidean distance and triangle based algorithm for magnetic field in door positioning system. *Alexandria Engineering Journal*, 3645-3655. Retrieved from <https://doi.org/10.1016/j.aej.2021.08.073>
- Arnita, Yani M, Marpaung F, Hidayat M, Widianto A. 2022. A comparative study of convolutional neural network and *k*-nearest neighbours algorithms for food image recognition. *XXVII*. doi:10.25743/ICT.2022.27.6.008
- Cardoso P, Amponsah-Mensah K, Barreiros JP, Bouhuys J, Cheung H, Davies A, Longhorn SJ. 2021. Scientists' warning to humanity on illegal or unsustainable wildlife trade. *Biological Conservation*, CCLXIII. Retrieved from <https://doi.org/10.1016/j.biocon.2021.109341>
- Eshimiakhe D, Lawal K. 2022. Application of K-means algorithm to Werner deconvolution solutions for depth and image estimations. *Heliyon*. Retrieved from <https://doi.org/10.1016/j.heliyon.2022.e11665>

- Febrian ME, Ferdinan FX, Sendani GP, Suryanigrum KM, Yunanda R. 2022. Diabetes prediction using supervised machine learning. *Procedia Computer Science*, 21-30. Retrieved from <https://creativecommons.org/licenses/by-nc-nd/4.0>
- Fukushima CS, Tricorache P, Toomes A, Stringham OC, Rivera-T'eliez E. 2021. Challenges and perspectives on tackling illegal or unsustainable wildlife trade. *Biological Conservation*. Retrieved from <https://doi.org/10.1016/j.biocon.2021.109342>
- Hitchens RT, Blakeslee AM. 2020. Trends in illegal wildlife trade: Analyzing personal baggage seizure data in the Pacific Northwest. Retrieved from <https://doi.org/10.1371/journal.pone.0234197>
- Ilbeigipour S, Albadvi A, Noughabi EA. 2022. Cluster-based analysis of COVID-19 cases using self-organizing map neural network and K-means methods to improve medical decision-making. *Informatics in Medicine Unlocked*. Retrieved from <https://doi.org/10.1016/j.imu.2022.101005>
- Jiao Y, Yeophantong P, Lee TM. 2021. Strengthening International Legal Cooperation to Combat the Illegal Wildlife Trade Between Southeast Asia and China. *IX*. doi:10.3389/fevo.2021.645427
- Kumar R, Puran R, Dhar J. 2011. Enhanced K-Means Clustering Algorithm Using Red Black Tree and Min-Heap. *International Journal of Innovation, Management and Technology*, II(1).
- Kumar YHS, Manohar N, Chethan HK. 2015. Animal Classification System: A Block Based Approach. *International Conference on Advanced Computing Technologies and Applications (ICACTA2015)*, 336 – 343. doi:10.1016/j.procs.2015.03.156
- Lee TE, Roberts DL. 2020. Moving Beyond Simple Descriptive Statistics in the Analysis of Online Wildlife Trade: An Example From Clustering and Ordination. *XIII*, 1-10. doi:10.1177/1940082920958401
- Melo de Sousa L, Lucanus O, Arroyo-Mora JP, Kalacska M. 2021. Conservation and trade of the endangered Hypancistrus zebra (Siluriformes, Loricariidae), one of the most trafficked Brazilian fish. *Global Ecology and Conservation*. Retrieved from <https://doi.org/10.1016/j.gecco.2021.e01570>
- Nijman V. 2009. An overview of international wildlife trade from Southeast Asia. *XIX*, 1101–1114. doi:10.1007/s10531-009-9758-4
- Roodpishi MV, Nashtaei RA. 2015. Market basket analysis in insurance industry. *Management Science Letters*. doi:10.5267/j.msl.2015.2.004
- Salamat MA, Delfin EB, Valguna RH, Reyes SS, Rabago LR, Boydon EC. 2021. CITESRANGER: Illegal Wildlife Trade Detection Tool. Retrieved from <https://www.researchgate.net/publication/331824364>
- Saputra DM, Saputra D, Oswari LD. 2020. Effect of Distance Metrics in Determining K-Value in KMeans Clustering Using Elbow and Silhouette Method. *CLXXII*. Retrieved from <http://creativecommons.org/licenses/by-nc/4.0/>
- Slabova M, Setlikova I, Vodrazkova M, Richtrova Z, Berec M. 2020. The number of shipments in the CITES Trade Database does not accurately reflect the volume of traded wildlife. *Biological Conservation*. Retrieved from <https://doi.org/10.1016/j.biocon.2020.108917>
- Syakur M, Khotimah B, Rochman E, Satoto B. 2018. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conf. Series: Materials Science and Engineering*. doi:10.1088/1757-899X/336/1/012017
- Tatiana K, Mikhail M. 2018. Market basket analysis of heterogeneous data sources for recommendation system improvement. *Procedia Computer Science*, 246–254.
- Uçar MK, Nour M, Sindi H, Polat K. 2020. The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets. (A. Beghdadi, Ed.) *Mathematical Problems in Engineering*. Retrieved from <https://doi.org/10.1155/2020/2836236>