

Taksonomi Pola Perdagangan Internasional: Implementasi *Unsupervised Learning* untuk Segmentasi Negara Berbasis Kinerja Impor

Aprilisa Arum Sari^{*1}, Oktalia Kumala Sari²

¹Program Studi Teknik Informatika, Universitas Duta Bangsa Surakarta

²Program Studi Magister Ilmu Komputer, Universitas Budi Luhur Jakarta

E-mail: ^{*1}aprilisa_arumsari@udb.ac.id, ²2311601187@student.budiluhur.ac.id

Abstrak

Perdagangan internasional, khususnya aktivitas impor, merupakan indikator vital dalam mengukur kesehatan ekonomi dan tingkat interkoneksi suatu negara di pasar global. Namun, penggunaan klasifikasi geopolitik statis konvensional seringkali gagal merefleksikan dinamika transaksi riil yang volatil. Penelitian ini bertujuan untuk mengekstraksi taksonomi pola perdagangan dunia menggunakan teknik data mining melalui pendekatan *unsupervised learning* (*K-Means Clustering*). Dengan memanfaatkan data panel historis Bank Dunia yang mencakup 7.763 observasi, penelitian ini melakukan *feature engineering* pada variabel rata-rata volume impor dan tingkat volatilitas historis untuk menangkap profil risiko dan kapasitas pasar secara simultan. Hasil analisis berhasil mengidentifikasi tiga kluster utama: (1) Raksasa Ekonomi/Pasar Utama (52 negara) dengan rata-rata impor mencapai \$USD\$ 204,5 miliar yang berfungsi sebagai pusat risiko global, (2) Pasar Berkembang (95 negara) dengan aktivitas impor menengah sebesar \$USD\$ 9,8 miliar, dan (3) Ekonomi Marginal (44 negara) dengan rata-rata impor minimal sebesar \$USD\$ 1,09 miliar. Taksonomi berbasis data ini memberikan landasan empiris yang krusial bagi otoritas perdagangan dalam merumuskan kebijakan yang adaptif serta memitigasi risiko volatilitas rantai pasok global di masa depan.

Kata kunci - Clustering, Data Mining, Ekonomi Global, Perdagangan Internasional, Volatilitas Impor.

1. PENDAHULUAN

Perdagangan internasional merupakan pilar utama yang menggerakkan roda ekonomi global, di mana nilai impor menjadi indikator vital untuk mengukur permintaan domestik, daya beli, dan interkoneksi antarnegara[1], [2]. Namun, permasalahan utama yang dihadapi saat ini adalah ketergantungan otoritas ekonomi pada klasifikasi geopolitik statis yang seringkali tidak mencerminkan dinamika transaksi aktual di lapangan. Berdasarkan laporan *Global Trade Update* yang dirilis oleh UNCTAD pada Desember 2025, perdagangan global berada pada jalur pemecahan rekor dengan total nilai yang diperkirakan menembus \$USD\$ 35 triliun pada akhir tahun 2025[3], [4]. Pertumbuhan ini didorong oleh ketahanan perdagangan jasa dan pemulihan sektor barang yang signifikan. Namun, UNCTAD menggarisbawahi bahwa di balik angka rekor tersebut, terdapat pergeseran struktural yang mendalam, di mana pola perdagangan kini lebih banyak dipengaruhi oleh kedekatan geo-ekonomi dan aliansi strategis dibandingkan sekadar jarak geografis murni[1], [5].

Dalam upaya memetakan kompleksitas data berskala besar tersebut, berbagai studi telah menunjukkan efektivitas penggunaan *unsupervised learning* untuk membentuk taksonomi yang objektif. Singh et al. (2026) mendecode disparitas sosio-ekonomi dengan menggabungkan Taksonomi Wroclaw dan *K-means clustering*, yang membuktikan bahwa pendekatan *data-driven* mampu mengungkap kesenjangan wilayah yang tidak tertangkap oleh analisis statistik deskriptif[6]. Relevansi taksonomi juga ditekankan dalam domain keamanan industri oleh

Zhukabayeva et al. (2025), yang menunjukkan bahwa struktur klasifikasi yang kuat adalah kunci bagi transparansi and akuntabilitas model *machine learning*[7]. Lebih lanjut, metodologi *clustering* kini telah berevolusi menjadi instrumen untuk regionalisasi spasial dan profil dimensi beban yang lebih akurat, sebagaimana ditunjukkan oleh Ke et al. (2026) maupun Michalakopoulos et al. (2026). Singha et al. (2026) turut memperkuat argumen bahwa integrasi *clustering* tetap menjadi fondasi kuat dalam pembelajaran dengan data yang heterogen[8][9].

Meskipun demikian, terdapat beberapa keterbatasan pada pendekatan klasifikasi yang selama ini digunakan. Klasifikasi negara berdasarkan kategori administratif atau geopolitik, seperti Income Group, kawasan geografis, dan status pembangunan, memberikan kategorisasi yang konsisten dan mudah digunakan, tetapi bersifat relatif statis sehingga tidak secara langsung mencerminkan perubahan perilaku perdagangan aktual. Di sisi lain, penelitian berbasis *clustering* telah membuktikan kemampuannya dalam berbagai domain, tetapi penerapan yang dirujuk dalam penelitian terdahulu masih berfokus pada disparitas sosial-ekonomi, regionalisasi spasial, profil energi, dan domain lainnya [6]–[9]. Dengan demikian, masih terdapat ruang penelitian untuk menerapkan *unsupervised learning* secara khusus dalam membangun taksonomi negara berdasarkan karakteristik perdagangan impor global. Selain itu, penggunaan ukuran volume perdagangan saja belum cukup untuk membedakan negara yang memiliki skala impor serupa tetapi menunjukkan tingkat fluktuasi perdagangan yang berbeda. Negara dengan rata-rata impor tinggi dan volatilitas rendah memiliki karakteristik yang berbeda dari negara dengan rata-rata impor tinggi tetapi volatilitas yang besar. Dimensi volatilitas tersebut penting karena dapat memberikan informasi tambahan mengenai tingkat ketidakstabilan atau kerentanan historis perdagangan suatu negara.

Research gap penelitian ini terletak pada belum terintegrasinya dimensi volume dan volatilitas historis dalam taksonomi negara berbasis *unsupervised learning*. Penelitian ini mengisi kesenjangan tersebut dengan menggunakan rata-rata nilai impor sebagai indikator volume dan standar deviasi historis sebagai indikator volatilitas dalam algoritma K-Means, sehingga negara dikelompokkan berdasarkan kemiripan karakteristik perdagangan aktual tanpa label yang ditentukan sebelumnya. Kebaruan penelitian terletak pada penerapan kombinasi kedua dimensi tersebut pada skala global yang mencakup 191 negara, serta interpretasi setiap kluster untuk mengidentifikasi profil perdagangan dan implikasi risiko yang berbeda. Inovasi utama dalam penggunaan pendekatan *unsupervised learning* pada konteks ini terletak pada kemampuannya untuk menangani sifat non-linier dan heterogenitas data perdagangan global yang sering kali terdistorsi oleh fluktuasi harga atau guncangan pasar[10], [11], [12]. Dibandingkan dengan metode pembobotan manual yang rentan terhadap bias subjektif, teknik *clustering* berbasis algoritma K-Means memungkinkan identifikasi struktur laten dalam data impor yang tidak segera terlihat melalui observasi statistik konvensional[13], [14].

Data dari *World Trade Organization* (WTO) pada awal tahun 2026 turut memperkuat konteks ini dengan mencatat bahwa meskipun tahun 2025 menunjukkan pertumbuhan volume sebesar 4,6%, proyeksi untuk tahun 2026 cenderung melambat ke angka 1,9%[1]. Fluktuasi ini dipicu oleh meningkatnya hambatan perdagangan non-tarif dan disrupsi rantai pasok di jalur-jalur logistik utama[15], [16]. Dalam perspektif *Data Mining*, kompleksitas data perdagangan yang masif and dinamis ini tidak lagi dapat diprediksi secara akurat hanya dengan menggunakan model statistik linier sederhana[17]. Dibutuhkan kemampuan untuk mengekstraksi pola tersembunyi dari variabel yang bersifat multidimensi, mencakup aspek temporal (waktu) dan spasial (geografis). Oleh karena itu, tujuan penelitian ini adalah membangun sebuah taksonomi empiris negara-negara di dunia dengan mengintegrasikan dimensi volume serapan pasar dan tingkat kerentanan historis ke dalam model *clustering* murni yang bersifat *data-driven*.

2. METODE PENELITIAN

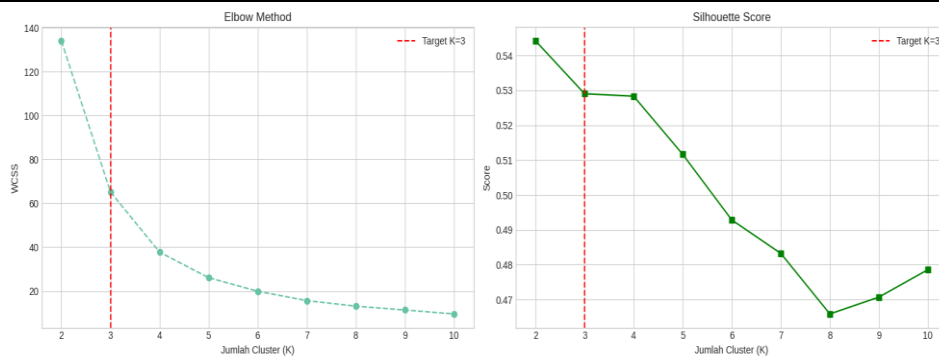
Penelitian ini mengadopsi metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang disusun secara sistematis untuk mengekstraksi pengetahuan dari data perdagangan internasional. Pada tahap *Business and Research Understanding*, penelitian

difokuskan pada kebutuhan untuk membangun taksonomi negara yang lebih adaptif terhadap karakteristik perdagangan aktual dibandingkan klasifikasi geopolitik konvensional. Tujuan utama tahap ini adalah merumuskan segmentasi negara berdasarkan karakteristik kinerja impor historis yang dapat digunakan sebagai dasar analisis kapasitas pasar dan profil risiko perdagangan.

Pada tahap *Data Understanding* dan *Data Preparation*, penelitian menggunakan dataset sekunder Bank Dunia yang terdiri atas 7.763 observasi. Dataset mencakup variabel utama *Country*, *Year*, dan *Imports_Current_USD*, serta atribut pendukung berupa *Continent*, *Subregion*, *Income_Group*, dan *Development_Status*. Tahap pra-pemrosesan dilakukan melalui penanganan nilai kosong dan pemeriksaan konsistensi data, sedangkan atribut geografis dan administratif digunakan sebagai informasi pendukung dalam interpretasi hasil kluster, bukan sebagai fitur utama pembentuk kluster. Selanjutnya, pada tahap *Feature Engineering*, data panel ditransformasikan menjadi representasi profil perdagangan masing-masing negara melalui dua fitur utama, yaitu rata-rata nilai impor tahunan (*Mean Import*) dan standar deviasi historis nilai impor (*Import Volatility*). Rata-rata impor digunakan untuk merepresentasikan skala atau kapasitas pasar suatu negara, sedangkan standar deviasi digunakan untuk menangkap tingkat fluktuasi perdagangan dari waktu ke waktu. Pemilihan kedua fitur tersebut dimaksudkan untuk membentuk taksonomi yang secara simultan mempertimbangkan dimensi skala perdagangan dan stabilitas historisnya. Agregasi data menghasilkan 191 entitas negara dari 7.763 observasi awal, sehingga unit analisis berada pada tingkat negara dan merepresentasikan karakteristik perdagangan dalam periode pengamatan.

Meskipun penggunaan dua fitur utama memberikan representasi yang terfokus dan mudah diinterpretasikan, pendekatan ini memiliki keterbatasan dalam menggambarkan kondisi ekonomi suatu negara secara menyeluruh. Faktor lain seperti nilai ekspor, pertumbuhan ekonomi, inflasi, nilai tukar, utang luar negeri, dan indikator makroekonomi lainnya belum dimasukkan sebagai fitur pembentuk kluster. Oleh karena itu, taksonomi yang dihasilkan dalam penelitian ini tidak dimaksudkan sebagai klasifikasi komprehensif kondisi ekonomi negara, melainkan sebagai segmentasi yang secara khusus menggambarkan karakteristik perdagangan impor berdasarkan dimensi volume dan volatilitas historis. Atribut *Income_Group*, *Development_Status*, *Continent*, dan *Subregion* tidak digunakan sebagai variabel pembentuk kluster untuk mencegah hasil segmentasi secara langsung mereplikasi klasifikasi administratif yang telah ada, tetapi digunakan pada tahap interpretasi untuk memberikan konteks terhadap karakteristik masing-masing kluster.

Pada tahap *Modeling*, algoritma *unsupervised learning* K-Means diterapkan terhadap dua fitur yang telah dinormalisasi. K-Means dipilih karena mampu mengelompokkan objek berdasarkan kedekatan karakteristik numerik dan menghasilkan centroid yang dapat digunakan untuk menginterpretasikan profil masing-masing kelompok. Namun, K-Means memiliki keterbatasan karena hasil pengelompokan dipengaruhi oleh penentuan jumlah kluster (K), inisialisasi centroid, serta karakteristik distribusi data. Oleh karena itu, pemilihan jumlah kluster tidak dilakukan secara arbitrer, tetapi dievaluasi menggunakan lebih dari satu pendekatan untuk memperoleh struktur kluster yang lebih dapat dipertanggungjawabkan. Penentuan jumlah kluster optimal dilakukan menggunakan *Elbow Method* dan *Silhouette Score*. *Elbow Method* digunakan dengan menghitung nilai within-cluster sum of squares atau inersia untuk beberapa nilai K. Titik ketika penurunan inersia mulai melandai digunakan sebagai indikasi jumlah kluster yang memberikan keseimbangan antara kesederhanaan model dan homogenitas dalam kluster. Sebagai validasi tambahan, *Silhouette Score* digunakan untuk mengukur kualitas pemisahan antar-kluster sekaligus kekompakan objek di dalam masing-masing kluster. Nilai *Silhouette Score* yang lebih tinggi menunjukkan bahwa suatu objek memiliki kedekatan yang lebih besar dengan klusternya sendiri dibandingkan dengan kluster lainnya. Penggunaan kedua metode tersebut memberikan evaluasi yang saling melengkapi dan *Elbow Method* berfokus pada perubahan inersia, sedangkan *Silhouette Score* mengevaluasi kekompakan dan separasi kluster.



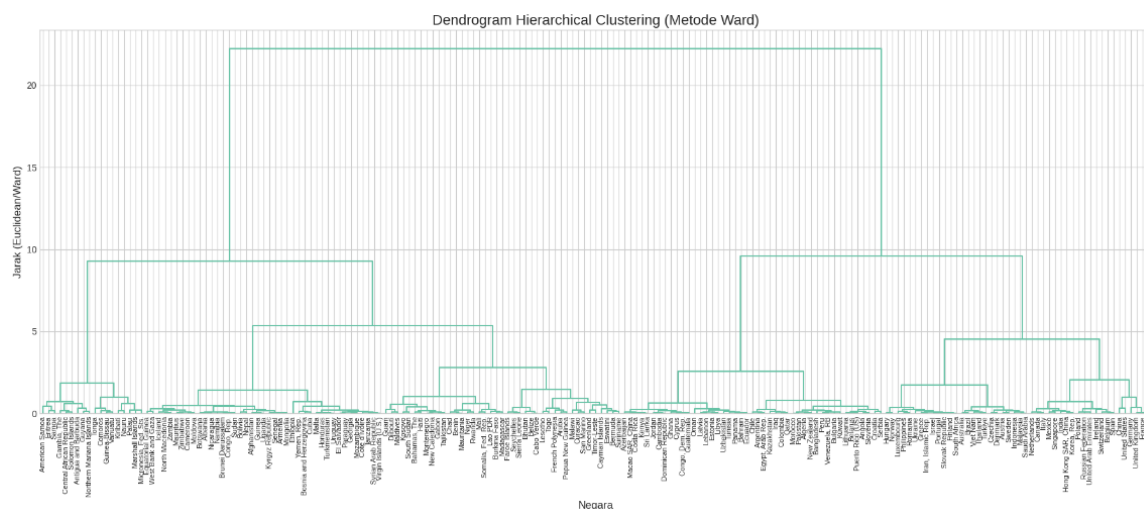
Gambar 1. Metode Elbow untuk Menentukan Jumlah Kluster Optimal ($K=3$)

Berdasarkan hasil evaluasi pada Gambar 1, kurva *Elbow* menunjukkan penurunan inersia yang tajam hingga $K=2$ dan mulai melandai pada $K=3$. Pada saat yang sama, *Silhouette Score* juga digunakan sebagai indikator tambahan untuk mengevaluasi kualitas pemisahan kluster. Hasil kedua evaluasi tersebut digunakan secara bersama-sama untuk menetapkan $K=3$ sebagai jumlah kluster yang digunakan dalam penelitian. Dengan demikian, penetapan tiga kluster tidak hanya didasarkan pada titik siku kurva inersia, tetapi juga mempertimbangkan kualitas pemisahan antar-kluster berdasarkan *Silhouette Score*. Pendekatan validasi ganda ini digunakan untuk mengurangi ketergantungan keputusan terhadap satu indikator evaluasi saja.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Analisis diawali dengan *Exploratory Data Analysis* (EDA) untuk mengidentifikasi karakteristik distribusi nilai impor global. Distribusi nilai impor pada skala asli menunjukkan pola positive skewness yang sangat kuat, dengan sebagian besar negara memiliki nilai impor relatif rendah dan hanya sebagian kecil negara memiliki volume impor yang sangat besar. Kondisi tersebut menunjukkan adanya ketimpangan skala perdagangan antarnegara. Transformasi logaritmik digunakan untuk mengurangi pengaruh nilai ekstrem dan menghasilkan distribusi yang lebih proporsional sebelum proses pengelompokan. Karakteristik ini menjadi dasar penting dalam penerapan K-Means karena perbedaan skala antarnegara dapat memengaruhi pembentukan kluster.



Gambar 2. Analisis Distribusi Nilai Impor Global Menunjukkan Pola Skewness Ekstrem

Gambar 2 menunjukkan struktur hierarki kemiripan negara berdasarkan metode Hierarchical Clustering dengan pendekatan Ward. Dendrogram memperlihatkan bahwa negara-negara membentuk kelompok berdasarkan tingkat kedekatan karakteristik perdagangan, dengan jarak penggabungan (linkage distance) yang semakin besar menunjukkan perbedaan karakteristik yang semakin tinggi. Pada tingkat jarak yang lebih rendah, terbentuk sejumlah kelompok kecil yang memiliki karakteristik relatif homogen, sedangkan pada jarak yang lebih tinggi kelompok-kelompok tersebut bergabung menjadi struktur yang lebih besar. Pola tersebut menunjukkan adanya hierarki kemiripan dalam karakteristik volume dan volatilitas impor antarnegara. Struktur dendrogram juga memberikan dukungan terhadap hasil K-Means yang membentuk tiga kelompok utama. Pemisahan pada tingkat jarak yang relatif tinggi menunjukkan adanya perbedaan karakteristik yang cukup jelas antar kelompok besar, sementara penggabungan pada tingkat jarak yang lebih rendah menunjukkan kemiripan yang lebih kuat di dalam kelompok. Dengan demikian, Hierarchical Clustering digunakan sebagai analisis pendukung untuk melihat struktur hierarkis data dan mengonfirmasi bahwa negara-negara tidak tersebar secara acak, melainkan memiliki pola kemiripan berdasarkan karakteristik perdagangan yang dianalisis. Namun, dendrogram ini tidak digunakan sebagai dasar utama penentuan jumlah klaster karena segmentasi akhir penelitian ditetapkan menggunakan K-Means dan dievaluasi melalui Elbow Method serta Silhouette Score.

Table 1. Hasil Clustering

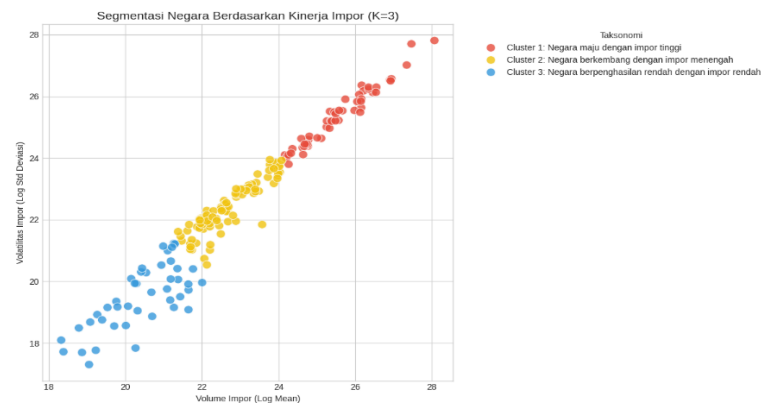
Klaster	Deskripsi Profil	Jumlah Negara	Rata-rata Impor Tahunan (USD)	Rata-rata Volatilitas (USD)
Klaster 1	Raksasa Ekonomi / Impor Tinggi	52	204.538.667.826	167.753.064.353
Klaster 2	Pasar Berkembang / Impor Menengah	95	9.885.098.594	7.610.050.420
Klaster 3	Ekonomi Marginal / Impor Rendah	44	1.095.361.842	844.674.846

Tabel 1 menunjukkan hasil pengelompokan 191 negara ke dalam tiga klaster berdasarkan rata-rata impor tahunan dan volatilitas impor. Klaster 1 terdiri atas 52 negara dengan rata-rata impor tahunan sebesar USD 204,54 miliar dan rata-rata volatilitas sebesar USD 167,75 miliar. Nilai tersebut menunjukkan bahwa kelompok ini memiliki aktivitas impor dengan skala paling besar dibandingkan dua klaster lainnya. Klaster 2 merupakan kelompok terbesar dengan 95 negara, memiliki rata-rata impor sebesar USD 9,89 miliar dan volatilitas sebesar USD 7,61 miliar, sehingga merepresentasikan negara dengan karakteristik perdagangan pada tingkat menengah. Sementara itu, Klaster 3 terdiri atas 44 negara dengan rata-rata impor sebesar USD 1,10 miliar dan volatilitas sebesar USD 0,84 miliar, yang menunjukkan tingkat aktivitas impor paling rendah.

Perbedaan nilai tersebut menunjukkan bahwa rata-rata impor menjadi karakteristik utama yang membedakan ketiga klaster, sedangkan volatilitas memberikan informasi tambahan mengenai variasi historis aktivitas perdagangan. Terlihat bahwa volatilitas meningkat seiring dengan meningkatnya rata-rata impor. Hal ini mengindikasikan bahwa negara dengan skala perdagangan yang lebih besar cenderung memiliki fluktuasi nilai impor yang lebih besar secara absolut. Oleh karena itu, tingginya volatilitas pada Klaster 1 tidak secara langsung dapat diartikan sebagai tingkat risiko ekonomi yang lebih tinggi, karena nilai volatilitas juga dipengaruhi oleh besarnya skala perdagangan. Dengan demikian, ketiga klaster lebih tepat dipahami sebagai segmentasi berdasarkan skala dan dinamika perdagangan impor, yaitu negara dengan impor tinggi, menengah, dan rendah.

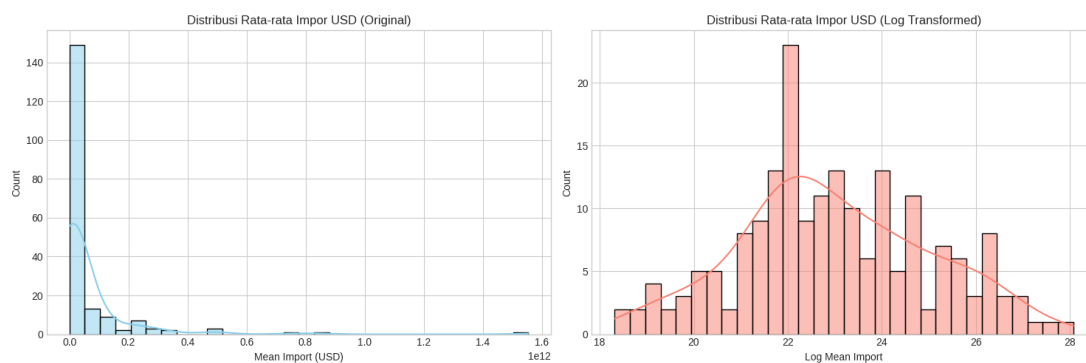
Implementasi algoritma K-Means menghasilkan tiga klaster yang distingtif berdasarkan fitur rata-rata impor dan standar deviasi (volatilitas). Secara kuantitatif, Klaster 1 terdiri dari 52 negara dengan rata-rata impor mencapai \$USD\$ 204.538.667.826 dan tingkat volatilitas sebesar \$USD\$ 167.753.064.353. Klaster 2 merupakan kelompok terbesar dengan 95 negara, mencatatkan rata-rata impor sebesar \$USD\$ 9.885.098.594 dengan volatilitas \$USD\$ 7.610.050.420. Sementara itu, Klaster 3 mencakup 44 negara dengan performa impor paling rendah, yakni rata-rata \$USD\$ 1.095.361.842 dan volatilitas sebesar \$USD\$ 844.674.846.

Ringkasan statistik ini disajikan pada Tabel 1, sedangkan sebaran spasial hasil pengelompokan dapat dilihat pada Gambar 3.



Gambar 3. Visualisasi Sebaran Negara Berdasarkan Rata-rata Impor dan Volatilitas

Gambar 3 menampilkan *scatter plot* yang memetakan posisi setiap negara. Terlihat pemisahan spasial yang sangat jelas; Klaster 1 menempati kuadran kanan atas yang menunjukkan dominasi volume dan volatilitas. Klaster 2 berada di area transisi tengah, sementara Klaster 3 menumpuk di titik rendah, mengonfirmasi profil interkoneksi global yang sangat terbatas.



Gambar 4. Perbandingan Rata-rata Impor dan Volatilitas Antar Klaster

Visualisasi diagram batang pada Gambar 4 mempertegas disparitas ekonomi yang tajam antar kelompok. Tinggi batang Klaster 1 mendominasi seluruh spektrum data, di mana rata-rata impornya mencapai ratusan kali lipat lebih besar dibandingkan Klaster 3. Hal ini memberikan bukti grafis bahwa kekuatan konsumsi dunia saat ini masih sangat terkonsentrasi pada segelintir pasar utama, sementara mayoritas negara lainnya memiliki profil risiko dan kapasitas yang jauh lebih rendah.

3.2. Pembahasan Penelitian

Hasil penelitian menunjukkan bahwa 191 negara dapat dikelompokkan ke dalam tiga klaster berdasarkan kombinasi rata-rata impor dan volatilitas historis. Klaster 1 terdiri atas 52 negara dengan karakteristik volume impor dan volatilitas absolut tertinggi, Klaster 2 mencakup 95 negara dengan karakteristik perdagangan menengah, sedangkan Klaster 3 terdiri atas 44 negara dengan tingkat impor dan volatilitas relatif rendah. Perbedaan tersebut menunjukkan bahwa skala aktivitas impor merupakan faktor utama yang membedakan posisi negara dalam klaster, sedangkan volatilitas memberikan informasi tambahan mengenai variasi historis aktivitas perdagangan. Dengan demikian, suatu negara cenderung berada pada Klaster 1 ketika memiliki volume impor dan volatilitas yang tinggi, sementara negara dengan kedua karakteristik pada tingkat menengah atau rendah cenderung masuk ke Klaster 2 atau Klaster 3. Pola ini menunjukkan

bahwa segmentasi yang dihasilkan merepresentasikan karakteristik perdagangan aktual dan tidak semata-mata mengikuti klasifikasi administratif seperti Income Group atau kawasan geografis.

Klaster 1 merepresentasikan kelompok negara dengan aktivitas impor berskala sangat besar, yang mencakup antara lain Argentina, Australia, Austria, Belgium, Brazil, Canada, China, France, Germany, Japan, United Kingdom, dan United States. Tingginya volatilitas pada kelompok ini perlu dipahami dalam konteks skala perdagangan. Karena volatilitas diukur menggunakan standar deviasi nilai impor dalam satuan absolut USD, negara dengan volume impor yang besar secara alami berpotensi memiliki nilai fluktuasi absolut yang lebih besar. Oleh karena itu, tingginya volatilitas Klaster 1 tidak secara langsung menunjukkan tingkat ketidakstabilan ekonomi yang lebih tinggi, melainkan menunjukkan besarnya dinamika perdagangan dalam skala absolut. Temuan ini memperlihatkan bahwa penggunaan dua fitur secara simultan dapat membedakan negara yang memiliki kapasitas perdagangan besar dari negara dengan aktivitas impor yang lebih terbatas.

Klaster 2 merupakan kelompok terbesar dengan 95 negara, antara lain Afghanistan, Albania, Algeria, Angola, Armenia, Azerbaijan, Bangladesh, Egypt, Indonesia, Philippines, dan Vietnam. Kelompok ini memiliki tingkat impor dan volatilitas yang berada di antara Klaster 1 dan Klaster 3. Posisi tersebut menunjukkan bahwa sebagian besar negara dalam dataset memiliki karakteristik perdagangan pada tingkat menengah. Sementara itu, Klaster 3 terdiri atas 44 negara, seperti Aruba, Belize, Benin, Bermuda, Bhutan, Burundi, Central African Republic, dan Gambia, dengan rata-rata impor dan volatilitas terendah. Karakteristik tersebut menunjukkan keterbatasan skala aktivitas impor dibandingkan kelompok lainnya. Namun, hasil clustering tidak dapat digunakan untuk menyimpulkan secara langsung penyebab struktural seperti rendahnya daya beli, keterisolasian ekonomi, atau ketergantungan terhadap komoditas karena faktor-faktor tersebut tidak digunakan sebagai variabel pembentuk klaster dalam penelitian ini. Interpretasi yang diberikan dibatasi pada karakteristik volume dan volatilitas impor yang terukur dalam dataset.

Pola pemisahan tersebut diperkuat oleh visualisasi scatter plot yang menunjukkan kecenderungan diagonal antara volume dan volatilitas impor. Negara dengan volume impor yang lebih tinggi cenderung memiliki volatilitas absolut yang lebih tinggi, sehingga kedua fitur tersebut secara bersama-sama membentuk struktur segmentasi yang relatif jelas. Selain itu, Hierarchical Clustering dengan metode Ward digunakan sebagai analisis pendukung untuk melihat struktur kemiripan antarnegara secara hierarkis. Dendrogram menunjukkan adanya pengelompokan negara berdasarkan tingkat kedekatan karakteristik perdagangan, sehingga memberikan perspektif tambahan terhadap struktur data yang dihasilkan K-Means. Meskipun demikian, K-Means tetap digunakan sebagai metode utama karena tujuan penelitian adalah membentuk segmentasi tiga kelompok berdasarkan karakteristik centroid yang dapat diinterpretasikan secara langsung.

Dari sisi validasi, penentuan jumlah klaster tidak hanya didasarkan pada pertimbangan interpretatif, tetapi juga menggunakan Elbow Method dan Silhouette Score sebagai evaluasi internal. Elbow Method digunakan untuk mengidentifikasi titik ketika penurunan inersia mulai melandai, sedangkan Silhouette Score digunakan untuk menilai kekompakan objek dalam klaster dan pemisahan antar-klaster. Penggunaan kedua indikator tersebut memberikan dasar yang lebih kuat dibandingkan hanya menggunakan satu metode dalam menentukan jumlah klaster. Namun, karena penelitian ini tidak menggunakan label kelas sebagai ground truth, hasil clustering tidak dinilai menggunakan metrik akurasi klasifikasi. Perbandingan kuantitatif yang lebih luas dengan algoritma lain, seperti Hierarchical Clustering atau DBSCAN, masih dapat dilakukan untuk menguji robustnes taksonomi yang dihasilkan.

Hasil penelitian ini memberikan kontribusi dalam membentuk taksonomi perdagangan berbasis data yang mempertimbangkan dua dimensi sekaligus, yaitu skala impor dan volatilitas historis. Berbeda dari klasifikasi administratif yang telah menetapkan kelompok berdasarkan kategori tertentu, pendekatan ini membentuk kelompok berdasarkan kemiripan karakteristik perdagangan aktual. Dengan demikian, taksonomi yang dihasilkan dapat digunakan sebagai dasar segmentasi negara berdasarkan profil perdagangan, meskipun interpretasinya tetap terbatas pada dua fitur yang digunakan. Selain itu, penelitian ini menggunakan agregasi data historis menjadi

satu profil untuk setiap negara, sehingga belum dapat menjelaskan perubahan keanggotaan klaster antarperiode waktu. Negara yang saat ini berada pada suatu klaster dapat mengalami perubahan karakteristik perdagangan pada periode berikutnya, tetapi dinamika tersebut belum dapat ditangkap oleh model agregat penelitian ini. Analisis longitudinal pada beberapa periode selanjutnya diperlukan untuk mengidentifikasi perpindahan negara antarklaster, stabilitas keanggotaan klaster, dan perubahan struktur perdagangan global dari waktu ke waktu.

4. KESIMPULAN

Penelitian ini menghasilkan taksonomi perdagangan internasional berbasis K-Means clustering yang mengelompokkan 191 negara ke dalam tiga profil berdasarkan rata-rata impor dan volatilitas historis, yaitu kelompok dengan volume impor tinggi, menengah, dan rendah. Hasil tersebut menunjukkan bahwa kombinasi volume dan volatilitas dapat memberikan perspektif yang lebih spesifik terhadap karakteristik perdagangan aktual dibandingkan pengelompokan administratif atau geografis. Secara praktis, taksonomi ini dapat digunakan sebagai dasar segmentasi kebijakan perdagangan, yaitu pemantauan eksposur dan kesinambungan rantai pasok bagi negara dengan volume impor tinggi, penguatan diversifikasi dan pengembangan pasar bagi kelompok menengah, serta peningkatan kapasitas perdagangan dan akses trade finance bagi kelompok dengan volume impor rendah. Namun, penelitian ini memiliki keterbatasan karena hanya menggunakan rata-rata impor dan volatilitas historis sebagai fitur utama, sehingga belum merepresentasikan kondisi ekonomi secara menyeluruh; K-Means juga memiliki ketergantungan terhadap jumlah klaster dan karakteristik data meskipun telah dievaluasi menggunakan Elbow Method dan Silhouette Score. Selain itu, agregasi data historis menjadi satu profil negara belum memungkinkan analisis perubahan keanggotaan klaster antarperiode. Penelitian selanjutnya dapat memperluas fitur dengan indikator ekonomi dan perdagangan lainnya, membandingkan beberapa algoritma clustering, serta melakukan analisis longitudinal untuk mengidentifikasi perpindahan dan stabilitas klaster dari waktu ke waktu.

5. SARAN

Sebagai pengembangan penelitian selanjutnya, prioritas utama diarahkan pada peningkatan validasi dan robustnes taksonomi melalui penggunaan beberapa metrik evaluasi clustering, seperti Silhouette Score, Davies–Bouldin Index, dan Calinski–Harabasz Index, serta perbandingan dengan algoritma alternatif seperti Hierarchical Clustering, DBSCAN, dan metode berbasis Deep Learning. Validasi juga perlu dilakukan menggunakan dataset perdagangan dari sumber atau periode yang berbeda untuk menguji konsistensi dan generalisasi hasil taksonomi. Selanjutnya, variabel makroekonomi tambahan, seperti nilai ekspor, utang luar negeri, dan tingkat inflasi, dapat diintegrasikan untuk menghasilkan profil perdagangan yang lebih komprehensif. Setelah model teruji secara robust, pengembangan dasbor pemantauan berbasis taksonomi dapat menjadi tahap lanjutan untuk membantu praktisi dan pembuat kebijakan memantau perubahan karakteristik perdagangan serta mendukung pengambilan keputusan yang lebih tepat sasaran.

DAFTAR PUSTAKA

- [1] O. K. Muhammad Romli, Ikmal Mumtahaen, “Penyelesaian Sengketa Perdagangan Internasional Melalui Dispute Settlement Body,” *J. Artif. Intell. Digit. Bus.*, Vol. 4, No. 2, 2025.
- [2] Nurhikmah And I. Rasulong, “Dampak Perang Dagang As-China Terhadap Perdagangan Internasional Dan Ekonomi Global,” *Researchgate*, No. January, 2025.
- [3] A. Giroud, J. Puck, And H. Pühr, “World Investment Report 2025: International Investment In The Digital Economy,” *J. Int. Bus. Policy*, 2026, Doi: 10.1057/S42214-025-00234-6.
- [4] I. Svyshchuk, “Concentration Of Exports And Imports Of Creative Industry Goods In

- Global Trade,” *Bull. Acad. Labor, Soc. Relations Tour. Ser. Econ. Psychol. Manag.*, No. 7, 2026, Doi: 10.54929/3041-2390-2026-07-01-01.
- [5] L. J. Falkenberg And P. W. S. Joyce, “Tracking The Coastal Bioinvasion By Mussels In Hong Kong,” *Mar. Pollut. Bull.*, Vol. 222, 2026, Doi: 10.1016/J.Marpolbul.2025.118678.
- [6] V. Singh, G. C. Hajare, S. A. Khan, And A. Kumari, “Decoding Socio-Economic Disparities In Uttar Pradesh: A Spatio-Temporal Analysis Using Wroclaw Taxonomy And K-Means Unsupervised Machine Learning Clustering,” *Math. Popul. Stud.*, 2026, Doi: 10.1080/08898480.2025.2612641.
- [7] T. Zhukabayeva *Et Al.*, “An Impact-Aware And Taxonomy-Driven Explainable Machine Learning Framework With Edge Computing For Security In Industrial Iot–Cyber Physical Systems,” *C. - Comput. Model. Eng. Sci.*, Vol. 145, No. 2, Pp. 2573–2599, 2025, Doi: 10.32604/Cmes.2025.070426.
- [8] V. Michalakopoulos *Et Al.*, “Exploring Dimensional Distinctions Of Residential Heat Load Profiles Using An Unsupervised Machine Learning Clustering Framework,” *Sustain. Energy, Grids Networks*, Vol. 45, 2026, Doi: 10.1016/J.Segan.2025.102117.
- [9] X. Ke *Et Al.*, “From Prediction To Regionalization: Enhancing Flash Flood Susceptibility Mapping Using Machine Learning And Geodetector,” *Geosci. Front.*, Vol. 17, No. 1, 2026, Doi: 10.1016/J.Gsf.2025.102213.
- [10] X.-J. Li, J.-M. Liu, Z.-B. Zhang, Y.-N. Wang, And Y.-Y. Li, “Advancing The Classification Of Gamma-Ray Bursts Using Machine Learning. I. Gamma-Ray Bursts With Extended Emission,” *Astrophys. J.*, Vol. 996, No. 1, 2026, Doi: 10.3847/1538-4357/Ae2274.
- [11] S. Naeem, A. Ali, S. Anam, And M. M. Ahmed, “An Unsupervised Machine Learning Algorithms: Comprehensive Review,” *Int. J. Comput. Digit. Syst.*, Vol. 13, No. 1, 2023, Doi: 10.12785/Ijcds/130172.
- [12] A. Catur Pamungkas, D. Wakhid Nur Sakti, F. Trisna Aqilla, K. Miza, And A. Arum Sari, “Clustering Produk Umkm Berdasarkan Penjualan Online Menggunakan K-Means,” *Pros. Semin. Nas. Teknol. Inf. Dan Bisnis*, 2025, Doi: 10.47701/Hgnnv193.
- [13] Hasim Azari, Dwi Hartanti, And Aprilisa Arum Sari, “Pengelompokan Produksi Padi Dan Beras Provinsi Jawa Timur Dengan Metode Agglomerative Hierarchical Clustering,” *Infotek J. Inform. Dan Teknol.*, Vol. 7, No. 2, 2024, Doi: 10.29408/Jit.V7i2.26016.
- [14] M. Rovi And H. Rustama, “Implementasi Algoritma K-Means Dalam Pengelompokan Gaya Belajar Mahasiswa Ubhi Menggunakan Model Vark,” *Infotech J.*, Vol. 12, No. 1, Pp. 25–32, Jan. 2026, Doi: 10.31949/Infotech.V12i1.16970.
- [15] R. A. Fauzianto And Supatman, “Analisis Sentimen Opini Masyarakat Terhadap Tech Winter Pada Twitter Menggunakan Natural Language Processing,” *J. Syntax Admiration*, Vol. 3, No. 9, 2023, Doi: 10.46799/Jsa.V3i9.909.
- [16] P. Ria Indah Kusuma, K. Dini, And A. Azminah, “Pemodelan Molekul Dan Molekular Dinamik Putik Bunga Safron (*Crocus Sativus L.*) Sebagai Kandidat Obat Antikanker Payudara,” *Pharmacoscrypt*, Vol. 6, No. 2, 2023, Doi: 10.36423/Pharmacoscrypt.V6i2.1391.
- [17] O. Kumala Sari And A. Arum Sari, “Prediksi Kunjungan Wisatawan Dengan Random Forest Berbasis Data Historis Dan Eksternal,” *Pros. Semin. Nas. Teknol. Inf. Dan Bisnis*, 2025, Doi: 10.47701/Kd7t9e88.
- [18] A. Wadanur And A. A. Sari, “Implementation Of Apriori And Fp-Growth Algorithms In Spare Parts Sales,” *Edumatic J. Inf. Educ.*, Vol. 6, No. 1, 2022.