

BAB I

PENDAHULUAN

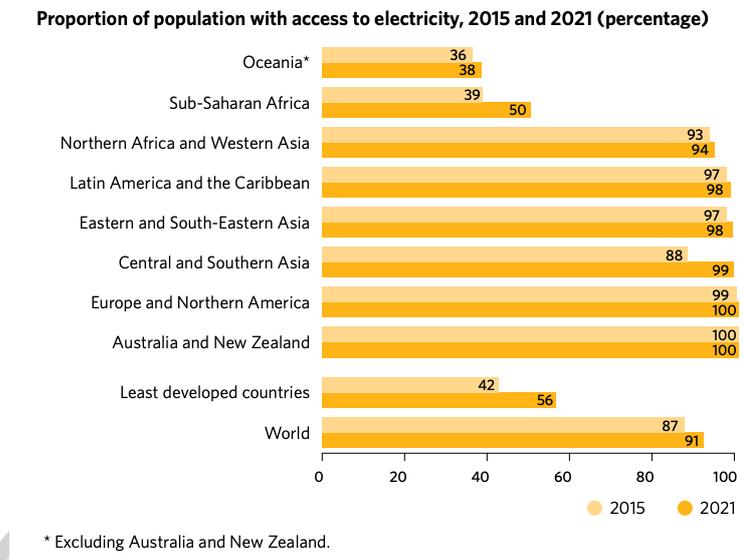
1.1 Latar Belakang

Setiap negara berupaya meningkatkan pertumbuhan ekonomi mereka. Hal tersebut dikarenakan pertumbuhan ekonomi pada suatu negara dapat berdampak pada berbagai aspek. Salah satu contoh aspek yang dapat terpengaruh oleh pertumbuhan ekonomi suatu negara adalah kesejahteraan dan kelangsungan hidup rakyat [1]. Semakin tinggi tingkat pertumbuhan ekonomi pada suatu negara, maka kesejahteraan rakyat semakin terjamin. Untuk menjaga kestabilan perekonomian suatu negara, identifikasi faktor-faktor yang dapat mendorong pertumbuhan dan penurunan ekonomi menjadi hal yang penting. Pada kondisi makro, telah ditemukan banyak faktor yang berkontribusi pada cepatnya pertumbuhan ekonomi, salah satunya adalah infrastruktur yang memiliki peran krusial sebagai salah satu faktor yang mendorong pertumbuhan ekonomi yang berkelanjutan [2].

Secara umum, infrastruktur dapat dibagi dalam beberapa klaster yaitu, air, transportasi, energi, bangunan dan pengelolaan limbah¹. Ketersediaan listrik merupakan salah satu infrastruktur di bidang energi. Definisi ketersediaan listrik secara umum dapat mencakup berbagai aspek listrik, misalkan jumlah listrik yang dikonsumsi, kapasitas pembangkitan, tingkat akses rumah tangga dan kualitas pasokan listrik. Kualitas ketersediaan listrik adalah aspek yang sangat penting bagi suatu negara. Sebagai contoh, jika pemadaman listrik dan lonjakan tegangan sering terjadi akibat ketersediaan listrik yang tidak stabil, maka hal ini dapat menyebabkan kerugian ekonomi yang besar bagi rumah tangga dan perusahaan [3]. Hal ini mengindikasikan bahwa ketersediaan listrik yang buruk dapat merugikan masyarakat dan perusahaan yang secara tidak langsung juga akan berdampak pada perekonomian negara. Menurut penelitian yang dilakukan oleh Stern *et al.*, ketersediaan listrik dan pertumbuhan ekonomi pada suatu negara memiliki hubungan yang sangat erat [4].

Berdasarkan data yang dilaporkan oleh PBB pada bulan Mei tahun 2023 terkait pencapaian *Sustainable Development Goals* (SDGs) berskala global, analisis data menyimpulkan bahwa terdapat 675 juta orang di dunia yang masih kekurangan akses listrik pada tahun 2021. Berikut merupakan grafik untuk proporsi penduduk yang mempunyai akses terhadap listrik, pada tahun 2015 dan

¹Sumber: money.kompas.com



Gambar 1.1: Proporsi Penduduk yang Memiliki Akses Listrik pada Tahun 2015 dan 2021 (%)

2021. Berdasarkan data statistik yang diperoleh, tingkat akses listrik global meningkat dari 87% pada tahun 2015 menjadi 91% pada tahun 2021. Hal ini mengindikasikan bahwa kurang lebih terdapat 800 juta orang yang terbantu dalam peningkatan akses listrik. Namun, 675 juta orang masih kekurangan akses listrik pada tahun 2021, dan sebagian besar populasi tersebut berada di negara yang kurang berkembang. Meskipun terjadi kemajuan yang stabil dalam enam tahun terakhir, tingkat pertumbuhan akses listrik setiap tahunnya antara tahun 2019 dan 2021 mengalami penurunan. Pada periode tersebut, tingkat pertumbuhan akses listrik hanya sebesar 0.6%, yang lebih rendah dibandingkan dengan tingkat pertumbuhan sebesar 0.8% yang terjadi pada periode 2015 hingga 2019. Mengacu pada Sub-Sahara, Afrika, pertumbuhan populasi telah menyebabkan stagnasi jumlah orang tanpa akses listrik sejak tahun 2010 hingga mencapai 567 juta orang pada 2021. Meskipun elektrifikasi dapat memberikan manfaat signifikan, termasuk peningkatan pendidikan, kesehatan, dan mendukung pengembangan pertanian, yang secara tidak langsung akan berpengaruh terhadap pertumbuhan ekonomi, namun jika laju pertumbuhan akan masih sama seperti saat ini, hal ini akan menyebabkan sekitar 660 juta orang masih akan kekurangan akses listrik pada tahun 2030. Maka dari itu, untuk mencapai akses universal pada tahun tersebut, diperlukan peningkatan tingkat akses sebesar 1% setiap tahunnya antara tahun 2021 dan 2030².

²Sumber: unstats.un.org

Berdasarkan data statistik yang diperoleh dari hasil laporan PBB terkait akses listrik populasi dunia, maka dilakukan penelitian dan analisis terkait sektor listrik di berbagai negara dan periode. Hal ini dilakukan untuk mendukung upaya yang telah dilakukan oleh PBB dalam peningkatan akses listrik bagi penduduk di dunia sehingga dapat tersedia data dan analisis yang mudah digunakan oleh pemerintahan negara-negara guna meningkatkan sektor listrik pada negara mereka. Penelitian dan pengembangan pada sektor listrik di suatu negara lebih mudah dilakukan jika mengetahui kesamaan karakteristik dari faktor-faktor yang digunakan untuk menganalisis ketersediaan listrik pada setiap negara. Oleh karena itu, perlu dilakukan analisis terhadap pola, tren, hubungan dan pengetahuan baru yang terkandung dalam data. Salah satu metode untuk mengidentifikasi pola, kelompok, atau hubungan yang ada dalam data adalah dengan menggunakan metode *clustering*. Metode ini merupakan suatu metode yang digunakan untuk mengelompokkan data ke dalam kelompok atau klaster, dengan tujuan agar data dalam satu klaster memiliki tingkat kemiripan yang tinggi dan data antar klaster memiliki tingkat kemiripan yang rendah [5].

Organisasi internasional yang bergerak dalam bidang perekonomian, pembangunan, dan kesehatan, pada umumnya mengelompokkan negara-negara berdasarkan dua kriteria utama, yaitu geografi dan pendapatan untuk mengalokasikan sumber daya dengan lebih efisien dan memberikan gambaran tentang kemajuan yang dicapai oleh kelompok-kelompok negara tersebut. Ketika akan dilakukan penelitian terhadap sektor tertentu yang lebih spesifik, sejumlah indikator yang relevan dapat menjadi dasar yang lebih tepat untuk mengelompokkan negara-negara, dibandingkan hanya menggunakan indeks pada sektor ekonomi atau geografi sebagai indikator.

Dalam penelitian yang berjudul: *Country Clustering Applied to the Water and Sanitation Sector: A Newtool with Potential Applications in Research and Policy* [6], Onda et al. melakukan pengelompokkan negara dengan menciptakan klasifikasi baru yang secara khusus diterapkan pada sektor air dan sanitasi (WatSan) berdasarkan sembilan indikator yang berkaitan dengan WatSan. Pada penelitian tersebut diperoleh lima klaster dari hasil *clustering* terhadap negara-negara, dengan analisis terhadap kelima klaster tersebut menunjukkan bahwa klaster-klaster tersebut lebih kompak dan terpisah dengan baik daripada kelompok negara yang diperoleh oleh PBB dan World Bank. Dengan kata lain, kelima klaster tersebut tidak dapat dijelaskan dengan baik oleh sektor pendapatan atau geografi, dan berbeda dengan kelompok negara yang digunakan dalam pembangunan internasional. Pada penelitian yang telah dilakukan oleh Onda et al.,

dapat disimpulkan bahwa ketika akan dilakukan penelitian terhadap sektor tertentu, pemilihan indikator yang relevan menjadi hal yang penting, dan hasil analisis menunjukkan bahwa melakukan *clustering* dapat memberikan pemahaman yang lebih baik tentang perbedaan dan kesamaan di antara negara-negara dalam konteks layanan air dan sanitasi.

Clustering terhadap negara-negara pada penelitian ini menggunakan sejumlah indeks sektor listrik yang relevan terhadap perkembangan dan ketersediaan layanan listrik di suatu negara. Seperti yang sudah dibahas pada paragraf sebelumnya, ketika akan melakukan analisis terhadap sektor tertentu, akan lebih baik jika indikator yang digunakan relevan dengan sektor yang diteliti. Maka dari itu, pada penelitian ini telah dipilih tujuh variabel yang berkaitan dengan sektor listrik, dan data diperoleh dari *World Bank*. Pada penelitian ini, *clustering* terhadap negara-negara menggunakan indeks sektor listrik akan dilakukan menggunakan tiga metode yang berbeda, yaitu *k-means*, *Gaussian mixture models* (GMM), dan *support vector machine* (SVM), dengan masing-masing metode memiliki pendekatan *machine learning* yang berbeda. *K-means* dan GMM merupakan metode *clustering* yang mengadopsi pendekatan *unsupervised learning*, sedangkan SVM merupakan metode yang berbasis *supervised learning*.

Metode *k-means* yang pertama kali diperkenalkan oleh Macqueen merupakan salah satu algoritma pengelompokan (*clustering*) yang paling sering digunakan. Metode ini tergolong sebagai pengelompokan partisional berbasis prototipe (*prototype-based*), dengan setiap data hanya dapat menjadi anggota dari satu klaster, sehingga dapat meminimalisir kemungkinan data yang bertumpang tindih pada klaster-klaster yang ada. *K-means* memiliki kompleksitas waktu yang dinyatakan dengan notasi $O(NKdT)$, untuk N adalah jumlah data, K adalah jumlah klaster, d adalah dimensi data dan T adalah jumlah iterasi yang dilakukan. Sering kali nilai K , d dan T jauh lebih kecil daripada nilai N . Hal ini mengindikasikan bahwa kompleksitas waktu *k-means* dapat dikatakan linier, yang berarti bahwa waktu yang dibutuhkan untuk melakukan *clustering* tumbuh secara proporsional dengan ukuran data N . Oleh karena itu, dapat dikatakan bahwa *k-means* memiliki kompleksitas waktu yang relatif efisien, sehingga menjadikan metode ini sebagai pilihan yang baik untuk mengelompokkan kumpulan data dengan skala besar seperti melakukan *clustering* pada banyak negara dengan variabel yang bervariasi [7].

Adapun metode *Gaussian mixture models* (GMM) merupakan metode *machine learning* yang banyak digunakan dalam berbagai macam area seperti pemodelan data, pengenalan pola, segmentasi gambar dan pemulihan sinyal.

Selain itu, salah satu fungsi metode GMM adalah untuk melakukan *clustering* pada data. GMM merupakan metode pengelompokan yang bersifat *soft clustering* dengan setiap data diberi probabilitas keanggotaan. Hal ini mengindikasikan bahwa GMM dapat memberikan gambaran yang lebih rinci mengenai sejauh mana data terkait dengan masing-masing kluster [8]. Ketika melakukan *clustering* terhadap banyak data dengan variabel yang bervariasi, melakukan estimasi pada parameter merupakan hal yang penting. GMM menggunakan algoritma *expectation-maximization* (EM) untuk mengestimasi parameter modelnya, sehingga masalah estimasi parameter dalam kasus variabel yang banyak dapat teratasi dengan efisien [9]. Pada penelitian yang dilakukan oleh Androniceanu *et al.* dengan tujuan untuk mengetahui posisi negara-negara di Uni Eropa dalam hal digitalisasi administrasi, metode GMM memberikan hasil *clustering* yang memuaskan. Hal ini karena hasil penelitian menunjukkan perbedaan yang signifikan antar kluster yang terbentuk mengenai keadaan administrasi pada setiap negara [10].

Selain metode *k-means* dan GMM yang merupakan metode *clustering* berbasis *unsupervised learning*, metode SVM juga seringkali digunakan untuk *clustering* atau prediksi. SVM merupakan metode *clustering* berbasis *supervised learning*, dengan algoritmanya adalah membentuk model pelatihan untuk menghasilkan pengelompokan yang diinginkan pada data uji. Dengan kata lain, SVM mencoba untuk mentransfer pengetahuan dari model pelatihan yang sudah diketahui ke data uji, sehingga algoritma dapat menghasilkan pengelompokan yang sesuai dengan data pelatihan yang diberikan [11]. Pada penelitian yang dilakukan oleh Xiahou, *et al.*, metode SVM digunakan untuk memprediksi kluster terhadap pelanggan yang akan berhenti menggunakan layanan suatu perusahaan *e-commerce*. Penelitian yang dilakukan menggunakan model prediksi kerugian yang berdasarkan pada kombinasi kluster pelanggan dengan metode *k-means* dan prediksi menggunakan SVM dan regresi logistik. Hasil penelitian menunjukkan bahwa akurasi prediksi setelah dilakukan segmentasi pelanggan mengalami peningkatan yang signifikan, membuktikan bahwa segmentasi kluster *k-means* diperlukan, dan hasil perbandingan metode prediksi juga menunjukkan bahwa akurasi prediksi menggunakan SVM lebih tinggi dari pada prediksi menggunakan regresi logistik [12].

Pada penelitian ini, *clustering* terhadap negara-negara dengan indeks sektor listrik dilakukan menggunakan tiga metode yang berbeda, seperti yang sudah dipaparkan pada paragraf-paragraf sebelumnya. Meskipun masing-masing metode memiliki keunggulannya tersendiri, namun perlu dilakukan pengujian dan analisis

untuk mengetahui metode yang paling optimal dalam melakukan *clustering* terhadap sektor listrik dengan membandingkan hasil klaster dari ketiga metode yang berbeda, dan juga membandingkan dengan acuan klaster yang diasumsikan benar. Informasi terkait hasil klaster terhadap negara-negara menggunakan indeks sektor listrik belum pernah dipublikasikan sebelumnya oleh institusi resmi. Maka dari itu, untuk melakukan analisis terhadap metode klaster yang paling optimal, digunakan hasil klasifikasi negara yang dilakukan oleh *World Bank* sebagai acuan hasil klaster pada penelitian ini. Hasil klasifikasi negara yang dilakukan oleh *World Bank* dilandaskan oleh indeks *gross national income* (GNI) per kapita, yang kemudian hasil klaster terbagi menjadi empat tingkatan pendapatan negara³. Pada penelitian ini, hasil klasifikasi negara oleh *World Bank* akan diasumsikan sebagai hasil klaster yang benar dan akurat sehingga dapat dijadikan acuan dalam menganalisis metode yang paling optimal berdasarkan hasil klaster yang diperoleh.

Perbandingan hasil klaster oleh *World Bank* dengan metode *k-means*, GMM, dan SVM dilakukan dengan menggunakan uji korelasi *chi-square*. Hal ini bertujuan untuk mengetahui apakah ada korelasi atau kecocokan antara kedua hasil klaster yang dibandingkan sehingga dapat ditentukan metode yang paling optimal dalam melakukan *clustering* menggunakan indeks sektor listrik. Hasil uji korelasi *chi-square* terhadap kedua hasil klaster semakin relevan ketika didukung oleh interpretasi hasil uji Spearman terhadap indeks sektor listrik dan GNI per kapita yang merupakan landasan indeks dalam pengklasifikasian negara oleh *World Bank*. Jika uji terhadap indeks sektor listrik dan indeks ekonomi menunjukkan tidak adanya korelasi, maka akan menjadi relevan jika uji antara klaster negara berdasarkan indeks ekonomi dan berdasarkan indeks sektor listrik menunjukkan korelasi yang tidak signifikan. Hal ini berlaku untuk sebaliknya. Oleh karena itu, uji korelasi yang dilakukan pada penelitian ini merupakan bagian yang penting.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas pada penelitian sebagai berikut.

1. Bagaimana melakukan *clustering* pada negara-negara dalam indeks sektor listrik?
2. Apakah ada keterkaitan antara indeks sektor listrik dan indeks GNI per kapita pada negara?

³Sumber: blogs.worldbank.org

3. Apakah metode terbaik untuk melakukan *clustering* pada negara-negara dalam indeks sektor listrik?

1.3 Tujuan Penelitian

Tujuan yang akan dicapai dari penelitian ini adalah sebagai berikut.

1. Mengaplikasikan metode *K-means*, GMM, dan SVM untuk dapat digunakan dalam melakukan *clustering* terhadap negara-negara dalam indeks sektor listrik.
2. Menentukan apakah ada keterkaitan antara indeks sektor listrik dan indeks GNI per kapita pada negara.
3. Menentukan metode terbaik untuk melakukan *clustering* pada negara-negara dalam indeks sektor listrik.

1.4 Batasan Masalah dan Asumsi

Berikut merupakan batasan masalah dan asumsi yang akan diterapkan pada penelitian ini.

1. Variabel yang digunakan pada penelitian merupakan variabel-variabel dalam indeks sektor listrik dan sosial ekonomi.
2. Data yang digunakan untuk mengelompokkan negara adalah data tahunan yang terdiri dari 117 negara dan 7 variabel pada sektor listrik.
3. *Clustering* negara akan dilakukan per 2 tahun, yaitu tahun 2010, 2012, dan 2014 sehingga terdapat 3 periode waktu.
4. Model yang digunakan diasumsikan berdistribusi Gaussian.
5. Data yang digunakan diasumsikan valid.
6. Perangkat lunak yang digunakan untuk penelitian ini adalah RStudio.
7. *Package* yang akan digunakan di RStudio adalah *mclust*, *cluster*, *openxlsx*, *e1071*, *dplyr*, dan *readxl*.

1.5 Manfaat Penelitian

Manfaat dari penelitian akan dibagi menjadi dua, yaitu manfaat teoritis dan manfaat praktis.

1.5.1 Manfaat Teoretis

Berikut adalah manfaat-manfaat yang didapat dari penelitian ini secara teoritis.

1. Menentukan apakah metode *clustering* dapat digunakan untuk mengelompokkan data.
2. Dapat mengaplikasikan metode *k-means* untuk melakukan *clustering*.
3. Dapat mengaplikasikan metode GMM untuk melakukan *clustering*.
4. Dapat mengaplikasikan metode SVM untuk melakukan *clustering*.
5. Dapat mengaplikasikan uji korelasi *chi-square* untuk membandingkan hasil klaster oleh *World Bank* dan hasil klaster oleh metode *k-means*, GMM, dan SVM.
6. Dapat mengaplikasikan uji korelasi Spearman untuk mengetahui korelasi antara indeks sektor listrik dan GNI per kapita.

1.5.2 Manfaat Praktis

Berikut adalah manfaat-manfaat yang didapat dari penelitian ini secara praktis.

1. Mengetahui metode yang digunakan untuk mengoptimalkan indeks sektor listrik untuk digunakan dalam *clustering* negara-negara.
2. Membantu negara-negara dalam melakukan penelitian, merencanakan investasi dan kebijakan serta membahas dan mengkomunikasikan kemajuan pada sektor ketenagalistrikan.

1.6 Sistematika Penulisan

Berikut merupakan penguraian dari beberapa poin yang akan dikaji pada bab-bab selanjutnya.

1. **BAB I - PENDAHULUAN**

Pada Bab I, dijelaskan latar belakang yang melandasi pembuatan penelitian serta pemetaan pada masalah dan rumusan masalah berdasarkan latar belakang. Selanjutnya, diuraikan tujuan dari penelitian, batasan masalah yang diterapkan untuk mencapai tujuan penelitian, manfaat dari penelitian ini baik secara teoretis maupun praktis dan sistematika penulisan sebagai gambaran besar skripsi ini.

2. **BAB II - LANDASAN TEORI**

Pada bab II, didefinisikan notasi dan terminologi yang digunakan pada penelitian ini. Berikutnya, diuraikan teori-teori yang digunakan dalam menyelesaikan penelitian ini. Teori yang akan dipaparkan meliputi metode *K-means clustering*, GMM, SVM, uji korelasi *chi-square* dan Spearman.

3. **BAB III - METODOLOGI**

Pada Bab III, dijelaskan mengenai langkah-langkah yang dilakukan dalam penelitian ini untuk mencapai tujuan penelitian. Disajikan sebuah diagram alur yang menjelaskan secara umum mengenai metodologi penelitian. Selain itu, setiap langkah penelitian akan dijelaskan secara terperinci dalam subbab-subbab yang ada dalam bab ini.

4. **BAB IV - ANALISIS DAN PEMBAHASAN**

Pada Bab IV, dibahas hasil *clustering* negara-negara menggunakan metode *K-means*, GMM, dan SVM. Selain itu, dijelaskan hasil analisis kinerja metode *k-means*, GMM, dan SVM, analisis keoptimalan kinerja metode *clustering*, dan dijelaskan analisis keterkaitan antara indeks sektor listrik dan indeks sektor ekonomi.

5. **BAB V - KESIMPULAN DAN SARAN**

Pada Bab V, dipaparkan kesimpulan dari setiap hasil, analisis dan pembahasan dalam penelitian ini. Selain itu, dijelaskan juga mengenai beberapa kelemahan yang terdapat dalam penelitian ini dan rekomendasi yang dapat digunakan sebagai pertimbangan dalam penelitian serupa di masa depan.