

Menganalisis Data Kesehatan Global : Pendekatan Analisis Data Eksplorasi Visual

Muhammad Zulfansyuri Siambaton¹, Amir Mahmud Husein^{2*}

¹Universitas Islam Sumatera Utara, Fakultas Teknik, Teknik Informatika

²Universitas Prima Indonesia, Fakultas Teknologi dan Ilmu Komputer, Teknik Informatika

¹zulfansyuri@ft.uisu.ac.id, ^{2*}amirmahmud@unprimdn.ac.id

Received: 5 Januari 2022

Accepted: 8 Januari 2022

Published: 12 Januari 2022



* Amir Mahmud Husein

Keywords: Data Kesehatan Global, Analisis Statistik, Analisis Data Eksplorasi, EDA.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Peningkatan derajat kesehatan masyarakat merupakan salah satu indikator utama kemajuan masyarakat, dimana analisis statistik untuk pemantauan kesehatan masyarakat sangat relevan untuk sejumlah sektor, seperti penelitian, pembuatan kebijakan, perawatan kesehatan, industri farmasi, asuransi dan lainnya. Makalah ini menyajikan analisis data kesehatan global dengan dengan tujuan menemukan variabel penting yang berpengaruh terhadap harapan hidup. Dari hasil observasi dengan menggunakan beberapa tahapan proses mulai dari pembersihan data, deteksi data nilai yang hilang, mendeteksi outlier ditemukan bahwa terdapat tujuh variabel penting yaitu *status, Adult Mortality, percentage expenditure, HIV, GDP, schooling, Income composition of resources*. Dalam penelitian ini juga disajikan upaya untuk menyusun dan menganalisis semua pertanyaan penting dalam bentuk Analisis Data Eksplorasi visualisasi secara keseluruhan dari kumpulan data.

PENDAHULUAN

Peningkatan derajat kesehatan masyarakat merupakan salah satu indikator utama kemajuan masyarakat. Data statistik untuk pemantauan kesehatan masyarakat sangat relevan untuk sejumlah sektor, seperti penelitian, pembuatan kebijakan, perawatan kesehatan, industri farmasi, asuransi dan lainnya [1]. Sementara data tersebut tersedia bahkan dalam skala global, misalnya di Observatorium Kesehatan Global (GHO) dari Organisasi Kesehatan Dunia (WHO) Perserikatan Bangsa-Bangsa. GHO terdiri lebih dari 50 kumpulan data yang berbeda, mencakup semua 198 negara anggota WHO dan diperbarui ketika data yang lebih baru atau direvisi tersedia atau ketika ada perubahan pada metodologi yang digunakan.

Meskipun telah banyak penelitian yang dilakukan di masa lalu tentang faktor-faktor yang mempengaruhi harapan hidup dengan mempertimbangkan variabel demografis, komposisi pendapatan dan tingkat kematian [2]. Akan tetapi, belum banyak penelitian yang dilakukan pada faktor-faktor seperti imunisasi, yang dapat menjadi sangat penting karena ada negara-negara di mana kesadaran untuk imunisasi kurang yang mungkin menyebabkan harapan hidup yang lebih rendah. Ditemukan bahwa pengaruh imunisasi dan indeks pembangunan manusia tidak diperhitungkan di masa lalu [3]. Juga, beberapa penelitian sebelumnya dilakukan dengan mempertimbangkan *regresi linier* [4], clustering [5], Bayesian [6] berdasarkan kumpulan data yang berbeda dan penerapan di berbagai negara berbeda. Oleh karena itu, hal ini memberikan motivasi untuk menyelesaikan dalam menentukan faktor-faktor penting yang dapat mempengaruhi hidup manusia dengan menggunakan Pendekatan Analisis Data Eksplorasi Visual dengan mempertimbangkan data dari periode 2000 hingga 2015 untuk semua negara. Imunisasi penting seperti Hepatitis B, Polio dan Difteri juga akan dipertimbangkan. Singkatnya, penelitian ini akan fokus pada faktor imunisasi, faktor kematian, faktor ekonomi, faktor sosial dan juga faktor kesehatan lainnya. Karena pengamatan dataset ini didasarkan pada

negara yang berbeda, akan lebih mudah bagi suatu negara untuk menentukan faktor prediksi yang berkontribusi terhadap nilai harapan hidup yang lebih rendah. Terakhir, hasil analisis ini akan membantu dalam menyarankan suatu negara daerah mana yang harus diberikan kepentingan untuk meningkatkan harapan hidup penduduknya secara efisien.

TINJAUAN LITERATUR

Exploratory Data Analysis (EDA) diperkenalkan oleh Tukey tahun 1961. EDA adalah proses pemeriksaan atau pemahaman data dan penggalan wawasan atau karakteristik utama dari data. EDA umumnya diklasifikasikan menjadi dua metode, yaitu analisis grafis dan analisis non-grafis [7]. EDA sangat penting karena merupakan praktik yang baik untuk terlebih dahulu memahami pernyataan masalah dan berbagai hubungan antara fitur data sebelum menerapkannya pada model. EDA adalah bagian dari proses data science. EDA menjadi sangat penting sebelum melakukan *feature engineering* dan modeling karena dalam tahap ini harus memahami datanya terlebih dahulu. EDA memungkinkan analyst memahami isi data yang digunakan, mulai dari distribusi, frekuensi, korelasi dan lainnya. Dalam prakteknya, curiosity sangat penting dalam proses ini, pemahaman konteks data juga diperhatikan, karena akan menjawab masalah masalah dasar [8]. Pada umumnya EDA dilakukan dengan beberapa cara:

1. Univariat Analysis merupakan analisis deskriptif dengan satu variabel.
2. Bivariat Analysis merupakan analisis relasi dengan dua variabel yang biasanya dengan target variabel.
3. Multivariat Analysis merupakan analisis yang menggunakan lebih dari atau sama dengan tiga variabel.

EDA adalah langkah penting sebelum memulai pembelajaran mesin atau pemodelan statistik untuk mengembangkan model yang sesuai dengan masalah yang dihadapi dan menyediakan konteks yang diperlukan untuk menginterpretasikan hasil dengan benar. EDA sangat penting untuk memastikan bahwa hasil yang dihasilkan oleh ilmuwan data valid, ditafsirkan dengan benar, dan dapat diterapkan pada konteks bisnis yang diinginkan [7].

BAHAN DAN METODE

Bahan

Kumpulan data yang terkait dengan penelitian ini disediakan oleh penyimpanan data *Global Health Observatory* (GHO) di bawah *World Health Organization* (WHO). Sebagian data ditunjukkan pada tabel 1.

Tabel 1 Sebagian dari Kumpulan Data

	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	...	Polio	Total expenditure	Diphtheria	HIV/AIDS
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279824	65.0	1154	...	6.0	8.16	65.0	0.1 584.
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	8.16	62.0	0.1 612.
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	8.13	64.0	0.1 631.
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0	2787	...	67.0	8.52	67.0	0.1 669.
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	7.87	68.0	0.1 63.

5 rows × 22 columns

Di antara semua kategori faktor yang berhubungan dengan kesehatan hanya faktor kritis yang dipilih yang lebih representatif. Dalam beberapa tahun terakhir ini, telah terjadi perkembangan besar di bidang kesehatan yang mengakibatkan peningkatan angka kematian manusia terutama di negara-negara berkembang dibandingkan dengan 30 tahun terakhir. Oleh karena itu, dalam penelitian ini kami telah mempertimbangkan data dari tahun 2000-2015 untuk 193 negara untuk analisis lebih lanjut. File data individual telah digabungkan menjadi satu kumpulan data, sebagian besar data yang hilang adalah untuk populasi, Hepatitis B dan PDB. Data yang hilang berasal dari negara-negara yang kurang dikenal seperti Vanuatu, Tonga, Togo,

Cabo Verde dan lainnya. Menemukan semua data untuk negara-negara ini sulit dan karenanya, diputuskan bahwa kami mengecualikan negara-negara ini dari kumpulan data model akhir. File gabungan terakhir (dataset akhir) terdiri dari 22 Kolom dan 2938 baris yang berarti 20 variabel prediksi. Semua variabel prediksi kemudian dibagi menjadi beberapa kategori besar: faktor terkait imunisasi, faktor kematian, faktor ekonomi dan faktor sosial.

Dataset ini berisi variabel seperti *country*, *year*, *status*, *life expectancy*, dan lainnya. Kolom *Adult Mortality* mewakili tingkat kematian orang dewasa dari kedua jenis kelamin, yang merupakan probabilitas kematian antara 15 dan 60 tahun per 1000 penduduk. Kolom *Infant deaths* menampilkan jumlah kematian bayi per 1000 penduduk. Kolom *Alcohol* menggambarkan total liter konsumsi alkohol murni, yang dicatat per kapita untuk usia 15 tahun ke atas. Variabel Hepatitis.B, Polio, dan Difteri menunjukkan persentase cakupan imunisasi anak usia 1 tahun untuk masing-masing penyakit tersebut. Kolom *Measles* menunjukkan jumlah kasus campak yang dilaporkan per 1000 penduduk. *kurus..1.19.tahun`* mewakili persentase ketipisan yang ada pada anak-anak mulai dari usia 10 hingga 19 tahun. Harapan hidup berkisar antara 44 hingga 89. Ini berarti bahwa orang-orang di negara dengan harapan hidup tertinggi hidup hampir dua kali lipat dari orang-orang di negara dengan harapan hidup terendah.

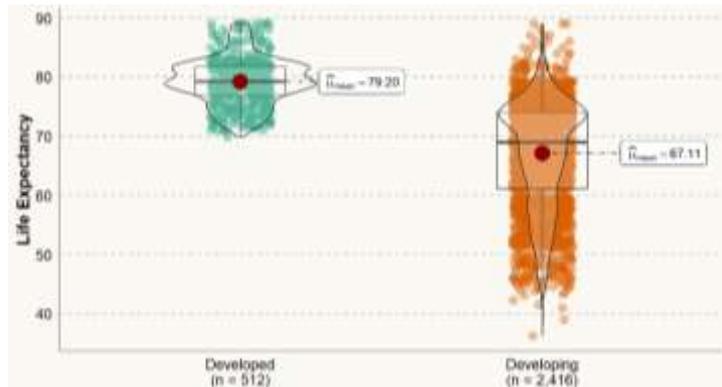
Metode

Analisis bisnis dalam penelitian ini adalah bertujuan untuk menentukan variabel yang paling signifikan untuk membuat model regresi linier yang optimal untuk memprediksi harapan hidup. Untuk mencapai hal ini, metode yang digunakan terdiri dari beberapa fase, yaitu:

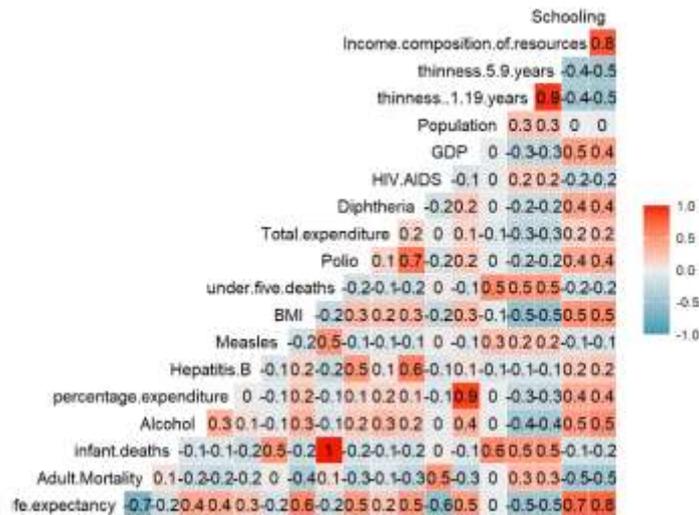
1. Pembersihan Data
2. Deteksi Data yang hilang
3. Deteksi data outlier
4. Data Exploration and Visualization
 - a. Berapa Harapan Hidup Berdasarkan Negara? Apa 10 negara teratas dengan populasi rata-rata tertinggi?
 - b. Bagaimana hubungan harapan hidup dengan kematian bayi ?
 - c. Bagaimana penyakit yang berbeda mempengaruhi harapan hidup di negara maju dan berkembang?
 - d. Bagaimana distribusi sekolah untuk negara maju dan berkembang dan Apa pengaruh Sekolah dan Alkohol terhadap Harapan Hidup?
5. Kesimpulan

HASIL PENELITIAN

Pada penelitian ini bertujuan untuk memahami dan menentukan fitur penting dalam dengan menggunakan bahasa pemrograman python 3 dengan jupyter notebook. Beberapa pustakan juga digunakan untuk kebutuhan manipulasi data, pengolahan data dan visualisasi data seperti numpy, pandas, matplotlib, seaborn dan lainnya. Untuk mendapatkan pemahaman tentang bagaimana status pembangunan suatu negara berkorelasi dengan harapan hidup, pada gambar 1 disajikan plot kotak visualisasi korelasi antara status pembangunan suatu negara dan harapan hidup.



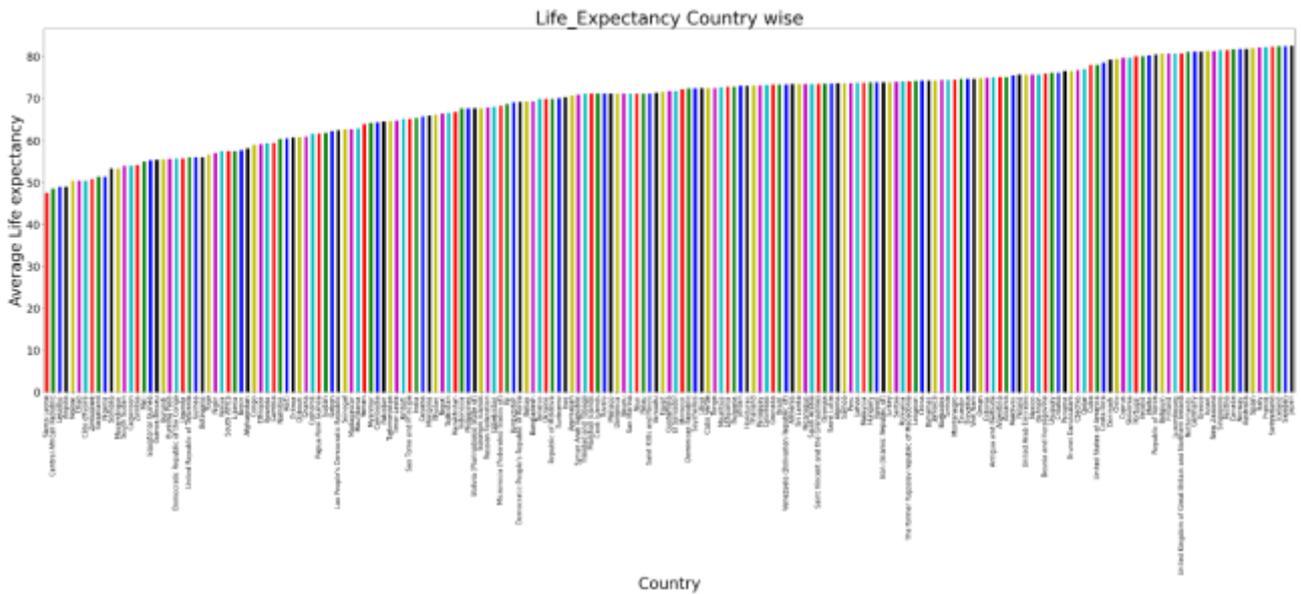
Gambar 1 Plot kotak korelasi antara status pembangunan suatu negara dan harapan hidup. Pada gambar 1 dapat dilihat bahwa ada perbedaan mencolok dalam harapan hidup di negara berkembang dan negara maju. Pertama-tama, rentang harapan hidup antarkuartil lebih rendah di negara maju daripada negara berkembang. Rata-rata dan median keduanya jauh lebih tinggi di negara maju. Harapan hidup minimum juga jauh lebih rendah di negara berkembang. Dengan demikian, status negara dapat mempengaruhi harapan hidup.



Gambar 2 Korelasi antara variabel

Analisis korelasi antar variabel bertujuan untuk lebih memahami keterhubungan antara semua variabel. Dari gambar 2 dapat dilihat hasil visualisasi korelasi semua variabel dimana variabel yang berhubungan dengan harapan hidup paling positif adalah: sekolah, dan komposisi pendapatan sumber daya. Variabel yang paling negatif terkait dengan harapan hidup adalah: kematian orang dewasa dan HIV/AIDS. Harapan hidup memiliki hubungan yang dapat diabaikan dengan Populasi, Total pengeluaran, kematian balita dan Campak. Dari pengamatan di atas, maka tahapan selanjutnya menjawab beberapa pertanyaan yang diajukan pada bagian sebelumnya.

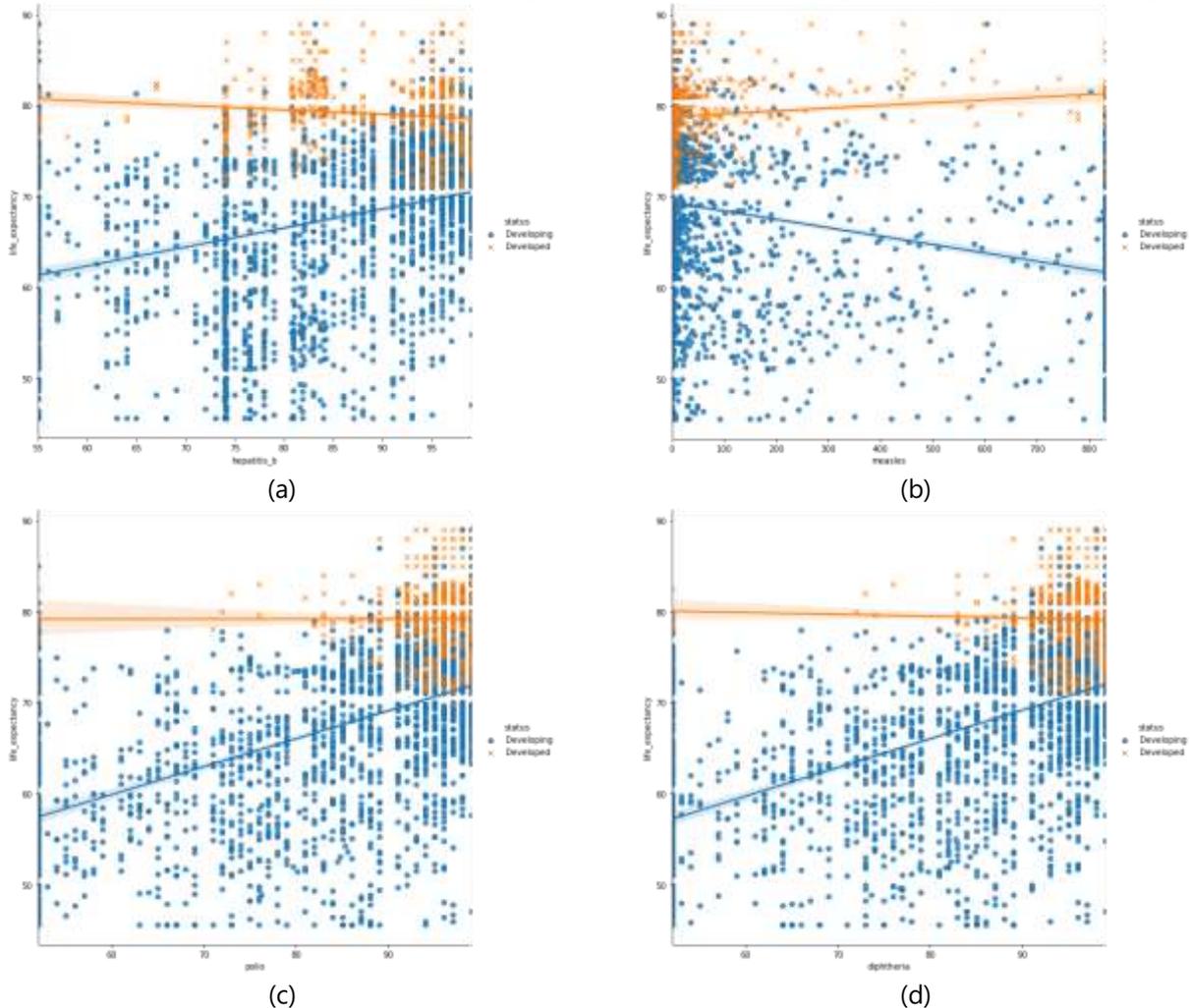
1. Berapa Harapan Hidup Berdasarkan Negara? Apa 10 negara teratas dengan populasi rata-rata tertinggi?

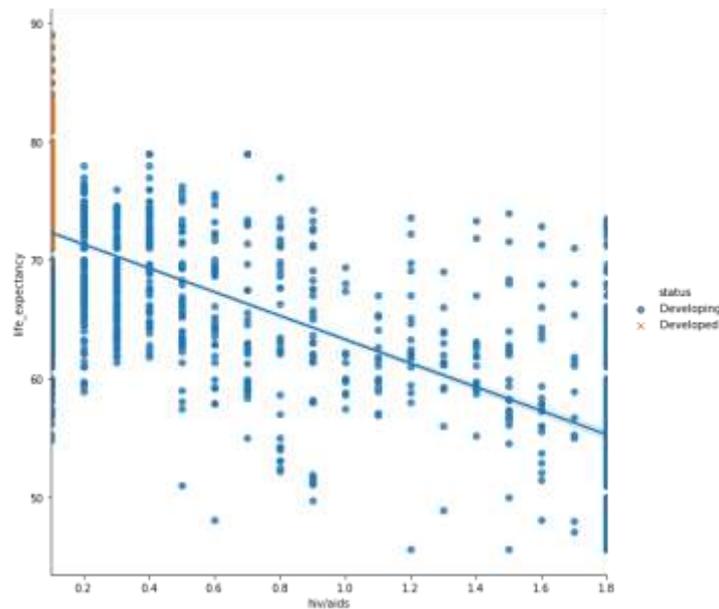


Gambar 3 Grafik Harapan Hidup Berdasarkan Negara

Jepang adalah negara dengan nilai Harapan Hidup tertinggi diikuti oleh Swedia dan Sierra Leone memiliki nilai Harapan Hidup terendah.

2. Bagaimana penyakit yang berbeda mempengaruhi harapan hidup di negara maju dan berkembang?

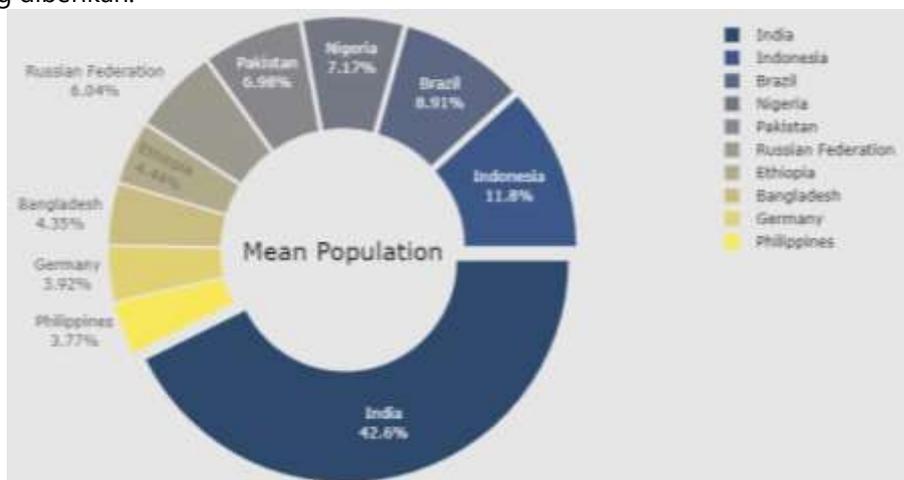




(e)

Gambar 4 Harapan hidup berdasarkan negara

Pada gambar 4 dapat dilihat pada bagian (a) terdapat sedikit penurunan nilai harapan hidup di negara maju sedangkan di negara maju nilai harapan hidup meningkat secara bertahap yang berarti bahwa negara berkembang mengambil langkah-langkah untuk menyiapkan vaksin hepatitis B. Dalam kasus Campak, menurut grafik pada gambar 3(b), negara maju tampaknya memiliki vaksin yang tersedia untuk mengatasi campak sedangkan nilai harapan hidup negara berkembang menurun dari hari ke hari mungkin karena kurangnya sumber daya untuk menangani campak. Negara-negara maju berhasil memberantas penyakit polio dan penyakit difteri seperti terlihat pada gambar 4 (c) dan (d), karena vaksin sedangkan di negara-negara berkembang nilai harapan awalnya rendah tetapi sekarang meningkat secara bertahap mungkin karena pemberian dosis yang tepat. Namun, pada gambar 4 (e) Grafik tersebut menunjukkan bahwa negara berkembang masih belum mampu menangani hiv/aids sama sekali karena nilai harapan hidup yang terus menurun dengan kisaran yang cepat. Hal ini dapat disebabkan oleh meningkatnya populasi dan tidak ada pendidikan yang diberikan.



Gambar 5 Grafik 10 negara tertinggi

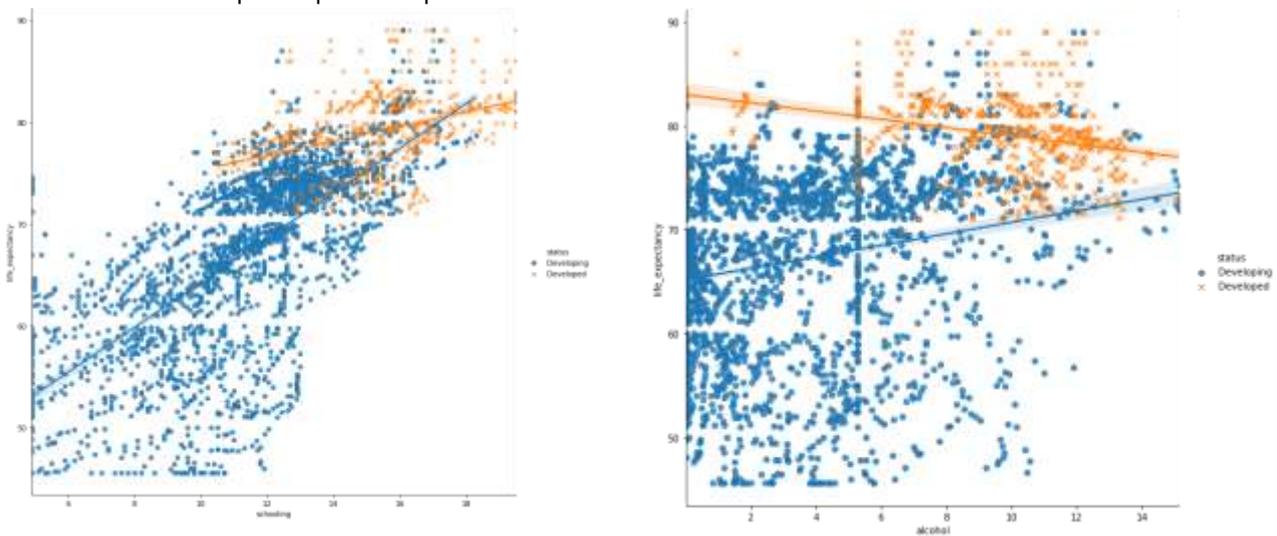
Pada gambar 5 merupakan visualisasi dalam bentuk pie (lingkaran) yang menjelaskan 10 negara tertinggi untuk populasi rata-rata tertinggi, terlihat bahwa negara India menempati posisi pertama, kemudian Indonesia dan yang terakhir negara Philippines. Dalam hal ini, dapat disimpulkan bahwa negara-negara di

benua Asia merupakan negara yang tingkat harapan hidupnya tertinggi di dibandingkan benua lainnya.
 3. Bagaimana hubungan harapan hidup dengan kematian bayi ?



Gambar 6 Visualisasi hubungan harapan hidup dengan kematian bayi

4. Bagaimana distribusi sekolah untuk negara maju dan berkembang dan Apa pengaruh Sekolah dan Alkohol terhadap Harapan Hidup?



Gambar 7 Visualiasi distribusi sekolah untuk negara maju dan berkembang dan pengaruh alkohol

Seperti yang terlihat pada gambar 7 bahwa sekolah dapat mempengaruhi harapan hidup lebih banyak di negara berkembang daripada negara maju. Hal ini mungkin karena pendidikan lebih mapan dan lazim di negara-negara kaya. Dapat diduga karena fakta bahwa hanya negara-negara kaya yang mampu membeli alkohol atau konsumsi alkohol lebih umum di antara populasi yang lebih kaya. Itulah sebabnya negara berkembang dan alkohol memiliki hubungan positif dan negara maju dan alkohol memiliki hubungan negatif.

Dari hasil analisis ini, sebagian besar melihat bagaimana negara-negara berkembang memiliki harapan hidup yang sangat sedikit ketika kami melihat penyakit seperti HIV/AIDS, polio dll dan bagaimana Sekolah memainkan peran besar dalam meningkatkan harapan hidup negara-negara berkembang karena orang menjadi jauh lebih terdidik dan membantu meningkatkan kesejahteraan dan perawatan kesehatan negara bersama dengan ekonomi Alkoholisme adalah masalah besar di negara maju di mana orang memiliki banyak uang untuk dibelanjakan dan ini menunjukkan betapa cerobohnya orang dalam hal kesehatan mereka dalam

hal alkoholisme. Terakhir variabel yang mempengaruhi harapan hidup adalah variabel status, *Adult Mortality, percentage expenditure, HIV, GDP, schooling, Income composition of resources*.

Diskusi

Beberapa hal penting menjadi pertimbangan di masa depan dari hasil sebelumnya, meskipun dataset ini dikumpulkan oleh WHO mengandung banyak nilai yang hilang dan terlihat bahwa sebagian besar nilai yang hilang berasal dari negara-negara dengan populasi yang sangat sedikit dan pengumpulan data adalah tugas yang sangat membosankan. Banyak outlier terdeteksi yang ditangani oleh Winsorization Jepang meskipun dihantam parah oleh perang dunia II kembali sangat kuat dan saat ini menjadi negara dengan angka harapan hidup tertinggi disusul oleh Swedia yang merupakan Prestasi besar.

Data WHO tentang harapan hidup dianalisis dengan menggunakan Explorasi Data Analysis sebagian dari proses tersebut melibatkan pemeriksaan untuk memastikan bahwa data tersebut benar-benar dapat dianalisis. Dan, variabel prediksi yang sebenarnya mempengaruhi harapan hidup adalah status, Kematian Orang Dewasa, persentase pengeluaran, HIV, PDB, sekolah, Komposisi pendapatan sumber daya, tingkat kematian Bayi tidak memiliki banyak hubungan dengan kematian bayi dan Dewasa mempengaruhi harapan hidup dan sekolah memiliki dampak terhadap umur manusia. Terakhir penyakit HIV masih merupakan salah satu penyakit yang memiliki tingkat kematian tertinggi di hampir semua negara.

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis Data Kesehatan Global dengan pendekatan Explorasi Data Analysis serta menemukan variabel penting yang dapat mempengaruhi harapan hidup manusia. Dari hasil analisis dengan beberapa tahapan yang dilakukan, maka data yang tersedia masih sangat perlu dilakukan proses pembersihan data, mendeteksi nilai yang hilang, deteksi outlier walaupun kumpulan data ini bersumber dari WHO, selanjutnya hasil observasi dari kumpulan data ini menyimpulkan bahwa cakupan vaksinasi Polio, Hepatitis, Difteri harus ditingkatkan Langkah-langkah harus diambil untuk memastikan keamanan pangan Langkah-langkah harus diambil untuk memberikan pendidikan dan mengurangi risiko kematian bayi. Sumber daya harus dimanfaatkan secara produktif, kampanye kesadaran AIDS harus diselenggarakan. Negara-negara maju harus membantu negara-negara berkembang dalam memberantas penyakit yang mempengaruhi kehidupan masyarakat dengan memberikan vaksinasi Pemerintah harus lebih fokus pada sekolah anak-anak yang akan menjadi wajah negara di masa depan dan memberi mereka makanan yang baik dan mendidik mereka dengan baik. Pemerintah negara berkembang harus meluncurkan berbagai skema untuk memotivasi masyarakat menyekolahkan anaknya Pemerintah harus menyelenggarakan kamp-kamp kesehatan gratis untuk memberikan vaksinasi gratis bagi orang-orang yang membutuhkan sehingga tetap sehat untuk memperlakukan keluarga mereka dengan baik tanpa mengeluarkan uang. Terakhir, WHO sebaiknya memberikan bantuan bersama negara-negara untuk membantu pemerintah negara-negara berkembang dalam menyediakan makanan dan pendidikan gratis dan menyelenggarakan kamp-kamp perawatan kesehatan

Supplementary Materials (optional)

Tidak tersedia.

Kontribusi Penulis

AM: Konseptualisasi,; metodologi,; perangkat lunak, validasi; analisis formal, visualisasi,; pengawasan; ZS: penyidikan, sumber daya, akurasi data, tulisan—persiapan draf asli; menulis—ulasan dan penyuntingan. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Konflik kepentingan

Para penulis menyatakan tidak ada konflik kepentingan.

REFERENCES

- [1] M. Luy, P. Di Giulio, V. Di Lego, P. Lazarevič, and M. Sauerberg, "Life Expectancy: Frequently Used, but Hardly Understood," *Gerontology*, vol. 66, no. 1, pp. 95–104, 2020, doi: 10.1159/000500955.
- [2] A. M. Duque *et al.*, "Analysis of the relationship between life expectancy and social determinants in a North-Eastern region of Brazil, 2010-2017," *Geospat. Health*, vol. 13, no. 2, pp. 345–352, 2018, doi: 10.4081/gh.2018.702.
- [3] J. M. Aburto, F. Villavicencio, U. Basellini, S. Kjærgaard, and J. W. Vaupel, "Dynamics of life expectancy and life span equality," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 10, pp. 5250–5259, 2020, doi: 10.1073/pnas.1915884117.
- [4] G. I. Castillo Tumaille, A. M. Guerra Tejada, and E. M. de L. Loaiza Massuh, "Life expectancy analysis from a multiple regression model," *Univ. Cienc. y Tecnol.*, vol. 25, no. 110, pp. 198–207, 2021, doi: 10.47460/uct.v25i110.492.
- [5] S. A. Paramita, C. Yamazaki, and H. Koyama, "Determinants of life expectancy and clustering of provinces to improve life expectancy: An ecological study in Indonesia," *BMC Public Health*, vol. 20, no. 1, pp. 1–8, 2020, doi: 10.1186/s12889-020-8408-3.
- [6] H. Al Azies and V. M. Dewi, "Factors Affecting Life Expectancy in East Java: Predictions with A Bayesian Model Averaging Approach," *J. Perenc. Pembang. Indones. J. Dev. Plan.*, vol. 5, no. 2, pp. 283–295, 2021, doi: 10.36574/jpp.v5i2.214.
- [7] M. Restori, "What is Exploratory Data Analysis | Tutorial by Chartio," 2019. <https://chartio.com/learn/data-analytics/what-is-exploratory-data-analysis/> (accessed Dec. 27, 2021).
- [8] A. Chandra, "Memahami Data Dengan Exploratory Data Analysis | by Andreas Chandra | Data Folks Indonesia | Medium," *Medium*, 2019. <https://medium.com/data-folks-indonesia/memahami-data-dengan-exploratory-data-analysis-a53b230cce84> (accessed Dec. 27, 2021).