



Analisis Faktor Tingkat Kebahagiaan Negara Menggunakan Data *World Happiness Report* dengan Metode Regresi Linier

Khoirunnisa Hamidah¹, Apriade Voutama²

¹ Prodi Sistem Informasi, Universitas Singaperbangsa Karawang, Karawang, Indonesia

² Prodi Sistem Informasi, Universitas Singaperbangsa Karawang, Karawang, Indonesia

email: ¹2010631250091@student.unsika.ac.id, ²apriade.voutama@staff.unsika.ac.id

INFO ARTIKEL

Sejarah Artikel:

Diterima 29 Mei 2023
Direvisi -
Disetujui 17 Juni 2023
Dipublikasi 18 Juni 2023

Katakunci:

Faktor Kebahagiaan
KDD
Regresi Linear

ABSTRAK

Kebahagiaan merupakan faktor yang penting dalam kehidupan. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi tingkat kebahagiaan di suatu negara dengan menggunakan data dari *World Happiness Report 2022*. Dalam upaya tersebut, data dikumpulkan melalui situs web Kaggle dengan menggunakan dataset *World Happiness Report 2022* yang tersedia secara public. Dataset ini berisi informasi tentang berbagai faktor yang mempengaruhi tingkat kebahagiaan seperti pendapatan per kapita, kesehatan, dukungan sosial, kebebasan, dan korupsi. Metode yang digunakan pada penelitian ini adalah metode KDD yaitu *Knowledge Discovery in Database* pada metode ini terdapat pemilihan data, pra pemrosesan, transformasi, data mining dan evaluasi. Hasil analisis menunjukkan bahwa variabel independen seperti pendapatan per kapita, kesehatan, dukungan sosial, kebebasan, dan persepsi korupsi berpengaruh signifikan terhadap tingkat kebahagiaan di suatu negara. Temuan ini dapat memberikan informasi penting bagi pemerintah dan organisasi untuk meningkatkan tingkat kebahagiaan masyarakat di suatu negara melalui kebijakan-kebijakan yang tepat. Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam pemahaman tentang faktor-faktor yang mempengaruhi tingkat kebahagiaan di suatu negara. Hasil penelitian ini dapat dijadikan acuan bagi pemerintah dan organisasi untuk mengembangkan program dan kebijakan yang mampu meningkatkan kesejahteraan dan kebahagiaan masyarakat di suatu negara.

ABSTRACT

Happiness is an important factor in life. This study aims to analyze the factors that influence the level of happiness in a country using data from the *World Happiness Report 2022*. In this effort, data was collected through the Kaggle website using the publicly available *World Happiness Report 2022* dataset. This dataset contains information about various factors that affect the level of happiness such as per capita income, health, social support, freedom, and corruption. The method used in this study is the KDD method, namely *Knowledge Discovery in Database*. In this method, there are data selection, pre-processing, transformation, data mining and evaluation. The results of the analysis show that independent variables such as per capita income, health, social support, freedom, and perceptions of corruption have a significant effect on the level of happiness in a country. These findings can provide important information for governments and organizations to increase the level of happiness of people in a country through appropriate policies. Overall, this research makes an important contribution to the understanding of the factors that influence the level of happiness in a country. The results of this study can be used as a reference for governments and organizations to develop programs and policies that can improve the welfare and happiness of people in a country.

©2023 diterbitkan oleh Prodi Teknik Informatika Universitas Yudharta Pasuruan

Keyword:

The Happiness Factor
KDD
Linear Regression

DOI Artikel:

10.35891/explorit.v15i1.3874

1. Pendahuluan

Kebahagiaan telah menjadi topik penelitian yang semakin populer di seluruh dunia. *World Happiness Report* adalah sebuah inisiatif yang diterbitkan oleh PBB yang telah dikaji oleh para ahli dan pakar terkemuka di bidang psikologi, ekonomi, dan statistik, yang bertujuan untuk mengukur tingkat kebahagiaan di negara-negara di seluruh dunia. Laporan ini memperkenalkan konsep kebahagiaan sebagai indikator kesejahteraan yang penting bagi suatu negara[1]. Kebahagiaan bukan hanya tentang pemenuhan kebutuhan dasar manusia, tetapi juga tentang kebebasan, keamanan, dan kesetaraan[2]. Dalam konteks ini, penelitian tentang faktor-faktor yang mempengaruhi tingkat kebahagiaan di suatu negara menjadi sangat penting untuk memahami kesejahteraan masyarakat di negara-negara tersebut. Tujuan kebahagiaan seseorang bukan hanya kepentingan pribadinya saja, tetapi kesadaran bahwa seseorang dapat menciptakan kebahagiaan bagi sekitarnya[3]. Oleh karena itu, penting bagi pemerintah untuk menciptakan kebahagiaan yang lebih luas bagi masyarakatnya.

Pada penelitian-penelitian sebelumnya, pendekatan yang umum digunakan dalam menentukan faktor-faktor yang mempengaruhi tingkat kebahagiaan adalah melalui pendekatan statistika. Statistik sering digunakan untuk menganalisis hubungan antara variabel-variabel yang relevan dengan kebahagiaan. Namun, ketika dihadapkan dengan data yang sangat banyak, pengolahan data menjadi lebih rumit dan kompleks.

Dalam analisis faktor kebahagiaan di seluruh dunia, penelitian sering dilakukan secara tahunan untuk melihat perubahan dan tren dalam tingkat kebahagiaan[1]. Oleh karena itu, penggunaan metode yang lebih canggih dan modern seperti *Knowledge Discovery in Databases* (KDD) dengan menggunakan algoritma regresi linear dapat membantu dalam menganalisis faktor-faktor yang mempengaruhi kebahagiaan di suatu negara.

Metode KDD memungkinkan peneliti untuk mengeksplorasi data yang besar dan kompleks[4], termasuk data dari laporan *World Happiness Report*, dengan menggunakan algoritma regresi linear untuk mengidentifikasi faktor-faktor yang signifikan dalam mempengaruhi tingkat kebahagiaan.

Pendekatan ini dapat membantu mengungkap pola dan hubungan yang lebih kompleks antara variabel-variabel yang berpotensi memengaruhi kebahagiaan, serta memberikan pemahaman yang lebih mendalam dalam konteks waktu.

Dengan menggunakan metode KDD dan algoritma regresi linear, penelitian dapat memberikan kontribusi yang lebih kuat dalam menganalisis faktor-faktor yang mempengaruhi kebahagiaan di negara-negara, memberikan wawasan yang lebih kaya, dan menghasilkan temuan yang lebih akurat untuk kebijakan dan intervensi yang bertujuan meningkatkan tingkat kebahagiaan masyarakat.

2. Kajian Teori

2.1 Penelitian Terdahulu

Penelitian sebelumnya terkait faktor kebahagiaan menunjukkan bahwa ada hubungan signifikan antara tiga faktor, yaitu jumlah penduduk, tingkat bebas korupsi, dan migrasi netto, dengan tingkat kebahagiaan suatu negara di dunia. Penelitian tersebut menyimpulkan bahwa ketika ketiga faktor tersebut dianalisis secara bersamaan, mereka memiliki pengaruh yang signifikan terhadap tingkat kebahagiaan negara. Hal ini menunjukkan bahwa populasi yang lebih besar, tingkat bebas korupsi yang tinggi, dan migrasi netto yang stabil dapat berkontribusi pada tingkat kebahagiaan yang lebih tinggi di suatu negara[5]. Pada tahun 2021 Sustainable Development Solutions Network mengevaluasi tingkat kebahagiaan di berbagai negara di seluruh dunia. Laporan ini memperhitungkan faktor-faktor seperti pendapatan, kesehatan, dukungan sosial, kebebasan, dan korupsi dalam mengukur tingkat kebahagiaan suatu negara. Salah satu tujuan dari laporan ini adalah untuk memberikan panduan bagi pemerintah dan masyarakat dalam meningkatkan kesejahteraan dan kebahagiaan. Pada masa itu juga bertepatan dengan adanya pandemic covid-19 pada penelitian tersebut tingkat kebahagiaannya didapatkan terjadi penurunan tingkat kebahagiaan. Hal ini dikaitkan dengan adanya ketidakpastian dan kecemasan terhadap kondisi kesehatan dan ekonomi, serta pengaruh dari kebijakan-kebijakan yang diambil oleh pemerintah[6][7].

2.2 Kebahagiaan

Kebahagiaan adalah impian setiap manusia. kebahagiaan merujuk pada perasaan kesenangan dan ketentraman dalam kehidupan, keberuntungan dan kemujuran, baik secara fisik maupun emosional. Kebahagiaan juga dapat diartikan sebagai hidup dalam damai, harmoni, ketenangan, sejahtera, mengutamakan kebaikan, dan hidup sesuai dengan prinsip-prinsip kebenaran[5].

2.3 Data Mining

Data mining adalah pendekatan yang menggunakan berbagai metode dan teknik untuk mengekstraksi wawasan berharga dan mengidentifikasi pola serta hubungan dari dataset yang besar[8]. Data mining melibatkan penggunaan algoritma statistik, Machine Learning, serta kecerdasan buatan untuk mengungkap informasi yang tersembunyi dalam data. Dalam proses data mining, ditemukan pola-pola yang dapat digunakan untuk membangun model yang mampu memprediksi perilaku atau mengidentifikasi karakteristik umum dari objek yang sedang diteliti[9]. Metodologi data mining memberikan kerangka kerja yang terstruktur untuk menjalankan langkah-langkah dalam proyek data mining dan memastikan pencapaian tujuan yang diinginkan. Data mining memiliki peran penting dalam pengambilan keputusan berbasis bukti dengan mengungkap fakta dan bukti dari data. Dengan kemampuan untuk mengekstraksi informasi berharga dari dataset yang besar, data mining memberikan wawasan yang berharga yang dapat mendukung pengembangan kebijakan baru, perencanaan aktivitas, dan merangsang kreativitas dalam bisnis[4].

2.4 KDD

KDD merupakan suatu kegiatan interdisiplin yang bertujuan untuk memperoleh pengetahuan yang berharga melalui data. Hal ini melibatkan penerapan metodologi[10]. KDD (Knowledge Discovery in Database) merupakan suatu pendekatan atau metode yang penting dalam mengungkapkan pengetahuan baru yang terkubur dalam data mentah. KDD merupakan proses untuk menemukan pola yang berharga atau pengetahuan baru dari data. KDD melibatkan serangkaian langkah yang dirancang untuk mengubah data kasar menjadi informasi yang lebih berguna dan bermakna[11]. Melalui tahapan-tahapan seperti seleksi, pra-pemrosesan, transformasi, data mining, dan evaluasi[12]. KDD memungkinkan kita untuk mengidentifikasi wawasan yang berharga dari data yang ada. KDD memberikan kerangka kerja yang sistematis dan terstruktur untuk menggali pengetahuan baru dari data dengan memanfaatkan teknik-teknik seperti kecerdasan buatan, algoritma, statistik, dan matematika. Dengan menggunakan metode KDD, kita dapat mengoptimalkan proses pengolahan dan analisis data, meningkatkan pemahaman tentang informasi yang ada, serta mendukung pengambilan keputusan yang lebih baik. KDD membuka peluang besar untuk memanfaatkan data secara efektif dan menghasilkan pengetahuan yang berharga. Dengan penerapan yang tepat, KDD dapat menjadi alat yang kuat dalam mendukung pengambilan keputusan, inovasi, dan pengembangan strategi yang lebih baik di berbagai bidang[13].

2.5 Linear Regresi

Regresi linier berganda adalah metode pengumpulan data yang digunakan untuk menganalisis hubungan antara variabel dependen dengan beberapa variabel independen[14]. Dalam analisis regresi linear, koefisien, konstanta, dan koefisien determinan memiliki peran penting. Koefisien menggambarkan pengaruh relatif dari setiap variabel independen terhadap variabel dependen dalam model regresi. Konstanta merupakan nilai tetap yang menunjukkan tingkat kebahagiaan yang diharapkan ketika semua variabel independen memiliki nilai nol. Sedangkan koefisien determinan mengindikasikan seberapa baik model dapat menjelaskan variasi dalam variabel dependen[15]. ketiga konsep ini dapat memperoleh pemahaman yang lebih baik tentang faktor-faktor yang mempengaruhi kebahagiaan suatu negara berdasarkan model regresi linear.

Dalam analisis regresi linier, kita dapat menggunakan beberapa metrik evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Square Error (RMSE) untuk mengevaluasi kualitas model regresi linear.

Mean Absolute Error (MAE) digunakan untuk menguji keakuratan hasil prediksi. Semakin kecil nilai MAE, semakin baik kualitas model regresi linear, karena hal ini menunjukkan bahwa selisih antara nilai prediksi dan nilai sebenarnya lebih kecil[16]. Persamaan dari MAE:

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

Mean Squared Error (MSE) adalah rata-rata dari perbedaan kuadrat antara nilai yang diamati (nilai sebenarnya) dan nilai yang diprediksi. Nilai MSE yang lebih kecil menunjukkan tingkat kesalahan yang lebih kecil dan kualitas prediksi yang lebih baik[15]. Persamaan dari MSE:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

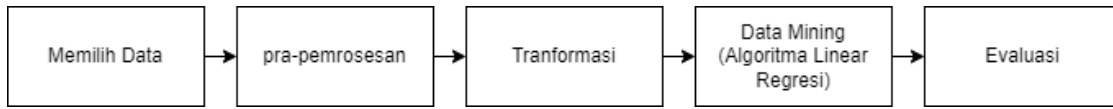
Root Mean Square Error (RMSE) merupakan metode pengukuran kesalahan pada prediksi yang menghitung akar kuadrat dari nilai rata-rata selisih antara nilai aktual dan nilai prediksi[15]. RMSE memberikan informasi yang lebih baik daripada MAE karena sensitif terhadap perbedaan nilai yang lebih besar antara nilai aktual dan nilai prediksi. Namun, perlu diingat bahwa RMSE juga dapat dipengaruhi oleh nilai ekstrim dalam data, sehingga penanganan nilai ekstrim perlu dipertimbangkan dengan hati-hati. Persamaan RMSE:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

Dalam evaluasi model regresi linear, penting untuk mempertimbangkan dan membandingkan nilai MAE, MSE, dan RMSE bersama-sama untuk mendapatkan gambaran yang lebih komprehensif tentang kualitas model dan akurasi prediksi.

2 Metodologi Penelitian

Dalam penelitian ini, metode atau tahapan yang digunakan adalah metode KDD (Knowledge Discovery in Database). Tahapan KDD terdiri dari 5 tahapan proses, yaitu seleksi, pra-pemrosesan, transformasi, Data Mining, dan evaluasi[11]. kelima tahapan tersebut mengikuti proses berikut:



Gambar 1. Metode Penelitian

1. Pemilihan Data

Pemilihan Data adalah tahap awal dalam KDD di mana data yang relevan dan berkualitas baik dipilih untuk dianalisis. Pemilihan data yang tepat sangat penting karena dapat mempengaruhi hasil akhir, pada tahap ini dilakukan proses seleksi atau pemilihan data yang tepat. Data yang akan digunakan adalah Data Numerik yang diambil dari kumpulan data *open source* yaitu Kaggle, dengan data yang digunakan adalah *World Happiness Report* khusus tahun 2022, dengan data yang ada sebanyak 147 baris dan terdapat 12 kolom.

2. Pra-pemrosesan

Pra-pemrosesan Data adalah tahap di mana data yang dipilih dari tahap sebelumnya dipersiapkan untuk analisis lebih lanjut. Pada tahap ini dilakukan pembersihan data, seperti menangani data yang kosong, menangani data yang duplikat, memperbaiki tipe data yang tidak tepat. Pada tahap ini dilakukan pengecekan data secara teliti agar tidak ada lagi permasalahan-permasalahan terhadap data yang dapat mengganggu permodelan data.

3. Tranformasi

Transformasi Data adalah tahap di mana data yang telah dipilih dan diproses dari tahap sebelumnya diubah menjadi bentuk yang lebih cocok untuk analisis. Transformasi data meliputi pemrosesan data skala, normalisasi, dan reduksi dimensi[16].

4. Data Mining

Data mining adalah sekumpulan proses untuk mencari pola, hubungan, mendapatkan nilai dari big data dan informasi berupa informasi, menemukan hubungan dan menyederhanakan data untuk mendapatkan informasi dan wawasan yang dapat dipahami dan bermanfaat melalui statistik serta matematika[8]. Pada Tahapan Data mining ini algoritma yang dipakai adalah linear regresi. Data Mining diimplementasikan dengan menggunakan bahasa pemrograman Python dan *library Scikit Learn*.

5. Evaluasi

Hasil dari permodelan regresi linear yang sudah dilakukan pada tahap data mining kemudian di evaluasi, apakah faktor-faktor yang ada dapat mempengaruhi kebahagiaan suatu negara atau tidak. Evaluasi Data adalah tahap terakhir di mana hasil dari proses KDD dievaluasi untuk memastikan hasil yang didapat sesuai dengan tujuan awal dari analisis data. Evaluasi data meliputi penilaian performa model, validasi, dan interpretasi hasil. Tahap ini penting untuk memastikan hasil analisis yang akurat dan berguna.[12]

3 Hasil Uji Coba Dan Pembahasan

3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini adalah Data *World Happiness Report* tahun 2022. Data asli yang diperoleh terdiri dari 147 baris dan 12 kolom. Namun, sebelum dilakukan analisis lebih lanjut, dilakukan pembersihan data untuk mengatasi missing value. Setelah dilakukan pembersihan data, penyeleksian data yang diperlukan, perubahan tipe data, maka didapatkan jumlah baris yang siap digunakan untuk pemodelan adalah sebanyak 146 baris dengan 8 kolom. Dengan melakukan pembersihan data, dapat memastikan keakuratan dan integritas data yang digunakan dalam analisis untuk menghasilkan hasil yang lebih valid dan reliabel. Berikut adalah gambaran dari keseluruhan data:

	Country	Happiness score	GDP per Capita	Social Support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
0	Finland	7.821	1.892	1.258	0.775	0.736	0.109	0.534
1	Denmark	7.636	1.953	1.243	0.777	0.719	0.188	0.532
2	Iceland	7.557	1.936	1.320	0.803	0.718	0.270	0.191
3	Switzerland	7.512	2.026	1.226	0.822	0.677	0.147	0.461
4	Netherlands	7.415	1.945	1.206	0.787	0.651	0.271	0.419
...
141	Botswana*	3.471	1.503	0.815	0.280	0.571	0.012	0.102
142	Rwanda*	3.268	0.785	0.133	0.462	0.621	0.187	0.544
143	Zimbabwe	2.995	0.947	0.690	0.270	0.329	0.106	0.105
144	Lebanon	2.955	1.392	0.498	0.631	0.103	0.082	0.034
145	Afghanistan	2.404	0.758	0.000	0.289	0.000	0.089	0.005

146 rows x 8 columns

Gambar 2. Dataset Word Hapiienes Report setelah dibersihkan

Tabel di atas berisi informasi mengenai variabel dalam dataset. Terdapat 146 baris data yang lengkap untuk setiap kolom. Kolom-kolom tersebut mencakup informasi penting terkait tingkat kebahagiaan di negara-negara yang terwakili dalam dataset.

Kolom "*Country*" menyimpan nama-nama negara yang diwakili dalam dataset. Kolom "*Happiness score*" berisi nilai tingkat kebahagiaan dari setiap negara.

Selanjutnya, terdapat kolom "*GDP per capita*" yang menyimpan kontribusi dari GDP per kapita terhadap tingkat kebahagiaan. Kolom "*Social support*" mencerminkan kontribusi dari dukungan sosial terhadap tingkat kebahagiaan.

Kolom "*Healthy life expectancy*" mengindikasikan kontribusi dari harapan hidup yang sehat terhadap tingkat kebahagiaan. Sementara itu, kolom "*Freedom to make life choices*" menunjukkan kontribusi dari kebebasan dalam membuat pilihan hidup terhadap tingkat kebahagiaan.

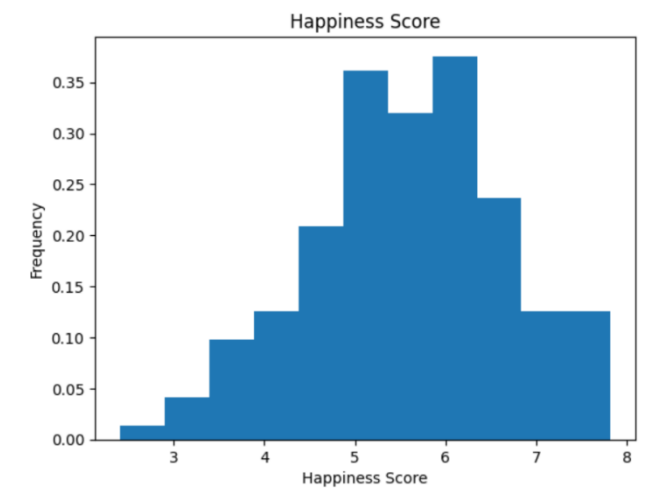
Ada pula kolom "Generosity" yang mencerminkan kontribusi dari tingkat kedermawanan terhadap tingkat kebahagiaan. Terakhir, kolom "Perceptions of corruption" menggambarkan kontribusi dari persepsi terhadap tingkat korupsi terhadap tingkat kebahagiaan.

Data dalam kolom-kolom ini memberikan informasi penting untuk melakukan analisis lebih lanjut dan pemodelan guna memahami faktor-faktor yang memengaruhi tingkat kebahagiaan di negara-negara yang tercakup dalam dataset ini.

Data ini menyediakan informasi sebagai berikut:

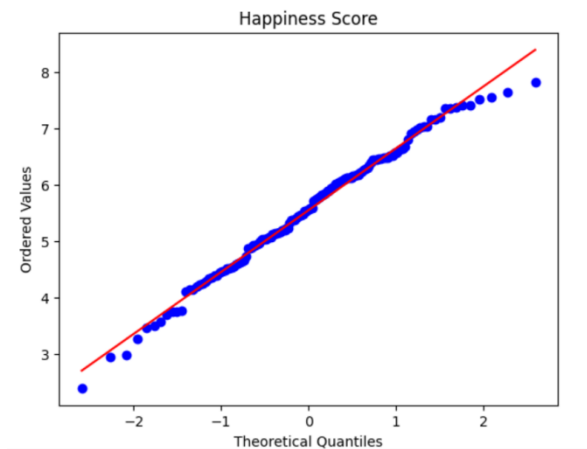
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 146 entries, 0 to 145
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Country                                146 non-null    object
1   Happiness score                        146 non-null    float64
2   GDP per Capita                        146 non-null    float64
3   Social Support                        146 non-null    float64
4   Healthy life expectancy                146 non-null    float64
5   Freedom to make life choices           146 non-null    float64
6   Generosity                             146 non-null    float64
7   Perceptions of corruption              146 non-null    float64
dtypes: float64(7), object(1)
memory usage: 10.3+ KB
```

Gambar 3. Info Dataset



Gambar 4. Histogram Happiness Score dengan Berdasarkan Frekuensi

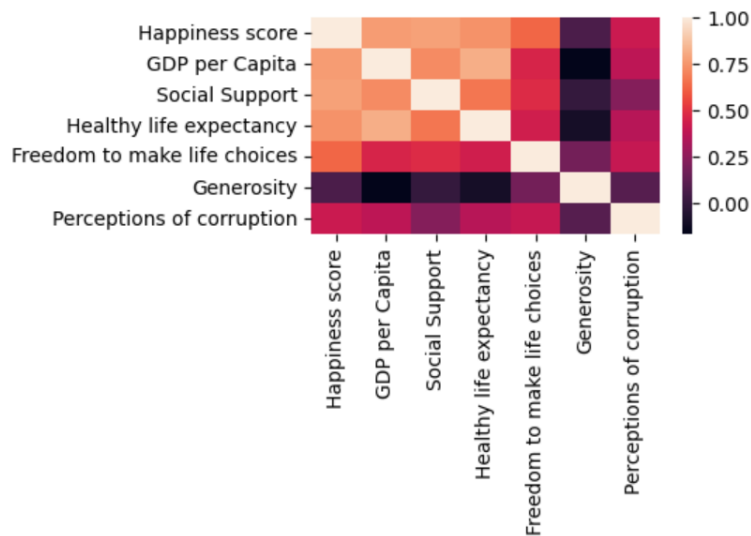
Grafik histogram digunakan untuk memvisualisasikan distribusi frekuensi dari variabel "Happiness Score". Dengan melihat histogram diatas, kita dapat memahami bagaimana data terdistribusi di sepanjang sumbu horizontal yang merepresentasikan nilai Happiness Score, serta sumbu vertikal yang menunjukkan frekuensi kemunculan nilai tersebut. Dalam konteks ini, histogram menunjukkan bahwa data terdistribusi secara normal, dikarenakan data memiliki pola yang mendekati kurva normal atau *bell-shaped curve*. Hal ini dapat diamati ketika data terkumpul di sekitar pusat histogram dan secara simetris menyebar ke kiri dan kanan.



Gambar 5. Plot Happiness Score

Grafik QQ plot digunakan untuk membandingkan distribusi data yang diamati, yaitu "Happiness Score", dengan distribusi yang diharapkan, dalam hal ini distribusi normal. Pada grafik ini, titik-titik data diamati dipetakan terhadap titik-titik pada garis referensi yang menggambarkan distribusi normal yang diharapkan. Jika titik-titik data hampir sejajar dengan garis referensi, itu menunjukkan bahwa data secara keseluruhan mengikuti pola distribusi

normal. Berdasarkan plot yang dihasilkan, jika titik-titik data dalam grafik QQ plot sejajar dengan garis referensi, dapat disimpulkan bahwa data "Happiness Score" mengikuti distribusi normal.



Gambar 6. Grafik Kolerasi

Berdasarkan Grafik korelasi di atas, mengenai hubungan antara variabel dalam dataset. Pertama, tingkat kebahagiaan (*Happiness Score*) memiliki korelasi positif yang kuat dengan GDP per kapita, dukungan sosial, dan harapan hidup yang sehat. Ini menunjukkan bahwa negara-negara dengan tingkat pendapatan yang tinggi, dukungan sosial yang kuat, dan kualitas hidup yang baik cenderung memiliki tingkat kebahagiaan yang tinggi.

Selanjutnya, tingkat kebebasan dalam membuat pilihan hidup juga memiliki korelasi positif yang moderat dengan tingkat kebahagiaan. Ini mengindikasikan bahwa kebebasan untuk mengambil keputusan dalam kehidupan memiliki pengaruh yang signifikan terhadap tingkat kebahagiaan di negara-negara tersebut.

Di sisi lain, korelasi antara tingkat kebahagiaan dan kedermawanan serta persepsi terhadap korupsi relatif lemah. Ini berarti bahwa tingkat kedermawanan masyarakat dan persepsi terhadap tingkat korupsi tidak memiliki pengaruh yang signifikan terhadap tingkat kebahagiaan secara keseluruhan.

Dengan demikian, faktor-faktor seperti GDP per kapita, dukungan sosial, harapan hidup yang sehat, dan kebebasan dalam membuat pilihan hidup memiliki peran yang lebih penting dalam menjelaskan tingkat kebahagiaan di negara-negara tersebut. Sementara itu, kedermawanan dan persepsi terhadap korupsi memiliki pengaruh yang lebih kecil.

3.2 Pembahasan

Data yang sudah siap digunakan maka dilakukan proses pemodelan menggunakan regresi linear untuk memprediksi tingkat kebahagiaan (*Happiness score*) berdasarkan faktor-faktor yang dijelaskan oleh variabel-variabel yang ada.

Pertama, data dibagi menjadi data latihan (X_{train} dan y_{train}) dan data uji (X_{test} dan y_{test}) dengan rasio 80:20. Kemudian, dilakukan *fitting* model menggunakan *Linear Regression*. Hasil dari pemodelan tersebut menunjukkan *nilai intercept* (nilai konstanta) sebesar 1.6455817838578142.

```
print(lm.intercept_)
1.6455817838578142
```

Gambar 7. Nilai Intercept

Intercept dalam pemodelan linear merupakan nilai yang menunjukkan tingkat kebahagiaan yang diharapkan ketika semua variabel penjelas (faktor-faktor yang diamati) memiliki nilai nol atau tidak memberikan pengaruh. Dalam konteks ini, *intercept* sebesar 1.6455817838578142 menunjukkan bahwa ketika semua faktor yang diamati tidak memberikan kontribusi, diperkirakan tingkat kebahagiaan dasar yang dapat dicapai adalah sekitar 1.645. Oleh karena itu, *intercept* dapat dianggap sebagai nilai awal atau titik referensi dari model, yang menunjukkan tingkat kebahagiaan yang dapat dicapai tanpa adanya faktor-faktor yang diamati.

	Coefficient
GDP per Capita	0.466575
Social Support	1.551701
Healthy life expectancy	1.266429
Freedom to make life choices	1.838195
Generosity	0.420611
Perceptions of corruption	0.675556

Gambar 8. Tabel Koefisien

Selanjutnya, koefisien yang diperoleh menunjukkan pengaruh masing-masing faktor terhadap tingkat kebahagiaan. Koefisien ini menggambarkan seberapa besar perubahan dalam tingkat kebahagiaan yang diharapkan ketika suatu faktor meningkat satu satuan, sementara faktor-faktor lainnya tetap konstan. Sebagai contoh, pada faktor GDP per kapita, peningkatan satu unit dalam GDP per kapita akan memberikan kontribusi sebesar 0.466575 terhadap tingkat kebahagiaan.

Koefisien determinan (*R-squared*) memiliki nilai sebesar 0.7736829763399227. Nilai ini mengindikasikan seberapa baik model linear mampu menjelaskan variasi dalam tingkat kebahagiaan. Semakin dekat nilai *R-squared* dengan 1, semakin baik model dalam menjelaskan variabilitas data. Dalam hal ini, sekitar 77.37% variabilitas dalam tingkat kebahagiaan dapat dijelaskan oleh faktor-faktor yang digunakan dalam model ini.

setelah model regresi dibuat maka dilakukan evaluasi terhadap model regresi linear, evaluasi dilakukan dengan mencari nilai *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE).

Mean Absolute Error (MAE) sebesar 0.433809262213575 menunjukkan bahwa rata-rata kesalahan prediksi model dalam memperkirakan tingkat kebahagiaan adalah sebesar 0.433809262213575. Artinya, secara rata-rata, prediksi model memiliki selisih sebesar 0.433809262213575 dari nilai sebenarnya.

Mean Squared Error (MSE) sebesar 0.27849009018351645 mengindikasikan bahwa rata-rata kesalahan prediksi model dalam bentuk selisih kuadrat dari nilai prediksi dan nilai sebenarnya adalah sebesar 0.27849009018351645. MSE mengukur penyebaran kesalahan prediksi secara keseluruhan.

Root Mean Squared Error (RMSE) sebesar 0.5277216029153217 adalah akar kuadrat dari MSE. RMSE memberikan estimasi rata-rata kesalahan prediksi dalam bentuk nilai sebenarnya. Dalam hal ini, RMSE sebesar 0.5277216029153217 menunjukkan tingkat kesalahan rata-rata sekitar 0.5277216029153217 dalam memprediksi tingkat kebahagiaan.

Interpretasi dari hasil tersebut adalah semakin rendah nilai MAE, MSE, dan RMSE, semakin baik performa model dalam memprediksi tingkat kebahagiaan. Dalam konteks ini, meskipun terdapat kesalahan prediksi, namun nilai-nilai MAE, MSE, dan RMSE yang relatif rendah menunjukkan bahwa model regresi linear ini memiliki kemampuan yang baik dalam memperkirakan tingkat kebahagiaan berdasarkan faktor-faktor yang diamati.

4 Kesimpulan

Berdasarkan model regresi linear yang telah dibuat, terdapat beberapa faktor yang mempengaruhi tingkat kebahagiaan suatu negara. Faktor-faktor tersebut adalah *GDP per capita*, *Social support*, *Healthy life expectancy*, *Freedom to make life choices*, *Generosity*, dan *Perceptions of corruption*. Dari hasil koefisien yang diperoleh, dapat diketahui bahwa faktor GDP per Kapita memiliki pengaruh positif sebesar 0.466575 terhadap tingkat kebahagiaan. Hal ini menunjukkan bahwa peningkatan dalam GDP per capita suatu negara berkontribusi terhadap peningkatan tingkat kebahagiaan. Selanjutnya, faktor *Social support* memiliki pengaruh positif yang lebih besar, yaitu sebesar 1.551701. Dukungan sosial yang ada dalam masyarakat memberikan kontribusi yang signifikan terhadap tingkat kebahagiaan. Faktor *Healthy life expectancy* juga memiliki pengaruh positif yang signifikan dengan koefisien sebesar 1.266429. Hal ini menunjukkan bahwa harapan hidup yang sehat berperan penting dalam meningkatkan tingkat kebahagiaan. *Freedom to make life choices* juga memiliki pengaruh positif yang kuat, dengan koefisien sebesar 1.838195. Kebebasan dalam membuat pilihan hidup memberikan dampak positif terhadap kebahagiaan suatu negara. Faktor *Generosity* memiliki pengaruh positif yang lebih rendah, yaitu sebesar 0.420611. Tingkat kedermawanan dalam masyarakat juga memberikan kontribusi terhadap tingkat kebahagiaan, meskipun tidak sebesar faktor-faktor lainnya. Terakhir, faktor *Perceptions of corruption* memiliki pengaruh positif sebesar 0.675556. Persepsi terhadap tingkat korupsi di suatu negara juga berperan dalam mempengaruhi tingkat kebahagiaan. Berdasarkan koefisien determinan (*R-squared*) bahwa model regresi linear yang digunakan mampu menjelaskan sekitar 77.37% variasi dalam tingkat kebahagiaan suatu negara berdasarkan faktor-faktor yang digunakan dalam model ini. Hal ini menunjukkan bahwa faktor-faktor yang telah dijelaskan dalam model, yaitu *GDP per capita*, *Social support*, *Healthy life expectancy*, *Freedom to make life choices*, *Generosity*, dan *Perceptions of corruption*, memiliki tingkat penjelasan yang cukup baik terhadap tingkat kebahagiaan suatu negara. Sebagian besar variasi dalam tingkat kebahagiaan dapat dijelaskan oleh faktor-faktor tersebut. Dari Hasil Analisa yang ada faktor-faktor ini bisa dijadikan pedoman bagi masyarakat ataupun instansi pemerintahan untuk berupaya meningkatkan kebahagiaan dan kesejahteraan suatu negara.

5 Daftar Pustaka

- [1] E. E. P. Unud *et al.*, "FakultasEkonomidanBisnisUniversitasUdayana (Unud), Bali , Indonesia PENDAHULUAN Studi mengenai konsep kebahagiaan telah banyak dilakukan melalui berbagai perspektif . Masing-masing perspektif menjelaskan berbagai hasil yang berbeda-beda mengenai apa yan," hal. 963–992.
- [2] J. F. Helliwell, R. Layard, dan J. D. Sachs, "Un Happiness Report," 2018.
- [3] Siska Wulandari dan Ami Widyastuti, "Faktor - Faktor Kebahagiaan Di Tempat Kerja," *J. Psikol. UIN Sultan Syarif Kasim Riau*, vol. 10, no. Juni, hal. 41–52, 2014.
- [4] V. Plotnikova, M. Dumas, dan F. Milani, "Adaptations of data mining methodologies: A systematic literature review," *PeerJ Comput. Sci.*, vol. 6, hal. 1–43, 2020, doi: 10.7717/PEERJ-CS.267.
- [5] D. A. Kumalasar dan I. G. W. M. Yasa, "Pengaruh Faktor-Faktor Yang Mempengaruhi Tingkat Kebahagiaan Negara Di Dunia," *E-Jurnal Ekon. Pambang. Univ. Udayana*, vol. 9, no. 5, hal. 963–992, 2020.

-
- [6] J. Sachs, C. Kroll, G. Lafortune, G. Fuller, dan F. Woelm, *Sustainable Development Report 2021*. 2021.
- [7] J. F. Helliwell, R. Layard, J. D. Sachs, dan J.-E. De Neve, "WellBeing International WellBeing International WBI Studies Repository WBI Studies Repository World Happiness Report 2021 World Happiness Report 2021," 2021, [Daring]. Tersedia pada: https://www.wellbeingintlstudiesrepository.org/hw_happiness.
- [8] A. Purwanto, A. Primajaya, dan A. Voutama, "Penerapan Algoritma C4.5 Dalam Prediksi Potensi Tingkat Kasus Pneumonia Di Kabupaten Karawang," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 4, hal. 390, 2020, doi: 10.26418/justin.v8i4.41959.
- [9] S. Alam, M. G. Resmi, dan N. Masripah, "Classification of Covid-19 vaccine data screening with Naive Bayes algorithm using Knowledge Discovery in database method," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 4, no. 2, hal. 177–185, 2022, doi: 10.47709/cnahpc.v4i2.1584.
- [10] P. S. T. Informatika dan U. I. Majapahit, "Pemanfaatan Knowledge Data Discovery(KDD) Pada Pola Permainan Atlet Bulutangkis," *Explor. IT J. Keilmuan dan Apl. Tek. Inform.*, vol. 11, no. 1, hal. 1–6, 2019, doi: 10.35891/explorit.v11i1.1467.
- [11] A. Yoga Pratama *et al.*, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, hal. 897–910, 2021.
- [12] Euis Saraswati, Yuyun Umaidah, dan Apriade Voutama, "Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19," *Gener. J.*, vol. 5, no. 2, hal. 109–118, 2021, doi: 10.29407/gj.v5i2.16125.
- [13] A. M. Abdulkadium, R. A. A. Shekan, dan H. A. Hussain, "Application of Data Mining and Knowledge Discovery in Medical Databases," *Webology*, vol. 19, no. 1, hal. 4912–4924, 2022, doi: 10.14704/web/v19i1/web19329.
- [14] S. Rath, A. Tripathy, dan A. R. Tripathy, "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 5, hal. 1467–1474, 2020, doi: 10.1016/j.dsx.2020.07.045.
- [15] D. Chicco, M. J. Warrens, dan G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, hal. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
- [16] A. A. Rizaldy *et al.*, "Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Penyebaran Vaksin Covid 19 di Kabupaten Cilacap," vol. 3, no. 2, hal. 43–50, 2021.