
Implementasi *Exploratory Data Analysis* (EDA) Untuk Menganalisis Berbagai Faktor Risiko Penyakit Jantung Di Amerika Serikat

¹Sherly Patricia, ²Christopher Parulian Marpaung, ³Lucky Raffael Wijaya,
⁴Manisha Arie Paramartha, ⁵Willyanto Dharmansyah Atmadja,
⁶Rahmi Yulia Ningsih

^{1,2,3,4,5,6}Teknik Informatika, Universitas Bina Nusantara

Email : [1sherly.patricia001@binus.ac.id](mailto:sherly.patricia001@binus.ac.id), [2christopher.marpaung@binus.ac.id](mailto:christopher.marpaung@binus.ac.id),
[3lucky.wijaya002@binus.ac.id](mailto:lucky.wijaya002@binus.ac.id), [4manisha.paramartha@binus.ac.id](mailto:manisha.paramartha@binus.ac.id),
[5willyanto.atmadja@binus.ac.id](mailto:willyanto.atmadja@binus.ac.id), [6rahmi.ningsih@binus.edu](mailto:rahmi.ningsih@binus.edu)

Corresponding Mail Author: sherly.patricia001@binus.ac.id

Abstract

Heart disease is still a serious health problem, especially in the United States. The data shows that heart disease as one of the primary causes of death in the United States. Several health factors are known to be the risk factors for heart disease. Therefore, this study aims to analyze the relationship between several risk factors for heart disease in the United States that will be presented based on the results of the implementation of Exploratory Data Analysis (EDA) using the Python programming language. The method used in this research is quantitative research. This method used to present the results of EDA implementation in the form of visualization and analysis related to the dataset about the risk factors for heart disease in the United States. The results showed that all health factors contained in the dataset gave expected results, except drinking alcohol, have a specific influence on the risk of heart disease. In addition, the conclusion shows that stroke is the factor that has the highest impact on the occurrence of heart disease.

Keywords: *Exploratory Data Analysis (EDA), Python, Heart Disease.*

Pendahuluan

National Center for Health Statistics (2022) menyatakan bahwa penyakit jantung telah menjadi salah satu penyebab utama kematian bagi sebagian besar ras di Amerika Serikat. Setiap tahun, terdapat sekitar 697.000 orang di Amerika Serikat meninggal akibat penyakit jantung (New York State, 2022). Setiap 33 detik, korban penyakit jantung mengalami kematian (National Center for Health Statistics, 2022). Selain itu, biaya yang dikeluarkan untuk layanan kesehatan, obat-obatan, dan hilangnya produktivitas karena kematian akibat penyakit jantung telah mencapai sekitar \$239,9 miliar setiap tahun dari 2018 hingga 2019 (National Center for Health Statistics, 2023). Dari data-data tersebut, dapat ditunjukkan betapa berpengaruhnya penyakit jantung terhadap tingkat kematian di Amerika Serikat.

Adapun berbagai faktor yang memicu risiko penyakit jantung. Kolesterol darah yang tinggi, tekanan darah yang tinggi, dan kebiasaan merokok merupakan faktor risiko utama penyakit jantung (Fryar et al., 2012). Selain itu, ada beberapa faktor lainnya yang

dapat memicu penyakit jantung. Faktor tersebut berkaitan dengan pola atau gaya hidup serta penyakit lain selain jantung dari para penderita misalnya, obesitas, diabetes, pola makan yang tidak sehat, penggunaan alkohol yang berlebihan, ketidakaktifan fisik, dan sebagainya (*Centers for Disease Control and Prevention, 2023*).

Berdasarkan uraian diatas, penyakit jantung masih menjadi permasalahan kesehatan yang serius, khususnya di Amerika Serikat. Kemudian, faktor-faktor risiko dari penyakit jantung memiliki peran yang penting dalam terjangkitnya penyakit jantung. Maka dari itu, faktor-faktor risiko akan dianalisis lebih lanjut dengan *Exploratory Data Analysis* (EDA). Bagaimanakah hubungan antara beberapa faktor risiko penyakit jantung di Amerika Serikat yang disajikan sesuai dengan implementasi *Exploratory Data Analysis* (EDA) menggunakan bahasa pemrograman *Python*? Kemudian, faktor risiko apakah yang berpengaruh paling besar dalam terjangkitnya penyakit jantung?

Penelitian ini bertujuan untuk menganalisis hubungan beberapa faktor risiko penyakit jantung di Amerika Serikat yang akan disajikan sesuai dengan hasil implementasi *Exploratory Data Analysis* (EDA) menggunakan bahasa pemrograman *Python*. Kemudian, penelitian ini juga bertujuan untuk mengetahui faktor risiko yang paling berpengaruh dalam terjangkitnya penyakit jantung.

Manfaat dari penelitian ini adalah dapat memberikan informasi kepada masyarakat mengenai faktor risiko yang paling berpengaruh terhadap terjadinya penyakit jantung melalui EDA yang telah dilakukan menggunakan pemrograman *python*. Melalui hasil EDA tersebut, pemahaman data yang sangat kompleks dapat menjadi suatu informasi yang lebih mudah dipahami.

Landasan Teori

***Exploratory Data Analysis* (EDA)**

Exploratory Data Analysis (EDA), yang merupakan salah satu metode menganalisis dan memvisualisasikan data yang memiliki tujuan untuk memperoleh pemahaman yang lebih baik terkait informasi yang ada pada data (Radhi et al., 2021). EDA memiliki beberapa peran penting. Pertama, EDA berperan dalam membantu pengumpulan informasi dan pemahaman terkait data dengan lebih baik. Kedua, EDA berperan dalam menghilangkan kekeliruan, mendeteksi *outliers*, serta nilai yang tidak diperlukan dari data. Ketiga, EDA dapat menyiapkan kumpulan data untuk dianalisis dan hasil yang lebih akurat. Selain itu, EDA juga lebih fokus pada grafik representasi data (Behrens, 1997). Dengan demikian, metode EDA dapat mempermudah penyajian visualisasi data sekaligus informasi terkait dataset yang meliputi faktor-faktor risiko penyakit jantung di Amerika Serikat. Selain itu, hasil implementasi dan analisis yang dilakukan juga mampu menghindari penafsiran dan asumsi dari pembaca yang dapat menimbulkan *biased*.

Python

Pada penelitian ini, metode EDA akan dilakukan dengan menggunakan bahasa pemrograman *Python*. *Python* memiliki kelebihan yang signifikan dibandingkan dengan bahasa pemrograman lainnya. Pertama, *Python* merupakan bahasa pemrograman yang memiliki rangkaian aturan kode yang mudah dipahami dan dipelajari. Kedua, tidak seperti bahasa lainnya, *Python* merupakan bahasa pemrograman yang ditafsirkan

(*interpreted language*) (Bogdanchikov et.al., 2013). Ketiga, *Python* memiliki *library* yang memadai dan mudah untuk digunakan. *Library-library* tersebut antara lain, *NumPy*, *SciPy*, *Pandas*, *Keras*, *Matplotlib*, *SciKit-Learn*, dan sebagainya (Dhruv et.al., 2022). Dengan demikian, *library* tersebut mempermudah para *programmer*, khususnya dalam menganalisis dan memvisualisasikan data. Selain itu, *Python* merupakan salah satu bahasa pemrograman yang penting dalam *machine learning*, pengembangan perangkat lunak, serta ilmu data (McKinney, 2022). Maka dari itu, *Python* menjadi bahasa pemrograman yang paling tepat untuk melakukan EDA.

Metode Penelitian

Jenis penelitian ini adalah penelitian kuantitatif. Populasi dari penelitian ini adalah dataset yang berjudul "*Personal Key Indicators of Heart Disease*" yang diambil dari *Kaggle*. Populasi tersebut merepresentasikan hasil survei tahunan 2020 dari Pusat Pengendalian dan Pencegahan Penyakit Amerika Serikat terhadap 400 ribu orang dewasa terkait status kesehatan mereka. Sampel pada penelitian ini yaitu jumlah data dari dataset yang digunakan yaitu sebanyak 319.795 data. Kemudian, jumlah variabel pada dataset ini adalah sebanyak 18 variabel.

Adapun alur dari penelitian ini dimulai dari menyiapkan dan mengumpulkan dataset yang akan dianalisis, kemudian mengolah dataset tersebut sesuai dengan implementasi *Exploratory Data Analysis* (EDA) hingga memperoleh informasi yang cukup terkait penelitian ini yaitu mengenai faktor-faktor yang mempengaruhi risiko penyakit jantung. Penelitian ini dilaksanakan secara *daring* menggunakan bahasa pemrograman *python* melalui media *google colab*. Selanjutnya, berikut adalah langkah-langkah yang perlu dilakukan dalam mengimplementasikan EDA:

Data Loading

Data loading pada penelitian ini merupakan proses memasukkan beberapa *library* yang terdapat dalam bahasa pemrograman *python* dengan cara memanggil *syntax import* atau *syntax from import*. *Syntax as* digunakan untuk mempersingkat penulisan *library* yang digunakan agar *code* lebih rapi dan efisien. Selanjutnya, mengunggah dan memperlihatkan dataset yang telah diunduh dari *kaggle*.

```
import pandas as pd
import seaborn as sns
import numpy as np
from matplotlib import pyplot as plt
from collections import Counter
from statsmodels.stats.proportion import proportions_ztest
from statsmodels.stats.weightstats import ztest as ztest
```

Gambar 1. Proses Memasukkan Beberapa *Library* Yang Akan Digunakan

```
df = pd.read_csv("heart_2020_cleaned.csv")
df
```

	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWalking	Sex	AgeCategory	Race	Diabetic	PhysicalActivity	GenHealth	SleepTime	Asthma
0	No	16.60	Yes	No	No	3.0	30.0	No	Female	55-59	White	Yes	Yes	Very good	5.0	Yes
1	No	20.34	No	No	Yes	0.0	0.0	No	Female	60 or older	White	No	Yes	Very good	7.0	No
2	No	26.58	Yes	No	No	20.0	30.0	No	Male	65-69	White	Yes	Yes	Fair	8.0	Yes
3	No	24.21	No	No	No	0.0	0.0	No	Female	75-79	White	No	No	Good	6.0	No
4	No	23.71	No	No	No	28.0	0.0	Yes	Female	40-44	White	No	Yes	Very good	8.0	No
...
319790	Yes	27.41	Yes	No	No	7.0	0.0	Yes	Male	60-64	Hispanic	Yes	No	Fair	6.0	Yes
319791	No	29.84	Yes	No	No	0.0	0.0	No	Male	35-39	Hispanic	No	Yes	Very good	5.0	Yes
319792	No	24.24	No	No	No	0.0	0.0	No	Female	45-49	Hispanic	No	Yes	Good	6.0	No
319793	No	32.81	No	No	No	0.0	0.0	No	Female	25-29	Hispanic	No	No	Good	12.0	No
319794	No	46.56	No	No	No	0.0	0.0	No	Female	60 or older	Hispanic	No	Yes	Good	8.0	No

319795 rows x 18 columns

Gambar 2. Proses Mengunggah Dan Memperlihatkan Dataset Yang Telah Diunduh Dari Kaggle

Data Cleaning

Data cleaning pada penelitian ini merupakan proses pemeriksaan nilai unik pada dataset serta penghapusan data yang tidak memiliki peran signifikan. Dari dataset ini, tidak ada satupun data yang bernilai *null* sehingga penghapusan data tidak diperlukan. Selain itu, proses pengubahan bentuk data menjadi sebuah angka juga dapat dilakukan. Salah satunya adalah mengubah *Yes* menjadi angka 1 dan *No* menjadi angka 0. Proses ini dilakukan untuk mempermudah pemrograman.

```
print(df.nunique())
```

HeartDisease	2
BMI	3604
Smoking	2
AlcoholDrinking	2
Stroke	2
PhysicalHealth	31
MentalHealth	31
DiffWalking	2
Sex	2
AgeCategory	13
Race	6
Diabetic	4
PhysicalActivity	2
GenHealth	5
SleepTime	24
Asthma	2
KidneyDisease	2
SkinCancer	2
dtype: int64	

Gambar 3. Hasil Yang Menunjukkan Nilai Unik Dari Setiap Variabel

```
print(df.isna().sum())
```

HeartDisease	0
BMI	0
Smoking	0
AlcoholDrinking	0
Stroke	0
PhysicalHealth	0
MentalHealth	0
DiffWalking	0
Sex	0
AgeCategory	0
Race	0
Diabetic	0
PhysicalActivity	0
GenHealth	0
SleepTime	0
Asthma	0
KidneyDisease	0
SkinCancer	0
dtype: int64	

Gambar 4. Hasil Yang Menunjukkan Nilai Null Yang Terdapat Pada Setiap Variabel

```
print(df.columns)
df = df[df.columns].replace({'Yes':1, 'No':0, 'Male':1, 'Female':0, 'No, borderline diabetes':0, 'Yes (during pregnancy)':1})
df["Diabetic"] = df["Diabetic"].astype(int)
df
```

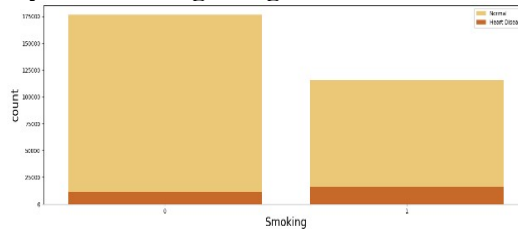
	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWalking	Sex	AgeCategory	Race	Diabetic	PhysicalActivity	GenHealth	SleepTime	Asthma	KidneyDisease
0	0	16.60	1	0	0	3.0	30.0	0	0	55-59	White	1	Very good	5.0	1	0	
1	0	20.34	0	0	1	0.0	0.0	0	0	80 or older	White	0	1 Very good	7.0	0	0	
2	0	26.58	1	0	0	20.0	30.0	0	1	65-69	White	1	1 Fair	8.0	1	0	
3	0	24.21	0	0	0	0.0	0.0	0	0	75-79	White	0	0 Good	6.0	0	0	
4	0	23.71	0	0	0	28.0	0.0	1	0	40-44	White	0	1 Very good	8.0	0	0	
...
319790	1	27.41	1	0	0	7.0	0.0	1	1	60-64	Hispanic	1	0 Fair	6.0	1	0	
319791	0	29.84	1	0	0	0.0	0.0	0	1	35-39	Hispanic	0	1 Very good	5.0	1	0	
319792	0	24.24	0	0	0	0.0	0.0	0	0	45-49	Hispanic	0	1 Good	6.0	0	0	
319793	0	32.81	0	0	0	0.0	0.0	0	0	25-29	Hispanic	0	0 Good	12.0	0	0	
319794	0	46.56	0	0	0	0.0	0.0	0	0	80 or older	Hispanic	0	1 Good	8.0	0	0	

319795 rows x 18 columns

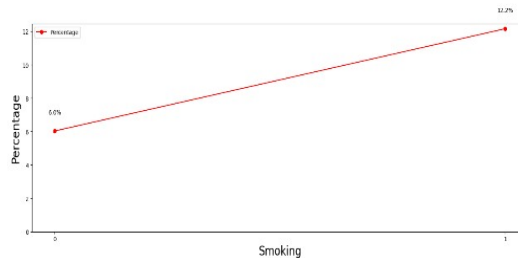
Gambar 5. Hasil Dari Proses Perubahan Bentuk Data

Hasil Dan Pembahasan

Hubungan Antara Penyakit Jantung Dengan Aktivitas Merokok



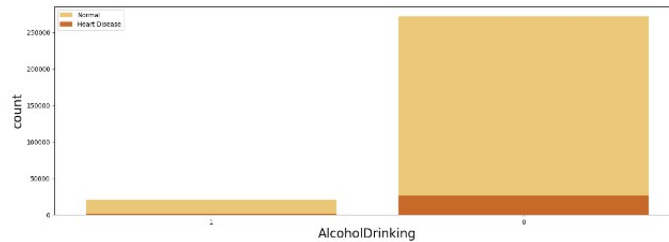
Gambar 6. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Aktivitas Merokok



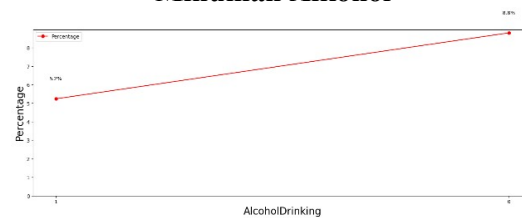
Gambar 7. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Aktivitas Merokok

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa persentase responden yang pernah merokok (setidaknya 5 bungkus atau 100 batang rokok) dalam hidupnya dan memiliki penyakit jantung adalah 12,2%, sedangkan persentase responden yang tidak pernah merokok (setidaknya 5 bungkus atau 100 batang rokok) dan memiliki penyakit jantung adalah 6,0%. Persentase tersebut menunjukkan bahwa responden yang pernah merokok (setidaknya 5 bungkus atau 100 batang rokok) memiliki risiko penyakit jantung sekitar 2,02 kali lebih tinggi dibandingkan dengan responden yang tidak pernah merokok (setidaknya 5 bungkus atau 100 batang rokok). Maka dari itu, dapat disimpulkan bahwa merokok dalam kadar tertentu merupakan salah satu faktor yang memiliki pengaruh yang signifikan dalam risiko penyakit jantung.

Hubungan Antara Penyakit Jantung Dengan Minuman Alkohol



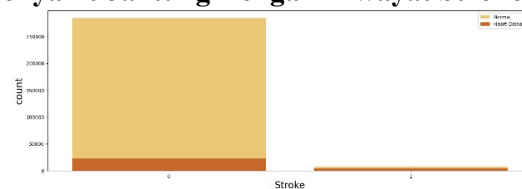
Gambar 8. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Minuman Alkohol



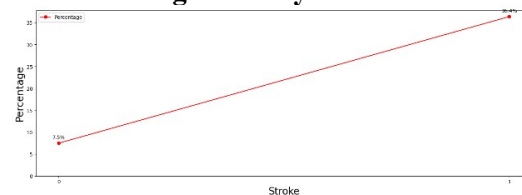
Gambar 9. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Minuman Alkohol

Berdasarkan pada gambar grafik di atas, ditemukan bahwa persentase responden yang memiliki penyakit jantung lebih sedikit di antara yang sering meminum alkohol (5,2%) daripada yang tidak sering meminum alkohol (8,8%). Namun, hal ini cukup tidak sesuai dengan ekspektasi sehingga dilakukanlah juga analisis statistik untuk memastikan pernyataan tersebut dengan melakukan perbandingan antara proporsi dua sampel (*z-test*). Dalam analisis tersebut, terdapatlah sebuah nilai *z* sebesar -18,141 dan nilai *p* sebesar 1.5×10^{-73} . Hasil ini menunjukkan bahwa data tersebut memiliki bukti yang cukup signifikan untuk menolak bahwa meminum alkohol dan penyakit jantung tidak saling berkaitan. Maka dari itu, dapat disimpulkan bahwa frekuensi meminum alkohol merupakan faktor yang memiliki pengaruh terhadap risiko penyakit jantung.

Hubungan Antara Penyakit Jantung Dengan Riwayat Stroke



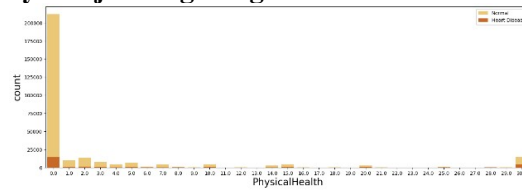
Gambar 10. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Stroke



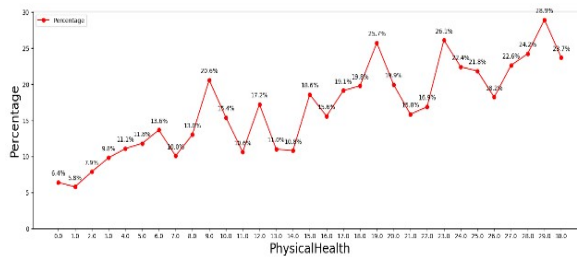
Gambar 11. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Stroke

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa persentase responden yang memiliki riwayat penyakit stroke dan memiliki penyakit jantung adalah 36,4%, sedangkan persentase responden yang tidak memiliki riwayat penyakit stroke dan memiliki penyakit jantung adalah 7,5%. Persentase tersebut menunjukkan bahwa responden dengan riwayat penyakit stroke memiliki risiko penyakit jantung sekitar 4,87 kali lebih tinggi dibandingkan dengan responden tanpa riwayat penyakit stroke. Maka dari itu, dapat disimpulkan bahwa penyakit stroke merupakan salah satu faktor yang memiliki pengaruh yang signifikan dalam risiko penyakit jantung.

Hubungan antara penyakit jantung dengan kondisi kesehatan fisik



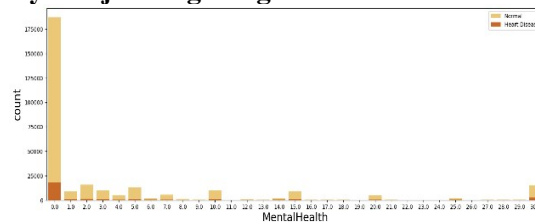
Gambar 12. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Kesehatan Fisik



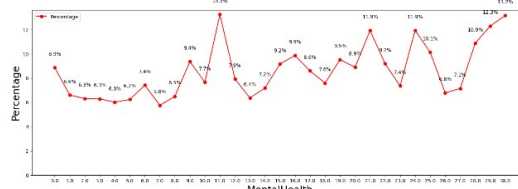
Gambar 13. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Kesehatan Fisik

Berdasarkan pada gambar grafik diatas, dapat diketahui bahwa grafik menunjukan fluktuasi atau keadaan naik turun dikarenakan kurangnya dataset dan faktor error pada data. Namun, kesehatan fisik yang meliputi penyakit dan cedera masih memiliki korelasi dengan terjadinya penyakit jantung, dengan kata lain kesehatan fisik menjadi salah satu faktor terjadinya penyakit jantung yang dibuktikan dari perbedaan rata-rata kesehatan fisik seseorang yang memiliki penyakit jantung dan dengan yang tidak memiliki penyakit jantung berbeda.

Hubungan antara penyakit jantung dengan kondisi kesehatan mental



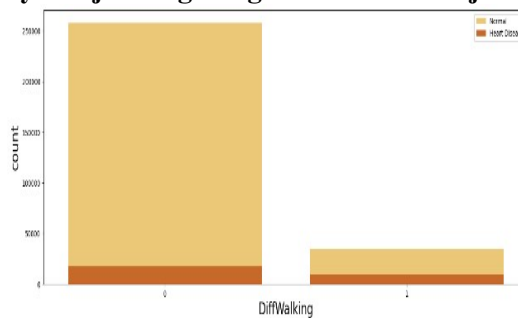
Gambar 14. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Kesehatan Mental



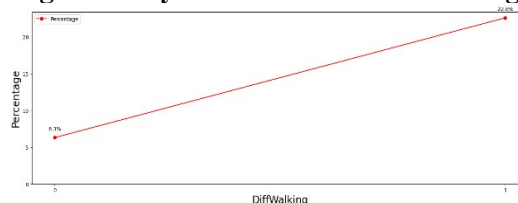
Gambar 15. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Kesehatan Mental

Berdasarkan pada gambar grafik di atas, dapat diketahui bahwa grafik menunjukkan fluktuasi atau keadaan naik turun yang menimbulkan kesulitan dalam menganalisis data melalui grafik saja. Hal ini menyebabkan analisis statistik (*z-test*) perlu dilakukan untuk membandingkan rata-rata hari dimana responden tidak merasa baik secara mental antara yang memiliki penyakit jantung (4,64 hari) dan dengan yang tidak memiliki penyakit jantung (3,82 hari). Dalam analisis tersebut, terdapatlah sebuah nilai *z* sebesar 16,174 dan nilai *p* yang lebih kecil daripada 7.6×10^{-59} . Maka dari itu, dapat disimpulkan bahwa kesehatan mental merupakan salah satu faktor yang dapat mempengaruhi penyakit jantung.

Hubungan antara penyakit jantung dengan kesulitan berjalan menaiki tangga



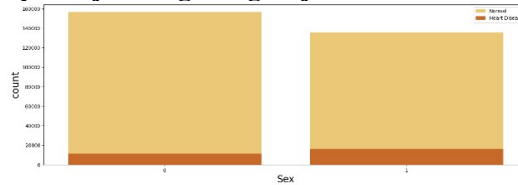
Gambar 16. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Kesulitan Menaiki Tangga



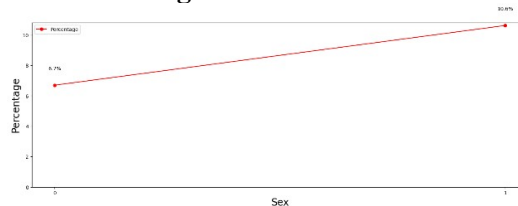
Gambar 17. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Kesulitan Menaiki Tangga

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa persentasi responden yang kesulitan menaiki tangga dan memiliki penyakit jantung adalah 6,3%, sedangkan responden tidak kesulitan menaiki tangga dan memiliki penyakit jantung adalah 22,6%. Persentase tersebut menunjukkan bahwa responden yang kesulitan menaiki tangga memiliki risiko penyakit jantung 3,59 kali lebih tinggi dibandingkan dengan responden yang tidak kesulitan menaiki tangga. Maka dari itu, dapat disimpulkan bahwa, kesulitan menaiki tangga merupakan faktor yang memiliki pengaruh yang signifikan dalam risiko penyakit jantung.

Hubungan antara penyakit jantung dengan jenis kelamin



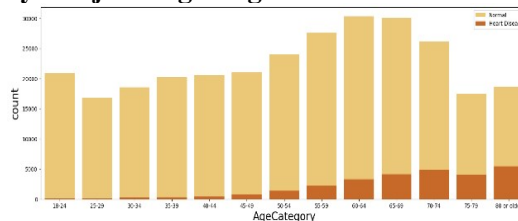
Gambar 18. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Jenis Kelamin



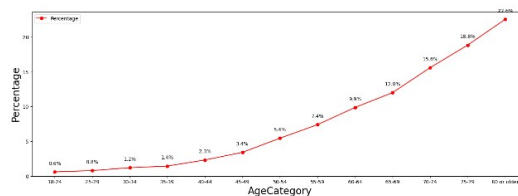
Gambar 19. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Jenis Kelamin

Berdasarkan pada gambar grafik diatas, dapat dilihat bahwa perbandingan jenis kelamin wanita dan pria responden yang mengalami penyakit jantung dan responden normal. Pada grafik tersebut, jenis kelamin wanita disimbolkan dengan angka 0, sedangkan pria disimbolkan dengan angka 1. Persentase responden wanita yang memiliki penyakit jantung adalah 6,7%, sedangkan persentase responden pria yang memiliki penyakit jantung adalah 10,6%. Persentase tersebut menunjukkan bahwa pria memiliki risiko penyakit jantung sekitar 1,59 kali lebih tinggi daripada wanita. Maka dari itu, dapat disimpulkan bahwa jenis kelamin merupakan salah satu faktor yang yang memiliki pengaruh yang cukup signifikan dalam risiko penyakit jantung.

Hubungan antara penyakit jantung dengan usia



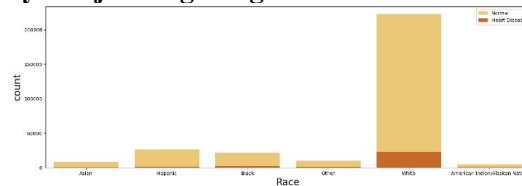
Gambar 20. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Usia



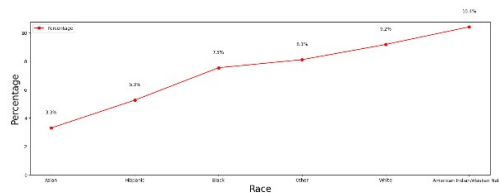
Gambar 21. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Usia

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa seiring bertambahnya usia, persentase yang menunjukkan jumlah penyakit jantung cenderung meningkat. Dari beberapa kategori usia tersebut, usia 80 atau 80 keatas merupakan usia dengan persentase jumlah penyakit jantung yang paling tinggi (22,6%) dibandingkan dengan usia dibawahnya. Maka dari itu, dapat disimpulkan bahwa usia merupakan salah satu faktor yang memiliki pengaruh yang signifikan dalam risiko penyakit jantung.

Hubungan antara penyakit jantung dengan ras



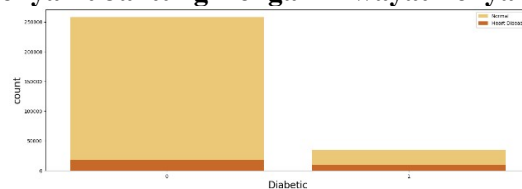
Gambar 22. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Ras



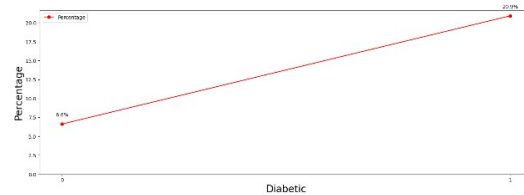
Gambar 23. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Ras

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa mayoritas ras dari responden adalah *White* (245.212 orang), sedangkan *American Indian/Alaskan Native* yang paling sedikit (5.202 orang). Walaupun demikian, sampel sebesar 5.202 orang tergolong cukup besar sehingga diasumsikan grafik persentase sudah cukup akurat. Didapatkan bahwa orang yang memiliki ras *American Indian/Alaskan Native* memiliki persentase yang paling tinggi (10.4%), sedangkan *Asian* memiliki persentase yang paling rendah (3.3%). Maka dari itu, dapat disimpulkan bahwa ras seseorang merupakan salah satu faktor yang memiliki pengaruh yang cukup signifikan dalam risiko penyakit jantung.

Hubungan Antara Penyakit Jantung Dengan Riwayat Penyakit Diabetes



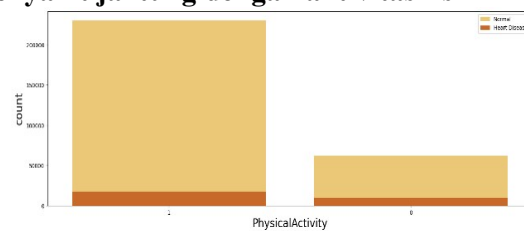
Gambar 24. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Riwayat Penyakit Diabetes



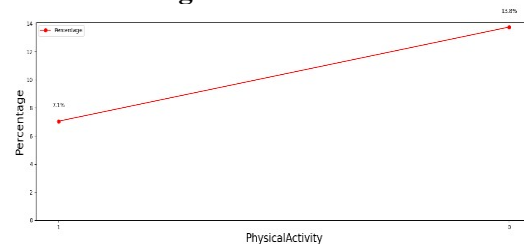
Gambar 25. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Diabetes

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa persentase responden dengan penyakit diabetes dan memiliki penyakit jantung adalah 20,9%, sedangkan persentase responden tanpa penyakit diabetes dan memiliki penyakit jantung adalah 6,6%. Persentase tersebut menunjukkan bahwa responden dengan penyakit diabetes memiliki risiko penyakit jantung sekitar 3,16 kali lebih tinggi dibandingkan dengan responden tanpa penyakit diabetes. Maka dari itu, dapat disimpulkan bahwa penyakit diabetes merupakan salah satu faktor yang memiliki pengaruh yang signifikan dalam risiko penyakit jantung.

Hubungan antara penyakit jantung dengan aktivitas fisik



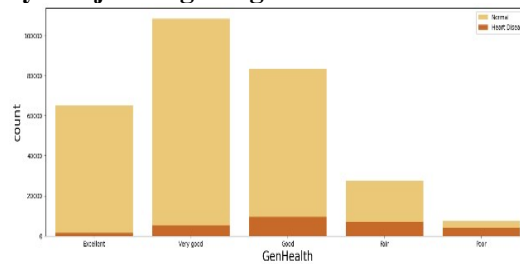
Gambar 26. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Aktivitas Fisik



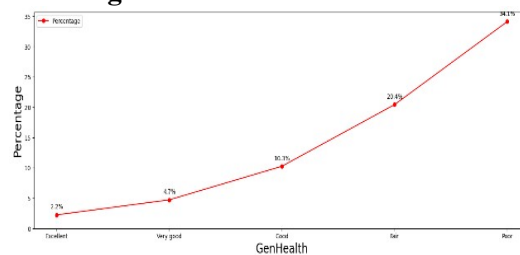
Gambar 27. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Aktivitas Fisik

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa persentase responden yang melakukan aktivitas fisik atau olahraga yang selama 30 hari terakhir, selain dari pekerjaan tetapnya adalah 7,1%. Kemudian, persentase responden yang tidak melakukan aktivitas fisik atau olahraga yang selama 30 hari terakhir, selain dari pekerjaan tetapnya adalah 13,8%. Persentase tersebut menunjukkan bahwa responden yang tidak melakukan aktivitas fisik memiliki risiko penyakit jantung sebesar 1,94 kali lebih tinggi dibandingkan dengan responden yang beraktivitas fisik (olahraga). Maka dari itu, dapat disimpulkan bahwa aktivitas fisik menjadi salah satu faktor yang memiliki pengaruh yang cukup signifikan dalam risiko penyakit jantung.

Hubungan antara penyakit jantung dengan kesehatan secara umum



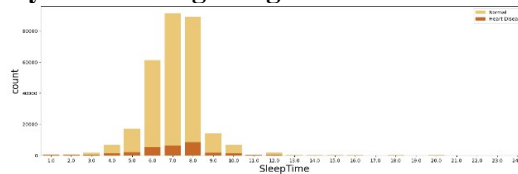
Gambar 28. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Kesehatan Secara Umum



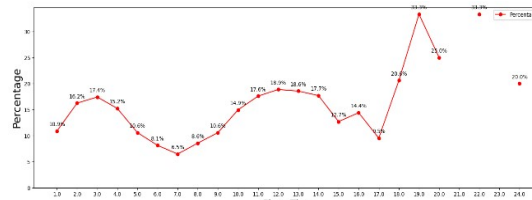
Gambar 29. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Kesehatan Secara Umum

Berdasarkan grafik batang di atas, responden dengan kesehatan secara umum yang sangat baik sekali, sangat baik, dan baik memiliki perbedaan angka yang cukup signifikan antara responden yang normal dan mengalami penyakit jantung. Secara keseluruhan, responden normal memiliki jumlah yang lebih tinggi dibandingkan responden yang mengalami penyakit jantung dengan kondisi kesehatan secara umum sangat baik sekali, sangat baik, baik, sedang/cukup, dan buruk. Namun, pada responden dengan kesehatan yang sedang/cukup dan buruk memiliki perbandingan jumlah yang kecil antara responden normal dan responden yang mengalami penyakit jantung. Selain itu, dapat dilihat juga bahwa persentase responden berdasarkan kondisi kesehatan secara umumnya meningkat seiring dengan buruknya kesehatan secara umum yang dimiliki oleh responden. Apabila diurutkan secara persentase responden yang mengalami penyakit jantung berdasarkan kesehatan secara umum, responden yang memiliki kesehatan secara umum sangat baik sekali memiliki persentase yang paling rendah (2,2%). Kemudian, diikuti oleh responden dengan kesehatan secara umum yang sangat baik (4,7%), baik (10,3%), sedang/cukup (20,4%), dan persentase yang paling tinggi adalah kesehatan secara umum yang buruk (34,1%). Maka dari itu, dapat disimpulkan bahwa kesehatan secara umum memiliki pengaruh yang signifikan pada risiko penyakit jantung. Semakin buruk kesehatan secara umum seseorang maka, semakin tinggi risiko seseorang mengalami penyakit jantung. Sebaliknya, semakin baik kesehatan secara umum seseorang maka, semakin rendah risiko seseorang mengalami penyakit jantung.

Hubungan Antara Penyakit Jantung Dengan Rata-Rata Waktu Tidur



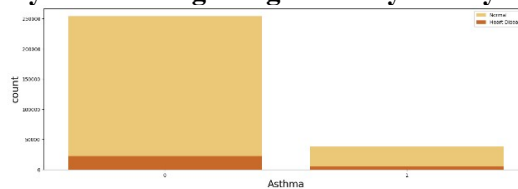
Gambar 30. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung



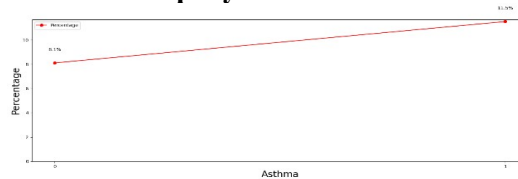
Gambar 31. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Waktu Tidur

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa rata-rata waktu tidur responden sangatlah beragam. Rata-rata waktu tidur 7 jam memiliki persentase risiko penyakit jantung yang paling rendah (6,5%), sedangkan rata-rata waktu tidur dibawah 7 jam maupun diatas 7 jam memiliki persentase yang lebih tinggi. Maka dari itu, dapat dianalisis bahwa rata-rata waktu tidur 7 jam merupakan rata-rata waktu tidur yang paling ideal pada penelitian ini. Namun, terdapat beberapa data yang janggal, yaitu rata-rata waktu tidur lebih dari 13 jam cenderung memiliki persentase yang lebih tinggi. Hal ini kemungkinan besar terjadi karena jumlah data yang cenderung sangat sedikit pada bagian rata-rata waktu tidur tersebut sehingga apabila responden dengan rata-rata waktu tidur lebih dari 13 jam memiliki penyakit jantung, maka persentase yang dihasilkan cenderung lebih tinggi. Selain itu, rata-rata waktu tidur lebih dari 17 jam juga memberikan kesan yang kurang logis sehingga persentase pada bagian tersebut dapat dihiraukan. Maka dari itu, dapat disimpulkan bahwa waktu tidur merupakan salah satu faktor yang memiliki pengaruh yang cukup signifikan dalam risiko penyakit jantung.

Hubungan Antara Penyakit Jantung Dengan Riwayat Penyakit Asma



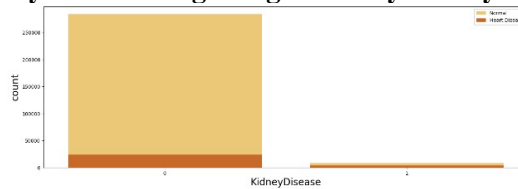
Gambar 32. Grafik batang mengenai hubungan antara penyakit jantung dengan penyakit asma



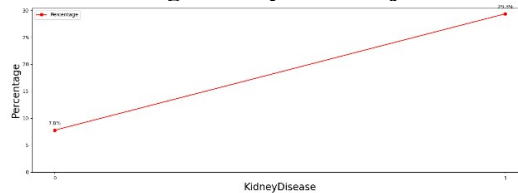
Gambar 33. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Penyakit Asma

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa persentase responden dengan penyakit asma dan memiliki penyakit jantung adalah 11,5%, sedangkan persentase responden tanpa penyakit asma dan memiliki penyakit jantung adalah 8,1%. Persentase tersebut menunjukkan bahwa responden dengan penyakit asma memiliki risiko penyakit jantung sekitar 1,42 kali lebih tinggi dibandingkan dengan responden tanpa penyakit asma. Maka dari itu, dapat disimpulkan bahwa penyakit asma merupakan salah satu faktor yang memiliki pengaruh yang cukup signifikan dalam risiko penyakit jantung.

Hubungan Antara Penyakit Jantung Dengan Riwayat Penyakit Ginjal



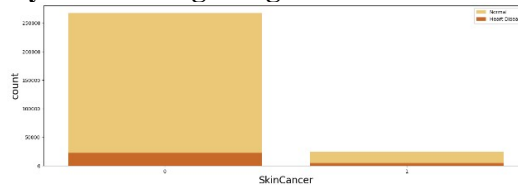
Gambar 34. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Penyakit Ginjal



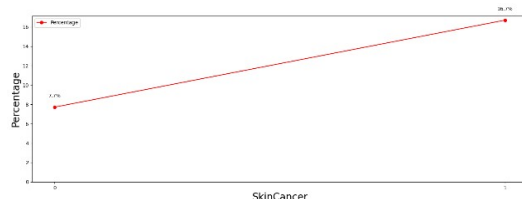
Gambar 35. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Penyakit Ginjal

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa persentase responden dengan penyakit ginjal dan memiliki penyakit jantung adalah 29,3%, sedangkan persentase responden tanpa penyakit ginjal dan memiliki penyakit jantung adalah 7,8%. Persentase tersebut menunjukkan bahwa responden dengan penyakit ginjal memiliki risiko penyakit jantung sekitar 3,78 kali lebih tinggi dibandingkan dengan responden tanpa penyakit ginjal. Maka dari itu, dapat disimpulkan bahwa penyakit ginjal merupakan salah satu faktor yang memiliki pengaruh yang signifikan dalam risiko penyakit jantung.

Hubungan Antara Penyakit Jantung Dengan Kanker Kulit



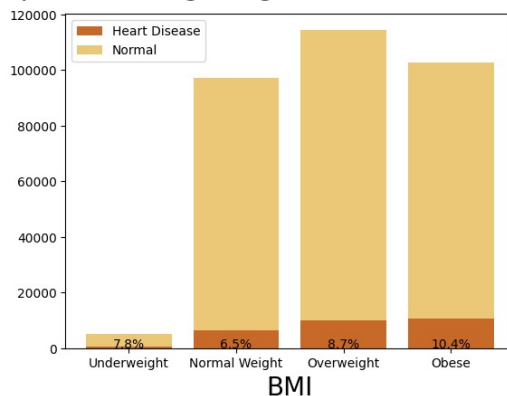
Gambar 36. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Kanker Kulit



Gambar 37. Grafik Persentase Mengenai Hubungan Antara Penyakit Jantung Dengan Kanker Kulit

Berdasarkan pada gambar grafik di atas, dapat dilihat bahwa persentase responden dengan penyakit kanker kulit dan memiliki penyakit jantung adalah 16,7%, sedangkan persentase responden tanpa penyakit kanker kulit dan memiliki penyakit jantung adalah 7,7%. Persentase tersebut menunjukkan bahwa responden dengan penyakit kanker kulit memiliki risiko penyakit jantung sekitar 2,16 kali lebih tinggi dibandingkan dengan responden tanpa penyakit kanker kulit. Maka dari itu, dapat disimpulkan bahwa penyakit kanker kulit merupakan salah satu faktor yang memiliki pengaruh yang signifikan dalam risiko penyakit jantung.

Hubungan Antara Penyakit Jantung Dengan Indeks Massa Tubuh



Gambar 38. Grafik Batang Mengenai Hubungan Antara Penyakit Jantung Dengan Indeks Massa Tubuh

Berdasarkan pada gambar grafik diatas, dapat dilihat bahwa Indeks Massa Tubuh (IMT) responden mempengaruhi risiko penyakit jantung. Grafik diatas menunjukkan responden dengan Indeks Massa Tubuh kurus (7,8%), kegemukan (8,7%), dan obesitas (10,4%) memiliki tingkat risiko penyakit jantung yang lebih tinggi dibandingkan dengan responden yang memiliki Indeks Massa Tubuh yang normal (6,5%). Maka dari itu, dapat disimpulkan bahwa Indeks Massa Tubuh menjadi salah satu faktor yang mempengaruhi risiko penyakit jantung. Semakin ideal suatu Indeks Massa Tubuh, semakin rendah pula risiko penyakit jantung seseorang.

Kesimpulan Dan Saran

Kesimpulan

Berdasarkan hasil implementasi *Exploratory Data Analysis* (EDA) yang telah dilakukan, responden yang pernah merokok (setidaknya 5 bungkus atau 100 batang rokok) memiliki risiko penyakit jantung sekitar 2,02 kali lebih tinggi dibandingkan

dengan responden yang tidak pernah merokok (setidaknya 5 bungkus atau 100 batang rokok). Frekuensi meminum alkohol merupakan faktor yang memiliki pengaruh terhadap risiko penyakit jantung. Responden dengan riwayat penyakit stroke memiliki risiko penyakit jantung sekitar 4,87 kali lebih tinggi dibandingkan dengan responden tanpa riwayat penyakit stroke. Responden dengan jenis kelamin pria memiliki risiko penyakit jantung sekitar 1,59 kali lebih tinggi daripada Wanita. Responden yang memiliki kesehatan secara umum sangat baik sekali memiliki persentase yang paling rendah (2,2%) dan persentase yang paling tinggi adalah kesehatan secara umum yang buruk (34,1%). Adanya perbedaan rata-rata kesehatan fisik seseorang yang memiliki penyakit jantung dan dengan yang tidak memiliki penyakit jantung menunjukkan bahwa Kesehatan fisik responden yang meliputi penyakit dan cedera memiliki korelasi dengan penyakit jantung. Indeks masa tubuh responden yang ideal memiliki risiko penyakit jantung yang lebih rendah maka pentingnya menjaga indeks masa tubuh agar tetap ideal untuk mencegah terjadinya penyakit jantung. Aktivitas fisik diluar dari pekerjaan tetap yang dilakukan responden menunjukkan risiko penyakit jantung yang lebih rendah dibandingkan dengan responden yang tidak melakukannya sekalipun selama 30 hari terakhir. Responden dengan usia 80 atau keatas memiliki persentase jumlah penyakit jantung yang lebih tinggi. Responden dengan penyakit diabetes memiliki risiko penyakit jantung yang lebih tinggi sebesar 3,16 kali dibandingkan dengan responden tanpa penyakit diabetes. Responden dengan rata-rata waktu tidur 7 jam memiliki risiko penyakit jantung yang lebih rendah dibandingkan dengan responden dengan rata-rata waktu tidur dibawah dan diatas 7 jam. Responden dengan penyakit asma memiliki risiko penyakit jantung lebih tinggi sebesar 1,42 dibanding dengan responden tanpa penyakit asma. Responden Amerika Serikat terutama yang memiliki ras *American Indian/Alaskan Native* memiliki persentase penyakit jantung yang 3,15 kali lebih tinggi dibandingkan dengan ras *Asian*. Responden yang memiliki jumlah hari dengan kesehatan mental yang lebih buruk dalam 30 hari terakhir juga mempunyai risiko penyakit jantung yang lebih tinggi. Responden yang memiliki kesulitan menaiki tangga juga mempunyai risiko penyakit jantung yang 3,59 kali lebih tinggi dibandingkan dengan responden yang tidak kesulitan menaiki tangga. Responden dengan penyakit ginjal memiliki risiko penyakit jantung sekitar 3,78 kali lebih tinggi dibandingkan dengan responden tanpa penyakit ginjal. Responden dengan penyakit kanker kulit memiliki risiko penyakit jantung sekitar 2,16 kali lebih tinggi dibandingkan dengan responden tanpa penyakit kanker kulit.

Maka dari itu, dapat disimpulkan bahwa penyakit stroke, kesulitan menaiki tangga, penyakit diabetes, kesehatan fisik, penyakit ginjal, merokok, penyakit kanker kulit, jenis kelamin, usia, Indeks Massa Tubuh (IMT), penyakit asma, aktivitas fisik, kesehatan mental, rata-rata waktu tidur, kesehatan secara umum, dan ras merupakan faktor risiko penyakit jantung. Kemudian, faktor yang berpengaruh paling besar dalam terjangkitnya penyakit jantung adalah penyakit stroke. Hal ini dibuktikan dari perbandingan persentasenya yang lebih tinggi dibandingkan dengan perbandingan persentase faktor lainnya, yaitu sekitar 4,87 kali lebih tinggi.

Saran

Dari penelitian yang telah dilakukan, saran yang dapat diambil adalah pentingnya menjaga kesehatan dan menerapkan pola hidup sehat untuk mengurangi

risiko penyakit jantung. Terdapat beberapa cara untuk menjaga kesehatan dengan menerapkan pola hidup sehat seperti, olahraga, makan makanan bergizi dan berserat, tidur cukup, tidak merokok, mengurangi konsumsi minuman beralkohol dan minuman berkadar gula tinggi, dan masih banyak lagi. Dari kebiasaan-kebiasaan pola hidup sehat tersebut, secara tidak langsung kesehatan fisik dan mental dapat terjaga. Kemudian, salah satu hal yang bisa menjadi indikasi bahwa kita sudah menerapkan pola hidup sehat adalah indeks massa tubuh (IMT) kita yang ideal. Dengan demikian, kita dapat meminimalisir risiko penyakit jantung.

Daftar Pustaka

- Behrens, J. T. (1997). Principles and Procedures of Exploratory Data Analysis. In *Psychological Methods*, 2(2), 131-160. <https://doi.org/10.1016/B978-0-08-044894-7.01327-0>
- Bogdanchikov, A., Zhaparov, M., & Suliyev, R. (2013). Python to learn programming. *Journal of Physics: Conference Series*, 423(1). <https://doi.org/10.1088/1742-6596/423/1/012027>
- Centers for Disease Control and Prevention. (2023). *Heart Disease Facts*. Diperoleh dari <https://www.cdc.gov/heartdisease/facts.htm>
- Dhruv, A. J., Patel, R., & Doshi, N. (2022). Python: The Most Advanced Programming Language for Computer Science Applications. *Proceedings of the International Conference on Culture Heritage, Education, Sustainable Tourism, and Innovation Technologies*, 1(), 292-299. <https://doi.org/10.5220/0010307900003051>
- Fryar, C. D., Chen, T.-C., & Li, X. (2012). Prevalence of uncontrolled risk factors for cardiovascular disease: United States, 1999–2010. NCHS data brief, no 103. Hyattsville, MD: National Center for Health Statistics.
- McKinney, W. (2022). *Python for Data Analysis, 3rd Edition*. Sebastopol, California: O'Reilly Media.
- National Center for Health Statistics. (2022). *Multiple Cause of Death by Single Race 2018-2021 on CDC WONDER Online Database*. Diperoleh dari <https://wonder.cdc.gov/wonder/help/mcd-expanded.html>
- National Center for Health Statistics. (2023). *Percentage of angina for adults aged 18 and over, United States, 2019–2021*. National Health Interview Survey. Diperoleh dari https://wwwn.cdc.gov/NHISDataQueryTool/SHS_adult/index.html
- New York State. (2022). *Heart Disease and Stroke Prevention*. Diperoleh dari https://www.health.ny.gov/diseases/cardiovascular/heart_disease/#:~:text=About%20697%2C000%20people%20die%20of,killing%20approximately%20382%2C820%20people%20annually.
- Radhi, M., Amalia, Ryan Hamonangan Sitompul, D., Hamonangan Sinurat, S., & Indra, E. (2021). ANALISIS BIG DATA DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA) DAN METODE VISUALISASI MENGGUNAKAN JUPYTER NOTEBOOK. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima*, 4(2). <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475>.