

Implementasi Algoritma *K-Means* Untuk Pengelompokan Pada Penyakit Menular Manusia Di Kabupaten Karawang

1st Yoga NugrahaUniversitas Buana Perjuangan Karawang
Karawang, Indonesiaif18.yoganugraha@mhs.ubpkarawang.ac.id2nd Amril Mutoi SiregarUniversitas Buana Perjuangan Karawang
Karawang, Indonesiaamrilmutoi@ubpkarawang.ac.id3rd RahmatUniversitas Buana Perjuangan Karawang
Karawang, Indonesiarahmat@ubpkarawang.ac.id

Abstract— Penyakit tuberkulosis, diare, kusta, dan pneumonia dapat dikategorikan sebuah penyakit menular yang patut diwaspadai. Terdapat permasalahan penyakit menular di daerah Kabupaten Karawang, hal ini ditandai dengan adanya kenaikan jumlah penderita penyakit menular pada setiap tahun nya. Hal ini ditandai penderita pneumonia dengan jumlah balita di Kabupaten Karawang sebanyak 239.792 perkiraan jumlah penderita penyakit pneumonia 10.133, jumlah kasus tuberkulosis anak tahun 2020 mengalami kenaikan yang cukup signifikan jumlah kasus dari 240 kasus menjadi 856 kasus, sedangkan jumlah pengidap diare yang ditemui serta ditangani Karawang tahun 2020 sebanyak 39.035 pengidap untuk seluruh usia ataupun 61%. Maka dari itu penelitian kali ini bertujuan untuk mengelompokkan penyakit menular disetiap puskesmas Karawang ke dalam beberapa kelompok menggunakan algoritma *k-means clustering*. Dari hasil penelitian Algoritma *K-Means Clustering* dalam melakukan pengelompokan diperoleh 1 data puskesmas pada *cluster 1*, 4 data puskesmas pada *cluster 2*, 45 data puskesmas pada *cluster 3* dari 50 data puskesmas. Adapun hasil evaluasi menggunakan metode *Elbow* untuk mencari nilai *K* terbaik, dimana hasil *K*=3 merupakan jumlah *cluster* terbaik. Kemudian hasil tersebut dievaluasi menggunakan *SSE (Sum of Square Error)*. Hasil evaluasi menggunakan *SSE (Sum of Square Error)* tingkat akurasi yang didapat, yaitu 1.949. Nilai jarak dari 2 *cluster* ke 3 *cluster* tersebut merupakan nilai yang mengalami penurunan paling signifikan atau paling besar sehingga 3 *cluster* merupakan *cluster* terbaik.

Kata kunci — Penyakit menular, Clustering, Elbow, SSE (Sum of Square Error).

I. PENDAHULUAN

Kemajuan teknologi telah terus menjadi tumbuh pesat di tiap lini kehidupan. Berbagai informasi yang didapatkan oleh teknologi data yang terus jadi mutakhir, mulai dari bidang industri, ekonomi, ilmu serta teknologi dan banyak sekali bidang kehidupan yang lain. Penerapan teknologi informasi dalam dunia kesehatan pula bisa membuat data yang berlimpah mengenai penyakit menular pada manusia. Kesehatan ialah menimpa yang berharga buat manusia sebab siapa saja bisa hadapi hambatan kesehatan, begitu pula pada manusia yang sangat rentan bermacam berbagai penyakit namun penyebabnya tidak bisa kita sadari[2]. Hambatan yang memunculkan sulitnya melaksanakan konsultasi penyakit oleh dokter disaat ini bisa diatasi dengan terdapatnya program komputer. Dalam Mengenai ini, data informasi dapat membantu pemecahan kasus terhadap penyakit dengan diberikan nasihat kepada pembaca serta memperoleh pemecahan terhadap berbagai kasus (Ade Bastian, et al., 2018).

Kesehatan ialah sesuatu keadaan supaya jiwa serta raga seorang, sehingga menjadikan seorang bisa produktif dalam melaksanakan seluruh kegiatan tiap hari. Bagi Undang- Undang No 23 Tahun 1992[3], Kesehatan ialah kondisi sejahtera dari tubuh, jiwa, serta sosial yang membolehkan tiap orang hidup produktif secara sosial dan murah[4]. Pada pesan edaran Menteri PPN/ Bappenas Nomor. B. 899/ Meter. PPN/ Ses/ PP. 03. 02/ 12/ 2019 bertepatan pada 20 Desember 2019 guna menunjang kenaikan mutu insan manusia di Indonesia, Kementerian Kesehatan memiliki misi yakni menguatkan upaya Kesehatan yang bermutu dan juga menjangkau segala penduduk Indonesia, memberdayakan rakyat dan mengutamakan pembangunan kesehatan dan mempertinggi pelayanan ketersediaan, pemerataan dan kualitas asal energi kesehatan. Hingga upaya pemerintah maksudnya dengan menaikkan pengendalian penyakit dan mempertinggi pelayanan kesehatan yang mencakup penguatan sistem kesehatan, pengawasan obat dan makanan (Kementerian Kesehatan, 2020).

Informasi Riskesdas (2007) mengatakan jika penyakit pneumonia menduduki peringkat kedua selaku pemicu kematian balita 23, 8% serta bayi 15, 5%. Bagi data Studi Kesehatan Dasar (Riskesdas) 2019 prevalensi pneumonia di Indonesia sebesar 2%. Jumlah balita di Kabupaten Karawang sebanyak 239.792 perkiraan jumlah penderita penyakit pneumonia 10.133. Jumlah kasus tuberkulosis anak tahun 2020 mengalami kenaikan yang cukup signifikan jumlah kasus dari dari 240 kasus menjadi 856 kasus. hal ini sama dengan lonjakan kasus tuberkulosis untuk orang dewasa dimana kasus ini terjadi karena adanya tambahan lapoaran kasus tuberkulosis anak oleh pihak rumah sakit yang ada di kabupaten Karawang yang tahun sebelumnya data hanya diambil dari 50 Puskesmas. Oleh karena itu butuh kepedulian yang lebih dari semua unsur yang terkait agar kasus tuberkulosis pada anak tidak semakin tinggi. Sementara itu guna Proporsi penderita tuberkulosis Anak diantara segala penderita tuberkulosis merupakan presentase penderita tuberkulosis anak (<15 tahun) diantara segala penderita tuberkulosis tercatat. Jumlah pengidap diare yang ditemui serta ditangani Karawang tahun 2020 sebanyak 39.035 pengidap buat seluruh usia ataupun 61%. Sementara itu guna permasalahan diare pada bayi dari sasaran pneumonia sebesar 39.966 dapat dilayani sebanyak 16.452 balita atau 41.2 %.

Sedangkan jumlah mendapatkan oralit untuk semua umur sebanyak : 12.052 atau 30,9 % dan untuk anak Balita sebanyak 4.056 atau 24,7 %. Maka dari itu dari hasil data diatas menunjukkan angka yang relatif mengkhawatirkan jika tidak adanya perhatian khusus dari pemerintah.

Adapun penelitian dalam pengelompokan data menggunakan metode *clustering* diantaranya oleh Ade Bastian dkk (2018) tentang pelaksanaan Algoritma *K-Means Clustering Analysis* pada penyakit menular manusia riset permasalahan Kabupaten Majalengka. Hasil riset yang diperoleh kalau Algoritma K-Means yakni salah satu tata metode data clustering non hirarki yang berupaya mempartisi data yang ada ke dalam bentuk satu maupun lebih cluster maupun kelompok, dari tata metode Algoritma K-Means Clustering data mining dapat digunakan buat pengendalian persediaan pada Puskesmas Pandanaran, sehingga apabila dicoba pengadaan persediaan obat pada tahun 2014, petugas dapat memandang catatan Puskesmas sangat banyak yang menderita penyakit menular[5]. Riset yang kedua dicoba oleh Dea Amelia dkk (2022) menimpa pengelompokan penyakit demam berdarah dengue (DBD) di Jawa Barat memakai Algoritma *K-Means* tata cara *Elbow*. Hasil penelitian yang dilakukan terdapat hasil Algoritma *K-Means Clustering* dengan menggunakan metode *Elbow* yang mampu menghasilkan 2 *cluster* optimal dalam pengelompokan penyakit demam berdarah dengue (DBD) di Jawa Barat pada tahun 2016 hingga 2021. Hasil *clustering* terbagi menjadi 2 *cluster* yaitu *cluster* 0 dengan kategori rendah yang terdiri dari 22 daerah kabupaten atau kota, dan *cluster* 1 dengan kategori tinggi yang terdiri dari 5 daerah kabupaten atau kota[6].

Berdasarkan riwayat penelitian yang telah dilakukan sebelumnya, penelitian ini bertujuan untuk pengelompokan data penyakit menular pada manusia di Kabupaten Karawang menggunakan algoritma *k-means clustering*. Metode evaluasi yang digunakan yaitu *Elbow* buat memastikan jumlah *cluster* terbaik dengan metode melihat presentase hasil perbandingan antara jumlah *cluster* yang hendak membentuk siku pada sesuatu titik. Buat memperoleh perbandingannya yakni dengan menghitung *SSE* (*Sum of Square Error*) dari tiap-tiap nilai *cluster* K maka nilai *SSE* akan semakin kecil (Irwanto, et al 2012). Hasil penelitian ini diharapkan untuk pengendalian persediaan obat pada Dinas Kesehatan Kabupaten Karawang.

II. METODOLOGI PENELITIAN

A. Data Mining

Data mining merupakan proses mendapatkan kolerasi, pola, serta trend baru yang bermanfaat dengan menambang beberapa repositori informasi secara massal, memakai teknologi semacam statistik serta metode semacam matematika. Informasi mining menyebar serta akhir- akhir ini tumbuh pesat, sebab kemampuannya buat menambang pola serta trend yang bermanfaat dari database yang telah ada[8]. Industri mempunyai miliaran dihabiskan buat pengumpulan informasi dalam megabyte ataupun terabyte namun tidak memperoleh fitur berharga di dalamnya, walaupun memiliki data berharga namun tersembunyi di repositori informasi (Larose, 2005).

B. Algoritma K-Means

Algoritma K-Means termasuk sebagai pengelompokan non hirarki, yang mana mampu membagi objek satu kelompok atau ke beberapa kelompok lainnya. Algoritma K-Means ialah salah satu algoritma yang bisa mengelompokan sesuatu informasi mejadi bermacam kelompok, dimana tiap kelompok memiliki ciri yang berbeda dengan kelompok lain, berikut merupakan langkah-langkah melaksanakan clustering dengan memakai tata cara k-means.

1. Tentukan banyaknya *cluster* k.
2. Tentukan titik pusat atau *centroid* dari masing-masing *cluster* secara acak atau *random*.
3. Pada setiap data dihitung jarak terdekat terhadap *centroid* menggunakan rumus Euclidean Distance.

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Keterangan:

D(i,j)= Jarak data i ke pusat *cluster* j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

4. Kemudian bentuk *cluster* baru berdasarkan jarak minimum terhadap *cluster*.
5. Ulang tahap 3 sampai anggota *cluster* tidak beralih ke *cluster* lainnya maka iterasi tersebut berhenti.

C. Penyakit Menular

Penyakit menular yakni penyakit yang dibuat oleh bakteri, virus, maupun parasit yang dapat ditularkan melalui media tertentu[12]. Penyakit menular ialah hasil dari perpaduan berbagai aspek berbeda yang bersama pengaruhi (Widoyono, 2008).

D. Metode *Elbow*

Bagi Madhulatha(2012) dalam Merliana, Ernawati serta Santoso (2015) disebutkan bila tata cara *elbow* ialah suatu tata cara yang digunakan guna menghasilkan data dalam membenarkan jumlah *cluster* terbaik dengan tata cara memandang presentase hasil perbandingan antara jumlah *cluster* yang hendak membentuk siku pada suatu titik[13]. Apabila nilai *cluster* dini dengan nilai *cluster* kedua membagikan sudut dalam grafik maupun nilainya natural pengurangan sangat besar sehingga nilai *cluster* tersebut yang terbaik. Buat memperoleh perbandingannya yakni dengan menghitung *SSE* dari tiap-tiap nilai *cluster* K hingga nilai *SSE* hendak terus menjadi kecil[14]. Rumus *SSE* pada *k-means* merupakan selaku persamaan berikut: (Irwanto, et al 2012)

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|^2$$

(2)

Keterangan:

K = jumlah *cluster*

xi = data ke i

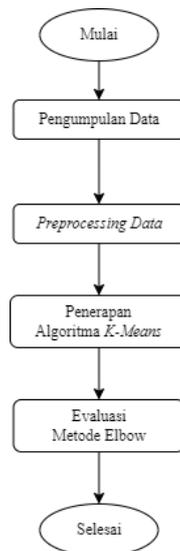
Ck = *centroid cluster*

E. Evaluasi

Pada sesi ini dicoba penilaian terhadap kelompok maupun *cluster* yang ada dengan metode melakukan optimasi jumlah *cluster*[15]. Dari sesi ini hendak dikenal jumlah *cluster* terbaik dengan mengenakan tata cara *elbow* yang mendapatkan perbandingan jumlah *cluster* dengan tata cara menghitung *SSE* (*Sum of Square Error*) dari tiap- masing- masing nilai *cluster*.

F. Prosedur Penelitian

Prosedur penelitian dimulai dengan pengumpulan data dengan rincian sebagai berikut:



Gambar 1. Prosedur Penelitian

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Sumber data utama yang diperoleh dari Dinas Kesehatan Kabupaten Karawang berbentuk file sehingga memudahkan dalam proses seleksi data atau cleaning data. Data tersebut yaitu tabel dengan kriteria data meliputi jumlah Kecamatan dengan jumlah keseluruhan 30 Kecamatan dan jumlah Puskesmas dengan jumlah keseluruhan 50 Puskesmas. Data tersebut digunakan untuk dataset penelitian yang terlampir pada tabel 1.

B. *Preprocessing Data*

Dalam penelitian ini tidak semua data penyakit menular yang terdapat dalam data dasar Dinas Kesehatan tahun 2020 digunakan. Data yang diambil terdapat beberapa penyakit menular yang terjangkit di Kabupaten Karawang. Dalam dataset ini terdapat 6 atribut, yaitu kecamatan, puskesmas, jumlah penyakit pneumonia, jumlah penyakit tuberkulosis, jumlah penyakit diare, jumlah penyakit kusta. Namun yang digunakan hanya 5 atribut untuk perhitungan manual dan python, yaitu puskesmas, jumlah penyakit pneumonia,

jumlah penyakit tuberkulosis, jumlah penyakit diare, jumlah penyakit kusta. Perhitungan manual dan python atribut-atribut yang akan digunakan sebagai berikut:

Tabel 1 Data yang akan digunakan

Puskesmas	Jumlah Penyakit Pneumonia	Jumlah Penyakit Tuberkulosis	Jumlah Penyakit Diare	Jumlah Penyakit Kusta
Pangkalan	4	42	193	0
Loji	23	66	74	0
Ciampel	14	32	170	1
Teluk Jambe	101	50	678	1
Wadas	97	49	600	4
Wanakerta	0	112	1.502	1
Klari	1.505	195	8.496	1
Anggadita	46	32	127	1
Curug	8	15	97	0
Cikampek	149	140	3.845	2
...
Pakisjaya	0	15	2.570	0

Sumber : Dinas Kesehatan Kabupaten Karawang

C. Penerapan Algoritma *K-Means*

1. Perhitungan Manual

Pada proses perhitungan *K-Means* dengan pembentukan $K=3$ menggunakan *excel* yang dilakukan 3 kali iterasi menghasilkan *cluster 3* (rendah) berjumlah 45 puskesmas, *cluster 2* (sedang) berjumlah 4 puskesmas, dan *cluster 1* (tertinggi) berjumlah 1 puskesmas.

2. Program *Python*

Berdasarkan perhitungan algoritma *K-Means* melalui sistem dengan menggunakan aplikasi *Jupyter Notebook* untuk menentukan *cluster* sebanyak $K=3$. Hasil *clustering* dapat dilihat pada gambar berikut:

```
In [32]: df3["Cluster"] = kmeans.labels_
df3

Out[32]:
```

	Pneumonia	Tuberkulosis	Diare	Kusta	Cluster
0	0.002658	0.211340	0.015532	0.000000	1
1	0.015282	0.335052	0.001423	0.000000	1
2	0.009302	0.159794	0.012805	0.071429	1
3	0.067110	0.252577	0.073038	0.071429	1
4	0.064452	0.247423	0.063789	0.285714	1
5	0.000000	0.572165	0.170737	0.071429	1
6	1.000000	1.000000	1.000000	0.071429	2
7	0.030565	0.159794	0.007707	0.071429	1
8	0.005316	0.072165	0.004150	0.000000	1
9	0.099003	0.716495	0.448542	0.142857	1
10	0.000000	0.453608	0.128646	1.000000	0
11	0.106977	0.231959	0.142874	0.000000	1
12	0.083056	0.768041	0.000000	0.142857	1
13	0.009302	0.144330	0.004743	0.142857	1
14	0.044518	0.030928	0.018141	0.214286	1
15	0.003322	0.123711	0.023358	0.071429	1
16	0.043854	0.190722	0.103865	0.071429	1

Gambar 1. Hasil Clustering

```
In [33]: df3['Cluster'].value_counts()

Out[33]:
```

1	45
0	4
2	1

Name: Cluster, dtype: int64

Gambar 2. Jumlah data percluster

D. Evaluasi

Setelah melakukan perhitungan manual selanjutnya adalah melakukan evaluasi dengan menggunakan metode *Elbow*. Dalam metode ini kita akan mencari nilai *cluster* yang paling optimum dengan menggunakan metode *SSE (Sum Of Square Error)*. Berikut adalah hasil evaluasi menggunakan metode *Elbow*:

1. Perhitungan *SSE* Manual

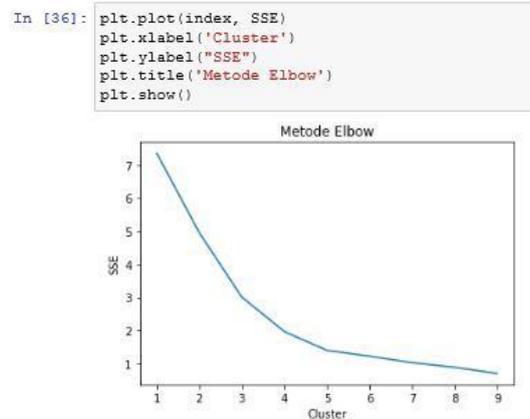
Perhitungan manual evaluasi *Sum Of Squared Error* terdapat nilai jarak *SSE* dari 2 *cluster* ke 3 *cluster* pada *elbow*, yaitu 1.949. Nilai jarak dari 2 *cluster* ke 3 *cluster* adalah nilai yang mengalami penurunan paling signifikan, sehingga 3 *cluster* adalah *cluster* terbaik. Dapat dilihat pada tabel 2.

Tabel 2. Hasil *SSE*

Cluster	Hasil Sum of Square Error	Selisih
K2	4.951	4.951
K3	3.002	1.949
K4	1.961	1.041
K5	1.395	0.566
K6	1.219	0.176
K7	1.022	0.217
K8	0.883	0.139
K9	0.696	0.187

2. Program Python

Berdasarkan evaluasi *elbow* menggunakan sistem didapatkan hasil *elbow* dan *SSE* seperti berikut.



Gambar 3. Grafik Metode *Elbow*

```
In [28]: SSE = []
index = range(1,10)
for i in index:
    kmeans = KMeans(n_clusters=i, random_state=24)
    kmeans.fit(df3)
    SSE.append(kmeans.inertia_)
    print(kmeans.inertia_)
```

```
7.354617527965813
4.951412448160909
3.002046072098161
1.9608625465871652
1.3948265252869196
1.21878009834195
1.0222402095409155
0.8830767532639672
0.6957654201670246
```

Gambar 4. *SSE (Sum Of Square Error)*

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian menghasilkan 3 *cluster*, yaitu tinggi, sedang, dan rendah terhadap penyakit menular manusia. Hasil algoritma *k-means* memperoleh hasil *cluster* 1 dengan kategori tinggi memiliki 1 puskesmas, *cluster* 2 dengan kategori sedang memiliki 4 puskesmas, *cluster* 3 dengan kategori rendah memiliki 45 puskesmas. Penerapan algoritma berdasarkan sistem yang dibuat dengan menggunakan bahasa pemrograman *python* mendapatkan hasil yang sama dengan perhitungan manual. Dari hasil evaluasi terdapat nilai *SSE* dari 2 *cluster* ke 3 *cluster* pada *elbow*, ialah 1.949. Nilai jarak dari 2 *cluster* ke 3 *cluster* tersebut ialah nilai yang alami pengurangan sangat signifikan ataupun sangat besar sehingga 3 *cluster* ialah *cluster* terbaik.

PENGAKUAN

Naskah ilmiah ini merupakan bagian dari penelitian Tugas Akhir milik Yoga Nugraha yang berjudul Penerapan Algoritma *K-Means Clustering* Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Karawang), yang dibimbing oleh Amril Mutoi Siregar, M.Kom dan Rahmat, M.Pd.

DAFTAR PUSTAKA

- [1] Amelia, Dea, Tesa Nur Padilah, dan Asep Jamaludin. (2022). *Optimasi Algoritma K-Means Menggunakan Metode Elbow dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD)*.
- [2] Arijaya, Ratri, Imam Cholissodin, dan Arief Soebroto. (2020). *Implementasi Metode Rough Set Dan Certainty Factor Untuk Deteksi Dini Penyakit Menular Seksual (Studi Kasus: Puskesmas Dinoyo Kota Malang)*. 1-14.
- [3] Fatmawati, Kiki dan Agus Perdana Windarto. (2018). *Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi*. *Computer Engineering, Science and System Journal*, 3.2, 173.
- [4] Fauzi, Muhammad dan Yudi Yudi. (2017). *Penerapan Algoritma K-Means Clustering Untuk Mendeteksi Penyebaran TBC (Studi Kasus: Di Kabupaten Deli Serdang)*. *JTIK (Jurnal Teknik Informatika Kaputama)*, 1.2, 1-7.
- [5] Kurniasari, Dita, Asroni, dan Slamet Aprilia Kurniasari. (2020). *The Implementation Of Clustering Method With K-Means Algorithm In Grouping Data Of Students Course Scores At Universitas Muhammadiyah Yogyakarta*.
- [6] Madhulatha, T.S. (2012). *An Overview On Clustering Methods*. *IOSR Journal Of Engineering*.
- [7] Nadilla Yuzar Dinda. (2020). *Penyakit Menular Dan Wabah Penyakit Covid-19*.
- [8] Purwaningrum, Oktania, Yudha Yunanto Putra, dan Amalia Anjani Arifiyanti. (2021). *Penentuan Kelompok Status Gii Balita Dengan Menggunakan Metode K-Means*.

- [9] Rustam, Suhardi, Heru Agus Santoso, dan Catur Supriyanto. (2018). *Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang. ILKOM Jurnal Ilmiah*, 10.3), 251-59.
- [10] Siregar, Amril Mutoi dan Adam Puphabuana. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan Rapidminer*.
- [11] Siregar, Amril Mutoi, Maulana Abdur Rofik, dan Dwi Sulisty Kusumaningrum. (2021). *Perbandingan Tingkat Kepuasan Siswa Terhadap Pelayanan Sekolah Menggunakan Algoritma K-Means dan K-Medoids*.
- [12] Sundari, Shinta Siti, dan Nida Ariani. (2019). *Penerapan Data Mining Untuk Pengelompokan Penyakit Dengan Algoritma Fuzzy C-Means (Studi Kasus: UPT Puskesmas Salawu)*. *Jurnal VOI (Voice of Informatics)*, 8.2, 63-76.
- [13] Tampubolon, S. S dan K Handoko. (2021). *Penerapan Data Mining Analisa Penyakit Menular Pada Manusia*. *Computer and Science*.