RESEARCH ARTICLE

**Clustering Tuberculosis Menggunakan Algoritma ….. dengan Python Di Makassar Tahun 2017-2019**

Muhammad Resha1, Anshar Suyuti2, Rhiza S Sadjad3, Muhammad Niswar4

1Department of Digital Business, IBK Nitro, Makasxsar, Indonesia
2,3,4 Department of Electrical Engineering, Faculty of Engineering, Hasanuddin University, Makassar, Indonesia

Corresponding author. E-mail: resha@nitromks.ac.id

Abstract

**BACKGROUND** :

**METHODS** :

**RESULTS** :

**CONCLUSION** :

**KEYWORDS** :

Introduction

Tuberculosis atau TB merupakan masalah kesehatan masyarakat yang menjadi tantangan global dan nasional[1]. Indonesia masih merupakan salah satu negara dengan beban TB tertinggi dan merupakan penyakit lama yang masih menjadi pembunuh terbanyak diantara penyakit menular[2].

Meningkatnya kasus TB menunjukkan bahwa ada penularan atau kontak penyakit di antara orang-orang yang terkait erat dalam masyarakat. Penelitian telah menunjukkan bahwa penularan atau kontak penyakit terjadi dari dalam rumah, rumah di sekitarnya, dan lingkungan kerja.[3]

Sumber penularan TB adalah penderita TB BTA positif, yang dapat menularkan kepada orang yang berada di sekelilingnya, terutama kontak erat. Penderita menyebarkan kuman ke udara dalam bentuk percikan dahak ( dropletnuclei) pada saat penderita itu batuk atau bersin. Droplet yang mengandung kuman dapat bertahan di udara pada suhu kamar selama beberapa jam[4]. Orang dapat terinfeksi kalau droplet tersebut terhirup ke dalam saluran napas. Setelah kuman TB masuk ke dalam tubuh manusia melalui pernapasan, ia dapat menyebar dari paru kebagian tubuh lainnya, melalui sistem peredaran darah, sistem saluran getah bening atau menyebar langsung ke bagian-bagian tubuh lainnya.[5]

Berdasarkan data yang diperoleh dari Bidang P2P Dinas Kesehatan Kota Makassar, kasus baru penderita TB Paru BTA (+) di Puskesmas dan Rumah Sakit tahun 2017 yaitu sebanyak 1.951 penderita dari 4117 perkiraan sasaran sehingga diperoleh Angka Penemuan Kasus Baru/Case Detection Rate (CDR) TB BTA (+) yaitu 47,39%. Angka ini menurun dibandingkan tahun 2016 yaitu 1.874 penderita dari 2373 perkiraan sasaran sehingga didapatkan Angka Penemuan Kasus Baru TB BTA (+) yaitu 78,97%.[6] Hal ini dikarenakan target penemuan kasus di tahun 2017 meningkat sehingga berpengaruh terhadap perhitungan perkiraan sasaran. Pada tahun 2015 angka penemuan kasus baru TB BTA(+) yaitu 1.928 penderita dari 2600 perkiraan sasaran sehingga didapatkan Angka Penemuan Kasus Baru TB BTA (+) yaitu 74,15%.

Prevalensi/ Case Notification Rate (CNR) semua kasus TB per 100.000 penduduk selama 3 tahun terakhir juga meningkat yaitu tahun 2017 sebesar 335/100.000 penduduk meningkat dari tahun 2016 yaitu 271/100.000 penduduk dan tahun 2017 yaitu 249/100.000 penduduk.

Methods

Data yang di peroleh dari kurun waktu triwulan I 2017 hingga triwulan III 2019 di proses awal dengan melihat atribut dari data yang memberikan gambaran adanya pola yang akan di cermati[7]. Dalam hal ini, beberapa atribut yang tidak diperlukan di hapus dari kandidat dataset yang akan digunakan. Penghapusan juga di lakukan secara horizontal terhadap data yang tidak mendukung baik dikarenakan data yang tidak lengkap atau kesalahan pada saat penginputan data[8] misalnya dalam status pengobatan terdapat 2 tupple data yang bernilai kosong dan juga memuat data kosong pada beberapa atribut lainnya.

Demikian juga pada hasil pengobatan terdapat data yang bernilai kosong, sementara yang bernilai null disebabkan oleh pengobatan yang baru mulai[9]. Karena yang akan di lihat adalah kondisi terapi maka yang bernilai kosong dan masa mulai terapi bukan dalam tahun 2019, akan dihapus sementara yang null tetap di pertahankan. Kondisi null dan nilai kosong pada hasil pengobatan di isi dengan kode yang sama dengan on-going (sedang berlangsung).

Sementara terdapat 1 data gagal terapi tetapi tidak mengisi tanggal terakhir pengobatan, sehingga di asumsikan pasien menjalani seluruh terapi proses selama 5 bulan. Modifikasi juga dilakukan pada tanggal terakhir terapi, terjadi kesalahan input misalnya 2019 di input sebagai 2009. Atribut hasil akhir mikroskopis juga tidak memiliki data, sehingga di hapus.

Proses transformasi dimana data nominal di jadikan data ordinal dengan mempertimbangkan tingkat perkembangan kejadian[10]. Atribut yang di pilih dalam proses seleksi tidak semuanya mengandung nilai ordinal, sehingga transformasi ini diperlukan seperti pada atribut jenis diagnose, kategori pengobatan dan berhubungan dengan umur, hasil mikroskopis bulan 2,3 dan 5, serta hasil dari pengobatan.

Tabel 1. Hasil Uji Mikroskopis

|  |  |  |
| --- | --- | --- |
| **No** | **Kejadian** | **Nilai** |
| 1 | Negatif | 2 |
| 2 | Pindah | 1 |
| 3 | Tidak dilakukan / kosong | 0 |
| 4 | Putus Berobat | -1 |
| 5 | Positif | -2 |
| 6 | Meninggal | -3 |

Pada Tabel 2 hasil pengobatan dikategorikan dalam 7 Kejadian nilai yang ditentukan sendiri.

Tabel 2. Hasil Pengobatan

|  |  |  |
| --- | --- | --- |
| **No** | **Kejadian** | **Nilai** |
| 1 | Sembuh | 3 |
| 2 | Pengobatan lengkap | 2 |
| 3 | On going / null / kosong | 1 |
| 4 | Pindah | 0 |
| 5 | Putus berobat | -1 |
| 6 | Gagal | -2 |
| 7 | Meninggal | -3 |

Transformasi juga dilakukan pada tanggal mulai berobat dan berakhir berobat, dimana lama pengobatan dihitung. Adapun perubahan hasil uji mikroskopis selama 5 bulan di lakukan perhitungan deviasi (eq 1) pada tiap pasien dimana apabila tidak terjadi perubahan maka didapatkan hasil 0[11].

$s=\sqrt{\frac{\sum\_{i=1}^{N}(x\_{i}-\overbar{x})}{N-1}}$ (eq. 1)

Rumus statistik yang dimanfaatkan dalam menentukan sebaran data dalam sampel dan seberapa dekat titik data individu ke rata-rata sampel. Saat perhitungan deviasi diperoleh nilai sampel data sama dengan nol maka menunjukkan bahwa semua nilai dalam data tersebut merupakan sama. Sehingga di dapatkan keadaan seperti berikut :

Tabel 3. Conoth Perubahan hasil uji mikroskopis

|  |  |  |  |
| --- | --- | --- | --- |
| Bulan 2 | Bulan 3 | Bulan 5 | StD |
| 2 | 0 | 0 | 1.154701 |
| 2 | 0 | 2 | 1.154701 |
| -2 | 0 | 2 | 2 |
| -1 | -1 | -1 | 0 |
| -2 | 0 | -1 | 1 |
| 0 | 0 | -1 | 0.57735 |
| 2 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0.57735 |
| 0 | 0 | 0 | 0 |
| -2 | -2 | 2 | 2.309401 |
| -2 | 2 | 2 | 2.309401 |

Tahap selanjutnya adalah melakukan proses memplot kondisi yang ditemukan berdasarkan umur dan lama terapi dengan indikasi hasil pengobatan, jenis kelamin, kategori pengobatan dan perubahan uji mikroskopis[12]. Data yang diperoleh pada tahap inilah yang akan digunakan.

Tabel 4. Contoh data hasil pre-proses

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gender | Diagnosa type | Kategori | Umur | Lama Terapi | HasilTerapi | STD |
| 1 | 1 | 3 | 65 | 173 | 2 | 1.154701 |
| 0 | -1 | 3 | 22 | 176 | 3 | 1.154701 |
| 1 | 1 | 3 | 60 | 167 | 2 | 1.154701 |
| 0 | -1 | 3 | 65 | 166 | 3 | 1.154701 |
| 0 | -1 | 3 | 43 | 170 | 3 | 2 |
| 0 | -1 | 3 | 27 | 162 | 2 | 1.154701 |
| 0 | -1 | 3 | 42 | 170 | 3 | 1.154701 |
| 1 | 1 | 3 | 19 | 42 | -1 | 0 |
| 0 | 1 | 3 | 20 | 162 | 2 | 1.154701 |
| 1 | 1 | 3 | 72 | 165 | 2 | 1.154701 |
| 1 | -1 | 3 | 54 | 21 | -1 | 1 |

Tabel 3 source code

import matplotlib.pyplot as plt

import csv

Xm=[]

Ym=[]

DiagTypem=[]

Catm=[]

Rsltm = []

Xf=[]

Yf=[]

DiagTypef=[]

Catf=[]

Rsltf=[]

X=[]

Y=[]

Result=[]

Label=[]

with open('D:\PreprosesB.csv') as cs:

 readCSV = csv.reader(cs, delimiter=',' )

 header = next(readCSV)

 for row in readCSV:

 X.append( float(row[4]))

 Y.append( float(row[3]))

 Result.append( int(row[5]))

 if int(row[5])==3:

 Label.append("Sembuh")

 elif int(row[5])==2:

 Label.append("Pengobatan Lengkap")

 elif int(row[5])==1:

 Label.append("On going")

 elif int(row[5])==0:

 Label.append("Pindah")

 elif int(row[5])==-1:

 Label.append("Putus Berobat")

 elif int(row[5])==-2:

 Label.append("Gagal")

 elif int(row[5])==-3:

 Label.append("Meninggal")

 if row[0] == 1:

 Xm.append( float(row[4]) )

 Ym.append( float(row[3]) )

 DiagTypem.append( int(row[1]))

 Rsltm.append(int(row[5])^2)

 Catm.append(int(row[2]))

 else:

 Xf.append( float(row[4]) )

 Yf.append( float(row[3]) )

 DiagTypef.append( int(row[1]) )

 Rsltf.append(int(row[5])^2)

 Catf.append(int(row[2]))

plt.figure(figsize=(12, 12))

plt.subplot(221)

plt.scatter(Xm,Ym,s=Catm,marker='v', c=Rsltm,alpha=0.5)

plt.scatter(Xf,Yf,s=Catf,marker='s', c=Rsltf,alpha=0.5)

plt.xlabel("Lama pengobatan")

plt.ylabel("Umur")

plt.title("Distribusi data")

fig, ax = plt.subplots()

for X,Y,Result, grp in zip(X,Y,Result,Label):

 ax.scatter(X, Y, c=Result, label=grp,

 alpha=0.3, edgecolors='none')

plt.xlabel("Lama pengobatan")

plt.ylabel("Umur")

plt.title("Cluster berdasar Hasil Pengobatan")

plt.show

import matplotlib

matplotlib.axes.Axes.scatter

matplotlib.pyplot.scatter

matplotlib.axes.Axes.legend

matplotlib.pyplot.legend

Results

Pada tahap ini di lakukan plot terhadap kondisi data untuk mengetahui pola penyebarannya.



Gambar Plot Sebaran Data

Dimana pada Gambar 1, hasil terapi di tunjukkan dengan warna. Warna hijau adalah hasil terapi =3 (sembuh)



Gambar Cluster berdasar Kepadatan

Gambar 2 menjelaskan

Discussion

Conclusion

Dengan menggunakan cluster, dapat dengan mudah mengidentifikasi pola keterhubungan dalam *treatment* pengobatan, lama pengobatan dan gender. Banyak pola lainnya yang bisa di hasilkan dengan menampilkan cluster yang berbeda dari atribut yang digunakan.

Acknowledgments

References

[1] A. Zumla, A. George, V. Sharma, R. H. N. Herbert, A. Oxley, and M. Oliver, “The WHO 2014 Global tuberculosis report-further to go,” *The Lancet Global Health*, vol. 3, no. 1. Elsevier Ltd, pp. e10–e12, 01-Jan-2015.

[2] D. Collins, F. Hafidz, and D. Mustikawati, “The economic burden of tuberculosis in Indonesia,” *Int. J. Tuberc. Lung Dis.*, vol. 21, no. 9, pp. 1041–1048, Sep. 2017.

[3] L. Grandjean *et al.*, “Tuberculosis in household contacts of multidrug-resistant tuberculosis patients,” *Int. J. Tuberc. Lung Dis.*, vol. 15, no. 9, pp. 1164–1169, Sep. 2011.

[4] C. Acuña-Villaorduña *et al.*, “Host determinants of infectiousness in smear-positive patients with pulmonary tuberculosis,” *Open Forum Infect. Dis.*, vol. 6, no. 6, Jun. 2019.

[5] E. A. Nardell, “Transmission and institutional infection control of tuberculosis,” *Cold Spring Harb. Perspect. Med.*, vol. 6, no. 2, Feb. 2016.

[6] Dinkes Kotas Makassar, *Profil kesehatan makassar 2017*. 2017.

[7] S.-J. Heo *et al.*, “Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers’ Health Examination Data,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 2, p. 250, 2019.

[8] A. L’Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, “Machine Learning with Big Data: Challenges and Approaches,” *IEEE Access*, vol. 5, pp. 7776–7797, 2017.

[9] O. Elezaj, S. Yildirim, and E. Kalemi, “Data-driven machine learning approach for predicting missing values in large data sets: A comparison study,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10710 LNCS, pp. 268–285.

[10] V. N. Gudivada, J. Ding, and A. Apon, “Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications View project Cognitive Computing: Theory and Applications View project Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations,” 2017.

[11] D. K. Lee, J. In, and S. Lee, “Standard deviation and standard error of the mean,” *Korean J. Anesthesiol.*, vol. 68, no. 3, pp. 220–223, Jun. 2015.

[12] S. G. Schumacher *et al.*, “Diagnostic Test Accuracy in Childhood Pulmonary Tuberculosis: A Bayesian Latent Class Analysis,” *Am. J. Epidemiol.*, vol. 184, no. 9, pp. 690–700, Nov. 2016.