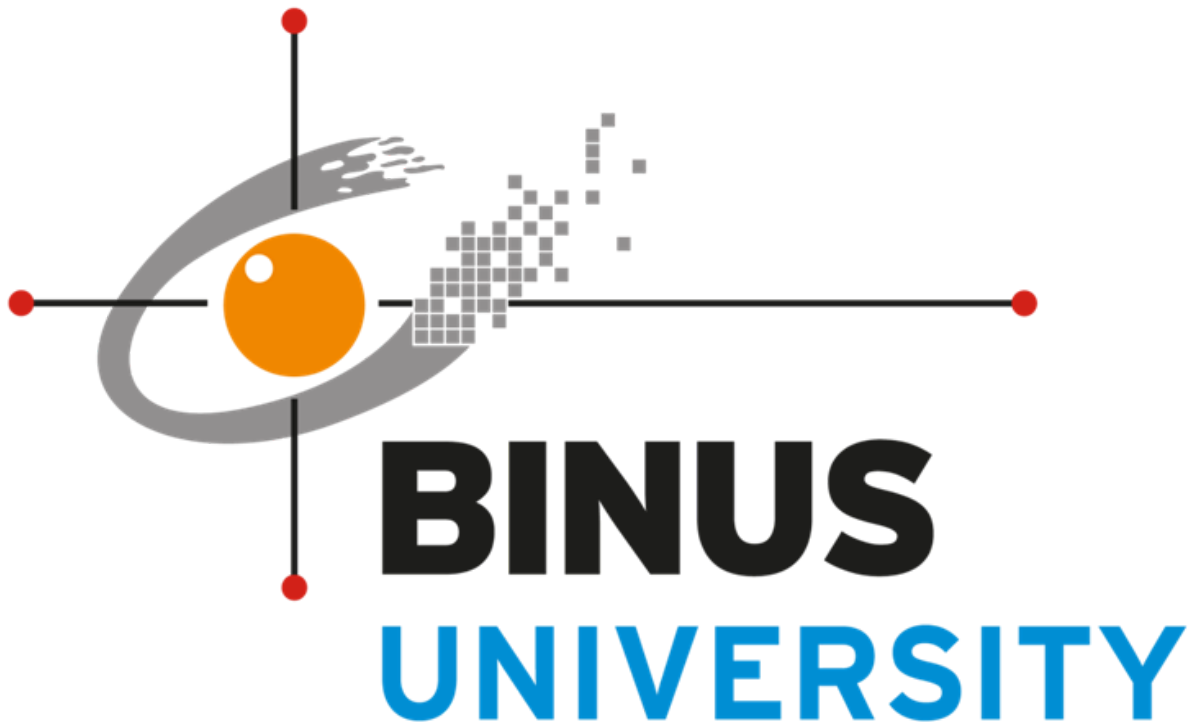


LAPORAN GROUP PROJECT

**Predictive Analytics terhadap Penyakit COVID-19 Berdasarkan Gejala
Menggunakan Algoritma Decision Tree**



Application of Predictive Analytics to Business Data

Kelas: LA11

Kelompok 1

2502047360 - Alif Prianda Bastian

2501960796 - Christopher Edmund Haryanto

2501983755 - Wesley Wilnio

2502024836 - Muhammad Fairuzzaky

2540128190 - Farrell Febrizal Rudy

Latar Belakang

Pandemi Covid-19 mendorong kita untuk mengembangkan model prediksi untuk membedakan apakah seseorang terinfeksi virus atau tidak. Dengan melibatkan , model ini dapat memberikan perkiraan yang lebih akurat terkait penyebaran virus dalam suatu populasi. Tujuan utama dari pembuatan model ini adalah untuk mendeteksi Covid-19 lebih dini dan memungkinkan tindakan pencegahan yang lebih awal serta juga mengurangi beban sistem kesehatan dengan memberi respon yang lebih efisien.

Dengan adanya model ini, pemerintah dan lembaga kesehatan dapat membuat rencana penanganan pandemi yang efektif. Respon terhadap pandemi akan lebih cepat serta pembatasan pada penyebaran virus akan lebih efektif. Model ini juga memberikan informasi dini kepada orang-orang yang mungkin terpapar sehingga mereka dapat lebih berhati-hati.

Dengan adanya model prediksi Covid-19, kita dapat meningkatkan efisiensi penggunaan sumber daya kesehatan, termasuk fasilitas perawatan dan distribusi vaksin. Ini tidak hanya akan membantu mengelola pandemi dengan lebih efektif, tetapi juga memberikan landasan yang kuat untuk menghadapi tantangan kesehatan masyarakat di masa depan.

Literature Review

Semenjak COVID-19 muncul pada akhir 2019 di China, sudah banyak penelitian yang menunjukkan gejala - gejala apa saja yang mungkin dapat dialami oleh pasien jika pasien tersebut mengalami COVID-19. Para dokter dan peneliti, mulai melakukan banyak penelitian guna mendeteksi seseorang yang diduga terjangkit COVID-19. Terdapat banyak teori yang bermunculan mengenai gejala apa saja yang menjadi pertanda bahwa seseorang terjangkit COVID-19. Beberapa gejala tersebut seperti demam, flu, tenggorokan kering, kesulitan bernafas, dll. Lalu, berdasarkan 3 penelitian yang telah kami kurasi, ada beberapa perbedaan pendapat mengenai gejala penentu seseorang terjangkit COVID-19.

Menurut penelitian dengan judul *Modeling the Onset of Symptoms of COVID-19* yang diterbitkan pada 2020 lalu, gejala yang paling menentukan apakah seseorang terjangkit COVID-19 adalah demam. Dengan persentase sebanyak 77% - 81% tergantung dari seberapa parah keadaan pasien tersebut. Lalu setelah melihat gejala demam, gejala lain yang muncul pada seseorang yang diduga terjangkit COVID-19 adalah batuk. Jika seseorang telah mengalami demam dan telah teridentifikasi sebagai pasien COVID-19, maka gejala selanjutnya yang akan muncul adalah batuk dengan persentase 98%.

Sementara itu, penelitian berjudul *Determine the Most Clinical Symptoms in COVID-19 Patients: A Systematic Review and Meta-Analysis*. memiliki hasil yang sama dengan penelitian sebelumnya. Pada penelitian ini, mereka menggunakan metode *systematic review* dari jurnal kesehatan yang telah diterbitkan dan juga metode *Meta-Regression Analysis*. Hasil dari penelitian mereka menjabarkan bahwa, gejala paling umum dari seseorang terjangkit COVID-19 adalah demam dengan persentase 81.2%. Lalu disusul dengan batuk sebesar 58.5%, lalu lemas dengan persentase sebesar 38.5%. Hasil tersebut, sama dengan hasil penelitian sebelumnya, yang memprediksi gejala COVID-19 paling umum dan keberlanjutan gejala.

Namun, di penelitian lain yang berjudul *Main Symptoms in Patients Presenting in the COVID-19 Period*, dengan menggunakan metode testing secara langsung, mereka menemukan bahwa gejala COVID-19 yang paling umum terjadi adalah batuk yang mempunyai persentase yaitu 56.6%. Lalu disusul dengan kehilangan indera perasa dengan persentase 35.7%

Dengan meningkatnya COVID-19 di seluruh dunia, virus yang dimana telah di dijelaskan oleh World Health Organization (WHO) sebagai darurat kesehatan masyarakat tingkat global dan juga sebagai pandemi global. Sangatlah penting untuk mengetahui gejala-gejala yang disebabkan oleh COVID-19, oleh karena itu kita membuat model prediktif dengan menggunakan model decision tree ke sebuah set data yang terdiri dari daftar orang-orang yang memiliki gejala-gejala COVID-19. Kami akan memprediksi apakah orang tersebut terkena COVID-19 atau tidak, sesuai dengan gejala-gejala yang mereka alami. Dengan menggunakan model prediktif tersebut, kami dapat memprediksi set data yang akan mendatang terhadap gejala-gejala yang dimiliki oleh orang-orang.

Sumber Data

<https://www.kaggle.com/datasets/hemanthhari/symptoms-and-covid-presence>
(Covid-19 patients symptom dataset)

* Mengambil 70% data pertama sebagai data training dan 30% sisanya untuk data testing.

Model

Decision tree

Data Processing

Mengubah data dari string menjadi float

Berikut adalah link untuk dataset kami sebelum data processing:

<https://www.kaggle.com/datasets/hemanthhari/symptoms-and-covid-presence>

Berikut adalah link untuk dataset kami setelah data processing:
https://docs.google.com/spreadsheets/d/1ySTj8cyPiWgV7AHWOaaSbKzqaJrca7r7z_pXaLmFi0E/edit?usp=sharing

https://docs.google.com/spreadsheets/d/1ySTj8cyPiWgV7AHWOaaSbKzqaJrca7r7z_pXaLmFi0E/edit?usp=sharing

Proses

Berikut adalah proses untuk coding kami :

```
1 # Load libraries
2 import pandas as pd
3 from sklearn.tree import DecisionTreeClassifier # Import Decision Tree Classifier
4 from sklearn.model_selection import train_test_split # Import train_test_split function
5 from sklearn import metrics #Import scikit-learn metrics module for accuracy calculation
6
```

```
1 # Ambil data dari drive
2 from google.colab import drive
3 drive.mount('/content/gdrive')
4 %cd /content/gdrive/MyDrive
```

```
1 # Memberikan nama untuk table
2 col_names = ['Breathing Problem','Fever','Dry Cough','Sore throat','Running 0se','Asthma','Chronic Lung Disease','Headache','Heart Disease','Diabet']
3 # load dataset
4 feature_cols = ['Breathing Problem','Fever','Dry Cough','Sore throat','Running 0se','Asthma','Chronic Lung Disease','Headache','Heart Disease','Diabet']
5 pima = pd.read_csv("/content/COVID-19 Data - Sheet1.csv", header=None, names=col_names)
6 pima[feature_cols] = pima[feature_cols].apply(pd.to_numeric, errors='coerce')
7 pima = pima.dropna()
```

```
1 #split dataset in features and target variable
2 X = pima[feature_cols] # Features
3 y = pima.COVID # Target variable
```

```
1 # Split dataset into training set and test set
2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1) # 70% training and 30% test
3
```

```
1 # Create Decision Tree classifier object
2 clf = DecisionTreeClassifier()
3
4 # Train Decision Tree Classifier
5 clf = clf.fit(X_train,y_train)
6
7 #Predict the response for test dataset
8 y_pred = clf.predict(X_test)
9
```

```
1 # Model Accuracy, how often is the classifier correct?
2 print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
3
```

```
1 # Install Library
2 pip install six
```

```

1 #Predict the response for test dataset
2 y_predict = clf.predict([[1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0]])
3 print(y_predict)

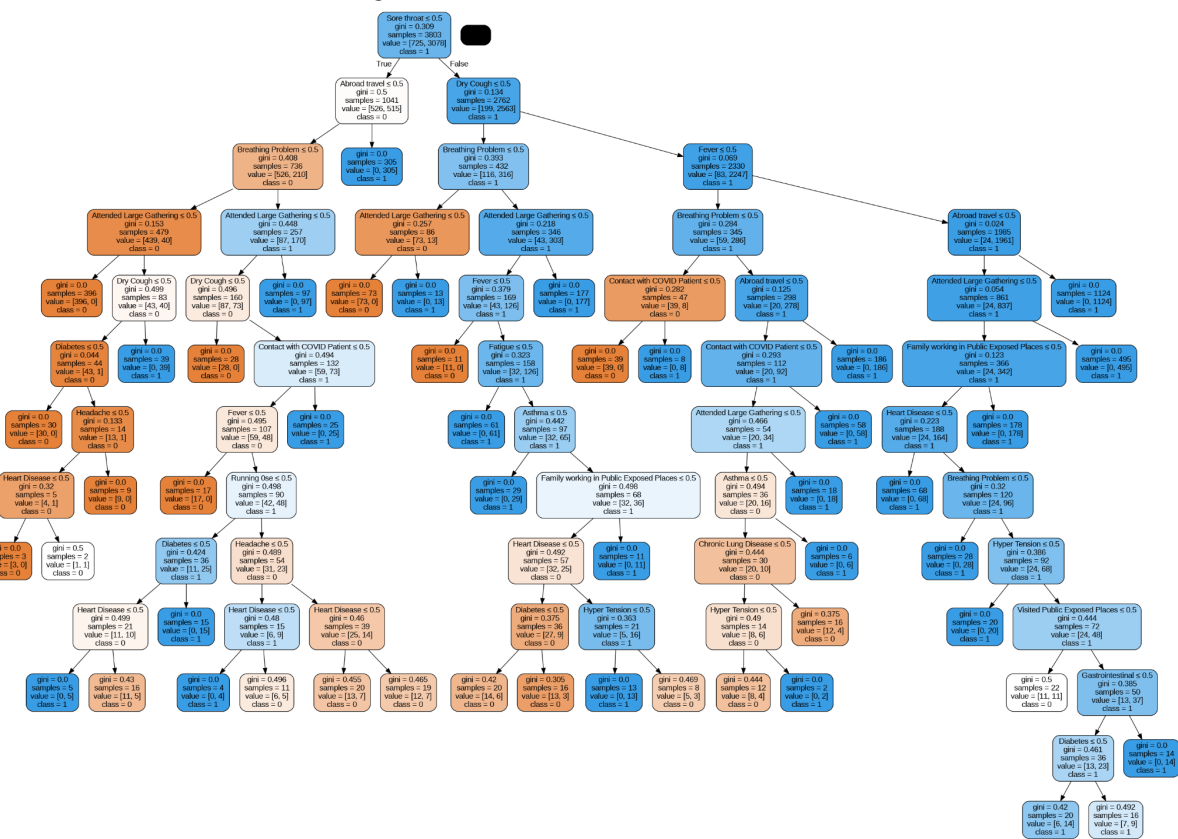
```

```

1 #generate Gambar Decision Tree
2 from sklearn.tree import export_graphviz
3 from six import StringIO # Import StringIO from the 'six' module
4 from IPython.display import Image
5 import pydotplus
6
7 dot_data = StringIO()
8 export_graphviz(clf, out_file=dot_data,
9                 filled=True, rounded=True,
10                 special_characters=True, feature_names=feature_cols, class_names=['0', '1'])
11 graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
12 graph.write_png('diabetes.png')
13 Image(graph.create_png())

```

Hasil : Tabel, Grafik, Penjelasan



Gambar diatas adalah hasil decision tree yang generate menggunakan coding python.

Berdasarkan *decision tree* diatas, dapat diketahui bahwa gejala paling menentukan dari virus COVID-19 adalah *sore throat*. Jika seseorang mengalami *sore throat*, maka kemungkinan besar, orang tersebut terkena virus COVID-19.

Lalu, setelah *sore throat*, gejala paling menentukan lainnya adalah *Abroad Travel* dan *Dry Cough*. Dapat diketahui bahwa jika seseorang mengalami *sore throat*, maka

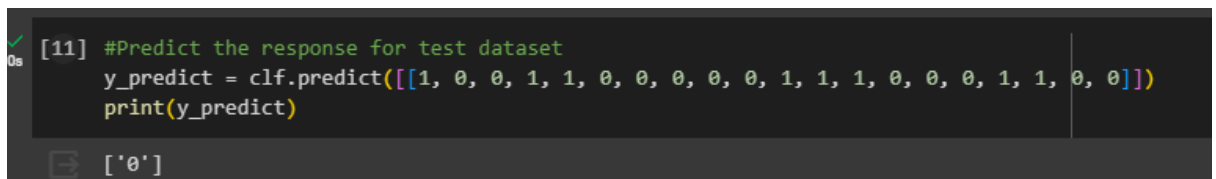
dia habis melakukan *Abroad Travel*. Lalu, jika user tidak mengalami *sore* maka user mengalami *dry cough*.

Untuk penjelasan lainnya dapat dilihat dari coding yang kami screenshot sebelumnya karena sudah tercantum (bagian-bagian coding untuk apa).

Kesimpulan

1. Algoritma

Kami menggunakan algoritma decision tree dan mendapatkan accuracy sebesar 0.9846719803801349 atau artinya accuracy dari algoritma tersebut adalah 98%. Kami juga mencoba memasukkan data untuk memprediksi apakah seseorang mengalami covid atau tidak. Berikut adalah contohnya :

A screenshot of a Jupyter Notebook cell. The cell contains the following code:

```
[11] #Predict the response for test dataset
y_predict = clf.predict([[1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0]])
print(y_predict)
```

 The output of the cell is `['0']`. The background of the notebook is dark, and the text is light green and yellow.

Dapat dilihat bahwa algoritma tersebut dapat memprediksi sesuai dengan data dummy yang kita berikan.

2. Tools

Tools yang kami gunakan untuk membuat algoritma ini adalah google collab. Alasan kami memilih tools itu adalah karena kami bekerja secara berkelompok sehingga lebih mudah bila kami semua dapat mengakses tugas kami bersama-sama.

3. Kesimpulan Penerapan Decision Tree

Kami menggunakan 20 atribut sebagai variabel independen untuk menghasilkan 1 variabel dependen yaitu Covid. Penerapan decision tree pada data tersebut menghasilkan accuracy sebesar 98%. Penerapan tersebut akan sangat bermanfaat untuk memprediksi apakah seseorang mengalami covid atau tidak berdasarkan 20 atribut tersebut. Jadi jika ingin mengetahui seseorang mengetahui covid atau tidak, kita dapat mengumpulkan data-data tersebut dan menggunakan model decision tree ini.

Referensi:

- Çalica Utku, A., Budak, G., Karabay, O., Güçlü, E., Okan, H. D., & Vatan, A. (2020). Main symptoms in patients presenting in the COVID-19 period. *Scottish Medical Journal*, 65(4), 127–132.
<https://doi.org/10.1177/0036933020949253>
- ALIMOHAMADI, Y., SEPANDI, M., TAGHDIR, M., & HOSAMIRUDSARI, H. (2020). *View of determine the most common clinical symptoms in COVID-19 patients: A systematic review and meta-analysis: Journal of Preventive Medicine and Hygiene*. View of Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis | Journal of Preventive Medicine and Hygiene.
<https://www.jpmmh.org/index.php/jpmmh/article/view/1530/777>
- Larsen, J. R., Martin, M. R., Martin, J. D., Kuhn, P., & Hicks, J. B. (2020, July 27). *Modeling the onset of symptoms of COVID-19*. Frontiers.
<https://www.frontiersin.org/articles/10.3389/fpubh.2020.00473/full>
- Harikrishnan, H. (2020, August 18). *Symptoms and COVID presence (May 2020 data)*. Kaggle.
<https://www.kaggle.com/datasets/hemanthhari/symptoms-and-covid-presence>