Jurnal Kesehatan Medika Saintika
e-ISSN:2540-9611
p-ISSN:2087-8508
Volume 13 nomor 2 (Desember 2022)| https://jurnal.syedzasaintika.ac.id
DOI: http://dx.doi.org/10.30633/jkms.v13i2.1683

ANALISIS KINERJA ALGORITMA C4.5 UNTUK PREDIKSI PASIEN COVID-19 DI SEMEN PADANG HOSPITAL

PERFORMANCE ANALYSIS OF C4.5 ALGORITHM FOR PREDICTION OF COVID-19 PATIENTS AT SEMEN PADANG HOSPITAL

Nurul Abdillah[1*], Herman Susilo[2], Muhammad Ihksan[3]
[1,2,3] STIKes Syedza Saintika
Email :Abdillahadik15@gmail.com

ABSTRAK

Virus Corona atau severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) adalah virus yang menyerang sistem pernapasan. Penyakit karena infeksi virus ini disebut COVID-19. Virus Corona bisa menyebabkan gangguan pada sistem pernapasan, pneumonia akut, sampai kematian. Pada tahun 2021 pasien covid masih mendoninasi di beberapa Rumah Sakit salah satunya adalah Semen Padang Hospital (SPH). Data rekam medis pasien covid-19 tersimpan di database Sistem Informasi Manajemen Rumah Sakit (SIMRS). Data rekam medis adalah catatan khusus pasien yang berisikan biodata, riwayat penyakit dan pengobatan, seringkali data rekam medis hanya menjadi data yang menumpuk dan tidak dilakukan penelusuran untuk menghasilkan pengetahuan yang berguna bagi rumah sakit. Penelitian ini bertujuan mengolah tumpukan data rekam medis khusus pasien covid-19 untuk mengklasifikasikan pasien covid-19 atau tidak yang terjadi di SPH. Metode yang digunakan dalam penelitian ini adalah metode klasifikasi dengan menggunakan algoritma C4.5. Atribut yang digunakan adalah bulan berobat, jenis kelamin, asal daerah, dan umur. Untuk atribut label tujuan sebanyak 2 kelompok yaitu diagnosa covid-19 atau tidak. Penilitian ini menghasilkan pohon keputusan. Dari penerapan algoritma C4.5 pada data pasien di SPH untuk mengklasifikasi pasien covid atau tidak, didapat hasil akurasinya kurang dari 70% yaitu adalah 51.97% dan error 48.03%, sehingga kesimpulannya adalah algoritma C4.5 memiliki kinerja yang kurang baik untuk mengklasifikasi pasien covid-19 di SPH.

Kata kunci: Data Mining, Klasifikasi, Rekam Medis, Covid-19, C4.5

ABSTRACT

Corona virus or severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a virus that attacks the respiratory system. The disease caused by this viral infection is called COVID-19. Corona virus can cause disorders of the respiratory system, acute pneumonia, to death. In 2021, Covid patients are still donating to several hospitals, one of which is Semen Padang Hospital (SPH). Covid-19 patient medical record data is stored in the Hospital Management Information System (SIMRS) database. Medical record data is a patient's special record that contains biodata, medical history and treatment, medical record data is often just piled up data and not traced to produce useful knowledge for the hospital. This study aims to process piles of medical record data specifically for Covid-19 patients to classify Covid-19 patients or not who occur at SPH. The method used in this study is a classification method using the C4.5 algorithm.

Jurnal Kesehatan Medika Saintika
e-ISSN:2540-9611
p-ISSN:2087-8508
Volume 13 nomor 2 (Desember 2022)| https://jurnal.syedzasaintika.ac.id
DOI: http://dx.doi.org/10.30633/jkms.v13i2.1683

Attributes used are month of treatment, gender, region of origin, and age. For the destination label attribute, there are 2 groups, namely Covid-19 diagnosis or not. This research produces a decision tree. From the application of the C4.5 algorithm to patient data at SPH to classify covid patients or not, the accuracy results are less than 70%, namely 51.97% and an error of 48.03%, so the conclusion is that the C4.5 algorithm has poor performance for classifying covid patients -19 at SPH.
Keywords: *Data Mining, Classification, Medical Records, Covid-19, C4.5*

PENDAHULUAN

Corona Virus Disease 2019 atau disingkat Covid-19 merupakan penyakit baru yang muncul di tahun 2019 dan dapat menyebabkan radang paruparu dan gangguan pernapasan (C. Long et al, 2020) Penyakit ini disebabkan oleh Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) dan dapat menyebabkan kematian (D. Susanna, 2020). Dunia tengah dilanda pandemi Covid-19 yang melumpuhkan banyak sektor penting kehidupan baik di bidang ekonomi, sosial, dan keamanan (C. Long et al, 2020). Tatanan kehidupan dan tantangan akibat krisis Kesehatan global ini melanda semua negara di dunia. Kondisi yang sama juga terjadi di Indonesia (D. Susanna, 2020). Sejak awal maret, Pemerintah Indonesia telah mengumumkan jumlah kasus Covid-19 yang dikonfirmasi, kasus yang pulih serta kasus kematian setiap harinya (D. Susanna, 2020).

Pada bulan April 2021, angka kematian akibat virus Corona di Indonesia merupakan yang tertinggi di Asia setelah China, dengan korban meninggal 41.815 orang. Jumlah total kasus virus
korona mencapai 1.537.967 kasus dengan 1.381.667 orang sembuh pada saat makalah ditulis. Angka ini terus merangkak naik. Mengingat wabah Covid19 merupakan masalah global di belahan dunia termasuk di Indonesia. Studi ini dilakukan sebagai bagian dari upaya Mitigasi terhadap penyebaran penyakit Covid-19 di Indonesia (M. Sukmana et al, 2020).

Kebanyakan orang yang terinfeksi virus Covid-19 akan mengalami penyakit pernapasan
ringan hingga sedang dan sembuh tanpa memerlukan perawatan. Orang tua, dan merekayang memiliki masalah medis seperti penyakit kardiovaskular, diabetes, penyakit pernapasan kronis, dan kanker lebih mungkin untuk mengembangkan penyakit serius. Penyakit ini telah menjadi pandemi krisis kesehatan global sejak Maret 2020 (P. Sharma et al, 2020).

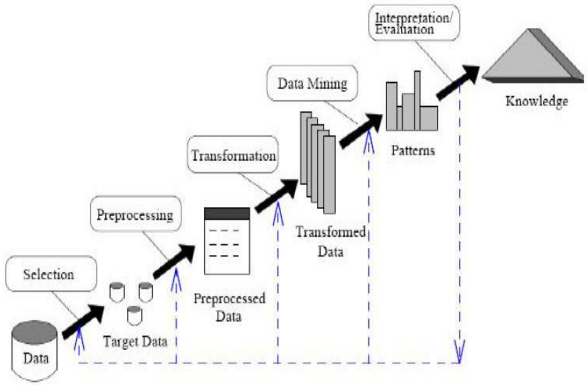
Banyaknya pasien yang di rumah sakit dan kapasitas tenaga medis menjadi permasalahan utama yang dihadapi di berbagai daerah. Pasien dengan tingkat kegawatan tinggi memerlukan prioritas penanganan dibanding pasien dengan gelaja sedang atau tanpa gejala. Tenaga medis memerlukan bantuan untuk mengklasifikasi status pasien berdasarkan data pasien secara otomatis untuk mengurangi kelelahan tenaga medis yang harus terus bertugas dan meminimalisir resiko penanganan yang terlambat terhadap pasien. Oleh karena itu dibutuhkan solusi teknologi berbasis data secara otomatis yang dapat membantu mengklasifikasikan status kegawatan berdasarkan data pasien (P. Sharma et al, 2020).

Knowledge discovery in database (KDD) merupakan proses untuk menemukan

Jurnal Kesehatan Medika Saintika
e-ISSN:2540-9611
p-ISSN:2087-8508
Volume 13 nomor 2 (Desember 2022)| https://jurnal.syedzasaintika.ac.id
DOI: http://dx.doi.org/10.30633/jkms.v13i2.1683

informasi yang berguna dalam database. Seluruh proses KDD biasanya terdiri dari langkah-langkah, yaitu memahami bidang aplikasi, membuat data target yang ditetapkan dari data mentah yang tersimpan dalam *database*, pembersihan data dan *preprocessing* data (Bastian et al, 2018).

Istilah *knowledge discovery in database* atau mencari pengetahuan dalam *database* atau KDD singkatnya, mengacu pada proses pencarian pengetahuan dalam data yang luas dan menekankan pada penerapan metode tingkat tinggi atau metode penambangan data tertentu. Ini menarik minat para peneliti dalam melakukan pengembang penelitian baik dalam bidang *machine learning* atau pembelajaran mesin (Bastian et al, 2018).



Gambar 1. Proses *Knowledge discovery in database*[2]

Data mining adalah teknik yang digunakan untuk membangun model pembelajaran mesin. Pembelajaran mesin (machine learning) adalah teknik kecerdasan buatan modern yang belajar membangun model dengan menggunakan data empiris. Data Mining digunakan untuk menemukan pola dalam kumpulan besar data mentah. Data Mining menerapkan teknik Machine Learning untuk menarik pengetahuan pada data. Dalam penelitian ini penulis menerapkan teknik data mining untuk mengklasifikasikan dataset Covid-19 menggunakan Algoritma C4.5 karena telah berhasil diterapkan dalam banyak tugas klasifikasi pada populasi data (Septiani et al, 2017).

Penelitian ini bertujuan memberikan solusi untuk mengklasifikasikan status pasien Covid-19 secara otomatis berdasarkan data rekam medis pasien. Klasifikasi dilakukan menggunakan algoritma C4.5 yang dikenal memiliki akurasi tinggi dengan pembelajaran tersupervisi berbasis distribusi dataset.

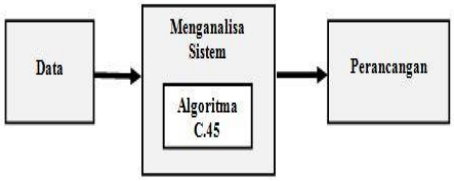
Kontribusi utama pada penelitian ini adalah menghasilkan model prediktif status pasien Covid19 Sehingga dapat disiapkan langkah-langkah yang cepat dan tepat untuk penanganan pasien. Algoritma C4.5 yang memiliki akurasi tinggi dan telah banyak diterapkan pada bermacam permasalahan prediksi, sehingga diharapkan dapat membantu memprediksi status pasien untuk mendapatkan penanganan yang tepat secara langsung oleh tim medis (Turnip et al, P. 2018).

BAHAN DAN METODE

Penelitian fokus pada proses menganalisis data rekam medis dengan algoritma C4.5 menggunakan program *RapidMiner Studio 7.5 (Tools Data Mining)* untuk memperoleh hasil klasifikasi. Ada 4 atribut yang dipakai dalam penelitian, yaitu: (1) umur yang dikelompokkan dalam kategori bayi/balita, anak-anak, remaja, dewasa, dan lansia; (2) jenis kelamin yang terdiri dari perempuan (P) dan laki-laki (L); (3) bulan yang terdiri dari Juli, Agustus, September dan Oktober; (4) diagnosa yang merupakan atribut tujuan yang terdiri dari 2 kelompok yaitu diagnose covid atau tidak.

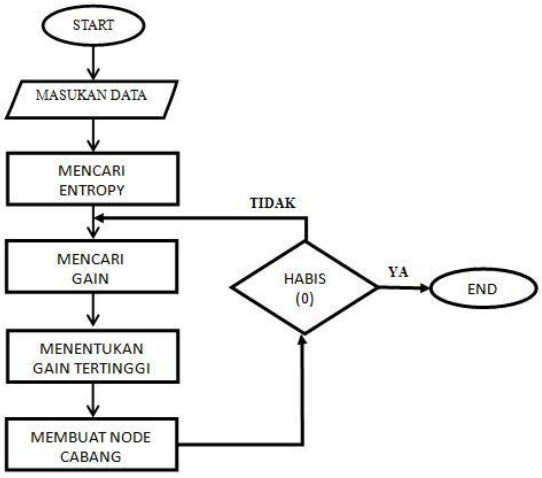
Jurnal Kesehatan Medika Saintika
e-ISSN:2540-9611
p-ISSN:2087-8508
Volume 13 nomor 2 (Desember 2022)| https://jurnal.syedzasaintika.ac.id
DOI: http://dx.doi.org/10.30633/jkms.v13i2.1683

Algoritma C4.5 dimulai dari proses memilih atribut dengan gain tertinggi sebagai akar pohon, kemudian membuat cabang untuk tiap-tiap nilai, lalu membagi kasus dalam cabang, setelah itu mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Untuk memudahkan dalam penerapan metodologi dan perancangan sistem maka dibuat bagan alir analisa dan perancangan seperti pada gambar 2 dibawah ini.



Gambar 2. Bagan Alir Analisa

Bentuk diagram alir (*flowchart*) dapat menggambarkan dengan jelas mengenai proses tahapan maupun langkah dalam klasifikasi menggunakan algoritma C4.5. Dapat dilihat dalam bentuk gambar 4.2 berupa *flowchart* sebagai berikut:



Gambar 3. Flowchart Proses pada Algoritma C4.5

Teknik Klasifikasi Algoritma C4.5 dimulai dengan processing dan transfromasi data agar data mentah yang digunakan untuk analisa adalah data dengan atribut yang lengkap dan dapat menghasilkan pohon keputusan , berikut urutan kerja untuk menentukan pohon keputusan

***Processing* dan Transformasi Data**

Tidak semua atribut dalam database rekam medis pasien digunakan dalam penelitian, atribut seperti: tanggal lahir, nama dokter penanggung jawab dan nomor rekam medis pasien tidak dibutuhkan dalam penelitian ini.

Pengolahan hanya memerlukan empat buah atribut, seperti: Jenis Kelamin, Kategori Umur, Bulan Berobat dan Kode Penyakit sebagai atribut tujuan dari penelitian ini, sehingga diperlukan proses transformasi data yang akan diolah, setelah dilakukan transformasi data maka nantinya akan dilakukan pengolahan menggunakan algoritma C4.5, Sampel data yang digunakan adalah sebanyak 356 data.

Pengolahan Data

Pengolahan data dimulai dengan mencari entropi total dari seluruh atribut dan kemudian menetukan gain tertinggi. Untuk mendapatkan nilai gain dalam pembentukan pohon keputusan, perlu menghitung dulu nilai informasi dalam satuan bits dari kumpulan objek.

Bentuk perhitungan untuk entropi adalah sebagai berikut :

$$Entropy(X) = \sum_{j=1}^{k} p_j * log_2 \frac{1}{p_j}$$

$$= -\sum_{j=1}^{k} p_j * log_2 p_j$$

dimana,

X : Himpunan Kasus

Jurnal Kesehatan Medika Saintika
e-ISSN:2540-9611
p-ISSN:2087-8508
Volume 13 nomor 2 (Desember 2022)| https://jurnal.syedzasaintika.ac.id
DOI: http://dx.doi.org/10.30633/jkms.v13i2.1683

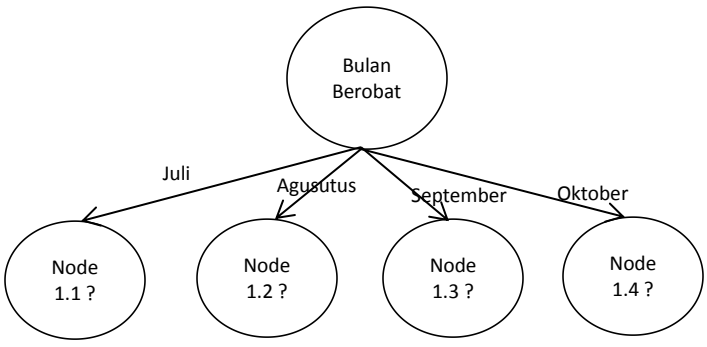
k : jumlah partisi X
p_j : Proporsi X_j terhadap X

Besar nilai *Entropy(X)* menunjukkan bahwa X adalah atribut yang acak. Nilai entropi mencapai nilai minimum 0, ketika semua p_j lain = 0 atau berada pada kelas yang sama. Pada kontruksi pohon C4.5, di setiap simpul pohon diisi oleh atribut dengan nilai *gain ratio* tertinggi, dengan rumus sebagai berikut:

$$Gain(a) = Entropy(X) - \sum_{j=1}^{k} \frac{|X_i|}{|X|} * Entropy(X_i)$$

Mencari *Entropy* Total dan *Gain* (*Root*)

Proses pencarian entropi total dan gain dilakukan dengan mengelompokan data dengan benar , kemudian menghitung data serta menggunakan rumus pencarian *entropy* dan *gain* pada masing-masing atribut data. Dari hasil perhitungan pada tabel diatas, dapat diketahui bahwa nilai *gain* terbesar yaitu pada atribut "Bulan Berobat" sebesar 1,346. Sehingga atribut "Bulan Berobat" menjadi node akar. Pada atribut "Bulan Berobat" terdapat 4 nilai atribut, yaitu Juli, Agustus, September dan Oktober, maka perlu dilakukan perhitungan lanjut. Dari proses tersebut maka dapat dihasilkan pohon sementara seperti berikut ini:

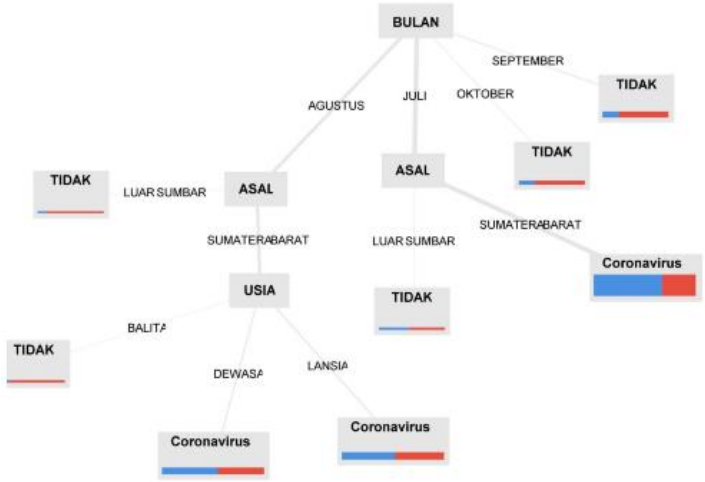


Gambar 4. Pohon Keputusan Sementara (*Root*)

HASIL

Dari perhitungan yang dilakukan menggunakan algoritma C4.5, hasil analisa menunjukan bahwa kinerja algoritma C4.5 kuranng baik untuk ditepakan dalam pengelompokan diagnose pasien covid berdasarkan data rekam medis pasien. Didapatkan nila akurasi kurang dari 70%, sehingga dapat dikatakan Algoritma C.45 kurang baik dalam mengklasifikasikan data pasien covid.

Hasil pengujian terhadap data yang telah dilakukan, didapatkan pohon keputusan seperti dibawah ini:



Gambar 5. Pohon Keputusan

PEMBAHASAN

Pada Implementasi dan Hasil akan dijelakan Implementasi atau pengujian untuk mendapatkan hasil kinerja algoritma C4.5 dengan membandingkan hasil dari perhitungan manual dengan hasil

Jurnal Kesehatan Medika Saintika
e-ISSN:2540-9611
p-ISSN:2087-8508
Volume 13 nomor 2 (Desember 2022)| https://jurnal.syedzasaintika.ac.id
DOI: http://dx.doi.org/10.30633/jkms.v13i2.1683

menggunakan software pendukung algoritma C4.5. Hal ini bertujuan untuk melihat data yang dianalisa dan diolah sudah benar atau belum. Software yang digunakan adalah Rapidminer Studio 7.5.3. Rapidminer Studio merupakan aplikasi Data Mining open source. Pada kasus memprediksi ketepatan waktu kelulusan mahasiswa ini, data yang akan digunakan pada Rapidminer berjumlah sebanyak 150 record. Pebandingan Tingkat Akurasi dan Error Algoritma C4.5 sebagai berikut :

1. Tingkat Akurasi C4.5

   Pada perhitungan akurasi C4.5 diperoleh akurasi sebesar 97,83 % karena menghasilkan 90 data yang diklasifikasikan secara benar.

accuracy: 51.97%

	true YA	true TIDAK	class precision
pred. YA	185	171	51.97%
pred. TIDAK	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 6. Tingkat Akurasi

2. Tingkat Error AlgoritmaC4.5

   Pada perhitungan error C4.5 diperoleh nilai error sebesar 2,17% karena menghasilkan 2 data yang diklasifikasikan secara tidak benar.

classification_error: 48.03%

	true YA	true TIDAK	class precision
pred. YA	185	171	51.97%
pred. TIDAK	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 7. Tingkat Error

KESIMPULAN DAN SARAN

Kesimpulan

Pengukuran tingkat akurasi kinejra metode klasifikasi C4.5 dan Naive Bayes menghasilkan nilai akurasi sebesar 51,97 %. Pengukuran tingkat error pada algoritma C4.5 menghasilkan tingkat error sebesar 48,03%. Dari hasi pengujian yang telah dilakukan, Algoritma C4.5 memiliki kinerja yang kurang baik untuk mengklasifikasikan pasien covid di SPH berdasarkan data rekam medis karena Algoritma C4.5 memilki nilai akurasi yang kurang dari 70%, semakin tinggi nilai akurasi maka pengklasifikasian data semakin mendekati benar

Saran

Disarankan untuk peneliti selanjutnya agar menggunakan atribut lebih banyak dalam mengklasifikasikan data pasien covid agar didapatkan akurasi yang lebih tinggi.

DAFTAR PUSTAKA

A. P. Fadillah, “Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ),” J. Tek. Inform. dan Sist. Inf., vol. 1, pp. 260–270, 2015

Jurnal Kesehatan Medika Saintika
e-ISSN:2540-9611
p-ISSN:2087-8508
Volume 13 nomor 2 (Desember 2022)| https://jurnal.syedzasaintika.ac.id
DOI: http://dx.doi.org/10.30633/jkms.v13i2.1683

Bastian, Ade, Harun Sujadi, Gigin Febrianto, Program Studi, Teknik Informatika, Universitas Majalengka, Jl Universitas, and Majalengka No. 2018. "Penerapan Algoritma K-Means Clustring Anlysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)." Jurnal Sistem Informasi (Jurnal Of Information System) 14(1):26–32.

D. Susanna, "When will the COVID-19 pandemic in indonesia end?," *Kesmas*, vol. 15, no. 4, pp. 160–162, 2020, doi: 10.21109/KESMAS.V15I4.4361.

C. Long et al., "Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT?," Eur. J. Radiol., vol. 126,p. 108961, May 2020, doi: 10.1016/j.ejrad.2020.108961.

M. Sukmana, M. Aminuddin, and D. Nopriyanto, "Indonesian government response in COVID-19 disaster prevention," *East Afrian Sch. J. Med. Sci.*, vol. 3, no. 3, pp. 81–6, 2020, doi: 10.36349/EASMS.2020.v03i03.025.

Rafiska R, Defit S, Nurcahyo, G. W. 2018. "Analisis Rekam Medis Untuk Menentukan Pola Kelompok Penyakit." 2(1) : 391 – 96.

Septiani, Wisti Dwi. 2017. "Komprasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis." 13(1):76–84.

S. K. Kar, S. M. Y. Arafat, P. Sharma, A. Dixit, M. Marthoenis, and R. Kabir, "COVID-19 pandemic and addiction: Current problems and future concerns," *Asian J. Psychiatr.*, vol. 51, p. 102064, 2020, doi: https://doi.org/10.1016/j.ajp.2020.102064.

Turnip, S. M, Silitonga, P. 2018. "Analisis Pola Penyebaran Penyakit Dengan Menggunakan Algoritma C4.5." 03(479):3–7.

Wijaya, Lalu, and Nur Arini Pratiwi. 2020. "Penerapan Algoritma K-Means Untuk Pendataan Obat Berdasarkan Laporan Bulanan Pada Dinas Kesehatan Kabupaten Lombok Timur." 3(2):64–73.

Word Healty Organization, "WHO Director Genera's Opening Remarks At the Media Breating On Covid-19–11 march 2020". 13 November 2020, [Online] Apalable: Https://www.who.int/director-general-s-opening-remarks-at-themedia-breating-on-covid-19-11-march-2020