

Clustering Tingkat Kedisiplinan Warga Bekasi Dalam Menjalankan Protokol Kesehatan di Masa Pandemi Covid-19 dengan Algoritme K-Means

Andri Dwi Noviandi¹, Tesa Nur Padillah², Yuyun Umaidah³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang Jl. H.S Ronggowaluyo Teluk Jambe Timur Kabupaten Karawang. 41361.

Email: andri.dwi17056@student.unsika.ac.id

Info Artikel

Sejarah Artikel:

Diterima: 30 Juli 2021

Direvisi: 19 Agustus 2021

Dipublikasikan: Agustus 2021

e-ISSN: 2089-5364

p-ISSN: 2622-8327

DOI: 10.5281/zenodo.5336446

Abstract:

Health protocols during the Covid-19 pandemic are very necessary because health protocols can speed up breaking the chain of spreading the Covid-19 virus. Violations that are often found in the Bekasi city environment are related to health protocols, namely maintaining distance, wearing masks and washing hands, or using hand sanitizer. There are still many who do not comply with the rules of the health protocol. The purpose of knowing the cluster level of discipline towards health protocols into five clusters spread by the number of respondents in various sub-districts in the city of Bekasi with the categories of discipline, somewhat disciplined, rarely disciplined, less disciplined, and undisciplined. Data mining is the process of extracting data to obtain new information. The technique used in this research is simple random sampling. This study using the CRIPS-DM methodology. This study calculates the k-means algorithm by obtaining a value of $k = 2$. The results of the test using the RapidMiner Studio 9.3 tools obtained two clusters or 2 categories of discipline levels against health protocols, namely cluster 0 with a percentage of 55.08% which is categorized as the most disciplined level, and cluster 1 with a percentage of 44.92% which is categorized as the least disciplined level. The results of clustering are evaluated by using the Silhouette Coefficient with the best cluster, $k = 2$ with a value of 0.926989, which is the best cluster.

Keywords: Data mining, Clustering, K-means Algoritme, Silhouette Coefficient

PENDAHULUAN

Penyebaran *Covid-19* di Indonesia sudah semakin bertambah banyak di masing-masing daerah. Corona adalah kumpulan virus yang bisa menginfeksi sistem pernapasan. Pada banyak kasus, virus ini hanya menyebabkan infeksi pernapasan ringan, seperti flu. Namun, virus ini juga bisa menyebabkan infeksi pernapasan berat, seperti infeksi paru-paru (pneumonia), *Middle-East Respiratory Syndrome* (MERS), *severe acute respiratory syndrome* (SARS). Virus ini dapat menyerang bayi, anak-anak, orang dewasa, lansia, dan ibu hamil. Wabah virus *Covid-19* membuat masyarakat merasa ketakutan dan panik. Begitu juga ada beberapa masyarakat yang tidak peduli/percaya mengenai wabah virus corona ini. Pandemi ini membuat masyarakat kita mengisolasi diri, menjaga jarak, mengenakan masker, dan kerja dari rumah (KEMENKES, 2020).

Menurut Kementerian Kesehatan Indonesia melakukan protokol kesehatan di masa pandemic *Covid-19* sangat perlu dilakukan, karena protokol kesehatan bisa mempercepat memutus mata rantai penyebaran virus *Covid-19* jika dilakukan dengan sadar, taat dan benar. Disiplin dalam menerapkan protokol kesehatan menjadi kunci pengendalian penularan *Covid-19*. Dari sisi masyarakat, hal ini membutuhkan kepatuhan untuk tertib dalam bermasker, mencuci tangan, dan menjaga jarak. Di sisi lain, masyarakat juga butuh teladan agar konsisten menjadikan perilaku hidup sehat tersebut sebagai sebuah identitas baru. Pelanggaran yang banyak dijumpai di lingkungan sekitar terkait protokol kesehatan yaitu menjaga jarak, memakai masker dan mencuci tangan atau menggunakan *handsanitizer*. Masih banyak yang tidak mematuhi aturan dari protokol kesehatan tersebut.

Penelitian ini dimaksudkan untuk mengetahui tingkat kedisiplinan

masyarakat di Kota Bekasi dalam mematuhi protokol kesehatan di masa pandemi saat ini. Penulisan ini tentu melihat penelitian sebelumnya yang menggunakan metode dan algoritme yang sama yaitu metode *Clustering* dan menggunakan Algoritme *K-means*. Mengapa menggunakan Metode *Clustering* dan Algoritme *K-means*? Menurut (Asroni, 2018) *clustering* adalah untuk meminimalkan fungsi objektif yang di set dalam suatu kelompok dan memaksimalkan variasi antar kelompok.

Berdasarkan permasalahan diatas *K-means* adalah salah satu bentuk pengelompokan yang paling sederhana. Prosedurnya sederhana dan mudah untuk mengklasifikasikan data yang diberikan melalui sejumlah *cluster* dan juga algoritme ini dapat membuat kelompok ataupun data set yang besar bisa menjadi mudah dengan algoritme tersebut.

METODOLOGI PENELITIAN

Objek Penelitian

Objek penelitian yang akan diteliti mengenai analisis tingkat kedisiplinan warga Bekasi dengan menggunakan teknik *clustering* dan algoritme *K-means* dalam menjalankan protokol kesehatan di masa pandemi *Covid-19*. Data diperoleh dari *kuesioner google form*. Metodologi *Data mining Life Cycle* (DMLC) yaitu *Cross Industry Standard Process for Data mining* (CRISP-DM) atau bisa disebut juga dengan tentang siklus hidup pengembangan *data mining*.

Metodologi Penelitian

Dalam penelitian ini metodologi *Data mining Life Cycle* (DMLC) yang digunakan adalah *Cross Industry Standard Process for Data mining* (CRISP-DM). Metodologi *CRISP-DM* merupakan standar dalam melakukan penelitian dengan bertemakan *data mining* dan *knowledge discovery*, karena metodologi ini banyak

digunakan dalam bisnis dan pengembangan *data mining* itu sendiri. *CRISP-DM* ini memiliki tahapan – tahapan sebagai berikut:

1. *Business Understanding*
2. *Data Understanding*
3. *Data Preparation*
4. *Modelling*
5. *Evaluation*
6. *Deployment*

Fase Pemahaman Bisnis (Business Understanding)

Tujuan dari penelitian ini yaitu melakukan pengelompokan berdasarkan karakteristik protokol kesehatan dan seberapa patuh dan disiplin masyarakat kota Bekasi dalam hal mengikuti arahan pemerintah dengan mematuhi aturan protokol kesehatan, kemudian pembagian *cluster* dengan menggunakan metode *K-means*.

Tahap awal peneliti menggunakan survei kuesioner untuk masyarakat sekitar kota Bekasi dengan mengirimkan sebuah link dan mengirim link tersebut ke sosial media *Whatsapp* dan lainnya. Survei dilakukan dari periode bulan Februari 2021 hingga April 2021. Teknik sampling dalam penelitian ini yaitu *Simple Random Sampling* yang digunakan untuk mendapatkan data set dari suatu responden dengan cara memilih secara acak dari jumlah penduduk yaitu 2.543.676 dari berbagai kecamatan yang ada di kota Bekasi dan dari responden yang secara acak itu akan mewakili wilayahnya untuk dijadikan data set yang akan digunakan dalam penelitian ini. Penelitian ini dilakukan untuk membantu memberikan masukan kepada masyarakat kota Bekasi dalam hal kedisiplinan terhadap protokol kesehatan dan dapat mengetahui apakah tingkat kedisiplinan tersebut ada peningkatan atau tidak. caranya dengan dilakukannya pengelompokan tingkat kedisiplinan warga Bekasi dengan cluster lalu dilakukan

perhitungan dengan algoritme yang dipilih untuk menentukan apakah meningkat atau menurun.

Fase Pemahaman Data (Data Understanding)

Pada penelitian ini sumber data yang akan digunakan data primer yaitu data dari survei kuesioner, yang dimana peneliti sudah memberikan *kuesioner google form* via *social media* yang sudah dilakukan sejak bulan februari 2021 sampai april 2021 tersebut yang telah diisi oleh beberapa masyarakat kota Bekasi yang bertujuan memudahkan dalam mengolah *dataset* yang telah diisi oleh masyarakat.

Fase Data Preparation

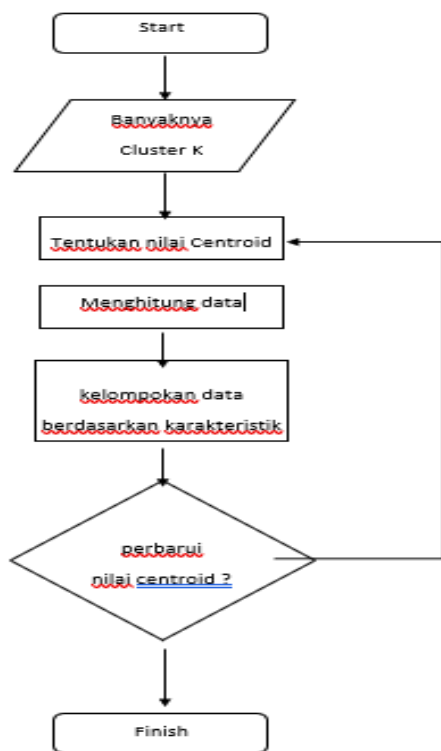
Tahap ini merupakan tahap dimana data yang sudah ada dari kuesioner tersebut dilakukan *preprocessing* untuk membuat data mentah menjadi data yang memiliki kualitas dan arti. Peneliti melakukan beberapa proses untuk dataset yang sudah didapatkan sebelumnya yang sudah diisi oleh masyarakat kota Bekasi dalam waktu beberapa bulan yaitu dari bulan februari sampai april 2021. Dalam fase ini yang harus dilakukan terhadap data tersebut antara lain:

1. *Data Cleaning* antara lain mengisi nilai yang hilang, mengidentifikasi adanya *outlier*, menangani *noise* pada data, mengoreksi data yang tidak konsisten, dan menyelesaikan masalah redudansi data yang disebabkan integrasi data.
2. *Data Transformation* yaitu melakukan normalisasi dan mengumpulkan data agar menjadi sama.
3. *Data Selection*, memilih atribut - atribut tertentu yang akan digunakan untuk melakukan proses *data mining*.

Fase Pemodelan (Modelling)

Tahapan ini merupakan tahapan implementasi dari data yang telah melalui

tahapan *preprocessing* dan sudah di-*import* dilakukan pemodelan menggunakan algoritme *K-means*. Peneliti melakukan proses *clustering* untuk mengetahui apakah tingkat kedisiplinan tersebut ada peningkatan atau tidak. *dataset* yang dipakai untuk pemodelan ini atribut yang digunakan setelah dilakukannya proses tahapan *preprocessing* yaitu atribut pemakaian masker, menjaga jarak, penggunaan *hand sanitizer* dan mencuci tangan yang sudah dilakukan peneliti di lakukan di kota Bekasi dari bulan february sampai april 2021. Dilihat pada **Gambar 1** siklus di fase modeling.



Gambar 1 Siklus Modelling

Fase Evaluasi (Evaluation)

Sebelum peneliti disebarkan mengenai hasil penelitian kepada masyarakat kota Bekasi, hasil yang didapatkan dari perhitungan algoritme *k-means* terhadap *dataset* yang sudah didapatkan dari kuesioner dari bulan february sampai april 2021 perlu

diadakannya tahap evaluasi. Tahap evaluasi ini dilakukan untuk menilai apakah kualitas dan efektifitas model yang sudah dirancang sesuai dengan tujuan dan masalah dalam penelitian ini. Tahap evaluasi ini menggunakan *silhouette coefficient* yakni pengukuran keakuratan hasil yang dicapai oleh model yang sudah digunakan dan mengetahui perhitungan yang sudah dilakukan menggunakan algoritme *K-means*.

Fase Penyebaran (Deployment)

Setelah peneliti melakukan evaluasi, tahap selanjutnya adalah tahap penyebaran yaitu hasil penelitian ini dapat digunakan untuk masyarakat Bekasi dari hasil *clustering* dengan menggunakan algoritme *k-means* dan keakuratan cluster dengan *silhouette coefficient* terhadap *dataset* dari kuesioner yang sudah didapatkan dari bulan february sampai april 2021. Lalu, dalam mencari informasi mengenai seberapa persentase patuh dan disiplin masyarakat dalam menerapkan protokol kesehatan di sekitar dengan menggunakan tools ArcGIS 10.8 untuk pemetaan daerah berdasarkan wilayah kecamatan, dan juga untuk dijadikan sebagai laporan dan artikel jurnal.

HASIL DAN PEMBAHASAN

Pada penelitian ini hasil yang didapatkan pada *dataset* yang sudah di analisis dari total survey kuesioner yang didapatkan berjumlah 423. Tabel 1 merupakan deskripsi *dataset*

Atribut	Tipe data	keterangan
Umur	numerik	Umur responden
Jenis Kelamin	nominal	Jenis kelamin responden
Pekerjaan	nominal	Pekerjaan responden
Memakai masker	nominal	Jumlah kedisiplinan proses
Mencuci tangan	nominal	Jumlah kedisiplinan proses
Menjaga jarak	nominal	Jumlah kedisiplinan proses
Hand sanitizer	nominal	Jumlah kedisiplinan proses

Tabel 1. Deskripsi Data

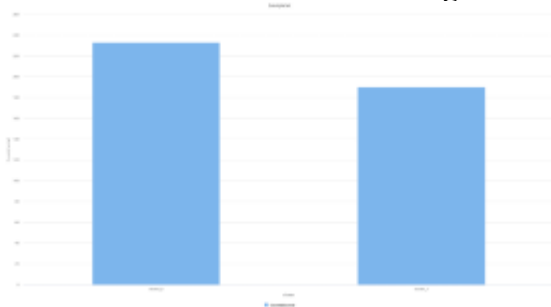
Lalu, atribut yang akan digunakan dalam proses data mining yang sudah ditransformasi yang sebelumnya tipe data nominal menjadi tipe data numerik. Atribut yang dipilih adalah atribut memakai masker, mencuci tangan, menjaga jarak dan *hand sanitizer*. **Tabel 2** dibawah ini adalah *dataset Final* yang akan digunakan.

Tabel 2. *Dataset Final*

Atribut	Tipe data	keterangan
Memakai masker	Numerik	Jumlah kedisiplinan proses
Mencuci tangan	Numerik	Jumlah kedisiplinan proses
Menjaga jarak	Numerik	Jumlah kedisiplinan proses
Hand sanitizer	Numerik	Jumlah kedisiplinan proses

Penelitian ini dilakukan menggunakan algoritme k-means dan *Davies Bouldin Index* untuk menentukan jumlah cluster yang optimal. *Tools* yang digunakan pada penelitian ini menggunakan RapidMiner Studio 9.3 untuk proses *data mining*. Hasil yang didapatkan dalam pengujian jumlah nilai K yang optimal dari K=2 sampai K=6, didapatkan jumlah K=2 adalah yang optimal dalam Gambar 2.

Gambar 2. Visual Hasil Clustering



Gambar 2 menunjukkan bahwa cluster 0 presentase sebesar 55,08 % dan cluster 1 presentase sebesar 44,92 %. Proses *k-means clustering* dapat mengetahui jumlah anggota tiap *cluster*. **Tabel 3** menunjukkan jumlah anggota setiap *cluster*.

Tabel 3. Jumlah Anggota Setiap *Cluster*

	<i>Cluster</i>	Anggota
	0	233
	1	190
Valid		423
Missing		0

Hasil pengelompokkan pada K=2 pada **Tabel 4** dan nilai rata – rata setiap *cluster/final centroid*.

Tabel 4. *Final Centroid*

Atribut	<i>Cluster 0</i>	<i>Cluster 1</i>
Pemakaian Masker	1.030	1.394
Mencuci Tangan	1.270	1.915
Menjaga Jarak	1.613	2.531
Hand Sanitizer	1.300	1.936

Proses selanjutnya adalah perhitungan kualitas cluster dengan menggunakan metode *Silhouette Coefficient* pada **Tabel 5**.

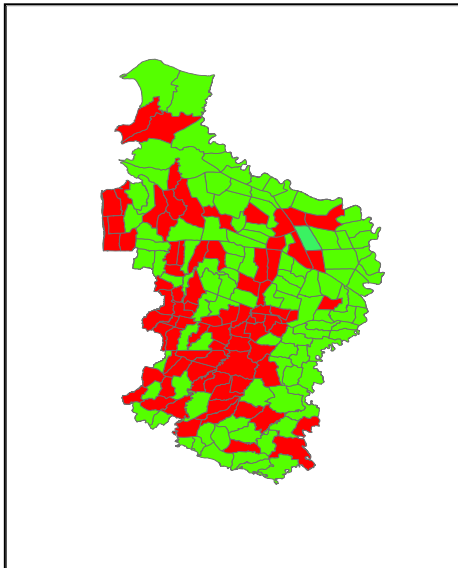
Tabel 5. Hasil Uji SC Jumlah *Cluster* Menggunakan *K-Means*

Data Uji	Nilai K	<i>Silhouette Coefficient</i>
423	2	0,926989
	3	0,512302
	4	0,474802
	5	0,039175
	6	0,121302
	Nilai SC	

Berdasarkan perhitungan pada kualitas cluster pada *dataset* tingkat kedisiplinan warga Bekasi dalam menjalankan protokol kesehatan dengan uji 423 data, pengujian kualitas cluster yang di uji yaitu nilai k=2 sampai k=6. Kualitas cluster paling baik diperoleh pada cluster 2 dengan nilai *Silhouette Coefficient* sebesar 0,926989.

Untuk pemetaan daerah dari hasil clustering sebelumnya menggunakan tools ArcGIS, hasilnya bisa dilihat pada **Gambar 3**.

Gambar 3. Pemetaan Hasil Clustering dengan software ArcGIS 10.8



Cluster tingkat kedisiplinan terhadap protokol kesehatan pada masa pandemik saat ini terbagi menjadi 2 *cluster* dan beberapa *cluster* tersebut terdapat beberapa wilayah per kecamatan yang ada di kota Bekasi.

Tabel 6. Hasil Cluster Dataset

<i>Cluster</i>	Kecamatan	Presentase	Kriteria	Warna
<i>Cluster 0</i>	Cibitung, Babelan, Tambelang, Cikarang Pusat, Cabangbungin, Cikarang Timur, Kedungwaringin, Muaragembong, Cibarusah, Karangbahagia, Pebayuran	55,08 %	Disiplin	Hijau
<i>Cluster 1</i>	Cikarang Selatan,	44,92 %	Tidak disiplin	Merah

	Tambun Selatan, Tambun Utara, Tarumajaya, Setu, Sukawangi			
--	---	--	--	--

Wilayah kecamatan yang masuk ke dalam cluster 0 adalah Cibitung, Babelan, Tambelang, Cikarang Pusat, Cabangbungin, Cikarang Timur, Kedungwaringin, Muaragembong, Cibarusah, Karangbahagia, dan Pebayuran.

Selanjutnya wilayah kecamatan yang masuk ke dalam cluster 1 adalah Cikarang Selatan, Tambun Selatan, Tambun Utara, Tarumajaya, Setu dan Sukawangi.

KESIMPULAN

1. Dari hasil *clustering* perhitungan $k=2$ terdapat lima titik *cluster* yang tersebar pada setiap wilayah Bekasi dengan kategori disiplin dan tidak disiplin dengan melihat dari warna *cluster* dari mulai Hijau dengan presentase sebesar 55,08 % yang paling disiplin dalam menjalankan protokol kesehatan dan warna merah dengan presentase sebesar 44,92 % adalah warna yang paling tidak disiplin dalam menjalankan protokol kesehatan.
2. Hasil *Silhouette Coefficient* dengan metode *K-Means* adalah sebesar 0,926989 menunjukkan bahwa kualitas *cluster* yang baik (*strong structure*) karena mendekati angka 1
3. Pemetaan wilayah tersebut digunakan untuk menginformasikan hasil yang sudah didapat ini dilakukan dengan mengirim ke berbagai sosial media, terutama ke *whatsapp* seperti kuesioner *dataset* yang dibuat sebagai jalan untuk menuju disiplin

dalam menjalankan protokol kesehatan yang lebih baik lagi.

DAFTAR PUSTAKA

- Abadi, S., The, K, S, M., Mohd. Nasir, B., Huda, M., L. Ivanova, N., Sari, T, I., Maseleno, A. Satria, F., Muslihudin, M. (2018) Application Model of K-means Clustering: Insights into Promotion Strategy of Vocational High School. *International Journal of Engineering and Technology*, 7(27), hal. 182–187.
- Asroni, H. Fitri, E.Prasetyo. (Mei 2018). Penerapan Metode *Clustering* dengan Algoritma *K-means* pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *SEMESTA TEKNIKA*, Vol. 21. No. 1, 60-64.
- Arieska, P. K., & Herdiani, N. (2018). Pemilihan Teknik Sampling Berdasarkan Perhitungan Efisiensi Relatif. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 6(2). 166-171.
- Darmi, Yulia., & Setiawan, A. (2016). Penerapan Metode Clustering K-means Dalam Pengelompokan Penjualan Produk. *Jurnal Media Infotama*. 2(12). 148-157.
- Fadillah, A. P. (2015). Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ). *JuTISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 1(3).
- G.D.Rembulan, T.Wijaya, D.Palullungan, K.N.Alfina, dan M.Qurthuby. (2020). Kebijakan Pemerintah Mengenai Coronavirus Disease (COVID-19) Di Setiap Provinsi Di Indonesia Berdasarkan Analisis Klaster. *Journal of Industrial Engineering and Management Systems*, Vol. 13, No. 2, 74-86.
- Hafeez, A., Ahmad, S., Siddqui, S, A., Ahmad, M., Mishra, S. (2020). A Review of COVID-19 (Coronavirus Disease-2019) Diagnosis, Treatments and Prevention. *Eurasian Journal of Medicine and Oncology*, 4(2), hal. 116–125.
- J.Han., M. Kamber., J. Pei. (2012). “Data Mining Concepts and Techniques Third Edition”. San Fransisco: Morgan Kauffman Publisher.
- Larose, D.T. (2005). *Discovering knowledge in data: an introduction to data mining*. John Willey & Sons, Inc.
- Liyas, J. N., & Primadi, R. (2017). Pengaruh disiplin kerja terhadap kinerja karyawan pada bank perkreditan rakyat. *Al-Masraf: Jurnal Lembaga Keuangan dan Perbankan*, 2(1), 17-26.
- Menteri Kesehatan Republik Indonesia. (2020). Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/382/2020 Tentang Protokol Kesehatan Bagi Masyarakat di Tempat dan Fasilitas Umum Dalam Rangka Pencegahan dan Pengendalian Corona Virus Disease 2019 (COVID 19), Jakarta: Departemen Kesehatan.
- Muliono, R., & Sembiring, Z. (2019). Data mining Clustering Menggunakan Algoritma K-means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), 272-279.
- Pramesti, D. F., Furqon, M. T. dan Dewi, C. (2017) “Implementasi Metode KMedoids Clustering Untuk

- Pengelompokan Data,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(9), hal. 723–732.
- Shafique, U., & Qaiser, A. (2014). A Comparative Study of Data mining Process Models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*. 12(1). 217-222
- Siska, S. T. (2016). Analisa Dan Penerapan Data mining Untuk Menentukan Kubikasi Air Terjual Berdasarkan Pengelompokan Pelanggan Menggunakan Algoritma K-means Clustering. *Jurnal Teknologi Informasi dan Pendidikan*, 9(1), 86-93.
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., Zer, F. I. R., & Hartama, D. (2020). Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia. *JurTI (Jurnal Teknologi Informasi)*, 4(1), 166-173.
- Shen, K., Yang, Y., Wang, T., Zhao, D., Jiang, Y., Jin, R., ... & Gao, L. (2020). Diagnosis, treatment, and prevention of 2019 novel coronavirus infection in children: experts’ consensus statement. *World journal of pediatrics*, 16(3), 223-231.