JURNAL LOCUS: Penelitian & Pengabdian
Volume 1 No. 6 September 2022
E-ISSN 2829-7334| P-ISSN 2829-5439
Hompage: https://locus.rivierapublishing.id/index.php/jl

KLASIFIKASI DETEKSI GEJALA AWAL COVID-19 DENGAN METODE *LOGISTIC REGRESSION, RANDOM FOREST CLASSIFIER* DAN *SUPPORT VECTOR MACHINE*

Fauzan Azimah[1], Kiky Rizky Nova Wardani[2]
Fakultas Ilmu Komputer, Universitas Bina Darma[1,2]
191410121@student.binadarma.ac.id[1], kikyrizkynovawardani@binadarma.ac.id[2]

Diterima: 02-09-2022 Review: 10-09-2022 Publish: 15-09-2022

Abstrak:

Coronavirus Disease 19 (COVID-19) merupakan virus baru yang menyebabkan infeksi saluran pernapasan. Virus ini berasal dari hewan yang dapat menular pada manusia dengan percikan air liur. Menurut data epidemiologi rata-rata pasien terjangkit virus ini berusia 15-80 tahun. Virus ini memiliki masa inkubasi 2-14 hari yang mempunyai gejala awal yaitu deman tinggi, sesak nafas, batuk pilek. Indonesia memiliki 2 kasus pertama pada 2 maret 2020. permasalahan yang diangkat dalam penelitian ini adalah bagaimana mengklasifiasi resiko terjangkit virus covid-19 dari gejala yang ditimbulkan. Tujuan penelitian ini untuk mengetahui tingkat resiko terjangkit virus covid-19 berdasarkan instrument yang digunakan dari metode AI Project Cycle yang terdiri dari 6 tahapan yaitu Problem Scoping, Data Acquisition, Data Exploration, Modelling, Evaluation dan Deployment. Dataset yang digunakan peneliti diambil dari web resmi kaggle.com. Penelitin ini menggunakan 3 (tiga) algoritma yaitu Logistic Regression, Random Forest Classfier dan Support Vector Machine. Nilai akurasi pada dataset dengan 6512 rows * 12 columns data pasien terjangkit covid-19 menggunakan algoritma Logistic Regression memperoleh 87%, Random Forest Classifier memperoleh 86% dan Support Vector Machine memperoleh 83%. Pada penelitian ini algoritma klasifikasi Logistic Regression memberikan nilai akurasi yang tertinggi.

Kata kunci: Covid-19; Klasifikasi, *Logistic Regression*; *Random Forest Classifier*; *Support Vector Machine*; *AI Project Cycle*; *kaggle.com*.

Abstract:

*Coronavirus Disease 19 (COVID-19) is a new virus that causes respiratory tract infections. This virus originates from which can be transmitted to humans through animal saliva. According to epidemiological data, the average patient contracted this virus at the age of 15-80 years. This virus has an incubation period of 2-14 days which has symptoms of shortness of breath, cough and runny nose. Indonesia had the first 2 cases on March 2, 2020. The problem raised in this study is how to classify the risk of contracting the Covid-19 virus from the symptoms it causes. The purpose of this study was to determine the level of risk of contracting the covid-19 virus based on the instrument used from the AI Project Cycle method which consists of 6 stages, namely Problem Scoping, Data Acquisition, Data Exploration, Modeling, Evaluation and Implementation. The dataset used by the researcher was taken from the official website kaggle.com. This study uses 3 (three) algorithms, namely Logistic Regression, Random Forest Classifier and Support Vector Machine. The accuracy value in the dataset with 6512 rows * 12 columns of data on patients infected with COVID-19 using the Logistic Regression algorithm was obtained 87%, the Random Forest Classifier was obtained 86% and the Support Vector Machine was obtained 83%. In this study the Logistics Regression classification provides the highest accuracy value.*

***Keywords**: Covid-19; Classification; Logistic Regression; Random Forest Classifier; Support Vector Machine; AI Project Cycle, kaggle.com.*

Corresponding: **Fauzan Azimah**
E-mail: 191410121@student.binadarma.ac.id

PENDAHULUAN

World Health Organization (WHO), memberitahukan kasus baru Pneumonia di kota Wuhan, Hubei, China yang mengidentifikasi jenis baru novel Corona Virus.Nama Corona virus Disease 2019 resmi ditetapkan oleh WHO (Adrian, Putra, Rafialdy, & Rakhmawati, 2021). Menurut ahli virologi dari

China, virus covid-19 ini merupakan virus yang berbeda dengan Severe Acute Respiratory Syndrome Associated Coronavirus (SARS COV2) yang muncul di Guandong, China tahun 2003 tetapi memiliki gajala yang sama (Adrian et al., 2021). Tingkat penyebaran Covid-19 lebih luas dibandingkan dengan SARS namun tingkat kematian SARS mencapai 9,6% dibanding tingkat kematian Covid-19 yang masih dibawah 5% (Adrian et al., 2021). Homologi Covid-19 mempunyai ciri-ciri DNA yang mirip hingga mencapai 85% dengan kelelawar SARS. Penularan virus ini dari hewan ke manusia disebut Transmisi Zoonosis dan dapat tertular dari manusia ke manusia dengan berkontak langsung atau terkena percikan liurnya (Anggraini, Akbar, Wijaya, Syaputra, & Sobri, 2021). Dari data pertama di wuhan 15% menunjukan kasus fatal usia diatas 80 tahun, 8,0% berusia 70 tahun, 1% anak dibawah 15 tahun. Sedangkan kasus ringan dan berat yang mempunyai penyakit bawaan 49,0% (Adrian et al., 2021). Gejala-gejala COVID-19 yang paling umum adalah demam, batuk kering, dan rasa Lelah (Linssen et al., 2020). Gejala lainnya yang lebih jarang dan mungkin dialami beberapa pasien meliputi rasa nyeri dan sakit, hidung tersumbat, sakit kepala, konjungtivitis, sakit tenggorokan, diare, kehilangan indera rasa atau penciuman, ruam pada kulit, atau perubahan warna jari tangan atau kaki. Gejala-gejala yang dialami biasanya bersifat ringan dan muncul secara bertahap. Beberapa orang menjadi terinfeksi tetapi hanya memiliki gejala ringan.Sebagian besar (sekitar 80%) orang yang terinfeksi berhasil pulih tanpa perlu perawatan khusus (Nafi'ah, 2021). Sekitar 1 dari 5 orang yang terinfeksi COVID-19 menderita sakit parah dan kesulitan bernapas. Orangorang lanjut usia (lansia) dan orang-orang dengan kondisi medis penyerta seperti tekanan darah tinggi, gangguan jantung dan paru-paru, diabetes, atau kanker memiliki kemungkinan lebih besar mengalami sakit lebih serius[WHO]. Namun, masyarakat tidak mengetahui perbedaan antara penyakit gejala seperti f biasa dengan penyakit flu indikasi Covid-19. Ada 3 gejala utama yang dapat muncul pada virus Covid-19 ini, yaitu : Demam tinggi, Batuk, Sesak nafas (Ramadhy & Sibaroni, 2022). Pasien juga bisa mengalami gangguan pengecapan atau penciuman, nyeri otot, sakit kepala, sakit tenggorokan, pilek, diare, mual, dan muntah. Namun, gejala ini tidak selalu terjadi pada pasien COVID-19. Pada kasus yang parah, infeksi virus Corona bisa menyebabkan komplikasi yang serius, seperti sindrom gangguan pernapasan akut, pneumonia (infeksi paru) yang berat, edema paru, dan kegagalan fungsi organorgan tubuh, misalnya ginjal. Gejala infeksi virus Corona yang berat ini lebih sering terjadi pada lansia dan orang yang memiliki kondisi medis tertentu (Covid-19, 2020).

Pada masa-masa seperti ini, gejala-gejala diatas mungkin memang sulit dibedakan. Jadi, ketika kamu merasa tidak enak badan dan mengalami gejala-gejala di atas, sebaiknya kamu segera melakukan tes pendeteksi COVID-19. Maka dari itu penulis berinisiatif untuk membuat program yang sederhana yaitu deteksi gejala awal Covid -19 berdasarkan gejala-gejala awal dengan menggunakan tiga algoritma, seperti Logistic Regression dengan akurasi 87%, Random Forest Classifier dengan akurasi 86%, dan Support Vector Machine dengan akurasi 83%. Dengan adanya fitur berbasis website sehingga masyarakat lebih mudah untuk mengeceknya hanya melalui website saja berdasarkan gejala-gejala awal Covid-19 dengan menggunakan metode pengembangan AI Project Cycle

METODE PENELITIAN

Pada penelitian ini, penulis membuat sistem pendeteksi gejala awal covid-19 dengan menggunakan metode AI Project Cycle untuk mendapatkan hasil prediksi apakah pasien terpapar covid-19 atau tidak dengan menggunakan tiga model yaitu Logistic Regression, Random Forest Classifier dan Support Vector Machine.

AI Project Cycle

AI Project Cycle adalah sebuah proses dalam membuat proyek AI secara utuh. Pada tahap penelitian peneliti ini menggunakan siklus AI Project Cycle untuk mengklasifikasi pasien yang terpapar Covid-19 atau tidak dengan mencari pola data pada dataset yang sudah di unduh di web resmi www.kaggle.com. Dibawah ini merupakan tahapan pelaksanaan dari siklus AI Project Cycle sebagai berikut:

Problem Scoping

Proses identifikasi atau memetakan batasan masalah yang ingin diselesaikan sehingga tujuan atau target menjadi semakin jelas dan lebih terarah serta akan lebih mudah untuk menemukan solusi [7]. Mengapa penting? Agar saat proses penelitian atau pengerjaan bisa lebih fokus sesuai tujuan dan rencana awal. Perlu diperhatikan apabila masalah yang diambil terlalu besar, biasanya cenderung sulit untuk dimulai atau diimplementasikan. Namun, apabila terlalu sempit akan sulit mencapai tujuan dan target [8]. Ada metode 4W untuk mempermudah proses problem scoping, yaitu:

A. Who : Siapa saja yang terlibat dalam masalah tersebut.
B. What : Apa masalah dan faktor pendukung masalah.
C. Where : Kondisi, Suasana atau tempat masalah yang diamati.
D. Why : Alasan mengapa masalah tersebut perlu diselesaikan dan apa manfaatnya

Data Acquisition

Proses mengumpulkan data-data yang dibutuhkan untuk membuat proyek AI. Hal ini merupakan dasar atau bahan yang selanjutnya diolah untuk dianalisis sesuai masalah dan diamati agar bisa menghasilkan solusi terbaik. Ada beberapa cara untuk mendapatkan sumber data tersebut, yaitu:

A. Tools/Alat : Kamera, Microphone dan Sensor.
B. Observasi : Survei, Penelitian.
C. Open Data : BPS, Kaggle, API (REST API, Twitter API, Youtube API).
D. Web Scraping/Crawling.

Data Exploration

Proses menjelajahi dataset untuk memahami isi, komponen dan karakteristiknya sehingga kita dapat mengetahui pola data tersebut. Exploratory Data Analysis (EDA) diperkenalkan oleh John Tukey dan mempunyai tujuan untuk mendorong ahli statistik dalam mengeksplorasi data dan merumuskan hipotesis (Tukey, 1977). Beberapa metode yang digunakan dalam eksplorasi data, yaitu:

A. Summary Descriptive Statistics, rangkuman properti (frekuensi, modus, mean, median, range, variance dan standar deviasi) dalam data.
B. Visualization, penyajian data dalam bentuk grafis (Bar Chart, Histogram, Box Plot, Scatter Plot, Star Plot, Chernoff Plot, Maps).
C. Clustering dan Anomaly Detection

Sederhananya menggunakan elemen visual, pembaca akan lebih memahami pola, tren bahkan lebih mudah menemukan outlier pada suatu data.

Modelling

Proses pembuatan algoritma dalam bahasa pemrograman sebagai metode pembelajaran mesin (training phase) yang digunakan untuk menemukan pola-pola dalam data sebagai bahan dasar

pengetahuan sistem untuk membuat keputusan atau melakukan prediksi (Mulajati, 2017). Pemodelan yang dibuat pada website ini menggunakan tiga model yaitu :

1. Logistik Regression Logistic Regression adalah algoritma klasifikasi machine learning yang digunakan untuk memprediksi probabilitas variabel dependen kategoris. Dalam Logistic Regression, variabel yang terikat adalah variabel biner yang berisi data berkode 1 (Ya) atau 0 (Tidak). Metode ini merupakan metode regresi linier umum untuk mempelajari pemetaan dari sejumlah variabelnumerik ke variabel biner atau probabilistic (Ramadhy & Sibaroni, 2022).
2. Random Forest Classifier Random Forest, merupakan sebuah metode yang dikembangkan dari metode CART (Classification and Regression Trees), yang juga merupakan metode atau algoritma dari teknik pohon keputusan (Adrian et al., 2021). Yang membedakan metode random forest dari metode CART adalah Random Forest menerapkan metode bootstrap aggregating (bagging) dan juga seleksi fitur random atau bisa disebut random feature selection (Purnomo, 2017). Random Forest adalah kombinasi dari masing masing teknik pohon keputusan yang ada, lalu kemudian digabung dan dikombinasikan kedalam suatu model. Ada tiga poin utama dalam metode Random Forest, tiga poin utama tersebut yaitu melakukan bootstrap sampling untuk membangun pohon prediksi, masing-masing pohon keputusan memprediksi dengan prediktor acak, kemudian Random Forest melakukan prediksi dengan mengombinasikan hasil dari tiap tiap pohon keputusan dengan cara majority vote untuk klasifikasi atau rata-rata untuk regresi.
3. Support Vector Machine Support Vector Machine atau SVM merupakan sekumpulan metode supervised learning yang membuat hyperlane atau sekumpulan hyperlane pada proses klasifikasi, regresi, dan outlier detection (Prabiantissa, 2021). Salah satu penggunaanya adalah dalam mengelompokkan text dan hypertext. Kelebihan pada SVM ini adalah efektif pada high dimensional space, efektif dalam kasus dengan jumlah dimensi yang lebih banyak daripada jumlah sampelnya, menggunakan subset titik pelatihan sehingga lebih memori efisien.

Evaluation

Proses pengkajian dan pemilihan model terbaik yang akan digunakan untuk membuat proyek AI. Salah satu metode yang digunakan yaitu Confusion Matrix menggunakan tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai actual (Nasir, Junita, & Ilham, 2014). Proses pemilihan model memperhatikan beberapa komponen, antara lain:

A. Accuracy/Akurasi : persentase nilai prediksi yang benar dari keseluruhan pengamatan. Rumus → (TP+TN)/(TP+FP+TN+FN)
B. Precision/Presisi : persentase kasus yang diprediksi AI dan memang terjadi berdasarkan realitanya. Rumus → (TP)/(TP+FP)
C. Recall : mengukur pecahan kasus yang terjadi dan diprediksi tepat oleh AI. Rumus → (TP)/(TP+FN)

Deployment

Proses implementasi AI pada sebuah aplikasi atau sistem sesuai dengan tujuan pembuatan produk sehingga diharapkan dapat memudahkan pekerjaan manusia untuk mengakses websitenya (Romindo et al., 2019). Ada banyak tempat penyedia untuk mendeployment/hosting sebuah website diantara: Heroku, Vercel, Github, Netlify.

COVID-19

COVID-19 adalah penyakit yang disebabkan oleh Novel Coronavirus (2019- n (Jongh, 2020)CoV), jenis baru coronavirus yang diidentifikasi untuk pertama kalinya di Wuhan, Cina, dinamai "penyakit coronavirus 2019" (COVID19) - " CO "untuk corona," VI "untuk virus dan" D "untuk penyakit dalam bahasa Inggris" (Dejongh, 2020).

COVID-19 adalah penyakit menular yang disebabkan oleh jenis corona virus yang baru ditemukan. Ini merupakan virus baru dan penyakit yang sebelumnya tidak dikenal sebelum terjadi wabah di Wuhan, Tiongkok, bulan Desember 2019. Gejala-gejala COVID-19 yang paling umum adalah demam, rasa lelah, dan batuk kering. Beberapa pasien mungkin mengalami rasa nyeri dan sakit, hidung tersumbat, pilek, sakit tenggorokan, diare, dan kehilangan indra penciuman. Gejala-gejala yang dialami biasanya bersifat ringan dan muncul secara bertahap. Beberapa orang yang terinfeksi tidak menunjukkan gejala apapun dan tetap merasa sehat. Sebagian besar (sekitar 80%) orang yang terinfeksi berhasil pulih tanpa perlu perawatan khusus. Sekitar 1 dari 6 orang yang terjangkit COVID19 menderita sakit parah dan kesulitan bernapas. Orang-orang yang lanjut usia dan orang-orang dengan kondisi medis yang sudah ada sebelumnya seperti tekanan darah tinggi, gangguan jantung atau diabetes, punya kemungkinan lebih besar mengalami sakit lebih serius. Mereka yang mengalami demam, batuk dan kesulitan bernapas sebaiknya mencari pertolongan medis. Orang dapat tertular COVID-19 dari orang lain yang terjangkit virus ini. COVID-19 dapat menyebar dari orang ke orang melalui percikan-percikan dari hidung atau mulut yang keluar saat orang yang terjangkit COVID-19 batuk atau mengeluarkan napas. Percikan-percikan ini kemudian jatuh ke benda-benda dan permukaan-permukaan di sekitar. Orang yang menyentuh benda atau permukaan tersebut lalu menyentuh mata, hidung atau mulutnya, dapat terjangkit COVID-19. Penularan COVID-19 juga dapat terjadi jika orang menghirup percikan yang keluar dari batuk atau napas orang yang terjangkit COVID-19. Oleh karena itu, penting bagi kita untuk menjaga jarak lebih dari 1 meter dari orang yang sakit. WHO terus mengkaji perkembangan penelitian tentang cara penyebaran COVID-19 dan akan menyampaikan temuan- temuan terbaru. [WHO, 2020].

KLASIFIKASI

Klasifikasi data adalah suatu proses yang menentukan property-property yang sama pada sebuah himpunan obyek di dalem sebuah basis data mengklasifikasikannya ke dalam kelas-kelas yang berbada menurut model klasifikasi yang ditetapkan . Tujuan dari klasifikasi adalah untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belom diketahui sebelumnya.

HASIL DAN PEMBAHASAN

Hasil Pengembangan Produk

Penelitian yang dilakukan penulis adalah sistem pendeteksi covid-19 berdasarkan gejala-gejala awal dengan pendekatan metode AI Project Cycle yang menghasilkan sebuah web untuk memprediksi berapa persen orang yang terpapar covid-19 ini. Adapun hasil penelitian yang diperoleh adalah sebagai berikut:

1. Problem Scoping

   Ada 4 langkah dalam menentukan problem scoping, yaitu:

Who (Siapa yang terkena dampak dari masalah tersebut) : Kami membuat sistem pendeteksi gejala awal covid-19 ini untuk semua masyarakat Indonesia karena ini sifatnya free website jadi kapanpun dan dimanapun user bisa mengaksesnya.

What (Apa permasalahan tersebut) : Berdasarkan latar belakang permasalahan, masyarakat banyak yang belum mengetahui penyakit gejala biasa seperti flu biasa dengan flu yang terindikasi covid-19. Jadi, ketika kamu merasa tidak enak badan dan mengalami gejala-gejala Covid-19, sebaiknya kamu segera melakukan tes pendeteksi COVID-19, sehingga kami berinisiatif membuat sistem pendeteksi gejala awal covid-19 yang sederhana ini.

Where (Lokasi Permasalahan) : Setelah membaca beberapa referensi dari jurnal dan diskusi kami mengangkat permasalahan ini berdasarkan kasus Covid-19 di Indonesia karena kasusnya terus meningkat dan berkembang.

Why (Mengapa permasalahan dibuat) : Untuk memudahkan masyarakat mengecek apakah user terpapar covid-19 atau tidak.

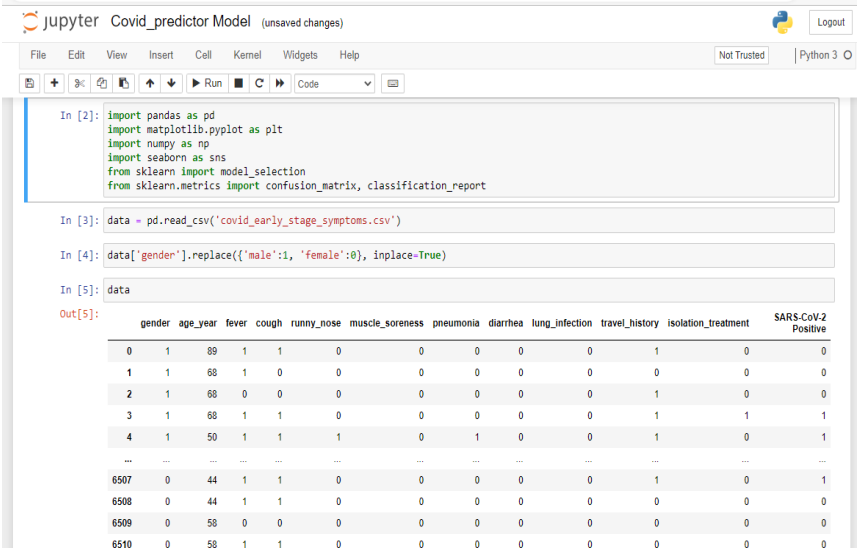
2. Data Acquistion

Proses mengumpulkan data-data yang dibutuhkan untuk membuat project AI. Hal ini merupakan dasar atau bahan yang selanjutnya diolah untuk dianalisis sesuai masalah dan diamati agar bisa menghasilkan solusi terbaik. Adapun tools yang gunakan untuk mendapatkan sumber data dalam proses pembuatan sistem pendeteksi covid-19 berdasarkan gejala-gejala awal yaitu dari situs resmi kaggle dengan mengambil dataset covid-19. Adapun link download dataset yang kami gunakan sebagai berikut :

Early stage symptoms of COVID-19 patient's | Kaggle

3. Data Exploration

Didalam proses data exploration, penulis memvisualisasikan dataset covid-19 ini dalam bentuk table dan grafik batang.

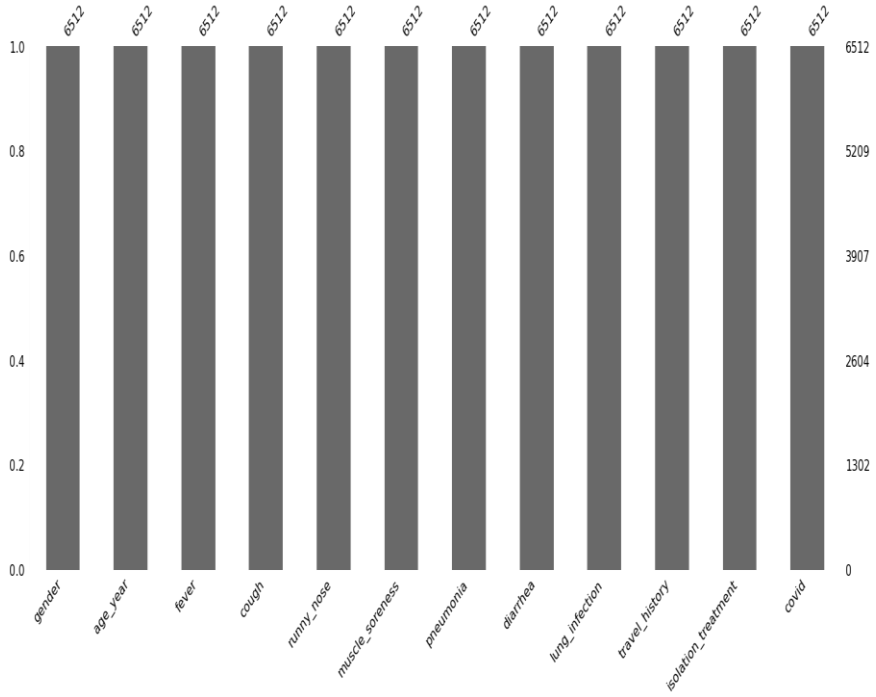


	gender	age_year	fever	cough	runny_nose	muscle_soreness	pneumonia	diarrhea	lung_infection	travel_history	isolation_treatment	SARS-CoV-2 Positive
0	1	89	1	1	0	0	0	0	0	1	0	0
1	1	68	1	0	0	0	0	0	0	0	0	0
2	1	68	0	0	0	0	0	0	0	1	0	0
3	1	68	1	1	0	0	0	0	0	1	1	1
4	1	50	1	1	1	0	1	0	0	1	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...
6507	0	44	1	1	0	0	0	0	0	1	0	1
6508	0	44	1	1	0	0	0	0	0	0	0	0
6509	0	58	0	0	0	0	0	0	0	0	0	0
6510	0	58	1	1	0	0	0	0	0	0	0	0

Gambar 3. 1 Menampilkan isi Dataset

Didalam dataset ini terdapat 6512 rows * 12 columns. Columns tersebut terdiri dari gender, age_year, fever, cough, runny_nose, muscle_soreness, pneumonia, diarhea,

lung_infection, travel_history, isolation_treatment dan covid. Adapun hasil visualisasi dengan grafik batang seperti pada gambar dibawah ini.

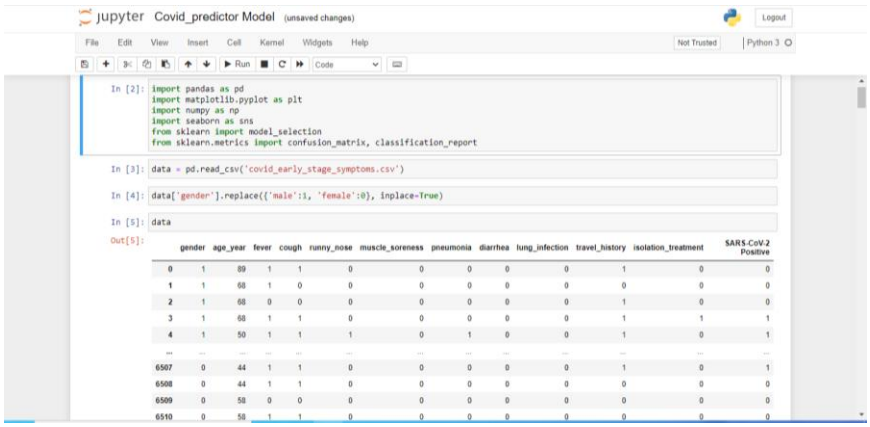


Gambar 3. 2 Grafik Batang

4. Modelling

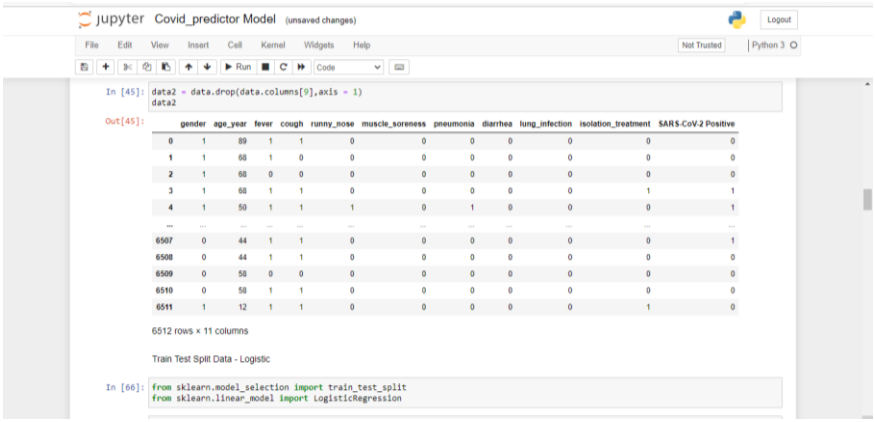
Ada 3 model yang penulis pilih untuk sistem pendeteksi covid-19 berdasarkan gejala-gejala awal yaitu : Logistic Regression, Random Forest Classifier dan Support Vector Machine. Adapun hasil dari tiga model tersebut yaitu :

A. Logistic Regression



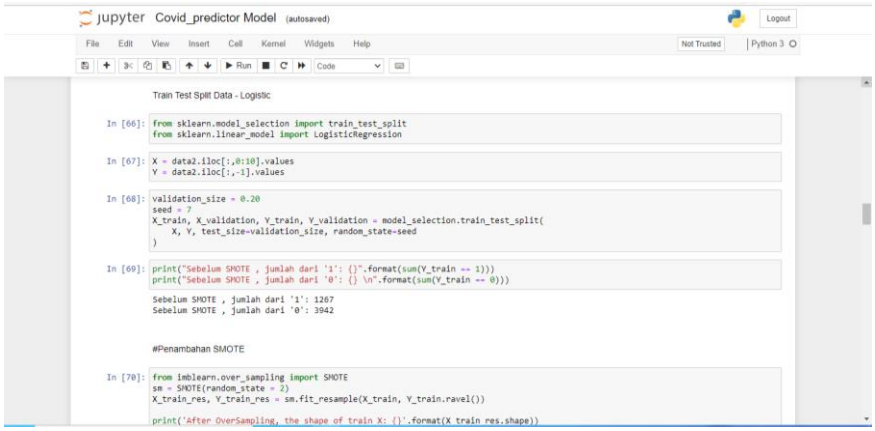
Gambar 3. 3 Membuat Label Gender 0 dan 1

Pertama kami membuat label pada column gender 0 artinya “female” dan 1 artinya “male” karena sifat pada model ini yaitu bilangan biner.



Gambar 3. 4 Drop Columns

Pada gambar ini kami menghapus data pada column 9 yaitu lung_infection karena tidak dipakai dalam pertanyaan web ini.



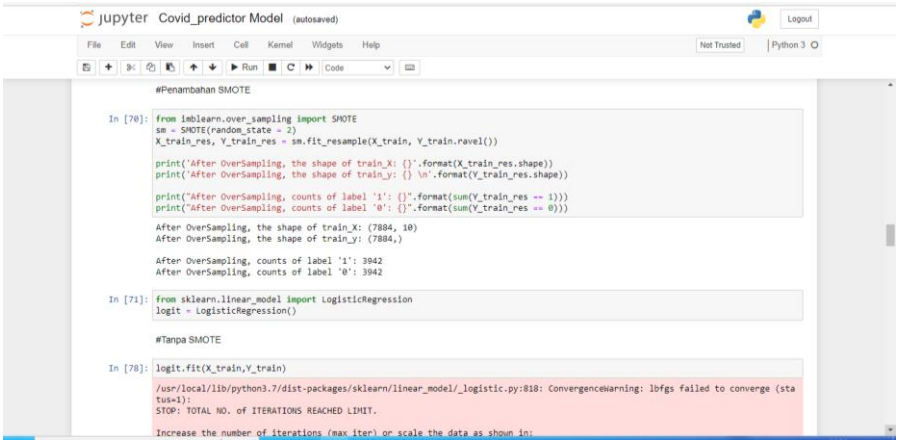
Gambar 3. 5 Train dataset

Pada gambar ini kami membagi data training X dan data test Y untuk melihat hasil accuracy. Sebelum di SMOTE hasilnya:

Table 3. 1 Hasil Sebelum Smote

Label	Gender	Jumlah
	Female	3942
	Male	1267

Karena label 0 dan 1 tidak sama jumlahnya maka dari itu kami melakukan bimbingan dengan mentor dan hasilnya adalah dilakukan penambahan SMOTE agar jumlah label 0 dan 1 sama.



Gambar 3. 6 Penambahan SMOTE

Pada gambar ini telah dilakukan penambahan SMOTE dengan mengimport library over_sampling yang di random_state dimulai dari 2, maka hasilnya :

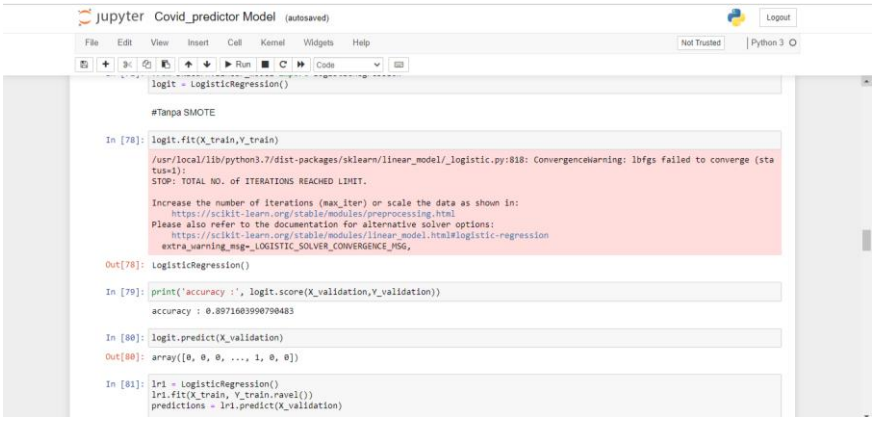
Table 3. 2 Hasil Setelah Smote

Over Sampling	Jumlah
Train_x	7884,10
Train_y	7884

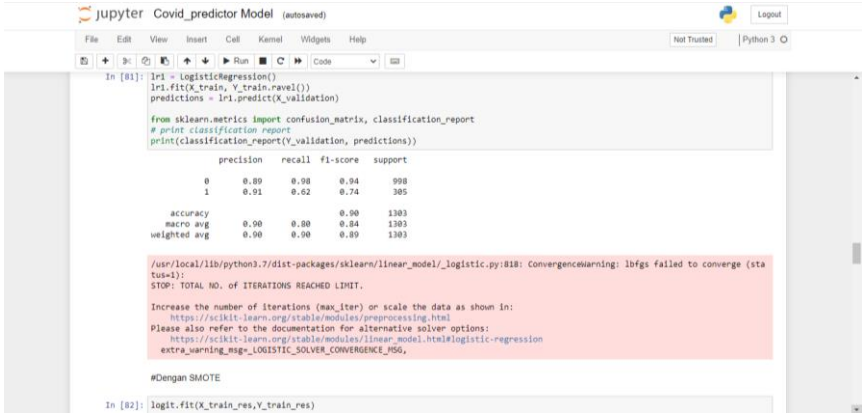
Setelah sama hasil train_x dan train y maka hasil dari label 0 dan 1 juga sama seperti table dibawah ini.

Table 3. 3 Hasil Sample

Label	Gender	Jumlah
	Female	3942
	Male	3942



Pada gambar ini, penulis menampilkan hasil accuracy tanpa SMOTE yang hasilnya 89%. Untuk hasil lebih lengkapnya lihat gambar dibawah ini.



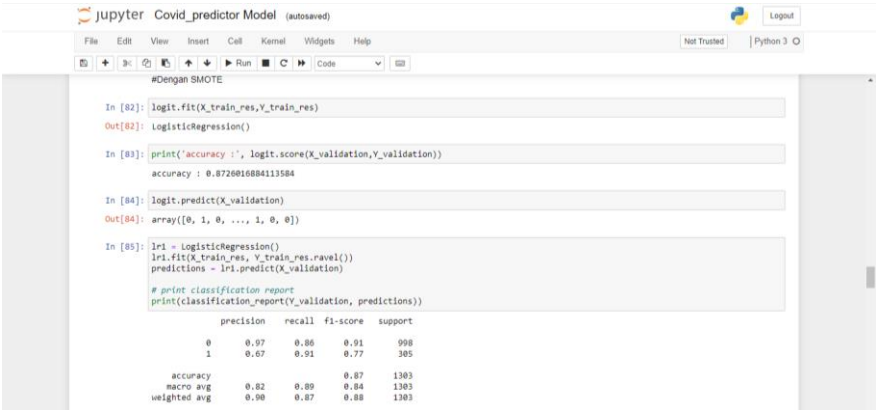
Gambar 3. 7 Hasil tanpa Smote

Mendapatkan hasil sebagai berikut :

Table 3. 4 Hasil accuracy tanpa Smote

Label	Precision	Recall	F1-score	Support
	0.89	0.98	0.94	998
	0.91	0.62	0.74	305

Label 0 : Precision 0.89, Recall 0.98, F1_score 0.94 dan support 998.
Label 1 : Precision 0.91, Recall 0.62, F1_score 0.74 dan support 305.
Perbandingannya sangat jauh untuk label 1 hasil Recallnya menurun sekitar 3% dan F1_score menurun sekitar 2%.
Jika dilakukan dengan SMOTE, maka hasilnya sebagai berikut:



Gambar 3. 8 Hasil pakai Smote

Hasil dari akurasinya setelah menambahkan SMOTE adalah 87% yang sedikit menurun 2% dari sebelum SMOTE 89%. Untuk hasil lebih lengkapnya lihat table dibawah ini.

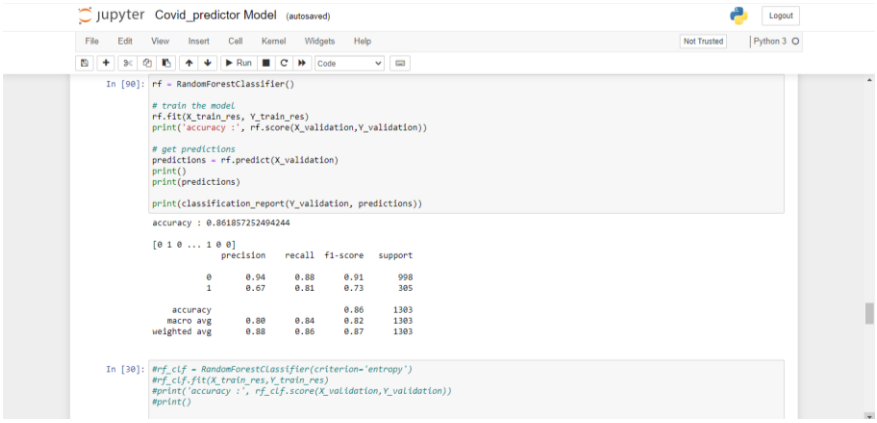
Table 3. 5 Hasil Accuracy dengan Smote

Label	Precision	Recall	F1-score	Support
	0.89	0.98	0.94	998
	0.91	0.62	0.74	305

Perbandingan sama dengan hasil sebelum di SMOTE, tetapi accuracynya menurun 2% menjadi 87%.

B. Random Forest Classifier

Setelah melakukan proses model logistic regression, penulis langsung train data_x dan train data_y untuk melihat accuracy pada model ini.



Gambar 3. 9 Train dataset

Hasil dari accuracy nya adalah : 86%. Untuk melihat hasil lengkapnya seperti table dibawah ini:

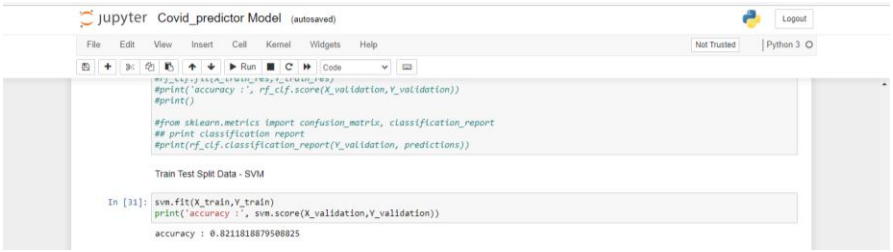
Table 3. 6 Hasil Accuracy

Label	Precision	Recall	F1-score	Support
	0.94	0.88	0.91	998
	0.67	0.81	0.73	305

Hasil perbandingannya pada label 1 precisionnya menurun hingga 3% yaitu 0.67, Recall menurun sedikit 0.81, F1_score menurun sekitar 2%. Hasil prediksi label 1 sedikit tidak akurat karena penurun yang signifikan.

C. Support Vector Machine

Setelah melakukan proses model Random Forest Classifier, penulis langsung train data_x dan train data_y untuk melihat hasil accuracy pada model ini.



Gambar 3. 10 Train dataset

Hasil dari accuracy model Support Vector Machine adalah 82%.

Hasil akurasi dari 3 (tiga) model seperti table dibawah ini:

Table 3. 7 Hasil Accuracy dari 3 model

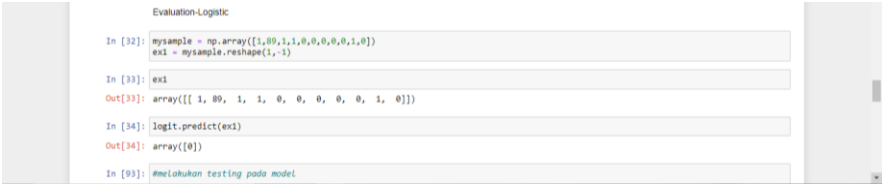
Modelling	Accuracy
Logistic Regression	87%
Random Forest Classifer	86%
Support Vector Machine	82%

5. Evaluation

Setelah melakukan proses pembuatan modelling, penulis melakukan evaluation pada tiga model tersebut dengan confision matrix untuk melihat apakah user positif covid-19 atau negatif covid-19. Adapun hasil dari evaluation tiga model tersebut, yaitu :

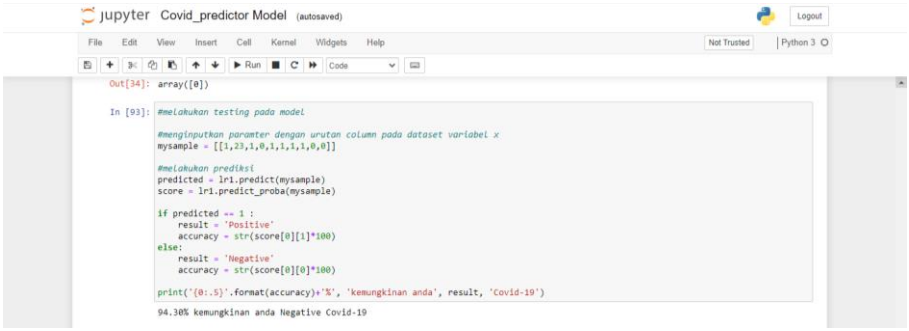
A. Logistic Regression

Untuk mengevaluasi model ini, penulis langsung mengambil beberapa sampel pada dataset.



Gambar 3. 11 sample data

Lalu kami melakukan testing pada model, seperti gambar dibawah ini:

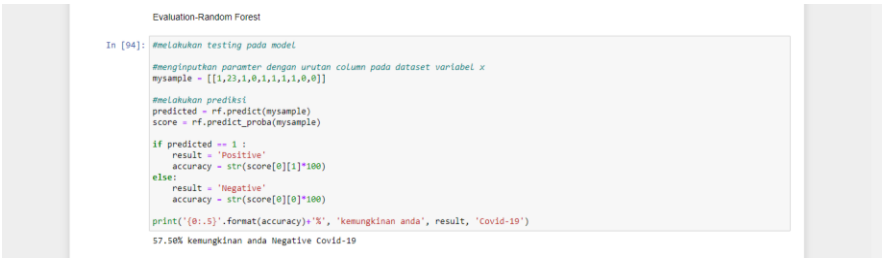


Gambar 3. 12 Hasil Prediksi Logistic Regression

Hasil prediksi yang diambil dari beberapa sample pada dataset adalah 94.30% kemungkinan anda Negative Covid-19.

B. Random Forest Classifier

Untuk mengevaluasi model ini, penulis langsung mengambil beberapa sampel pada dataset.

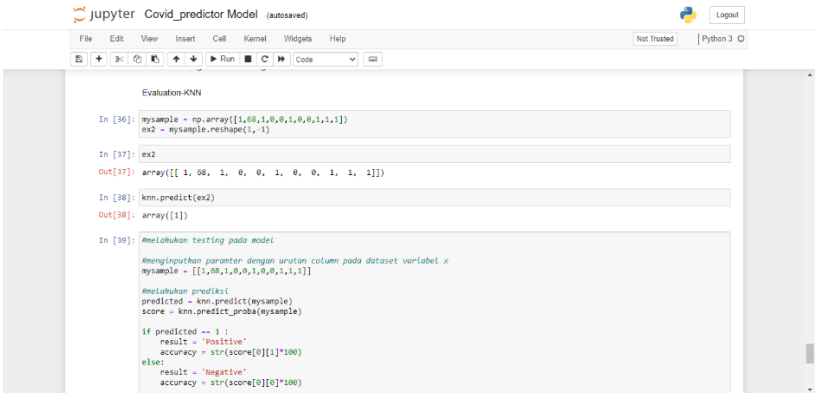


Gambar 3. 13 Hasil Prediksi Random Forest

Penulis, melakukan testing pada model, hasilnya adalah 57.50% kemungkinan anda Negative Covid-19.

C. Support Vector Machine

Untuk mengevaluasi model ini, penulis langsung mengambil beberapa sampel pada dataset.



Gambar 3. 14 Hasil prediksi SVM

Lalu kami melakukan testing pada model, hasilnya adalah 60% kemungkinan anda Positive Covid-19.

6. Deployment

Model yang telah dibuat nantinya akan deploy dalam bentuk website dengan tujuan agar model ini bisa diakses dengan mudah oleh user. Deployment dimulai dengan mendesign UI/UX dari website project akhir kami dengan mengambil beberapa referensi sehingga nantinya user dapat menggunakannya dengan lebih mudah dan dipahami. Tahap selanjutnya adalah proses development website di mana kami memanfaatka tiga Bahasa pemrograman yaitu HTML, JS, dan CSS. Proses ini ditujukan agar website yang dibangun dapat memfasilitasi user apakah user tersebut terkana covid - 19 atau tidak. Proses selanjutnya adalah pembuatan back-end dan integrasi. Proses ini menggunakan flask dan proses ini ditujukan agar website dapat berjalan secara local. Proses terakhir yaitu menghubungkan website yang sebelumnya berjalan secara local menjadi berjalan secara online atau cloud dengan memanfaatkan Heroku.

KESIMPULAN

Kesimpulan dari penelitian ini adalah: 1. Klasifikasi Covid-19 menggunakan Metode AI Project Cycle dengan menggunakan 3 Model yaitu Logistic Regression, Random Forest Classifier dan Support Vector Machine dapat membantu untuk mendeteksi penyakit covid-19. 2. Klasifikasi Covid-19 menggunakan metode logistic regression, random forest classifier dan support Vector Machine memiliki tingkat akurasi lebih dari 80% disertai precision dan recall yang tinggi dengan menambahkan smote dan melakukan drop variabel pada preprocessing. 3. Klasifikasi Covid-19 dapat diimplementasikan secara online menggunakan heroku untuk digunakan secara publik.

DAFTAR PUSTAKA

Adrian, Muhammad Rivza, Putra, Muhammad Papuandivitama, Rafialdy, Muhammad Hilman, & Rakhmawati, Nur Aini. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB. *Jurnal Informatika Upgris*, *7*(1).

Anggraini, Suci, Akbar, Muhamad, Wijaya, Alex, Syaputra, Hadi, & Sobri, Muhammad. (2021). Klasifikasi Gejala Penyakit Coronavirus Disease 19 (COVID-19) Menggunakan Machine Learning. *Journal of Software Engineering Ampera*, *2*(1), 57–68.

Covid-19, Satuan Tugas Penanganan. (2020). Gejala Umum | Satgas Penanganan Covid-19.

Linssen, Joachim, Ermens, Anthony, Berrevoets, Marvin, Seghezzi, Michela, Previtali, Giulia, Russcher,

Henk, Verbon, Annelies, Gillis, Judith, Riedl, Jürgen, & de Jongh, Eva. (2020). A novel haemocytometric COVID-19 prognostic score developed and validated in an observational multicentre European hospital-based study. *Elife*, *9*, e63195.

Mulajati, Muhammad. (2017). *IMPLEMENTASI TEKNIK WEB SCRAPING DAN KLASIFIKASI SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DAN ASOSIASI TEKS (Studi Kasus: Data Ulasan Penumpang Maskapai Penerbangan Garuda Indonesia Pada Situs TripAdvisor)*.

Nafi'ah, H. (2021). AI Project Cycle.

Nasir, Azwir, Junita, Mega, & Ilham, Elfi. (2014). *Pengaruh profitabilitas, pertumbuhan aset, operating leverage, dan ukuran perusahaan terhadap struktur modal studi empiris pada perusahaan food and beverages yang terdaftar di bursa efek indonesia periode 2010-2012*. Riau University.

Prabiantissa, Citra Nurina. (2021). Klasifikasi pada Dataset Penyakit Hati Menggunakan Algoritma Support Vector Machine, K-NN, dan Naïve Bayes. *Prosiding Seminar Nasional Teknik Elektro, Sistem Informasi, Dan Teknik Informatika (SNESTIK)*, *1*(1), 263–268.

Purnomo, Rakhmat. (2017). Penerapan Greedy Forward Selection dan Bagging pada Logistic Regression untuk Prediksi Cacat Perangkat Lunak. *Jurnal Karya Ilmiah*, *17*(2), 1–11.

Ramadhy, Izzan Faikar, & Sibaroni, Yuliant. (2022). Analisis Trending Topik Twitter dengan Fitur Ekspansi FastText Menggunakan Metode Logistic Regression. *JURIKOM (Jurnal Riset Komputer)*, *9*(1), 1–7.

Romindo, Romindo, Muttaqin, Muttaqin, Saputra, Didin Hadi, Purba, Deddy Wahyudin, Iswahyudi, M., Banjarnahor, Astri Rumondang, Kusuma, Aditya Halim Perdana, Effendy, Faried, Sulaiman, Oris Krianto, & Simarmata, Janner. (2019). *E-Commerce: Implementasi, Strategi dan Inovasinya*. Yayasan Kita Menulis.

Tukey, John W. (1977). *Exploratory data analysis* (Vol. 2). Reading, MA.