

MAKALAH BABAK PENYISIHAN LOMBA DATA SCIENCE

Analisis Dampak COVID-19 Terhadap Kehidupan Sosial Ekonomi Masyarakat Indonesia

Shenzhen Tech

MATHEMATICAL CHALLENGE FESTIVAL INSTITUT TEKNOLOGI BANDUNG 2022



Disusun oleh :

Dharma Setiawan

Jefferson Joseph Tedjojuwono

Fikri Diva Sambasri

DAFTAR ISI

BAB I : PENDAHULUAN	3
1.1 Latar Belakang.....	3
1.2 Rumusan Masalah.....	4
1.3 Tujuan.....	4
BAB II : PEMBAHASAN	4
2.1 Data Cleaning.....	4
2.2 Exploratory Data Analysis.....	4
2.3 Machine Learning Deployment.....	15
2.4 Hasil Prediksi.....	21
BAB III : PENUTUP	21
3.1 Kesimpulan dan Saran.....	21
Daftar Pustaka	23

BAB I : PENDAHULUAN

1.1 Latar Belakang

COVID-19 adalah penyakit menular yang disebabkan oleh jenis coronavirus yang baru ditemukan. Coronavirus sendiri merupakan suatu kelompok virus yang dapat menyebabkan penyakit pada hewan atau manusia. Beberapa jenis coronavirus diketahui menyebabkan infeksi saluran nafas pada manusia. Coronavirus merupakan jenis baru yang ditemukan menyebabkan penyakit COVID-19. Virus baru dan penyakit yang disebabkan ini tidak dikenal sebelum mulai nya wabah di Wuhan, Tiongkok, bulan Desember 2019.

Penyebaran COVID-19 di Indonesia mulai didata pada tanggal 1 Maret 2020. Virus ini menyebar dengan sangat cepat ke seluruh provinsi di Indonesia. Jakarta, sebagai pusat ibukota, terkena dampak yang paling parah dari penyebaran virus ini. Pemerintah Indonesia telah melakukan tindakan untuk mencegah penyebaran virus ini salah satunya adalah PSBB (Pembatasan Sosial Berskala Besar).

Penyebaran COVID-19 ini memiliki berbagai macam dampak dalam bidang sosial dan ekonomi untuk seluruh masyarakat Indonesia. Menurut Yamali (2020), dampak-dampak COVID-19 pada bidang ekonomi di Indonesia termasuk PHK besar-besaran (lebih dari 1,5 juta pekerja dirumahkan), penurunan impor pada triwulan I, penurunan *PMI Manufacturing Indonesia* pada Maret 2020, dan juga 6 ribu hotel mengalami penurunan okupansi sampai 50% yang menyebabkan kehilangan devisa pariwisata. Nasution (2020) juga menyatakan bahwa pandemi COVID-19 menyebabkan jumlah investor di pasar menurun yang berakibat membawa pasar ke arah negatif.

Pandemi COVID-19 juga mempengaruhi dalam bidang sosial. Menurut Aeni (2021), salah satu contoh dampak COVID-19 dalam aspek sosial dapat dilihat dengan kondisi kemiskinan di kabupaten Pati pada tahun 2020. Aeni menemukan bahwa jumlah penduduk miskin di Pati meningkat sebanyak 7% di tahun 2020 dibandingkan tahun 2019. Proporsi penduduk miskin terhadap penduduk secara keseluruhan di tahun 2020 adalah 10.08%, peningkatan sebanyak 0.52% dari tahun sebelumnya. COVID-19 juga menyebabkan layanan pendidikan menjadi terbatas yang dapat mengakibatkan turunnya tingkat pendidikan di Indonesia.

Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, penelitian ini memiliki tujuan untuk menganalisa lebih dalam dampak COVID-19 terhadap masyarakat Indonesia dalam berbagai bidang menggunakan data yang disediakan oleh Badan Pusat Statistik (BPS). Selain itu penelitian-penelitian sebelumnya belum ada yang menggunakan machine learning untuk memprediksi data-data seperti umur harapan hidup. Oleh karena itu penelitian ini juga akan mencoba menggunakan machine learning untuk memprediksi salah satu data masyarakat yaitu Umur Harapan Hidup.

1.2 Rumusan Masalah

1. Bagaimana prediksi tingkat Umur Harapan Hidup (UHH) 2021 masyarakat Indonesia pasca COVID-19?
2. Bagaimana keadaan sosial masyarakat diukur dari pertumbuhan ekonomi, Umur Harapan Hidup masyarakat, sektor UMKM, dan tingkat kemiskinan di Indonesia?

1.3 Tujuan

Adapun tujuan dalam pembuatan makalah ini adalah sebagai berikut:

1. Mencari model regresi yang terbaik untuk memprediksi Umur Harapan Hidup 2021 menggunakan data yang disediakan.
2. Mengetahui dampak-dampak COVID-19 terhadap aspek sosial dan ekonomi di Indonesia.

BAB II: PEMBAHASAN

2.1 Data Cleaning

- Tidak ada data yang memiliki missing value
- Tidak ada data yang memiliki invalid value
- Tidak ada data yang memiliki duplicated value

2.2 Exploratory Data Analysis

1. Bagaimana tingkat pertumbuhan ekonomi masyarakat terdampak COVID-19? (Dharma)

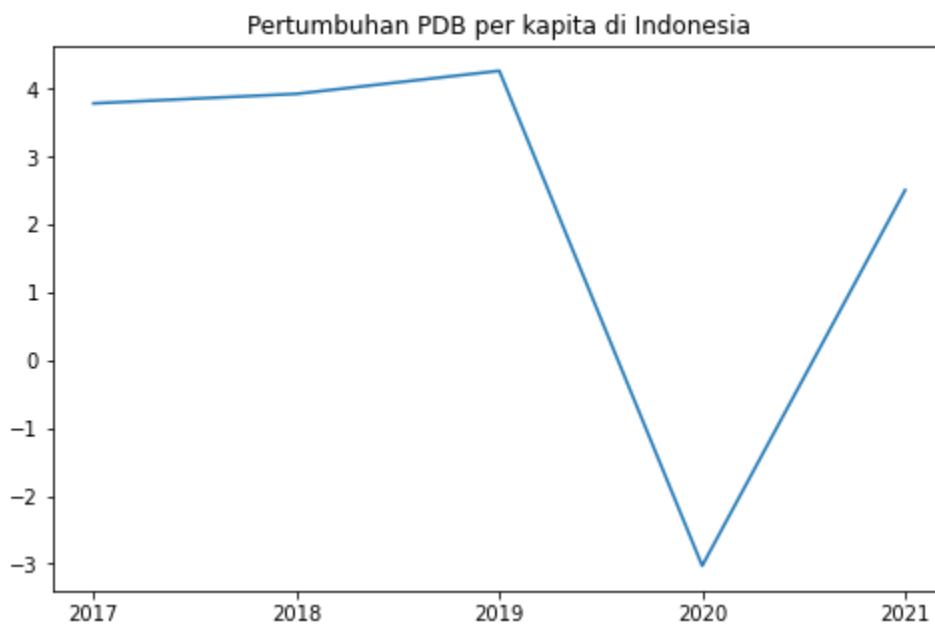
Menurut Shaïd (2022), “pertumbuhan ekonomi adalah kenaikan nilai dan jumlah produksi barang atau jasa dalam kurun waktu tertentu.” Komponen yang digunakan untuk

mengukur pertumbuhan ekonomi adalah PDB (Produk Domestik Bruto) atau PNB (Produk Nasional Bruto).

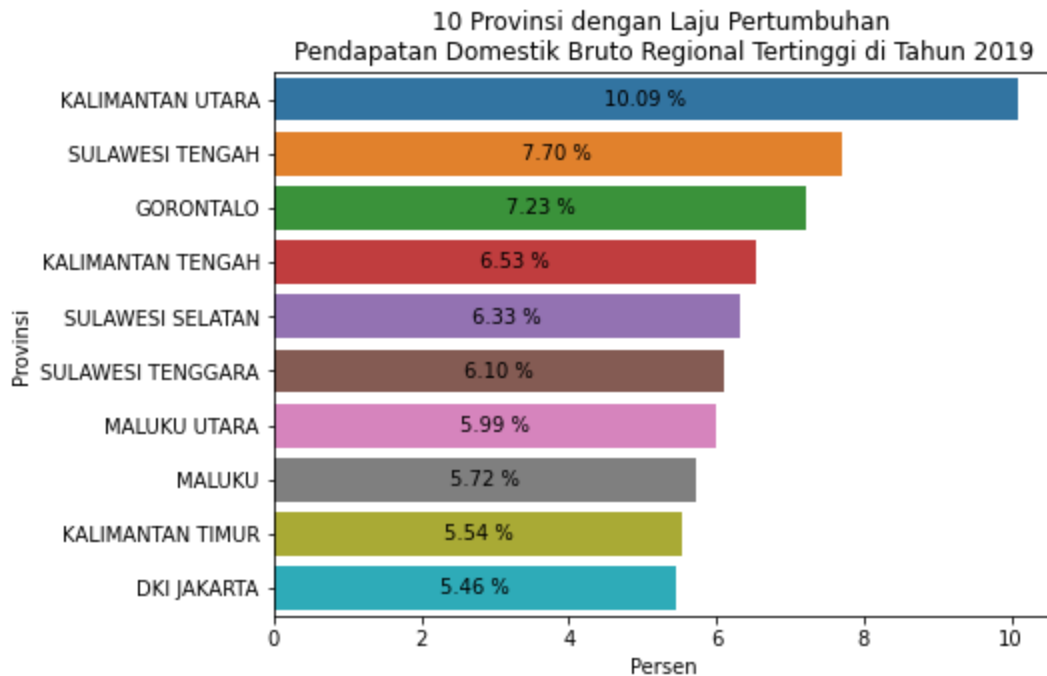
Pendekatan perhitungan PDB dapat dihitung dari pendekatan pendapatan, pendekatan pengeluaran, dan pendekatan produksi. Secara konsep ketiga pendekatan tersebut menghasilkan kesimpulan yang sama.

Pada penelitian ini kami menggunakan data laju pertumbuhan Produk Domestik Bruto Per Kapita Atas Dasar Harga Konstan. Harga Konstan kami pilih karena Harga Konstan tidak terpengaruh dengan inflasi yang menyebabkan harga dapat naik turun.

Pertumbuhan ekonomi di Indonesia berdasarkan PDB per kapita berdasarkan harga konstan mengalami peningkatan dari tahun 2017-2019. Di tahun 2020, pertumbuhan ekonomi Indonesia mengalami penurunan yang signifikan. Pertumbuhan ekonomi mulai menanjak di tahun 2021 pasca pemulihan ekonomi.



Pada tahun 2019, pertumbuhan ekonomi tertinggi terjadi di provinsi Kalimantan Utara. Hal ini masih terbilang sangat wajar karena Kalimantan Utara merupakan provinsi termuda di Indonesia. Pertumbuhan ekonomi yang tinggi bisa terlihat setiap kali ada pemekaran wilayah.



DKI Jakarta sebagai pusat ibu kota masih menyumbang pertumbuhan PDRB (Produk Domestik Regional Bruto) yang signifikan bagi pertumbuhan ekonomi di Indonesia.

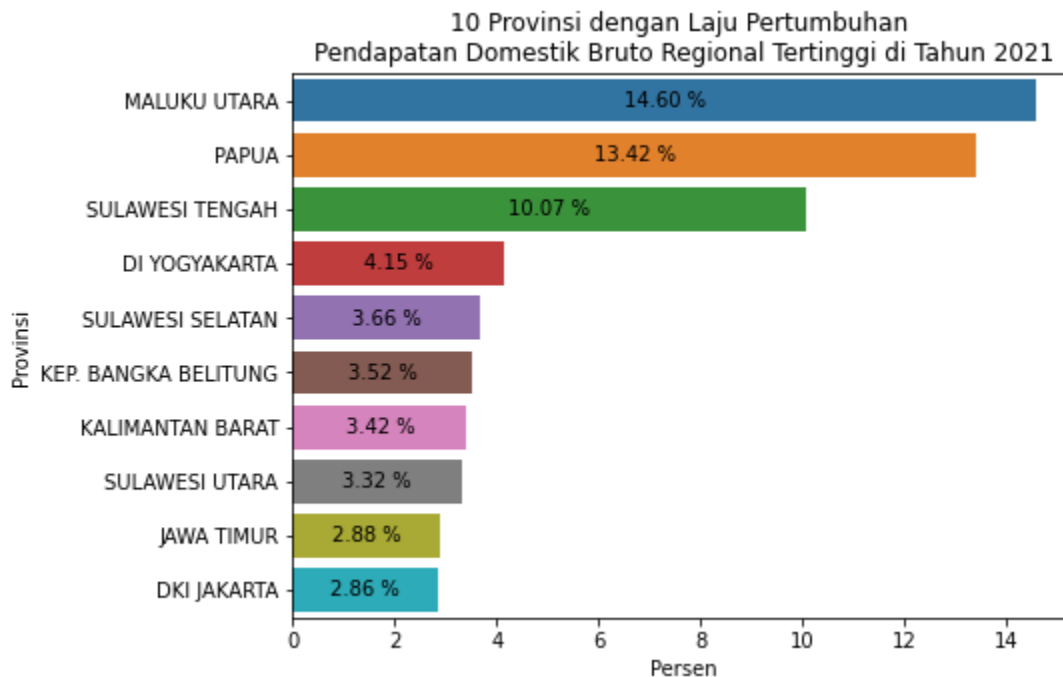
Beberapa provinsi di Indonesia mengalami pertumbuhan PDRB yang negatif. Pola ini dapat dilihat pada 3 provinsi yaitu Papua, Papua Barat, dan Kepulauan Riau. Provinsi Papua mengalami pertumbuhan PDRB yang paling parah. Hal ini dapat menjadi indikasi bahwa Papua menjadi provinsi yang kurang diperhatikan oleh pemerintah dan dapat dikatakan tertinggal dibanding provinsi lain.

Laju Pertumbuhan PDRB Terendah 2019	
Papua	-16.36%
Papua Barat	-0.13%
Kepulauan Riau	-0.08%

Tahun 2020 menjadi tahun terberat bagi Indonesia karena hanya 13 provinsi yang mengalami pertumbuhan ekonomi yang positif.

Meskipun demikian, kondisi berbeda terjadi pasca COVID-19. Ada 3 provinsi di bagian Timur Indonesia yang mengalami pertumbuhan lebih dari 10 persen. Ketiga provinsi tersebut adalah Maluku Utara, Papua, dan Sulawesi Tengah.

Papua yang pada tahun 2019 mengalami penurunan yang paling signifikan mengalami pertumbuhan kedua tertinggi di tahun 2021. Pertumbuhan ekonomi di Papua disebabkan oleh pembangunan infrastruktur yang signifikan di Papua. Pekan Olahraga Nasional (PON) 2020 juga diadakan di Papua membuat PDRB Papua meningkat signifikan.



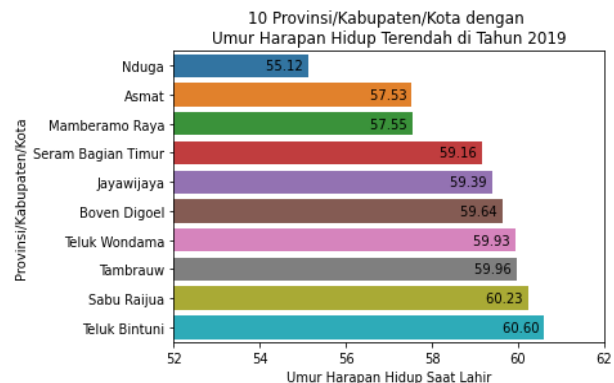
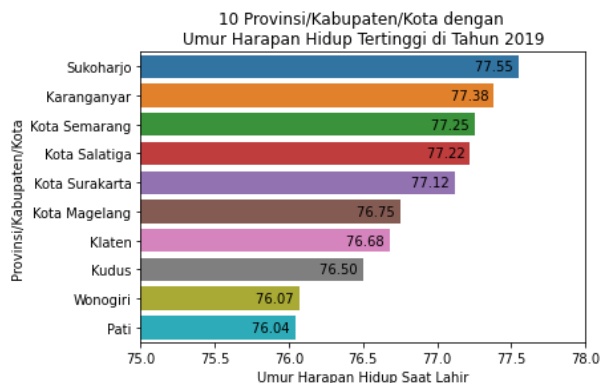
Provinsi yang paling terdampak adalah Bali. Bali yang hanya mengandalkan industri pariwisatanya mengalami penurunan PDRB yang signifikan. Aturan lockdown di banyak negara menyebabkan Bali kehilangan pendapatan utamanya. Papua Barat dan Kepulauan Riau masih menjadi provinsi yang belum berkembang.

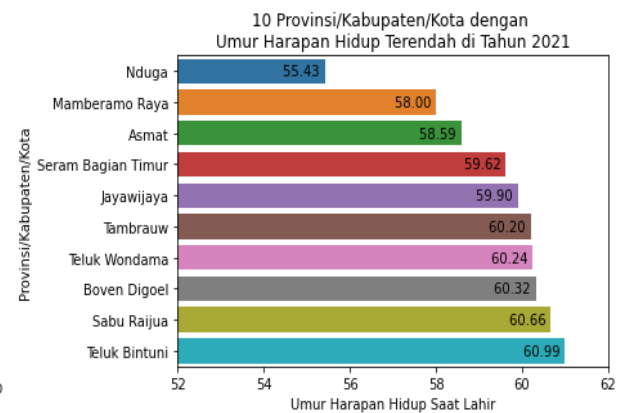
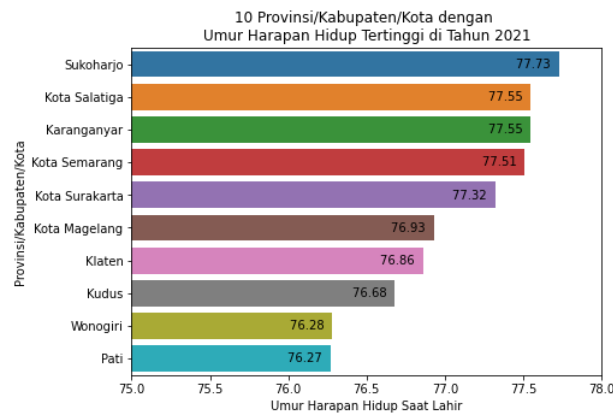
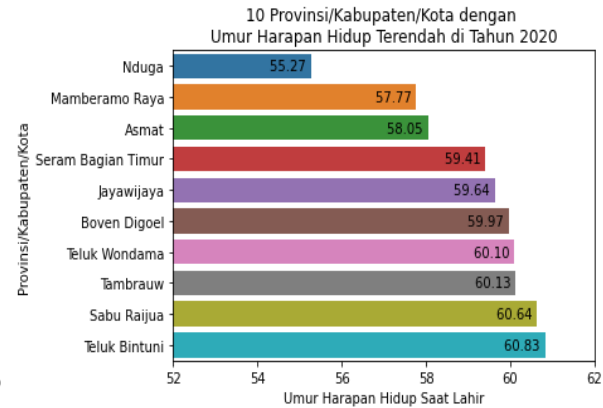
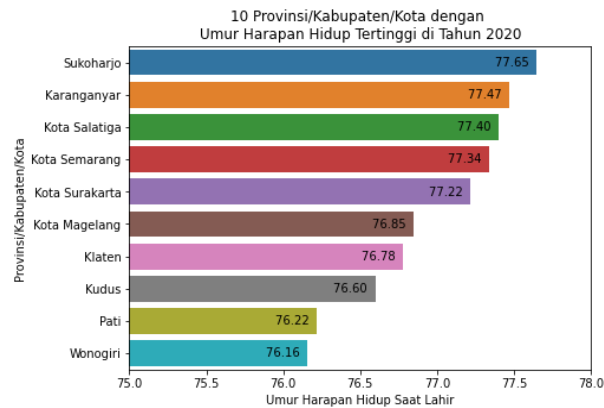
Laju Pertumbuhan PDRB Terendah 2021	
Bali	-3.64%
Papua Barat	-2.77%
Kepulauan Riau	-0.49%

2. Bagaimana harapan hidup masyarakat imbas COVID-19?



Untuk melihat dampak COVID-19 terhadap harapan hidup masyarakat Indonesia, kami mengambil data umur harapan hidup saat lahir seluruh Indonesia dari BPS selama 5 tahun ke belakang, dari tahun 2017 sampai 2021. Dilihat dari grafik diatas bahwa harapan hidup masyarakat Indonesia meningkat tiap tahunnya, termasuk pada tahun dimana pandemi COVID-19 terjadi yaitu pada 2019 sampai 2021. Umur harapan hidup seluruh Indonesia di tahun 2019 adalah 71.34, yang merupakan kenaikan dari 71.2 di tahun 2018. Lalu di tahun 2020 dan 2021 mengalami kenaikan juga dengan umur harapan hidup 71.47 di tahun 2020 dan 71.57 di tahun 2021. Maka dapat disimpulkan bahwa COVID-19 tidak mempengaruhi perkembangan harapan hidup masyarakat Indonesia dan harapan hidup masyarakat cenderung meningkat tiap tahun.





Dapat dilihat dari grafik di atas bahwa 10 Provinsi/ Kabupaten/ Kota dengan umur harapan hidup terendah dan tertinggi selalu sama selama tahun 2019, 2020, dan 2021, yang berarti COVID-19 juga tidak mempengaruhi secara signifikan peringkat umur harapan hidup diantara Provinsi/ Kabupaten/ Kota.

3. Apa UMKM yang bisa disupport oleh pemerintah di saat New Normal pasca COVID-19?

Menurut kementerian koordinator bidang perekonomian Republik Indonesia pandemi COVID-19 telah memberikan dampak negatif bagi mayoritas UMKM. Dampak negatif bagi UMKM berupa hal-hal seperti kehilangan pelanggan, tidak bisa membuka usaha/toko dikarenakan PSBB, dan hal-hal lainnya.

Di Indonesia, sektor UMKM dapat dibagi menjadi 17 sektor yaitu:

No	Sektor UMKM
1	Pertanian, Perburuan dan Kehutanan
2	Perikanan
3	Pertambangan dan Penggalian
4	Industri Pengolahan

5	Listrik, gas dan air
6	Konstruksi
7	Perdagangan Besar dan Eceran
8	Penyediaan akomodasi dan makan minum
9	Transportasi, pergudangan dan komunikasi
10	Perantara Keuangan
11	Real Estate, Usaha Persewaan, dan Jasa Perusahaan
12	Administrasi Pemerintahan, Pertahanan dan Jaminan Sosial Wajib
13	Jasa Pendidikan
14	Jasa Kesehatan dan Kegiatan Sosial
15	Jasa Kemasyarakatan, Sosial Budaya, Hiburan dan Perorangan lainnya
16	Jasa Perorangan yang Melayani Rumah Tangga
17	Badan Internasional dan Badan Ekstra Internasional Lainnya

Dalam masa New Normal yang telah diterapkan oleh pemerintah pasca COVID-19, UMKM yang dapat berkembang adalah perusahaan yang memiliki 4 karakteristik bisnis yaitu *Hygiene, Low-Touch, Less-Crowd*, dan *Low-Mobility*. Pemerintah juga telah memberi dukungan untuk UMKM dengan program Pemulihan Ekonomi Nasional (PEN) di tahun 2020. Sudah terealisasi dukungan sebanyak Rp. 122,84 triliun untuk lebih dari 30 juta UMKM di tahun 2020 dalam bentuk penundaan angsuran dan subsidi bunga. Ada 3 sektor krusial yang harus dibantu oleh pemerintah di masa New Normal seperti sektor pedagang besar dan eceran, pertanian, dan peternakan.

Menurut Staf Khusus Menteri Koperasi dan UKM, Agus Santoso mengungkapkan, berbagai sektor usaha UMKM terdampak pandemi COVID-19, namun yang paling terdampak adalah pedagang besar dan eceran (40.92%). Menurut data BPS, sektor perdagangan besar dan eceran menempati peringkat pertama dalam hal posisi kredit bank umum tertinggi di tahun 2020. Pemerintah perlu memberikan bantuan kepada sektor ini karena sektor ini merupakan sektor UMKM yang paling signifikan.

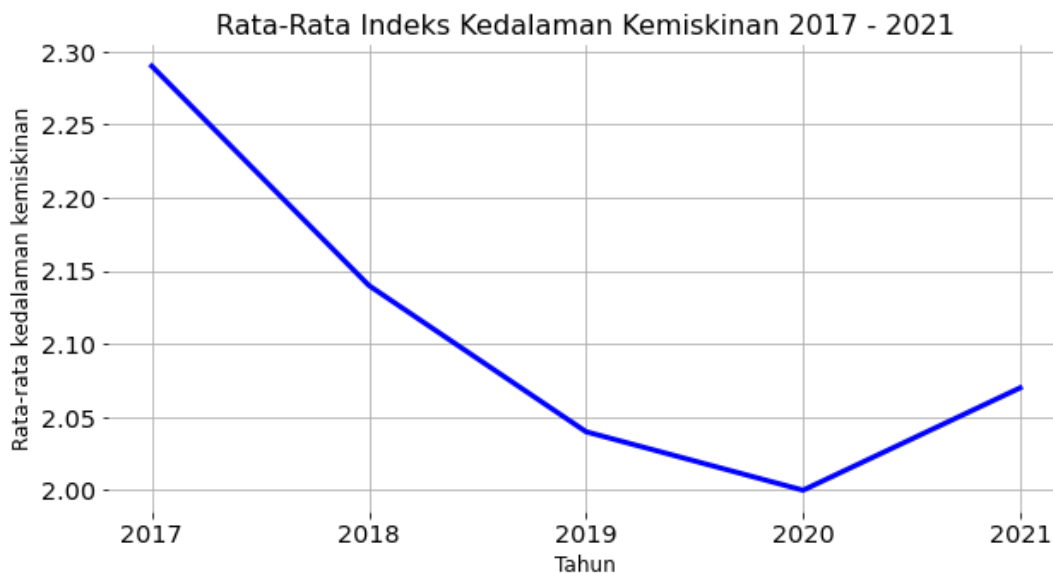
Selain itu, pemerintah juga perlu untuk memberikan bantuan kepada sektor pertanian. Menurut Anggota Komisi IV DPR Charles Meikyansah, sektor pertanian perlu mendapat perhatian dari pemerintah di tengah pandemi COVID-19 ini. Sektor ini merupakan yang terbesar karena berpengaruh terhadap hajat hidup masyarakat Indonesia. Indonesia yang dikenal sebagai

negara agraris dan negara penghasil beras tertinggi ketiga di dunia masih belum mampu untuk mencukupi kebutuhan pangan dalam negeri dan masih mengimpor dari negara lain. Pemerintah dapat lebih memperhatikan para petani sehingga kebutuhan padi dalam negeri bisa tercukupi.

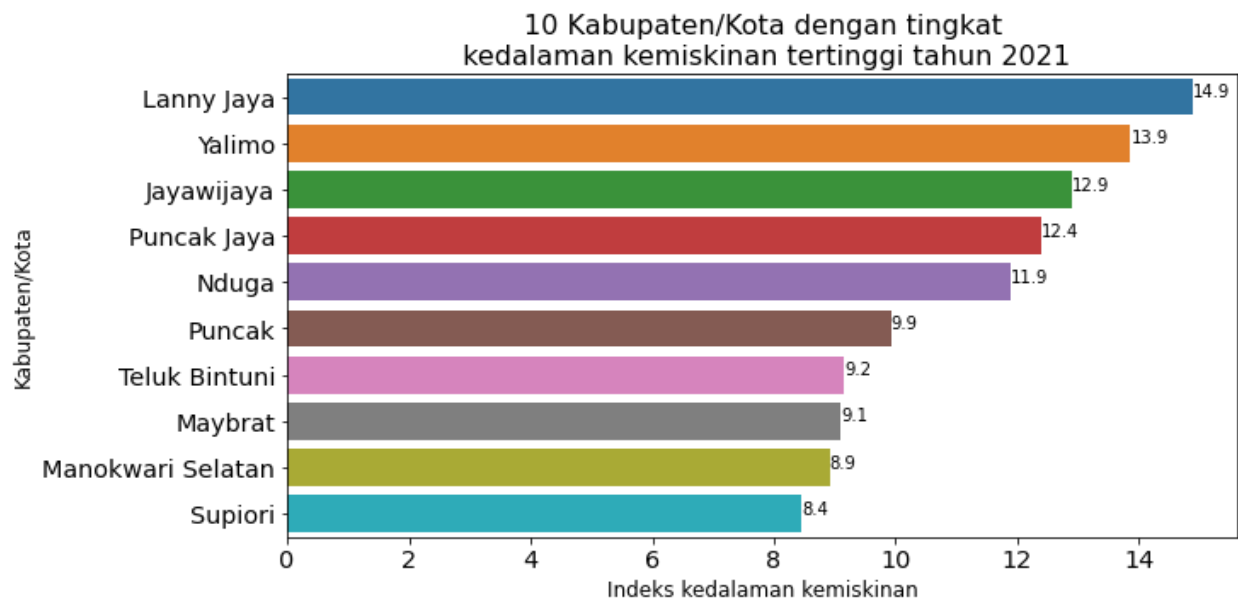
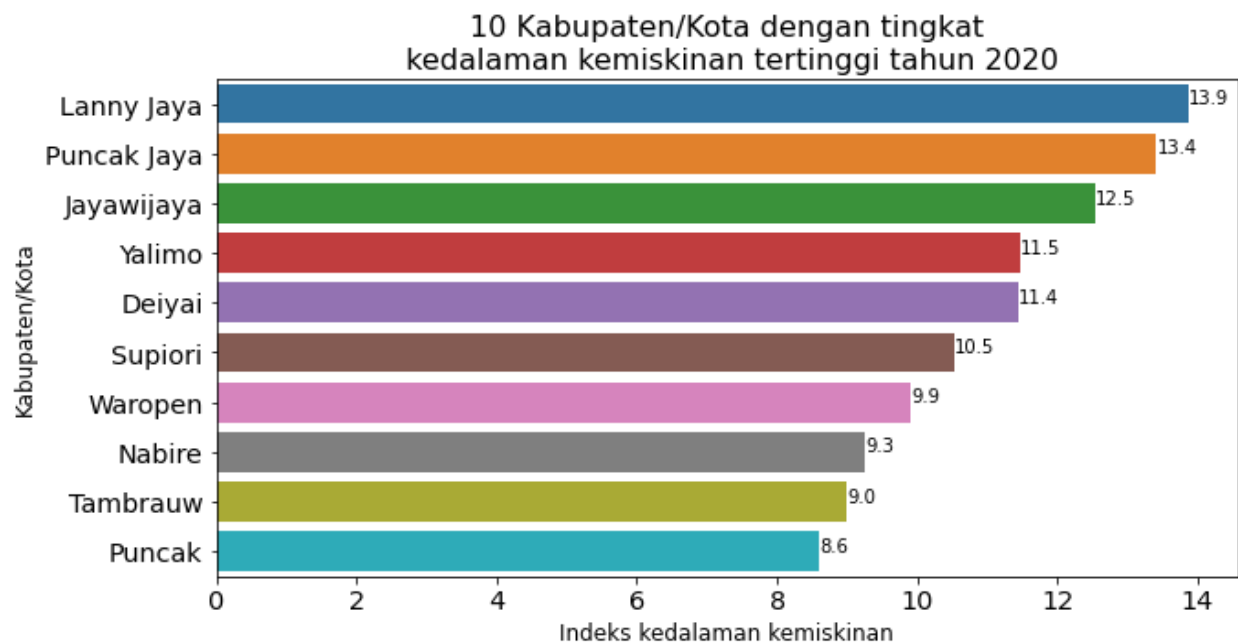
Dapat disimpulkan bahwa semua 17 sektor UMKM dapat disupport oleh pemerintah melalui penundaan angsuran dan keringanan bunga yang dihasilkan oleh program pemulihan ekonomi nasional. Akan tetapi, pemerintah harus memfokuskan support mereka kepada sektor perdagangan besar dan eceran serta pertanian, perburuan, dan kehutanan karena mereka sektor yang sangat penting dan juga memiliki jumlah kredit pada bank umum terbesar dibanding sektor lainnya.



4. Bagaimana tingkat kemiskinan masyarakat terdampak COVID-19? (Fikri)



Untuk mengetahui tingkat kemiskinan masyarakat terdampak COVID-19 di Indonesia. Kami mengambil data tambahan tentang Indeks Kedalaman Kemiskinan dari 2017 - 2021 dari website BPS Indonesia. Alasan mengambil data dari tahun 2017 - 2021 dikarenakan ingin mengetahui laju kedalaman kemiskinan dari sebelum COVID-19 hingga COVID-19 terjadi. Dari visualisasi diatas dapat disimpulkan kalau dari tahun 2017 - 2020 rata-rata indeks kedalaman kemiskinan mengalami penurunan namun di tahun 2020 rata-rata indeks kedalaman kemiskinan mulai naik hingga tahun 2021.



Dapat dilihat dari kedua grafik diatas untuk tahun 2020 hingga tahun 2021 tingkat kedalaman kemiskinan tertinggi berdasarkan berdasarkan kabupaten/kota adalah pada kabupaten/kota Lanny Jaya yang mana kabupaten/kota ini masuk ke dalam provinsi papua. Provinsi Papua masih mengalami tingkat kemiskinan yang parah dengan 5 kabupaten/kota teratas berada di Provinsi Papua.

Exploratory Data Analysis Data Survey Sosial dan Kependudukan

Exploratory Data Analysis (EDA) adalah proses analisis untuk memahami karakteristik data, dan hal-hal yang perlu kami lakukan agar data tersebut dapat digunakan untuk proses pembelajaran model. EDA sangat diperlukan untuk mencari insight-insight menarik dari data set yang telah disediakan. EDA dapat dibagi menjadi 3 bagian yaitu descriptive statistic, univariate analysis, dan multivariate analysis.

Pada data Survei Sosial dan kependudukan, target dari data testing percobaan ini adalah UHH 2021. Pada data ini ditemukan bahwa tidak terdapat nilai yang missing, persebaran datanya cukup linear, dan setiap data numeriknya memiliki outlier.

Kami melakukan analisis descriptive statistic dan ditemukan bahwa semua kolom sudah sesuai dengan datatype nya. Tidak ada data yang hilang. Hal ini dapat kita lihat dari kolom dengan count = jumlah row = 407.

Data columns (total 26 columns):			
#	Column	Non-Null Count	Dtype
0	Id	407 non-null	int64
1	Kabupaten/Kota	407 non-null	object
2	Harapan Lama Sekolah (2019)	407 non-null	float64
3	Harapan Lama Sekolah (2020)	407 non-null	float64
4	Harapan Lama Sekolah (2021)	407 non-null	float64
5	Pengeluaran per Kapita Disesuaikan (Ribu Rupiah/Orang) (2019)	407 non-null	int64
6	Pengeluaran per Kapita Disesuaikan (Ribu Rupiah/Orang) (2020)	407 non-null	int64
7	Pengeluaran per Kapita Disesuaikan (Ribu Rupiah/Orang) (2021)	407 non-null	int64
8	Rata-rata Lama Sekolah (2019)	407 non-null	float64
9	Rata-rata Lama Sekolah (2020)	407 non-null	float64
10	Rata-rata Lama Sekolah (2021)	407 non-null	float64
11	Umur Harapan Hidup Saat Lahir (UHH) (2019)	407 non-null	float64
12	Umur Harapan Hidup Saat Lahir (UHH) (2020)	407 non-null	float64
13	Umur Harapan Hidup Saat Lahir (UHH) (2021)	407 non-null	float64
14	Indeks Pembangunan Manusia (2019)	407 non-null	float64
15	Indeks Pembangunan Manusia (2020)	407 non-null	float64
16	Indeks Pembangunan Manusia (2021)	407 non-null	float64
17	Peringkat Indeks Pembangunan Manusia (2019)	407 non-null	object
18	Peringkat Indeks Pembangunan Manusia (2020)	407 non-null	object
19	Peringkat Indeks Pembangunan Manusia (2021)	407 non-null	object
20	Indeks Keparahatan Kemiskinan (P2) Menurut Kabupaten/Kota (2019)	407 non-null	float64
21	Indeks Keparahatan Kemiskinan (P2) Menurut Kabupaten/Kota (2020)	407 non-null	float64
22	Indeks Keparahatan Kemiskinan (P2) Menurut Kabupaten/Kota (2021)	407 non-null	float64
23	Indeks Kedalaman Kemiskinan (P1) Menurut Kabupaten/Kota (2019)	407 non-null	float64
24	Indeks Kedalaman Kemiskinan (P1) Menurut Kabupaten/Kota (2020)	407 non-null	float64
25	Indeks Kedalaman Kemiskinan (P1) Menurut Kabupaten/Kota (2021)	407 non-null	float64

dtypes: float64(18), int64(4), object(4)
memory usage: 82.8+ KB

Terdapat kejanggalan pada kolom (Peringkat indeks pembangunan manusia 2019 - 2021) yang berbentuk objek. Biasanya data peringkat seharusnya berbentuk numerik.

Dilihat dari persebarannya, terdapat beberapa kolom yang positively skewed seperti Indeks Keparahana Kemiskinan (P2) Menurut Kabupaten/Kota (2019-2021), Indeks Kedalaman Kemiskinan (P1) Menurut Kabupaten/Kota (2019-2021). Kolom yang lain masih dapat dikategorikan tergolong distribusi normal.

Kami juga membagi kolom-kolom data berdasarkan kolom numerik dan kategorik. Kolom kategorik pada dataset ini adalah Kabupaten/Kota, Peringkat Indeks Pembangunan Manusia (2019), Peringkat Indeks Pembangunan Manusia (2020), Peringkat Indeks Pembangunan Manusia (2021). Kolom-kolom yang lain termasuk kolom numerik. Pada kolom-kolom di dataset ini juga tidak terdapat kolom yang termasuk boolean dan tidak ada nilai yang negatif.

Tahap EDA selanjutnya adalah univariate analysis. Di tahap ini, kami mencoba mencari pola persebaran data. Pola persebaran data yang kami pakai dengan menggunakan boxplot dan violin plot. Distribusi boxplot dapat digunakan untuk mencari tahu outlier, nilai median, nilai minimum, dan nilai maksimum. Pada dataset ini ditemukan banyak sekali outlier. Distribusi violin plot digunakan untuk mengetahui persebaran data baik normal, positively skewed, maupun negatively skewed. Fitur-fitur ini yang nantinya akan dilakukan log dan inverse log transformation.

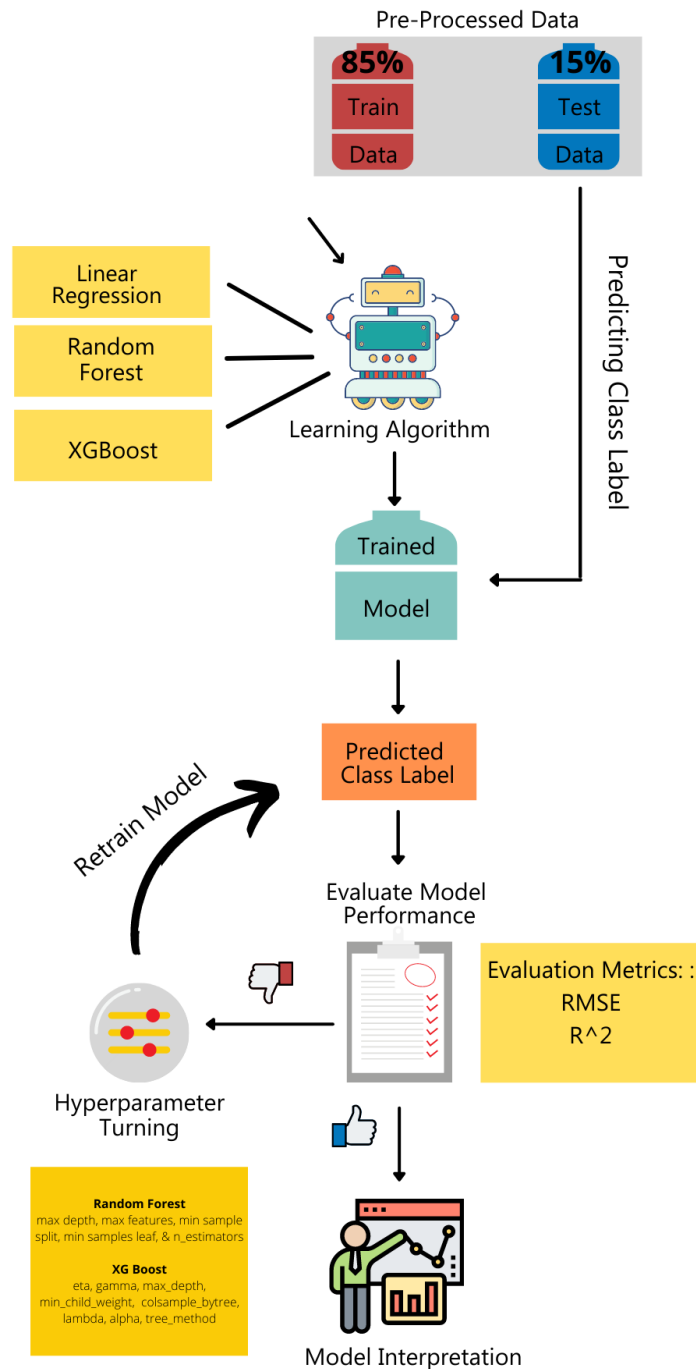
Kami juga melakukan uji korelasi feature-feature dengan target kolom UHH 2021 untuk menentukan fitur mana yang akan dipakai pada saat Machine Learning modelling.

Pada tahap ketiga, kami mencoba melakukan multivariate analysis. Di tahap ini kami mencoba mencari korelasi antar fitur-fitur dengan membandingkan korelasi heatmap. Fitur-fitur yang sama, tetapi berbeda tahun (misalnya UHH 2019, UHH 2020, dan UHH 2021) memiliki korelasi yang positif kuat satu dengan yang lainnya.

Di tahap ini, kami juga mengecek hubungan antara category column dan target column. Peringkat Indeks Pembangunan Manusia merupakan fitur yang menarik untuk ditelusuri lebih lanjut.

2.3 Machine Learning Deployment

MODELING



1. Data Processing

Outliers	Tidak menghapus nilai outlier untuk setiap data dikarenakan datanya tidak terlalu besar dan akan mengurangi banyak datanya
Selection	<ul style="list-style-type: none"> - Menghapus “Id”, “Kabupaten/Kota”, “Peringkat Indeks Pembangunan” dikarenakan memiliki nilai unique yang besar - Menghapus “Indeks Keparahen Kemiskinan (P2) Menurut Kabupaten/Kota” dikarenakan korelasinya tidak begitu kuat dengan target
Normalization	Normalisasi data untuk fitur yang bertipe numerik

2. Hasil Modeling

Algoritma yang kami pakai dalam modeling adalah Linear Regression, Random Forest, dan XGBoost. Ketiga algoritma tersebut memiliki kelebihan dan kekurangannya masing-masing. Setiap algoritma juga memiliki feature importance masing-masing.

Linear Regression				
Nama Percobaan	RMSE Score Train	RMSE Score Test	R ² Score Train	R ² Score Test
Feature Importance	0.06844	0.06583	0.99960	0.99960
Tanpa menggunakan peringkat indeks pembangunan manusia dan melakukan standarisasi	0.04625	0.27808	0.99982	0.99285
Tanpa menggunakan peringkat indeks pembangunan manusia & indeks keparahan kemiskinan	0.04740	0.03935	0.99981	0.99986
Melakukan Log pada data Indeks Kedalaman Kemiskinan (P1), menghilangkan feature peringkat indeks pembangunan manusia 2019 - 2021 dan indeks keparahan kemiskinan 2019 - 2021 dan	0.04752	0.04011	0.99981	0.99985

tidak melakukan proses standarisasi				
Random Forest				
Feature Importance	0.06901	0.08947	0.99960	0.99926
Tanpa menggunakan peringkat indeks pembangunan manusia dan melakukan standarisasi	0.09172	0.35063	0.99929	0.98863
Tanpa menggunakan peringkat indeks pembangunan manusia & indeks keparahan kemiskinan dengan standarisasi	0.09816	0.34574	0.99919	0.98894
Dengan tuning parameter model Tanpa menggunakan peringkat indeks pembangunan manusia & indeks keparahan kemiskinan dengan standarisasi	0.56659	0.77396	0.97305	0.94459
XGBoost				
Feature Importance	0.08050	0.13596	0.99945	0.99829
Tanpa menggunakan peringkat indeks pembangunan manusia dan melakukan standarisasi	0.07441	0.30777	0.99953	0.99124
Tanpa menggunakan peringkat indeks pembangunan manusia & indeks keparahan kemiskinan dengan standarisasi	0.07629	0.30950	0.99951	0.99114
Dengan tuning parameter model Tanpa menggunakan peringkat indeks pembangunan manusia & indeks keparahan kemiskinan dengan standarisasi	0.45271	0.54118	0.98279	0.97291
Neural Network				
Tanpa menggunakan peringkat indeks pembangunan manusia & indeks keparahan kemiskinan	4.71905	3.50397	-0.86923	-0.13573

dengan melakukan standarisasi dan membuat 4 layer				
---	--	--	--	--

Model pertama yang kami gunakan adalah model linear regression. Linear Regression adalah model regresi yang dapat digunakan jika data yang dimiliki memiliki korelasi yang tinggi dan bersifat linear. Linear Regression menghubungkan data-data menggunakan 1 garis lurus yang meminimalisasikan jarak antara semua data point dengan garis tersebut. Kami menggunakan ukuran split data train 0.15 dan random state 32. Ukuran split data train hanya 0.15 karena jumlah data di dataset train hanya sekitar 400 dan model ini perlu menggunakan mayoritas (0.85 atau 85%) data untuk training dan sisanya baru digunakan untuk testing. Tidak ada feature encoding yang kami lakukan karena kolom kategorikal di dataset ini hanya Kabupaten/Kota dan Peringkat IPM 2019-2021.

Pertama kami melakukan *feature importance regression* untuk mengetahui fitur-fitur mana yang memiliki korelasi yang tinggi. Kami menggunakan semua fitur yang memiliki nilai fitur importance positif. Didapatkan RMSE dari data test sebesar 0.065 dan RMSE dari data train 0.068. Setelah menggunakan regression dengan feature importance kami melakukan beberapa percobaan untuk mencoba mendapatkan nilai RMSE dari data train yang lebih rendah.

Pada percobaan linear regression pertama kami menstandarisasi data dan menggunakan semua fitur kecuali peringkat indeks pembangunan manusia 2019 - 2021. Fitur ini tidak digunakan karena banyak data kosong dan bersifat objek bukan numerik. Model ini mengeluarkan output RMSE dari data test sebesar 0.278 dan RMSE dari data train sebesar 0.046. Dalam percobaan ini model cenderung overfit.

Pada percobaan kedua sama seperti percobaan pertama tetapi fitur yang digunakan lebih sedikit dengan fitur indeks keparahan kemiskinan 2019 - 2021 tidak digunakan karena memiliki korelasi yang negatif dan rendah. Ditemukan RMSE dari data test sebesar 0.039 dan RMSE dari data train sebesar 0.047. Dalam percobaan ini model cenderung underfit.

Pada percobaan ketiga sama dengan percobaan sebelumnya tetapi kami melakukan proses logaritma pada data indeks kedalaman kemiskinan karena data tersebut bersifat skew positif. Pada percobaan ini didapatkan RMSE dari data test sebesar 0.04011 dan RMSE dari data training sebesar 0.04752. Dalam percobaan ini model bersifat overfit sedikit tetapi karena

perbedaan RMSE dari data test dan data train sangat kecil maka dapat dianggap bahwa model ini tidak underfit atau overfit.

Model kedua yang kami gunakan adalah Random Forest Regressor. Random Forest Regressor adalah model regresi yang menjalankan beberapa *decision tree* secara bersamaan dan mengembalikan rata-rata dari semua kelas sebagai prediksi model. Random Forest dapat menghandle data baik linier maupun non-linear.

Sama seperti linear regression pertama kami melakukan feature importance regression untuk mengetahui fitur-fitur yang penting. Lalu fitur-fitur yang memiliki nilai feature importance diatas 0.1 digunakan dalam model regresi. Ditemukan nilai RMSE dari data test sebesar 0.089 dan RMSE dari data train sebesar 0.069. Selanjutnya kami melakukan percobaan yang sama pada saat menggunakan model linear regression kecuali percobaan transformasi log karena Random Forest Regressor sudah robust dan bisa menangani data yang tidak berdistribusi normal.

Pada percobaan Random Forest Regressor pertama kami menstandarisasikan data dan menggunakan semua fitur kecuali fitur peringkat indeks pembangunan 2019 - 2021 dikarenakan banyak data kosong dan hanya berisi peringkat melainkan nilai. Ditemukan RMSE dari data test sebesar 0.350 dan RMSE dari data train 0.091. Dalam percobaan ini model cenderung overfit.

Pada percobaan kedua sama seperti percobaan pertama tetapi tidak menggunakan fitur indeks keparahan kemiskinan 2019 - 2021. Ditemukan RMSE dari data test sebesar 0.345 dan RMSE dari data train sebesar 0.098. Dalam percobaan ini model cenderung overfit.

Pada percobaan ketiga kami melakukan percobaan yang sama dengan percobaan kedua, tetapi mengganti tuning parameter Forest Regressor. Kami menggunakan fungsi GridSearch untuk menemukan parameter Forest Regressor yang paling tepat. Parameter yang digunakan untuk forest regressor adalah bootstrap, max depth, ,max features, min samples leaf, min samples split, dan n estimators. Bootstrap adalah parameter untuk menentukan metode sampling (*with or without replacement*). Max depth adalah jumlah level maximum dalam tiap *tree*. Max features adalah jumlah fitur maksimum saat *node splitting*. Min sample leaf adalah jumlah data minimum untuk sebuah *leaf node*. Min samples split adalah jumlah data minimum dalam sebuah node sebelum membelah/*split* node tersebut. N estimators adalah jumlah *tree*/pohon dalam forest regressor. Dalam percobaan ketiga kami menemukan RMSE dari data test sebesar 0.773 dan RMSE dari data train sebesar 0.566. Dalam percobaan ini model cenderung overfit.

Model ketiga yang kami gunakan adalah *XGBoost Regression*. XGBoost Regression atau Extreme Gradient Boosting Regression adalah model regression yang menggunakan *decision trees* dan peningkatan gradien untuk mengurangi bias dan varians.

Seperti model-model sebelumnya pertama kami melakukan feature importance regression untuk mengetahui fitur-fitur mana yang penting menurut model ini. Kami melakukan regresi dengan fitur-fitur memiliki feature importance lebih dari 0.1. Ditemukan RMSE dari data test sebesar 0.135 dan RMSE dari data train 0.080. Selanjutnya kami melakukan percobaan yang sama dengan kedua model sebelumnya.

Pada percobaan pertama, kami menstandarisasikan data dan menggunakan semua fitur kecuali fitur peringkat indeks pembangunan 2019 - 2021 dikarenakan banyak data kosong dan hanya berisi peringkat melainkan nilai. Ditemukan RMSE dari data test sebesar 0.307 dan RMSE dari data train 0.074. Dalam percobaan ini model cenderung overfit.

Pada percobaan kedua sama seperti percobaan pertama tetapi tidak menggunakan fitur indeks keparahan kemiskinan 2019 - 2021. Ditemukan RMSE dari data test sebesar 0.336 dan RMSE dari data train sebesar 0.097. Dalam percobaan ini mode cenderung overfit.

Pada percobaan terakhir sama seperti random forest regressor kami melakukan tuning parameter untuk model XGBoost. Parameter yang kami ubah adalah max depth, min child weight, gamma, tree method, colsample bytree, eta, lambda, dan alpha. Max depth adalah level maximum dalam tiap pohon/tree. Min child weight adalah jumlah semua weight minimum dalam sebuah anak node. Gamma menentukan pengurangan loss minimum yang diperlukan untuk membuat sebuah node split. Tree method untuk menentukan metode tree apa yang digunakan. Colsample bytree menentukan fraksi kolom yang digunakan sebagai sampel untuk tiap pohon. Eta untuk menentukan *learning rate*. Alpha dan lambda adalah istilah regulasi L1 dan L2 untuk fungsi loss. Dalam percobaan ketiga ditemukan RMSE dari data test sebesar 0.541 dan RMSE dari data train sebesar 0.452.

Untuk percobaan terakhir kami mencoba deep learning dengan cara menggunakan model Neural Network. Pada percobaan inisial kami menemukan score R^2 bernilai negatif yang menandakan model tidak cocok dengan data yang digunakan. RMSE dari data train dan test yang kami temukan juga sangat besar (>1) oleh karena itu kami tidak melakukan percobaan lebih lanjut menggunakan Neural Network.

Dikarenakan persebaran datanya cukup linier maka model machine learning yang paling bagus digunakan untuk kasus ini adalah model Linear Regression dikarenakan memiliki RMSE yang paling rendah dibandingkan model lainnya yang kami coba yaitu score RMSE 0.003935 dan score R^2 0.99986.

2.4 Hasil Prediksi

Berdasarkan hasil Machine Learning Deployment diatas, model regresi yang memiliki RMSE paling rendah adalah Linear Regression. Dari ketiga percobaan linear regression, percobaan ketiga menghasilkan RMSE yang paling kecil yaitu 0.04861 pada dataset test. Performa model ini sudah bagus karena terdapat nilai RMSE testing dan training 0.04740 dan 0.03935 yang memiliki perbedaan sangat kecil maka percobaan model ini tidak termasuk overfitting dan underfitting.

BAB III: PENUTUP

3.1 Kesimpulan dan Saran

1. Kesimpulan

Dari pembahasan diatas dapat disimpulkan bahwa COVID-19 menyebabkan pertumbuhan ekonomi turun dan juga tingkat kemiskinan masyarakat naik, tetapi tidak mempengaruhi harapan hidup masyarakat. Pemerintah juga harus memberi perhatian lebih banyak untuk Papua Barat dan Kepulauan Riau karena pertumbuhan ekonomi daerah tersebut cenderung negatif dari tahun ke tahun. Pemerintah dapat mensupport UMKM kepada sektor perdagangan besar dan eceran serta pertanian, perburuan, dan kehutanan untuk membantu perkembangan ekonomi meningkat lagi.

Berdasarkan prediksi Machine Learning Deployment, didapatkan rmse sebesar 0.04861. Dari 4 model regresi yang kami coba, Linear, Random Forest, XGBoost, dan Neural Network, Linear Regression memberi nilai RMSE yang paling rendah dari semua hasil percobaan sedangkan Neural Network memberi nilai RMSE yang paling tinggi. Dari semua percobaan untuk tiap model, Linear Regression memberi nilai RMSE data test terendah 0.03935, Forest Regressor memberi nilai RMSE terendah 0.08947, XGBoost memberi nilai RMSE terendah 0.13596, dan Neural Network memberi RMSE terendah 3.50397. Oleh karena itu dapat disimpulkan model algoritma yang terbaik untuk data ini adalah Linear Regression.

2. Saran

Untuk saran penelitian lebih lanjut dapat dilakukan dengan data pasca COVID-19 yang lebih banyak dan detail berhubung dengan penelitian ini dilakukan hanya 2 tahun setelah pandemi COVID-19 maka data pasca COVID-19 belum cukup banyak. Masih banyak sektor UMKM yang dapat dianalisis dan dapat dibantu oleh pemerintah.

Untuk pemodelan lebih lanjut dapat digunakan algoritma yang lain dan dapat mencoba fitur-fitur yang lain. Penggunaan metode deep learning juga dapat dipelajari lebih lanjut.

Daftar Pustaka

Aeni, N. (2021). Pandemi COVID-19: Dampak Kesehatan, Ekonomi, & Sosial. *Jurnal Litbang: Media Informasi Penelitian, Pengembangan Dan IPTEK*, 17(1), 17–34. <https://doi.org/10.33658/jl.v17i1.249>

Badan Pusat Statistik (BPS) diakses dari <http://www.bps.go.id/> , diakses pada tanggal 10 Februari 2022

Herman. (2021, Februari 2021). *Pedagang Besar dan Eceran Paling Terdampak Pandemi*. Beritasatu.com. <https://www.beritasatu.com/ekonomi/738857/pedagang-besar-dan-eceran-paling-terdampak-pandemi>

Kementerian Koordinator Bidang Perekonomian Republik Indonesia. (2021, April 04). *Dukungan Pemerintah Bagi UMKM Agar Pulih di Masa Pandemi* [Siaran Pers]. Diambil dari <https://www.ekon.go.id/publikasi/detail/2939/dukungan-pemerintah-bagi-umkm-agar-pulih-di-masa-pandemi>

Komisi IV DPR RI. (2020, Juni 07). *Pertanian jadi Sektor Terbesar yang Terdampak Pandemi*. Diambil dari <https://www.dpr.go.id/berita/detail/id/29040/t/Pertanian+jadi+Sektor+Terbesar+yang+Terdampak+Pandemi>

Nasution, D., Erlina, E., & Muda, I. (2020). Dampak Pandemi COVID-19 terhadap Perekonomian Indonesia. *Jurnal Benefita*, 5(2), 212-224. Doi: <http://dx.doi.org/10.22216/jbe.v5i2.5313>

Shaid, N. J. (2022, Januari 8). *Pertumbuhan Ekonomi: Pengertian, Ciri dan Faktor yang Mempengaruhinya*. Kompas.com. <https://money.kompas.com/read/2022/01/08/080854626/pertumbuhan-ekonomi-pengertian-ciri-dan-faktor-yang-mempengaruhinya?page=all>

Yamali, F. R., & Putri, R. N. (2020). Dampak Covid-19 Terhadap Ekonomi Indonesia.
Ekonomis: Journal of Economics and Business, 4(2), 384. <https://doi.org/10.33087/ekonomis.v4i2.179>