

DATA SCIENCE: MACHINE LEARNING

**OPTIMASI PERFORMA DINAMIS DAN EFISIENSI PROSES
DARI PENGIRIMAN DOORDASH DI AREA METROPOLITAN
LOS ANGELES BERDASARKAN WAKTU DAN PENGIRIMAN
MENGUNAKAN METODE RANDOM FOREST.**

ISYE6301011

Submitted to:

D6127 - Dr. Ir. RIENNA OKTARINA, S.T., M.T.

Lecturer of Data Science: Machine Learning

Bina Nusantara University



Oleh:

Bayu Askha Wijaya Nasoetion 2602172611

Marc Anthony Samuel 2602193982

Samuel Yuda Lampe 2602185375

JURUSAN TEKNIK INDUSTRI

FAKULTAS TEKNIK

BINUS UNIVERSITY

JAKARTA

2024

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

DoorDash adalah perusahaan teknologi yang berkecimpung di sektor teknologi yang menawarkan jasa pada bisnis-bisnis kecil yang berada ada di 30 negara. Didirikan pada tahun 2013 oleh tiga orang yaitu Tony Xu, Stanley Tang, Andy Fang, dan Evan Moore. DoorDash membangun layanan untuk membantu bisnis kecil untuk tumbuh, berinovasi, dan membangun pasar lebih luas. DoorDash dibuat untuk memberikan service yang dimana perusahaan menawarkan kemudahan untuk memenuhi demand dengan menawarkan service pickup atau delivery untuk pelanggan. Model bisnisnya berfokus untuk pengiriman yang cepat dan efisien, selain itu DoorDash juga turut membuka lapangan kerja bagi penduduk-penduduk di negara yang dimana DoorDash beroperasi.

Era digital mengalami perkembangan yang signifikan dan menjadi sektor pertumbuhan yang tumbuh pesat. DoorDash sebagai perusahaan yang bergerak di bidang teknologi berupaya memenuhi demand konsumen yang tinggi dengan harapan akan kecepatan sesuai kemampuan layanan. DoorDash mengalami tantangan dalam berbisnis dengan mengeluarkan inovasi pelayanan yang cepat dengan mengeluarkan biaya yang minimal. Persaingan DoorDash dengan kompetitor lainnya semakin kuat, alasan tersebut menjadi kewajiban perusahaan untuk bersaing dengan kompetitor dengan mengeluarkan inovasi dan strategi yang tetap kompetitif diantara pesaing. Cara agar DoorDash dapat mengungguli kompetitornya adalah dengan cara pengelolaan manajemen kurir, rute, dan waktu pemesanan dapat ditingkatkan.

Efisiensi dalam perusahaan berpengaruh penting dalam menjalankan bisnis seperti DoorDash dikarenakan semakin cepat pelayanan semakin tinggi kepuasan konsumen terhadap perusahaan dan meningkatkan keuntungan secara finansial. Penggunaan teknologi di machine learning dapat membantu dalam waktu pengiriman dan jenis pesanan pelanggan agar data diolah untuk mencari langkah terbaik untuk DoorDash.

1.2 Ruang Lingkup

Laporan ini ditulis untuk menganalisa berapa lama waktu pengiriman makanan, kepuasan pelanggan, dan performa kurir di DoorDash. Ruang lingkup dari laporan ini mencakup beberapa poin-poin.

1. Subjek Penelitian : Perusahaan DoorDash difokuskan menjadi perusahaan yang praktikan gunakan untuk mengamati proses pelayanan.
2. Data : Data yang praktikan peroleh untuk membuat laporan ini didapatkan melalui *website* kaggle yang mencakup data pengiriman DoorDash. Dataset yang praktikan teliti berisi waktu dan jam keterlambatan pengiriman yang mempengaruhi.
3. Aspek Yang Dikaji: Penulisan laporan ini menilai dari beberapa aspek seperti waktu mulai pengiriman, waktu selesai pengiriman, tipe jenis makanan yang dipesan, jumlah makanan yang dipesan, total waktu kerja, dan total durasi pemesanan secara keseluruhan

1.3 Tujuan

Laporan yang telah dilakukan oleh praktikan mengenai *final report* ini diharapkan dapat menjawab usulan rumusan masalah yang sudah tertera. Tujuan laporan dilakukan dengan menganalisa perusahaan DoorDash.

1. Mengetahui performa *dashers* dari waktu sibuk hingga tidak sibuk
2. Mengetahui dinamika dan efisiensi proses
3. Mengetahui pengaruh dari protokol pesanan

BAB 2

TINJAUAN PUSTAKA

2.1 Landasan Teori

Machine learning suatu gagasan yang menggunakan data-data yang diprogramkan untuk membantu dalam memproses dan memahami data. *Machine learning* menggunakan algoritma *supervised learning* dan *Unsupervised learning* untuk menentukan *output*. Fokus dari *machine learning* adalah sistem yang akan mengelompokkan datanya berdasarkan data-data yang diinput ke dalam sistem. Sistem *machine learning* menyatukan statistika, matematika, dan *data mining* untuk menganalisa suatu kejadian. *Machine learning* memiliki sistem sendiri yang mampu mempelajari data serta pemahaman dari data yang diinput oleh *user* (Baihaqi, Sarmini, Kusuma, & Honimah, 2022).

Unsupervised learning merupakan sebuah metode yang terdapat pada *machine learning* dengan tujuan mengelompokkan data, menemukan pola, dan mencari hubungan dalam data. Metode ini juga bertujuan untuk mengeksplorasi data dan juga memahami data secara mendalam. Terdapat 2 metode *algorithm* yang dapat digunakan untuk menggunakan metode *unsupervised learning*, yaitu *clustering algorithms* dan *informative data transformations*. *Clustering algorithms* merupakan sebuah metode yang digunakan untuk *unsupervised learning*, metode *clustering algorithms* ini dapat menemukan struktur pengumpulan data tanpa harus mempelajari data nya secara detail. Dalam *clustering alghoritms* ini bertujuan untuk mengukur persamaan ataupun ketidaksamaan antara yang berbeda dalam subjek pengumpulan data. Metode *clustering algorithms* ini memiliki 3 poin utama, yaitu *distance measure* untuk melihat persamaan dan pertidaksamaan data, selanjutnya ada *additional distance measure* untuk melihat perbedaan antar *cluster* ataupun antar *cluster* dengan subjek, dan yang terakhir ada *computer algorithm* yang tujuan nya adalah memaksimalkan kesamaan antar sebuah *cluster* dan ketidaksamaan antar sebuah *cluster* (Valkenborg, Rousseau, Geubbelmans, & Burzykowski, 2023).

Supply chain merupakan sebuah system untuk mengontrol dan mengintegrasikan seluruh entitas yang ada dalam proses produksi barang seperti

supplier ke *manufacture*, *manufacture* ke *wholesaler*, *wholesaler* ke *retailer*, dan *retailer* ke *end customer*. Teknik *machine learning* berguna untuk mengintegrasikan model simulasi dengan tujuan mengefisiensikan *control process* yang biasanya membutuhkan tenaga manusia. *Supply chain* merupakan sebuah sistem yang rumit, jadi harus tetap mencari cara agar sistem *supply chain* tersebut bisa berjalan dengan lancar. Untuk mengambil keputusan seperti penjadwalan, *lead time*, *routing*, dan *truck loading* biasanya akan di ambil alih oleh *manager* (Ali, et al., 2022).

Metode *machine learning* ini dapat menciptakan solusi praktis untuk sistem *supply chain* pada industri. Pendekatan dengan cara melakukan simulasi sudah banyak dilakukan untuk digunakan menyelidiki masalah-masalah yang terjadi pada internal pabrik terutama untuk bagian logistik. Model simulasi menyelidiki gudang efisiensi dan mengurangi biaya logistik internal. Ketidakmampuan untuk menyelidiki potensi bahaya serta mengembangkan mitigasi untuk potensi bahaya yang kemungkinan besar terjadi dapat diatasi dengan menerapkan teknik *machine learning* (Ali, et al., 2022).

2.2 Penelitian Terkait

Bisnis *delivery* dalam lingkup *supply chain* telah menjadi bagian dari perusahaan DoorDash dan kompetitornya yang melibatkan jaringan yang kompleks dengan menghubungkan restoran, *dasher*, dan pelanggan. Tantangan yang mengakibatkan bisnis DoorDash kurang berkembang adalah biaya operasional yang tinggi dan biaya aplikasi yang memberatkan *dasher*. Semakin luasnya layanan pengantaran *dasher* makan peningkatan konsumsi energi yang menimbulkan dampak negatif terhadap, sehingga isu ini perlu ditangani untuk mengevaluasi data untuk menilai dan mengkategorikan strategi yang dinilai efektif untuk bisnis DoorDash (Li, et al., 2024). Dampak keterlambatan di dalam bisnis menghambat pertumbuhan perusahaan, *supply chain* sangat berpengaruh di dalam proses bisnis model DoorDash dikarenakan kelancaran *supply chain* dan input data yang akurat memungkinkan DoorDash untuk memperbaiki sistem yang kurang efisien menjadi efisien dibandingkan kompetitor (Carvalho, Naghshineh, Govindan, & Machado, 2022).

Data yang ingin di observasi berasal dari situs www.kaggle.com yang membahas tentang *dataset* mengenai kota dimana DoorDash beroperasi, waktu estimasi waktu restoran menerima pesanan dari DoorDash, berapa lama waktu barang diterima konsumen, kategori jumlah makanan yang dipesan, harga termahal dan termurah produk yang dipesan, total *shift dasher* bekerja, jarak order dalam radius 10 *miles*, dan jumlah total secara keseluruhan order konsumen. Data-data tersebut akan diolah untuk mencari cara kemampuan DoorDash untuk mencari cara agar waktu pemesanan bisa dipersingkat agar efektif dan efisien.

BAB 3

DATA & *PREPROCESSING*

3.1 Deskripsi Data

Pencarian data yang berkaitan dengan rantai pasok dari sebuah industri perusahaan. Berdasarkan kecocokan data dan minat praktikan, praktikan memilih dataset *doordash* yang terdiri dari pencacatan waktu dan tanggal pemesanan.

Dataset yang digunakan merupakan *dataset* publik yang dapat diakses oleh khalayak umum. Pencarian dilakukan melalui situs www.kaggle.com yang merupakan sebuah situs kumpulan *dataset* dan analisis.

3.2 Pembersihan Data (*Data Cleaning*)

Pembersihan data dilakukan sebagai tindakan awal dari eksplorasi data, pembersihan data ini ditujukan agar *dataset* yang diberikan dapat dianalisa secara akurat dan menghindari terjadinya *bug* atau *error*. Pembersihan data dapat dilihat pada Gambar 3.1 *Data Cleaning Code*.

Gambar 3.1 *Data Cleaning Code*

```
1 import pandas as pd
2 from IPython.display import display
3
4 class DataWrangler:
5     def __init__(self, url):
6         self.url = url
7         self.df = None
8         self.features = None
9
10    def load_data(self):
11        # Load the dataset
12        self.df = pd.read_csv(self.url)
13        print("Raw Dataset:")
14        display(self.df.head())
15
16    def preprocess_data(self):
17        # Convert datetime columns to datetime objects
18        self.df['created_at'] = pd.to_datetime(self.df['created_at'])
19        self.df['actual_delivery_time'] = pd.to_datetime(self.df['actual_delivery_time'])
20
21        # Print the data types to verify the conversion
22        print("\nData Types after Conversion:")
23        display(self.df.dtypes)
24
25        # Feature engineering
26        self.df['Delivery Duration'] = (self.df['actual_delivery_time'] - self.df['created_at']).dt.total_seconds()
27
28        # Print the first few rows after feature engineering
29        print("\nDataset after Feature Engineering:")
30        display(self.df.head())
31
32    def select_features(self):
33        # Select relevant columns for clustering
34        self.features = self.df[['subtotal',
35                                'total_items',
36                                'num_distinct_items',
37                                'min_item_price',
38                                'max_item_price',
39                                'total_onshift_dashers',
40                                'total_busy_dashers',
41                                'total_outstanding_orders',
42                                'estimated_order_place_duration',
43                                'estimated_store_to_consumer_driving_duration',
44                                'Delivery Duration']]
45
46        # Handle missing values if any
47        self.features = self.features.dropna()
48
49        # Print the first few rows after handling missing values
50        print("\nDataset after Handling Missing Values:")
51        display(self.features.head())
52
53    def get_features(self):
54        return self.features
55
56    # Usage
57    url = 'https://raw.githubusercontent.com/marcanthonny/projectDSML/refs/heads/main/historical_data.csv'
58    data_wrangler = DataWrangler(url)
59    data_wrangler.load_data()
60    data_wrangler.preprocess_data()
61    data_wrangler.select_features()
62    features = data_wrangler.get_features()
```

Pembersihan data dilakukan sebagai tindakan awal dari eksplorasi data, pembersihan data ini ditujukan agar *dataset* yang diberikan dapat dianalisa secara akurat dan menghindari terjadinya *bug* atau *error*. Hasil dari pembersihan data dapat dilihat pada Tabel 3.1 *Data Cleaning Output*.

Tabel 3.1 *Data Cleaning Output*

No.	market_id	created_at	actual_delivery_time	store_id	store_primary_category	order_protocol	total_items	subtotal	num_distinct_items
0	1.0	2/6/2015 22:24	2/6/2015 23:27	1845	american	1.0	4	3441	4
1	2.0	2/10/2015 21:49	2/10/2015 22:56	5477	mexican	2.0	1	1900	1
2	3.0	1/22/2015 20:39	1/22/2015 21:09	5477	NaN	1.0	1	1900	1
3	3.0	2/3/2015 21:21	2/3/2015 22:13	5477	NaN	1.0	6	6900	5
4	3.0	2/15/2015 2:40	2/15/2015 3:20	5477	NaN	1.0	3	3900	3

Tabel 3.1 *Data Cleaning Output (Lanjutan)*

min_item_price	max_item_price	total_onshift_dashers	total_busy_dashers	total_outstanding_orders	estimated_order_placement_duration	estimated_store_to_consumer_driving_duration
557	1239	33.0	14.0	21.0	446	861.0
1400	1400	1.0	2.0	2.0	446	690.0
1900	1900	1.0	0.0	0.0	446	690.0
600	1800	1.0	1.0	2.0	446	289.0
1100	1600	6.0	6.0	9.0	446	650.0

3.3 Eksplorasi Data

Eksplorasi data dilakukan untuk mengetahui lebih lanjut tentang struktur dan distribusi data yang ada pada dataset, mengidentifikasi pola dan hubungan, mendeteksi anomali dan nilai ekstrem, dan dapat memilih fitur-fitur yang relevan. Data hasil eksplorasi ditunjukkan dalam Tabel 3.2 Eksplorasi Data

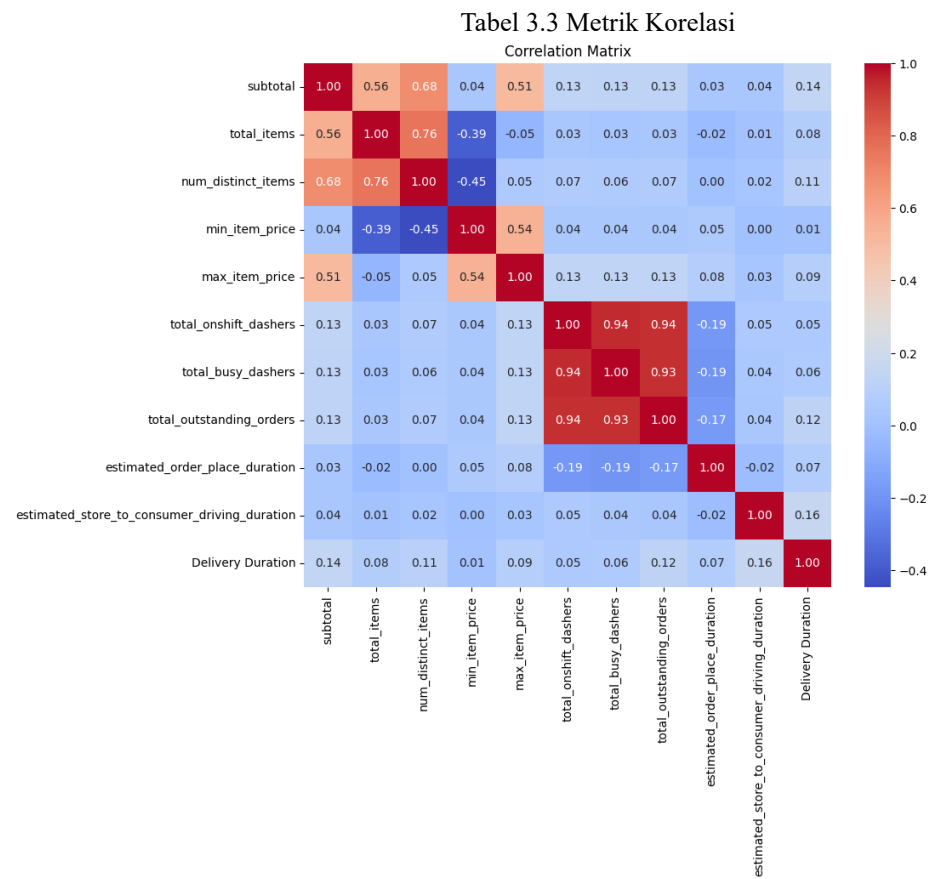
Tabel 3.2 Eksplorasi Data

No.	subtotal	total_items	num_distinct_items	min_item_price	max_item_price	total_onshift_dashers
0	3441	4	4	557	1239	33.0
1	1900	1	1	1400	1400	1.0
2	1900	1	1	1900	1900	1.0
3	6900	6	5	600	1800	1.0
4	3900	3	3	1100	1600	6.0

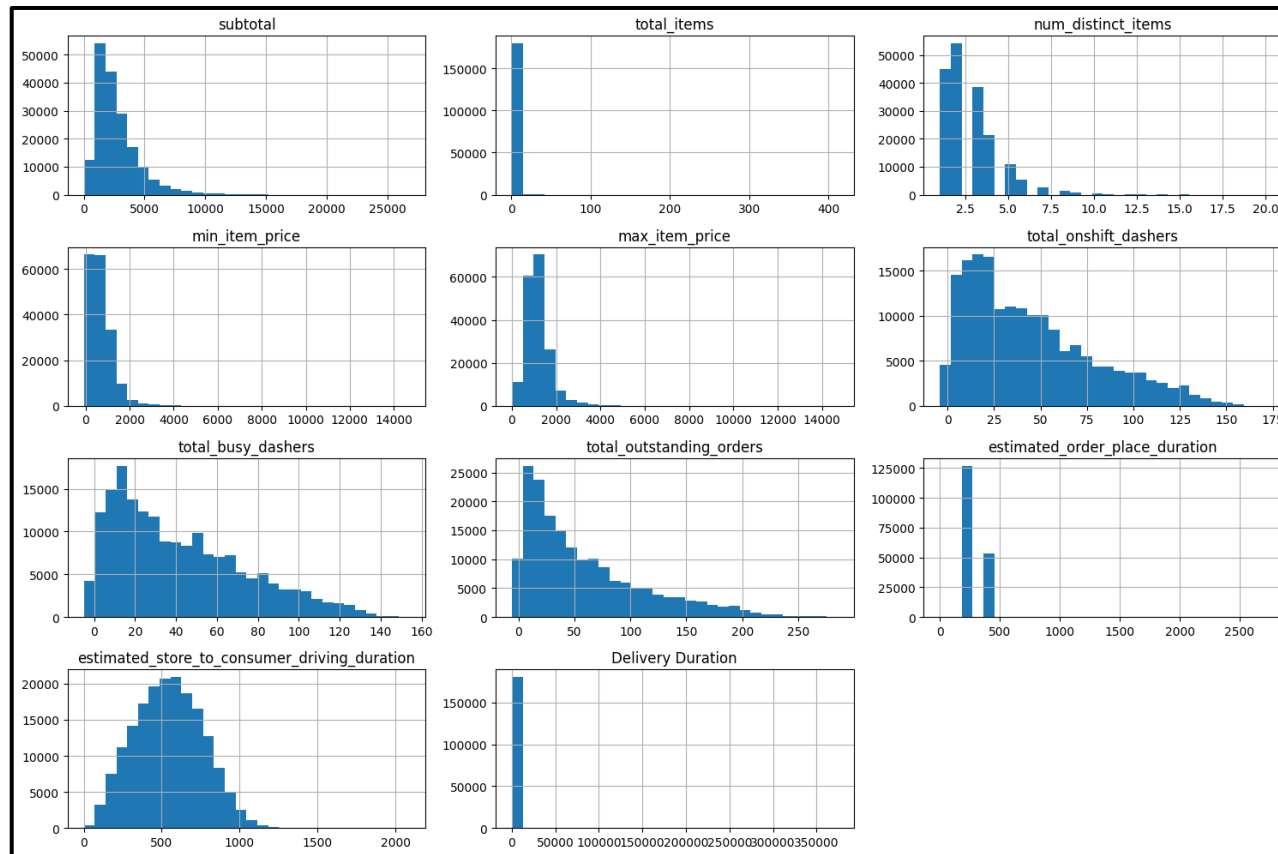
Tabel 3.2 Eksplorasi Data (Lanjutan)

total_busy_dashers	total_outstanding_orders	estimated_order_place_duration	estimated_store_to_consumer_driving_duration	Delivery Duration
14.0	21.0	446	861.0	3779.0
2.0	2.0	446	690.0	4024.0
0.0	0.0	446	690.0	1781.0
1.0	2.0	446	289.0	3075.0
6.0	9.0	446	650.0	2390.0

Eksplorasi data dilakukan untuk mengetahui lebih lanjut tentang struktur dan distribusi data yang ada pada dataset, mengidentifikasi pola dan hubungan, mendeteksi anomali dan nilai ekstrem, dan dapat memilih fitur-fitur yang relevan dengan menggunakan **tabel metrik korelasi**. Berikut adalah Tabel 3.3 Metrik Korelasi. Tabel korelasi menunjukkan bahwa terdapat koefisien korelasi 0,94 antara total_onshift_dashers dengan total_busy_dashers. Koefisien 0,93 antara total_onshift_dashers dengan total_outstanding_orders, dan koefisien 0,93 antara total_busy_dashers dengan total_outstanding_orders.



Eksplorasi data dilakukan untuk mengetahui lebih lanjut tentang struktur dan distribusi data yang ada pada dataset, mengidentifikasi pola dan hubungan, mendeteksi anomali dan nilai ekstrem, berikut adalah Gambar 3.2 Distribusi Variabel yang menunjukkan distribusi variabel yang akan digunakan dalam penelitian ini.



Gambar 3.2 Distribusi Variabel

Histogram menunjukkan bahwa sebagian besar pesanan memiliki subtotal kurang dari 5000 unit dan biasanya berisi antara 0 dan 50 item. Waktu transmisi umumnya sekitar 5000 detik, yang terbukti merupakan proses yang efisien. Harga produk minimum dan maksimum

biasanya kurang dari 2000 unit, yang menunjukkan konsistensi dalam penetapan harga produk. Jumlah petugas polisi yang bertugas dan jumlah perintah sibuk serta perintah yang belum diselesaikan biasanya kurang dari 50, yang mencerminkan kerja yang baik dan manajemen komando. Misalnya, jumlah item berbeda dalam satu pesanan biasanya antara 2,5 dan 10, yang menunjukkan kecenderungan untuk melakukan beberapa pesanan. Proses pemesanan biasanya cepat, dengan sebagian besar pesanan diselesaikan dalam waktu 500 detik, sehingga meningkatkan kenyamanan pelanggan. Waktu tempuh dari toko ke konsumen juga biasanya sekitar 500 detik, yang menunjukkan waktu pengiriman yang cepat dan konsisten, kemungkinan besar karena rute yang efisien dan kedekatannya dengan lokasi pengiriman. Secara umum, data ini menyoroti sistem pemesanan dan pengiriman yang dikelola dengan baik dan efisien. Waktu pengiriman yang singkat dan rendahnya jumlah pesanan yang belum diselesaikan menunjukkan sistem manajemen dan pengiriman yang efektif. Proses yang efisien ini, disertai dengan pemilihan yang baik dan penetapan harga yang konsisten, menunjukkan layanan yang berorientasi pada pelanggan dan operasi yang efektif. Analisis histogram ini membantu mengidentifikasi area yang memerlukan optimasi lebih lanjut dan memastikan kepuasan pelanggan yang tinggi secara berkelanjutan.

3.4 Transformasi Data

Transformasi data bertujuan untuk menyiapkan data agar siap digunakan dalam model dengan mengatasi kualitas data, dan membuat data lebih informatif/relevan untuk meningkatkan performa prediksi komputatif melalui *feature engineering*. Hasil transformasi data dapat dilihat pada Tabel 3.4 Hasil Transformasi Data.

Tabel 3.4 Hasil Transformasi Data.

No.	market_id	created_at	actual_delivery_time	store_id	store_primary_category	order_protocol	total_items	subtotal	num_distinct_items	min_item_price
0	1.0	2/6/2015 22:24	2/6/2015 23:27	1845	american	1.0	4	3441	4	557
1	2.0	2/10/2015 21:49	2/10/2015 22:56	5477	mexican	2.0	1	1900	1	1400
2	3.0	1/22/2015 20:39	1/22/2015 21:09	5477	NaN	1.0	1	1900	1	1900
3	3.0	2/3/2015 21:21	2/3/2015 22:13	5477	NaN	1.0	6	6900	5	600
4	3.0	2/15/2015 2:40	2/15/2015 3:20	5477	NaN	1.0	3	3900	3	1100

Tabel 3.4 Hasil Transformasi Data. (Lanjutan)

max_item_price	total_onshift_dashers	total_busy_dashers	total_outstanding_orders	estimated_order_placement_duration	estimated_store_to_consumer_driving_duration	Delivery Duration
1239	33.0	14.0	21.0	446	861.0	3779.0
1400	1.0	2.0	2.0	446	690.0	4024.0
1900	1.0	0.0	0.0	446	690.0	1781.0
1800	1.0	1.0	2.0	446	289.0	3075.0
1600	6.0	6.0	9.0	446	650.0	2390.0

Transformasi data menghasilkan beberapa wawasan tentang kinerja kurir dengan menganalisa *dashers* yang *on shift* dan yang *busy*. Nilai pesanan dengan durasi pengiriman untuk dinamika dan efisiensi proses. Protokol pesanan yang mempengaruhi durasi dan kualitas pengiriman.

BAB 4

PEMODELAN *MACHINE LEARNING*

4.1 Pemilihan Algoritma

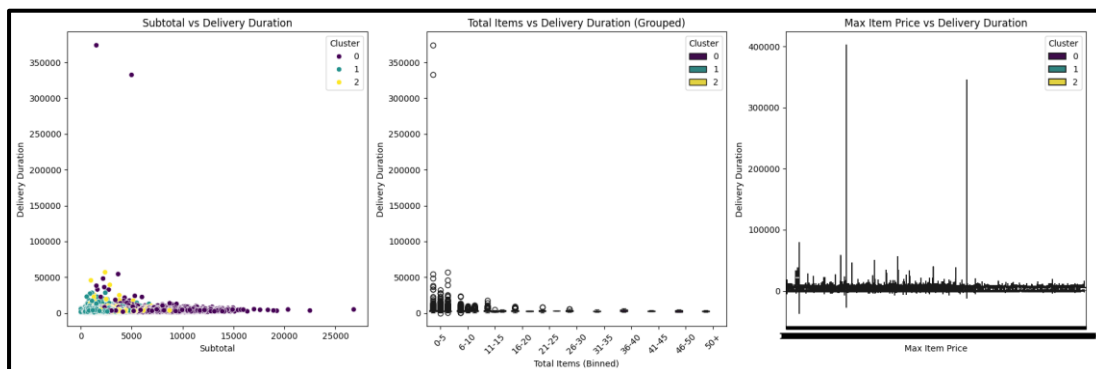
Algoritma terpilih merupakan *clustering* untuk mengklasifikasikan poin-poin data yang memiliki kemiripan khusus yang dapat dikelompokkan. Pertimbangan ini didasari dengan usaha dalam menemukan kelompok tertentu dalam sebuah *dataset* yang dapat dijadikan model pelatihan lebih lanjut. Terdapat 3 variabel yang dapat dilakukan model algoritma *clustering* yaitu

1. *Subtotal vs Delivery Duration*
2. *Total Items vs Delivery Duration*
3. *Max Item Price vs Delivery Duration*

Model selanjutnya menggunakan model *random forest* untuk membuat prediksi yang akurat, mengurangi *overfitting*, dan mengidentifikasi fitur yang paling penting (*feature importance*). *Random forest* juga dapat *handle* data *numerical* dan data kategorikal. Model *training* ini dapat menunjukkan keakuratan sistem dan mesin dalam mempelajari pengklasifikasian.

4.1 Clustering

Clustering dilakukan terhadap *dataset doordash* menggunakan 3 model *chart*, *scatter plot chart*, *boxplot chart*, dan *violin chart* dikarenakan kecocokannya dengan variabel yang ada. Hasil ketiga *clustering* variabel dapat dilihat melalui Gambar 4.1 Hasil *Clustering* 3 Variabel.



Gambar 4.1 Hasil *Clustering* 3 Variabel

4.3 Training dan Validasi Data

Training dan validasi data digunakan dan berfungsi untuk mempelajari model dan mengoptimalkan parameter yang ada sehingga model mampu memberikan prediksi yang lebih akurat. Tahap *training* model akan mempelajari pola-pola dari data untuk menyesuaikan parameter agar meminimalkan *error* sekecil mungkin. Validasi data digunakan untuk mengevaluasi kinerja model pada data yang belum pernah di tes atau dianalisa sebelumnya. Validasi membantu mendeteksi *overfitting* dan memastikan model cocok pada data dan dapat digeneralisasikan ke dalam data baru. Berikut adalah Gambar 4.2 Hasil *Training Data*.

```
Training and Testing Set Shapes:
X_train: (144541, 11), X_test: (36136, 11)
y_train: (144541,), y_test: (36136,)

Model Training Completed

Predictions:
[0 1 1 2 1 0 2 1 1 1]

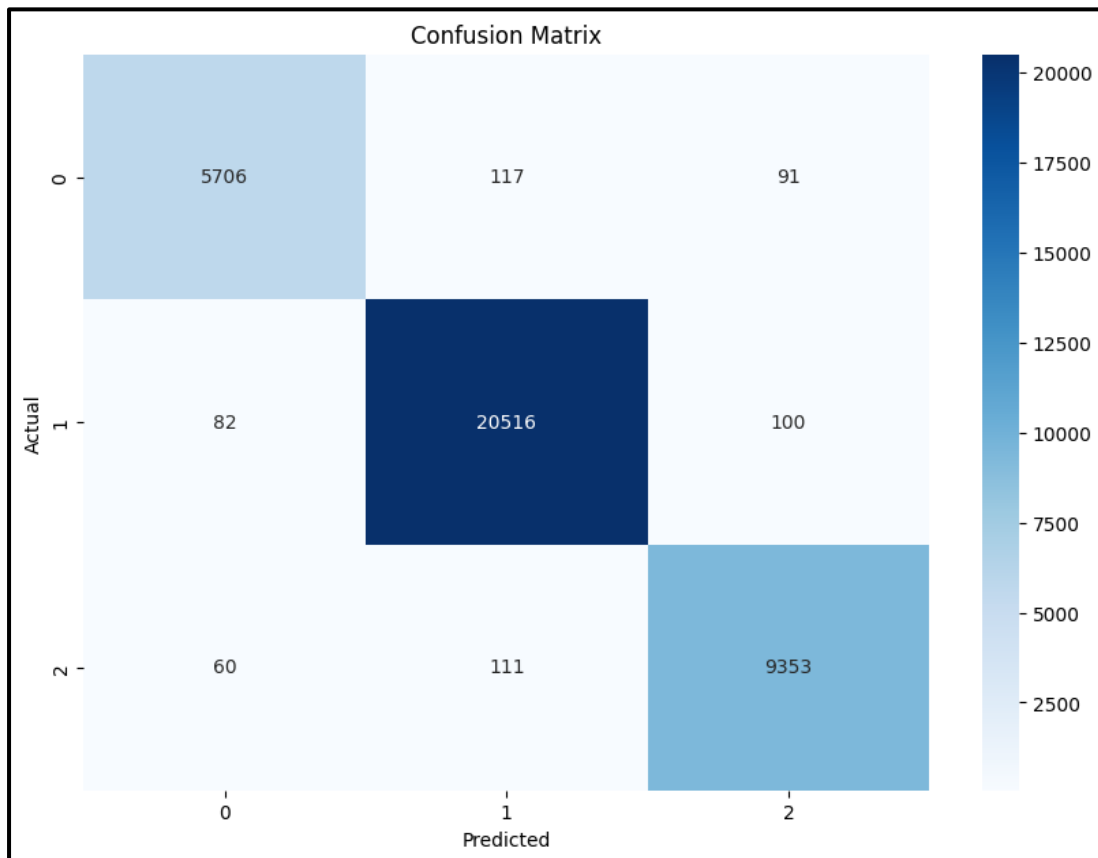
Classification Report:
              precision    recall  f1-score   support

     0           0.98       0.96       0.97       5914
     1           0.99       0.99       0.99      20698
     2           0.98       0.98       0.98       9524

 accuracy              0.98              36136
 macro avg           0.98       0.98       0.98       36136
 weighted avg       0.98       0.98       0.98       36136
```

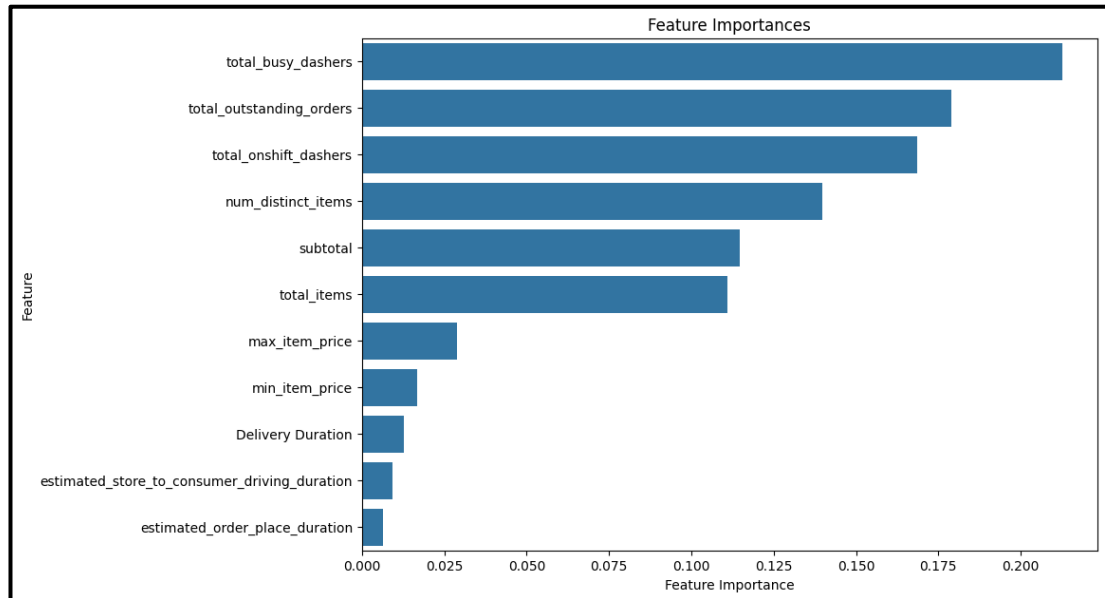
Gambar 4.2 Hasil Training Data

Berdasarkan *confusion matrix*, akurasi *training* ini sudah tinggi dan sangat baik, dengan jumlah prediksi yang tinggi untuk semua *class*. Jumlah kesalahan klasifikasi memang ada namun jumlahnya sangat relatif kecil. *Class* dengan performa terbaik adalah *class* 1 dengan 20516 *instances* dan kesalahan minim. *Confusion matrix* dapat dilihat pada Gambar 4.3 *Confusion Matrix*.



Gambar 4.3 Confusion Matrix

Hasil *training* data menunjukkan akurasi sebesar 98% dengan performa dihasilkan sangat baik dengan performa yang *balance*. Beberapa variabel dapat ditentukan dalam prioritas dalam *training* model ini, berikut adalah Gambar 4.4 *Feature Importance*.



Gambar 4.4 Feature Importance.

DAFTAR PUSTAKA

- Ali, N., Ghazal, T. M., Ahmed, A., Abbas, S., Khan, M. A., Alzoubi, H. M., . . . Khan, M. A. (2022). *Fusion-Based Supply Chain Collaboration Using Machine Learning Techniques*. Korea: Tech Science Press.
- Baihaqi, W. M., Sarmini, Kusuma, B. A., & Honimah, L. H. (2022). *LEVEL DASAR PEMBELAJARAN MESIN DENGAN PYTHON DAN IOT DENGAN C*. Jakarta: Zahira Media Publisher.
- Carvalho, H., Naghshineh, B., Govindan, K., & Machado, V. C. (2022). Computers & Industrial Engineering. *The resilience of on-time delivery to capacity and material shortages empirical investigation in the automotive supply chain*.
- Li, L., Zhang, Z., Xixi, L., Su, J., Jiang, Y., Cao, J., & Zhou, F. (2024). Heliyon . *Mining the sustainability of takeaway businesses in online food*.
- Valkenborg, D., Rousseau, A. J., Geubbelmans, M., & Burzykowski, T. (2023). Unsupervised learning. *STATISTICS AND RESEARCH DESIGN*.

LAMPIRAN

DoorDash ETA Prediction BY Kaggle:

<https://www.kaggle.com/code/dharun4772/doordash-eta-regression-prediction-eda>

Data Cleaning:

<https://github.com/marcanthonny/projectDSML/blob/1df39bcd3881adccafbacf8ec9ead849ad676a4a/Data%20Cleaning%20%26%20Exploration.ipynb>

Data Clustering:

<https://github.com/marcanthonny/projectDSML/blob/1df39bcd3881adccafbacf8ec9ead849ad676a4a/Data%20Clustering.ipynb>

Data Training dan Random Forest Tree

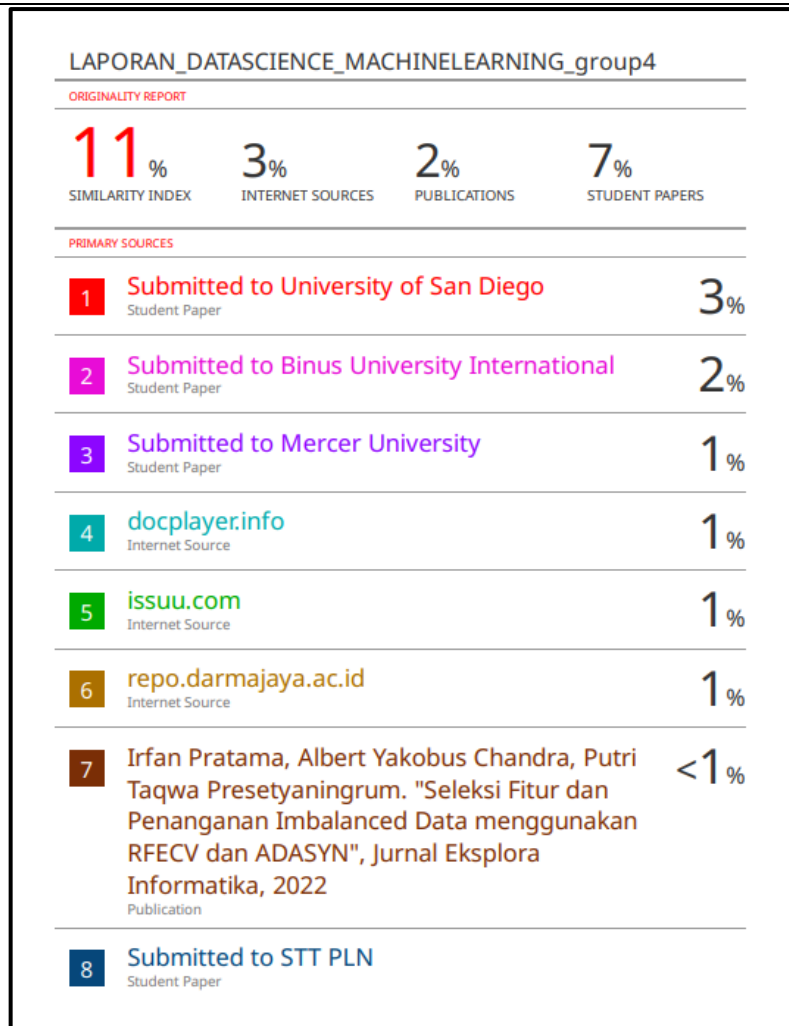
<https://github.com/marcanthonny/projectDSML/blob/1df39bcd3881adccafbacf8ec9ead849ad676a4a/Data%20Training.ipynb>

PPT Project:

https://www.canva.com/design/DAGUNMBLiOc/UqZ5HOQnQTeK_GqzvFZngA/edit?utm_content=DAGUNMBLiOc&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Table 1 *Worksheet Data Science: Machine Learning*

No	Bab	Anggota Kelompok		
		Marc Anthony Samuel	Bayu Askha Wijaya Nasoetion	Samuel Yuda Lampe
1	Pendahuluan		x	x
2	Tinjauan Pustaka		x	x
3	Data Preprocessing	x		
4	Pemodelan Machine Learning	x		
5	Interpretasi Hasil			
6	Kesimpulan dan Rekomendasi			
7	Refrensi		x	x
8	Lampiran	x	x	x



Gambar 1 Hasil Turnitin Laporan Group 4