



Improving Fitbit Retention Through Data Analysis



Dokumen Laporan Final Project Kelompok 4

Anggota:

Bintang Phylosophie
Asty Nabilah 'Izzaturrahmah
Putri Sausan Azzahra Alindra

David Bryan Christiansen
Tofan Pratama

Dataset

Dataset yang digunakan pada project ini ada 2 yaitu: **dailyActivity_merged.csv** dan **minuteMETsNarrow_merged.csv**

Dataset ini terdiri dari beberapa atribut yang berhubungan dengan aktivitas fisik harian 32 orang pengguna, dengan 14 kolom dan 940 row. dengan meta data seperti berikut:

Id: Identitas pengguna.

TotalSteps: Jumlah langkah total yang diambil pengguna.

TotalDistance: Jarak total yang ditempuh (dalam kilometer).

TrackerDistance: Jarak yang dilacak oleh perangkat.

LoggedActivitiesDistance: Jarak yang tercatat dari aktivitas yang dilacak secara manual.

VeryActiveDistance: Jarak yang ditempuh selama aktivitas fisik yang sangat aktif.

ModeratelyActiveDistance: Jarak yang ditempuh selama aktivitas moderat.

LightActiveDistance: Jarak yang ditempuh selama aktivitas ringan.

SedentaryActiveDistance: Jarak yang ditempuh saat aktivitas ringan.

Dataset

VeryActiveMinutes: Jumlah menit yang dihabiskan untuk aktivitas sangat aktif.

FairlyActiveMinutes: Jumlah menit yang dihabiskan untuk aktivitas cukup aktif.

LightlyActiveMinutes: Jumlah menit yang dihabiskan untuk aktivitas ringan.

SedentaryMinutes: Jumlah menit dalam kondisi duduk atau tidak bergerak.

Calories: Target output, yaitu jumlah kalori yang dibakar oleh pengguna.

ActivityDate: berisi tanggal dalam format string yang dapat diabaikan dalam perhitungan numerik karena tidak relevan untuk prediksi kalori.

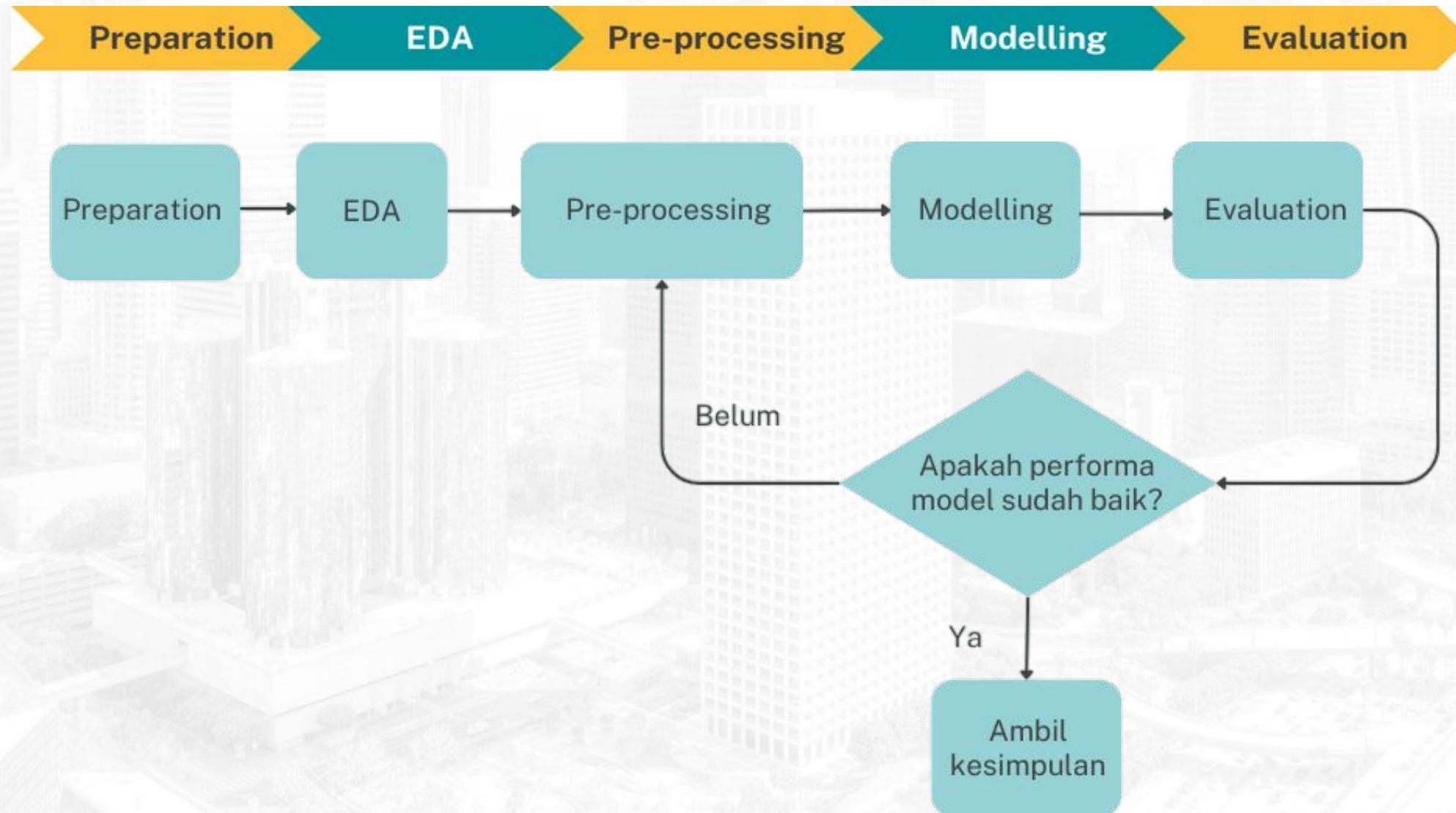
AvgMET: Rata-rata MET dalam 1 hari

TotalMET: Total MET dalam 1 hari

Notes:

MET: Metabolic equivalents (rasio tingkat energi yang dikeluarkan saat aktif : saat istirahat)

Project Flow



Presentation Agenda



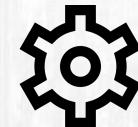
1. Preparation



2. EDA & Insight



3. Data Preparation



4. Modelling & Evaluation

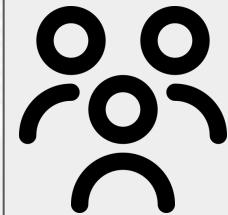


5. Business Recommendation



Preparation

Latar Belakang Masalah



38.5 M

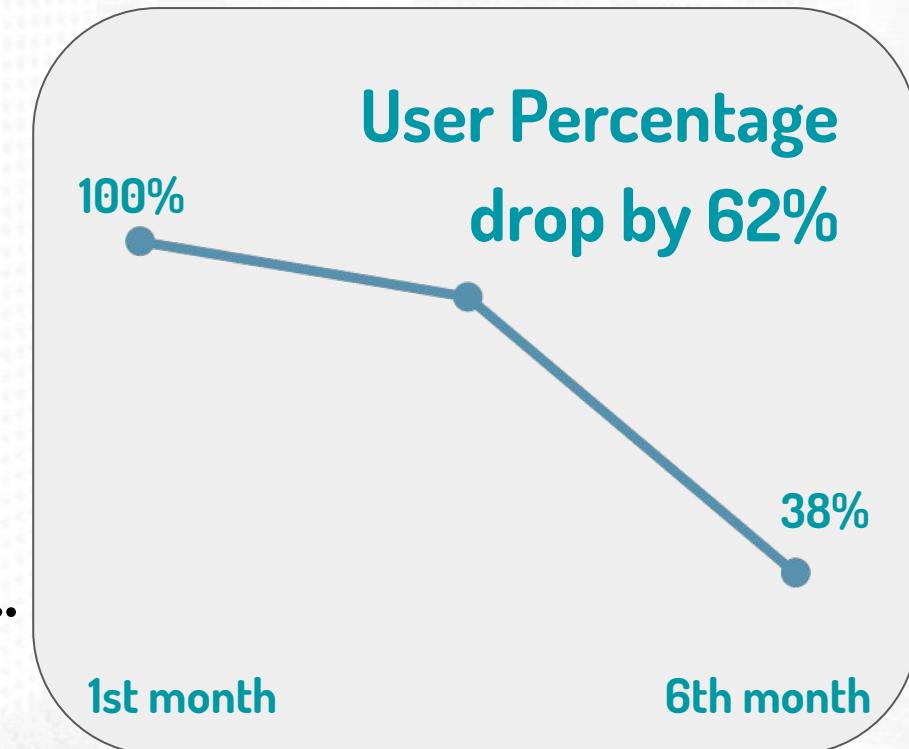
User around the world*

But...



Kebosanan dan kurangnya personalisasi wawasan yang relevan.**

Because...



*Statista(2023)

**Gartner(2022)

Bagaimana Agar User Tetap Bertahan Dalam Jangka Panjang?

Solusi yang Dapat Kembangkan*:

Fitur Personalisasi



6/10

Gamifikasi



6/10

Notifikasi Motivasi



5/10



namun perlu dilakukan analisis mendalam dan pemodelan untuk memastikan strategi yang diambil sesuai dengan kebutuhan.

*Varecol(2023)

Fitur Personalisasi

Reddy, et al. (2023)

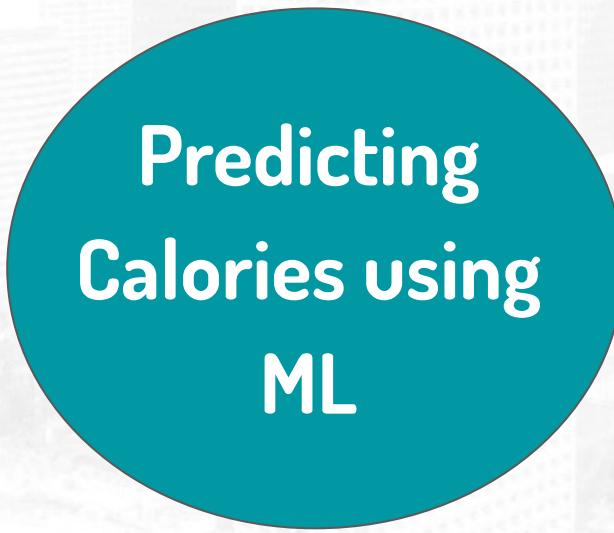
“Measurement of exact calories burned can be difficult”

Laranjo, et al. (2020)

“individuals tracking caloric expenditure are more likely to achieve weight loss or maintenance, up to **16% greater weight loss**”

Laranjo, et al. (2020)

“ML-based models predict calories burned **more accurately** compared to traditional formulas”



Predicting
Calories using
ML

Laranjo, et al. (2020)

“Predicting calorie expenditure based on real-time activity **empowers users to set realistic health goals**”

Edward, et al. (2018)

“Calorie prediction can make an app significantly improves engagement rates, often **increasing active daily use by 20%-25%**”

Edward, et al. (2018)

“ML model **can improve over time** by analyzing trends in user behavior and refining predictions”

Latar Belakang Masalah

Solusi yang Kami Kembangkan:

- Memberikan business recommendation dari data hasil eksplorasi data
- Tim data science kami merancang **model prediksi yang mempersonalisasi wawasan kesehatan**, bertujuan untuk memberikan rekomendasi yang lebih tepat dan relevan bagi pengguna.
- Salah satu fitur utama model ini adalah **prediksi jumlah kalori yang terbakar setiap hari**, yang dapat membantu pengguna mengelola kesehatan mereka dengan lebih efektif.

Dampak yang Diharapkan:

- memberikan rekomendasi yang lebih relevan, **meningkatkan efektivitas dan penggunaan aplikasi dalam jangka panjang**.

Tujuan Jangka Panjang:

- Melalui upaya ini, kami berharap Fitbit bukan hanya dapat **meningkatkan keterlibatan pengguna**, tetapi juga secara signifikan, serta **mengurangi risiko penyakit terkait gaya hidup tidak aktif**.
- Model ini diharapkan akan membantu Fitbit menjadi **alat yang lebih efektif** dalam mencapai tujuan kesehatan global, membantu lebih banyak orang hidup lebih sehat dan aktif.

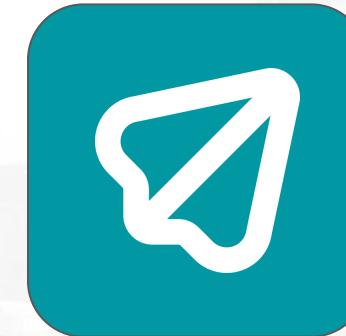


Goal



Meningkatkan keterlibatan jangka panjang dan berkontribusi secara signifikan dalam mencapai tujuan kesehatan yang lebih luas.

Objective



- **Memprediksi jumlah kalori yang akan dibakar**
- Mendapatkan **rekomendasi bisnis**

Goal

Meningkatkan keterlibatan dan kepuasan jangka panjang pengguna melalui **prediksi kalori yang dipersonalisasi** dapat membantu pengguna dalam mengelola dan memantau kalori yang terbakar. Dengan demikian, Fitbit berpotensi **berkontribusi secara signifikan** dalam mengurangi risiko penyakit yang terkait dengan gaya hidup tidak aktif, serta menjadi alat yang lebih efektif dalam mencapai tujuan kesehatan yang lebih luas.

Objective

Dalam peran kami sebagai data scientist, objective kami adalah:

- Mengembangkan machine learning yang mampu memprediksi jumlah kalori yang akan dibakar pengguna setiap hari, berdasarkan rata-rata aktivitas harian mereka
- Mendapatkan business insight dan rekomendasi bisnis berdasarkan eksplorasi data dan hasil modelling

Business Metrics

Churn rate: Menurunkan persentase user yang berhenti menggunakan fitbit

$$Churn\ Rate = \frac{\text{Pelanggan yang Berhenti}}{\text{Total Pelanggan Awal}} \times 100\%$$



Exploratory Data Analyst, Insight, & Visualization

DESCRIPTIVE STATISTICS

Pada dataset terdapat 14 kolom, 13 numeric dan 1 date dan row sejumlah 940.

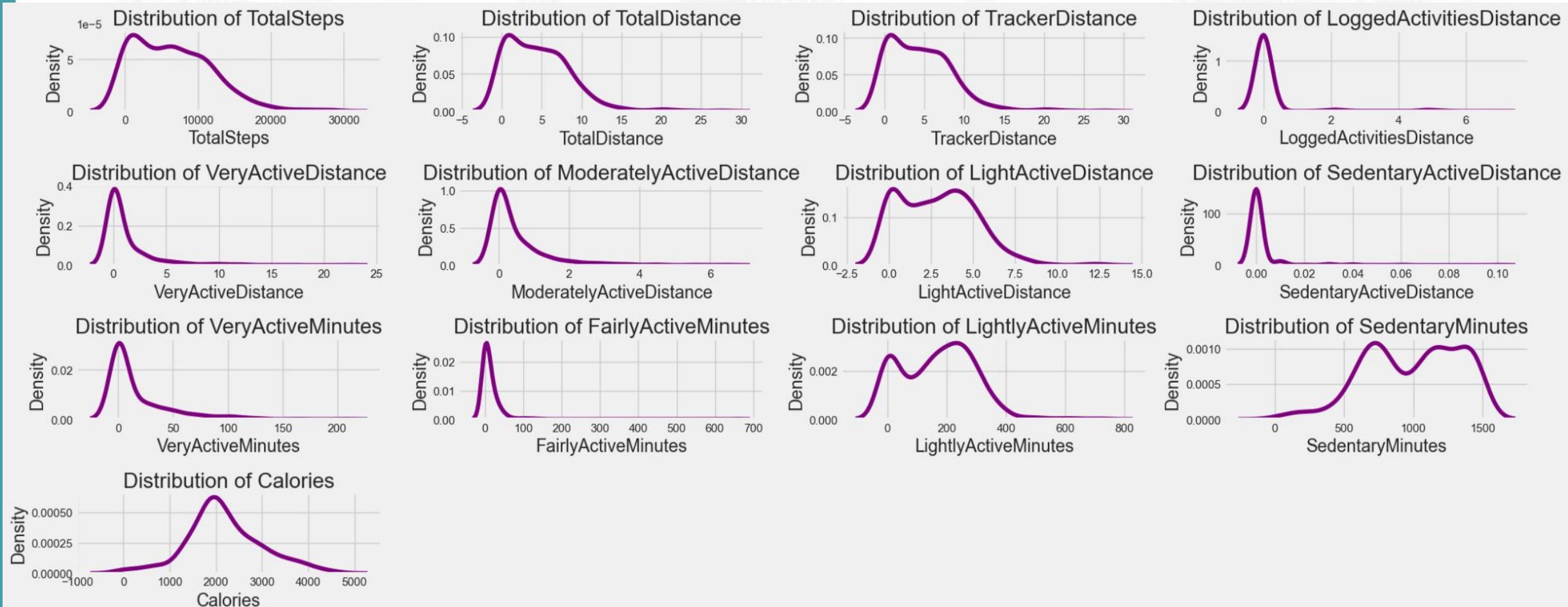
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 940 entries, 0 to 939
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   Id               940 non-null    int64  
 1   ActivityDate     940 non-null    object  
 2   TotalSteps       940 non-null    int64  
 3   TotalDistance    940 non-null    float64 
 4   TrackerDistance  940 non-null    float64 
 5   LoggedActivitiesDistance  940 non-null  float64 
 6   VeryActiveDistance  940 non-null  float64 
 7   ModeratelyActiveDistance 940 non-null  float64 
 8   LightActiveDistance 940 non-null  float64 
 9   SedentaryActiveDistance 940 non-null  float64 
 10  VeryActiveMinutes 940 non-null    int64  
 11  FairlyActiveMinutes 940 non-null  int64  
 12  LightlyActiveMinutes 940 non-null  int64  
 13  SedentaryMinutes 940 non-null    int64  
 14  Calories          940 non-null    int64  
dtypes: float64(7), int64(7), object(1)
memory usage: 110.3+ KB
```

	data.describe()						
	Id	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	
count	9.400000e+02	940.000000	940.000000	940.000000	940.000000	940.000000	940.000000
mean	4.855407e+09	7637.910638	5.489702	5.475351	0.108171	1.502681	
std	2.424805e+09	5087.150742	3.924606	3.907276	0.619897	2.658941	
min	1.503960e+09	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2.320127e+09	3789.750000	2.620000	2.620000	0.000000	0.000000	0.000000
50%	4.445115e+09	7405.500000	5.245000	5.245000	0.000000	0.210000	
75%	6.962181e+09	10727.000000	7.712500	7.710000	0.000000	2.052500	
max	8.877689e+09	36019.000000	28.030001	28.030001	4.942142	21.920000	
<hr/>							
	data.describe()						
8:	ActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMinutes	Calories
	940.000000	940.000000	940.000000	940.000000	940.000000	940.000000	940.000000
	3.340819	0.001606	21.164894	13.564894	192.812766	991.210638	2303.609574
	2.040655	0.007346	32.844803	19.987404	109.174700	301.267437	718.166862
	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	1.945000	0.000000	0.000000	0.000000	127.000000	729.750000	1828.500000
	3.365000	0.000000	4.000000	6.000000	199.000000	1057.500000	2134.000000
	4.782500	0.000000	32.000000	19.000000	264.000000	1229.500000	2793.250000
	10.710000	0.110000	210.000000	143.000000	518.000000	1440.000000	4900.000000

Activate Windows

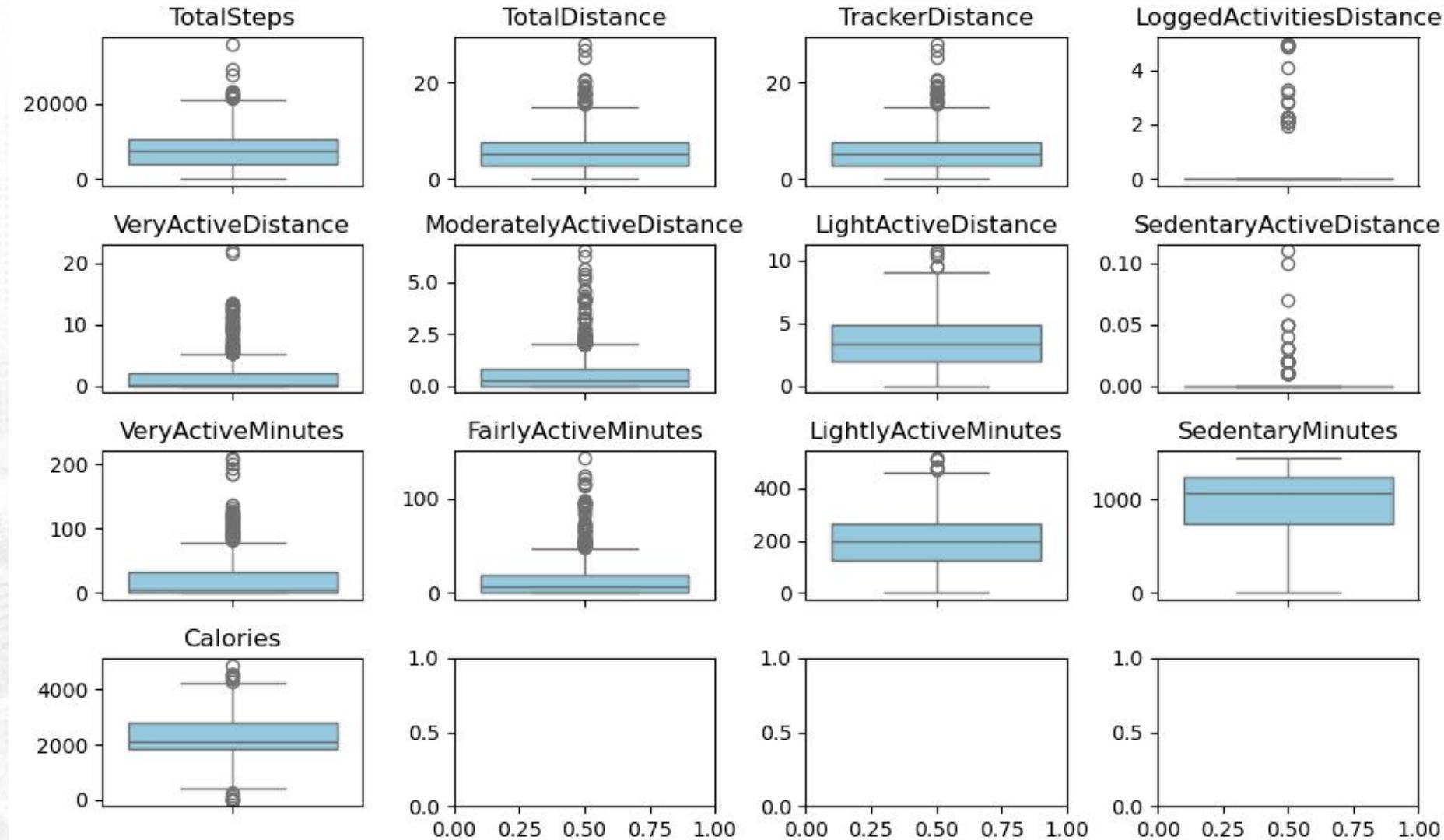
UNIVARIATE ANALYSIS

Pada Analisis Kdeplot, kolom *Calories* terdistribusi cukup normal, beberapa kolom lainnya cenderung *skew positive* dan *bimodal*.

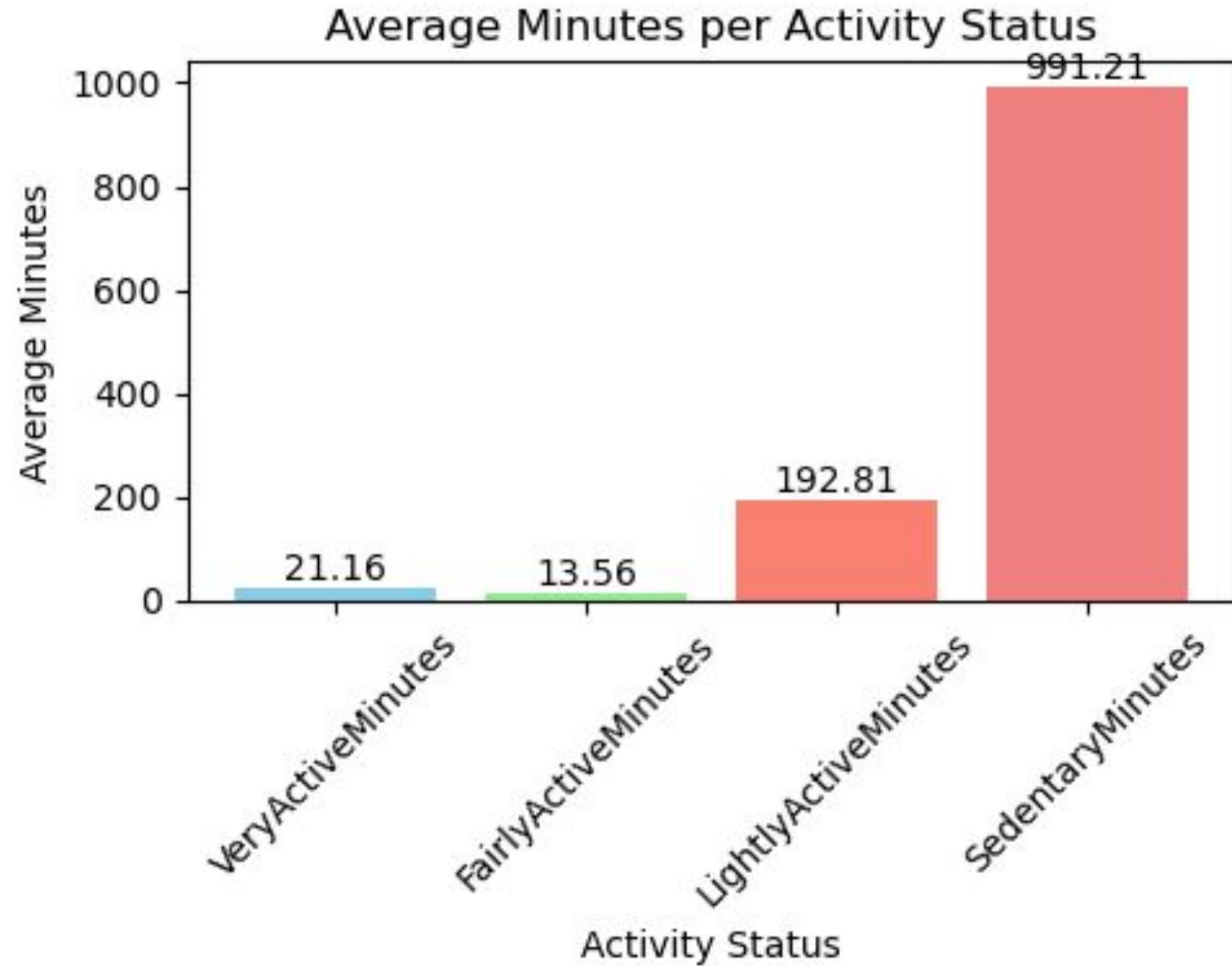


EDA

UNIVARIATE ANALYSIS



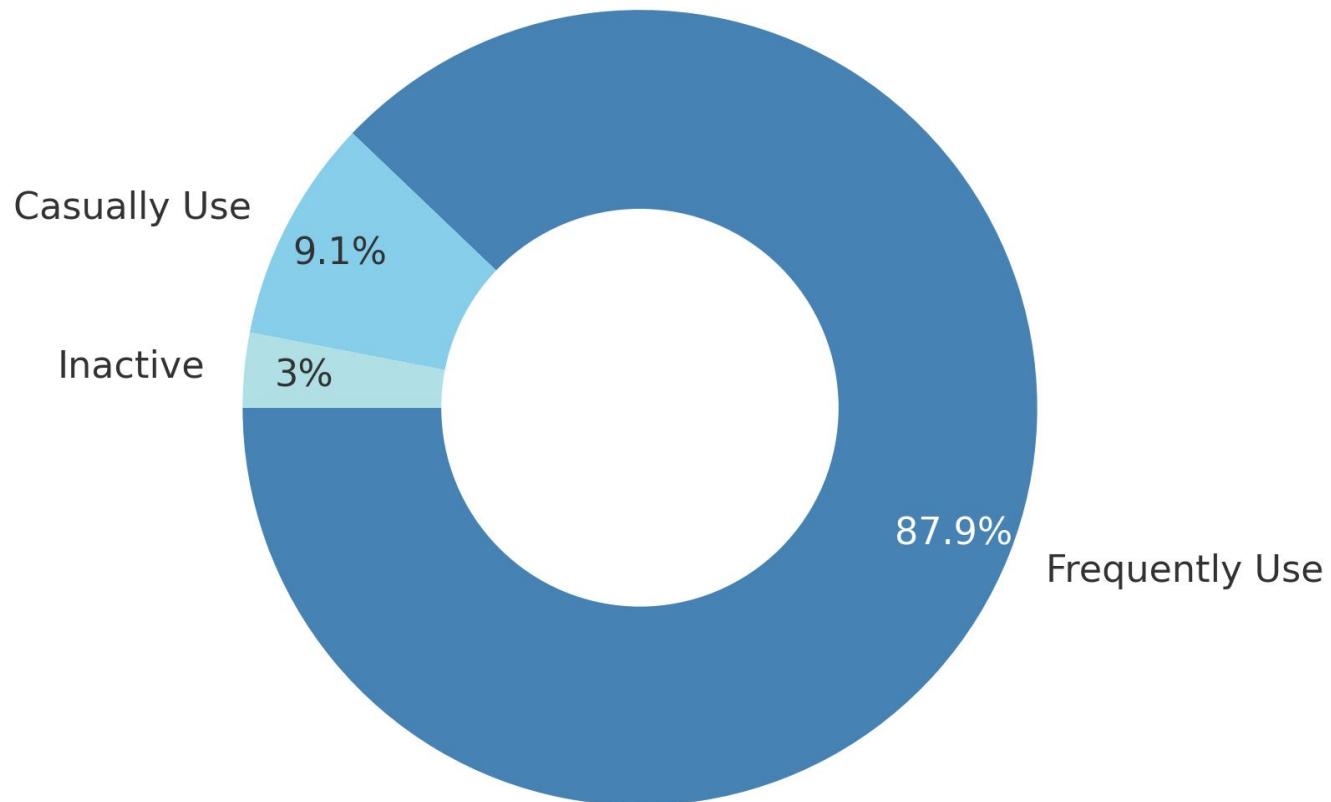
Sedentary - Very Active: Bagaimana User Menghabiskan Hari Mereka?



Fitbit: Dari Gadget Menjadi Gaya Hidup,

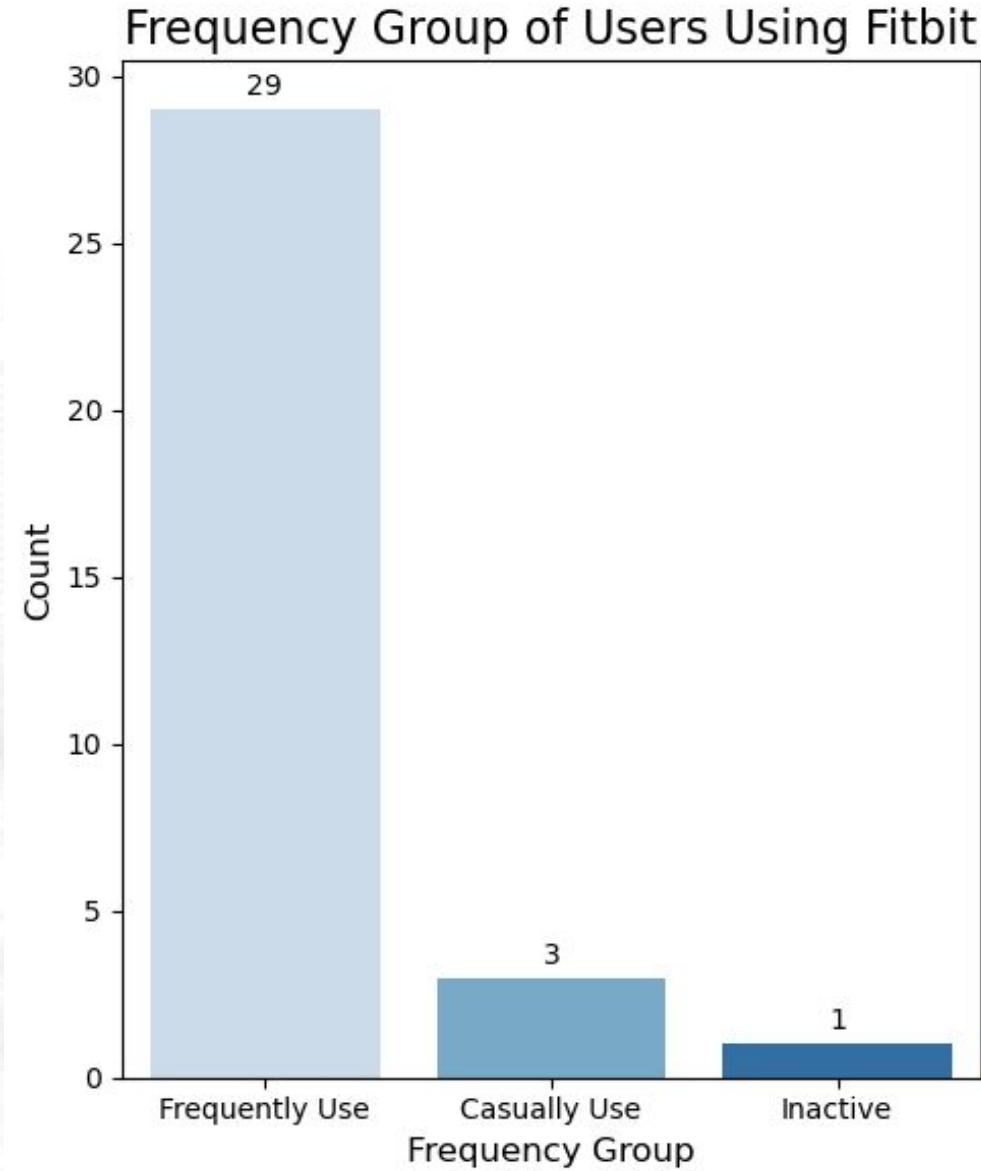
**Sebagian Besar Pengguna
Aktif Memakai Fitbit
Sepanjang Bulan**

Frequency Group of Users Using Fitbit



Fitbit: Dari Gadget Menjadi Gaya Hidup,

Sebagian Besar
Pengguna Aktif
Memakai Fitbit
Sepanjang Bulan



Insight

BUSINESS INSIGHT

Melihat Lebih Dekat: Pola Aktivitas Pengguna Fitbit yang Sering Digunakan

Pengelompokan kategori aktivitas (Tudor, et al.,2008)

Sedentary (Tidak Aktif): 0 - 4999 langkah

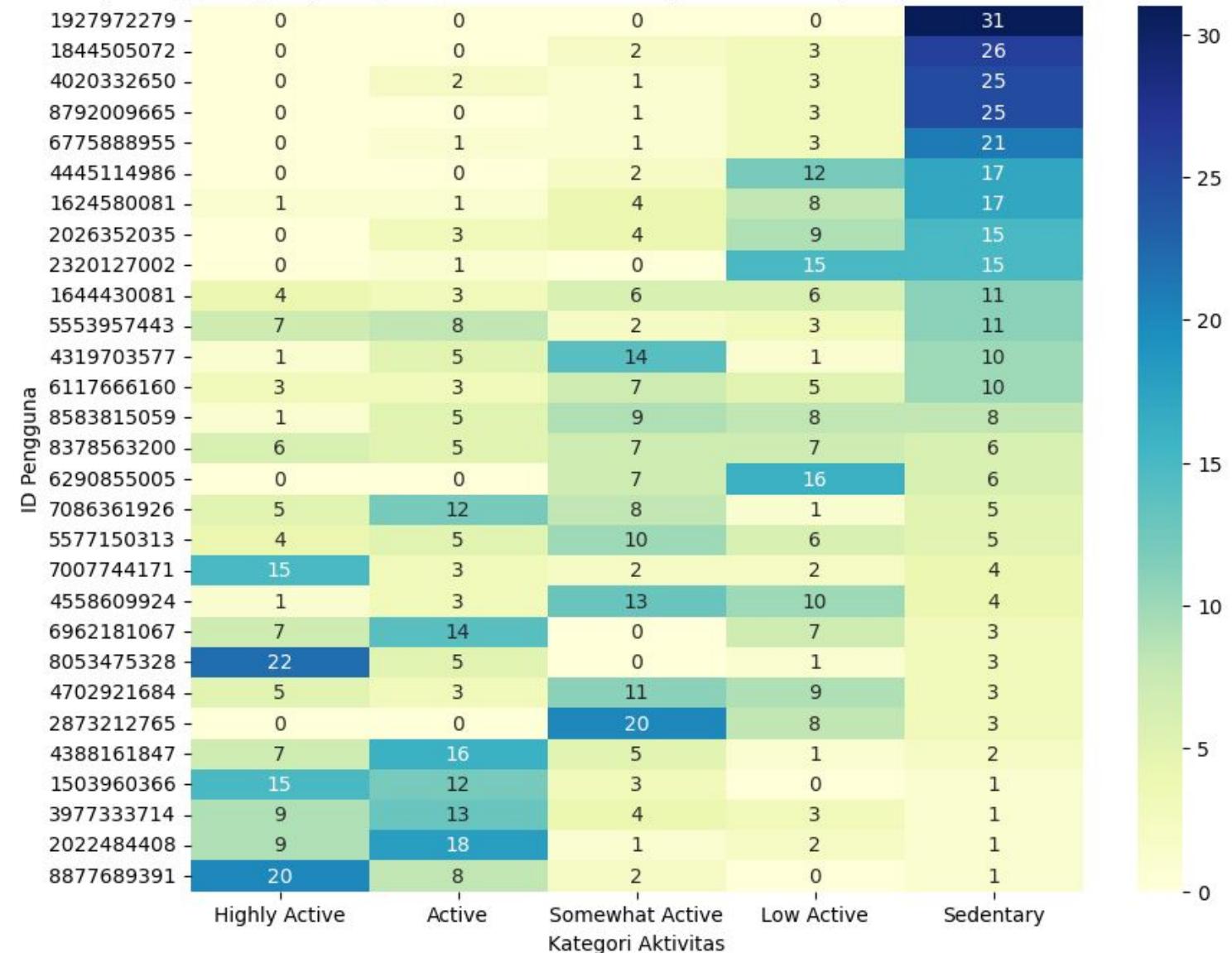
Low Active (Aktif Rendah): 5000 - 7499 langkah

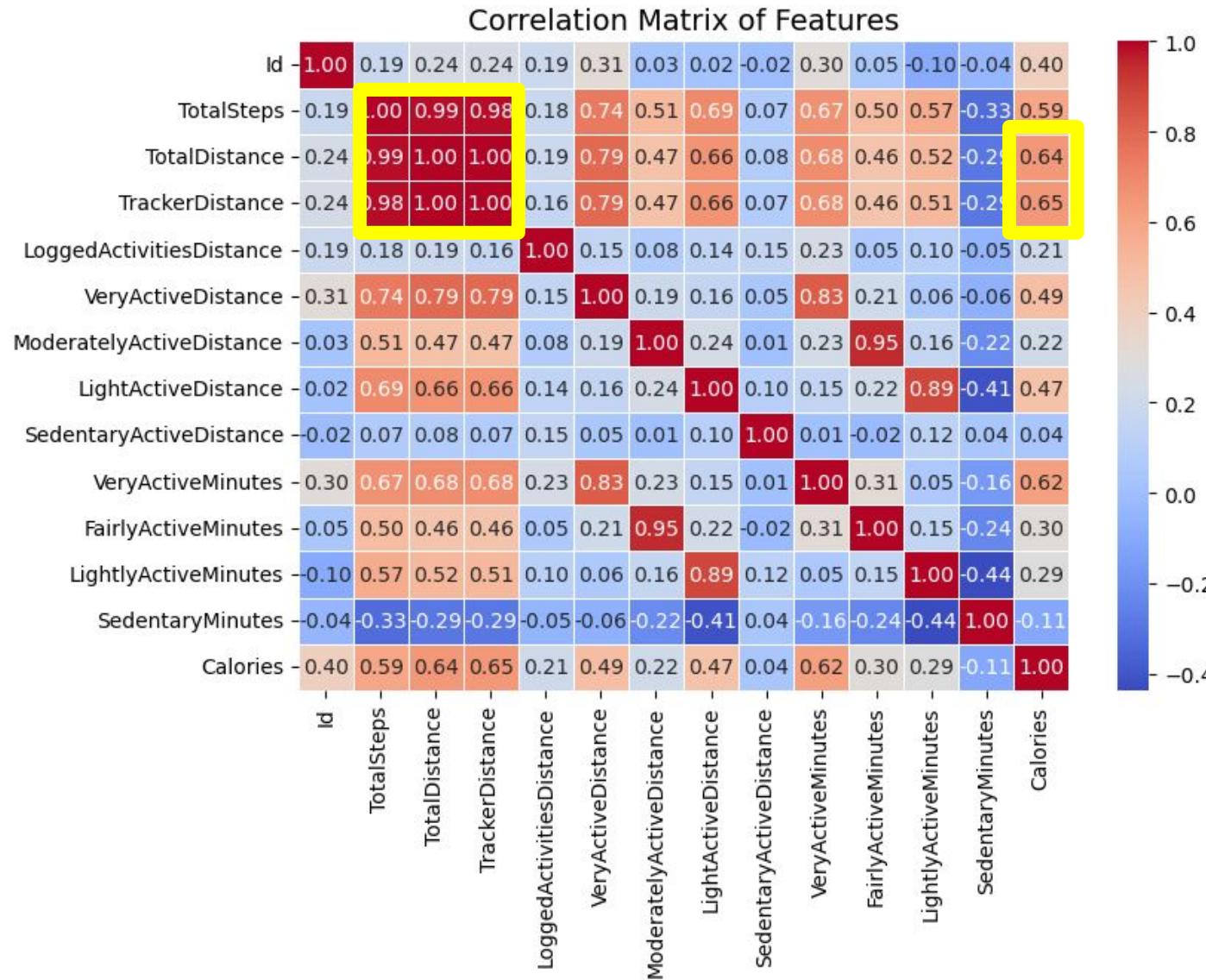
Somewhat Active (Aktif Sedang): 7500 - 9999
langkah

Active (Aktif): 10,000 - 12,499 langkah

Highly Active (Sangat Aktif): 12,500 langkah ke atas

Heatmap Pengguna (Frequently use) berdasarkan Kategori Aktivitas (Diurutkan berdasarkan Sedentary)





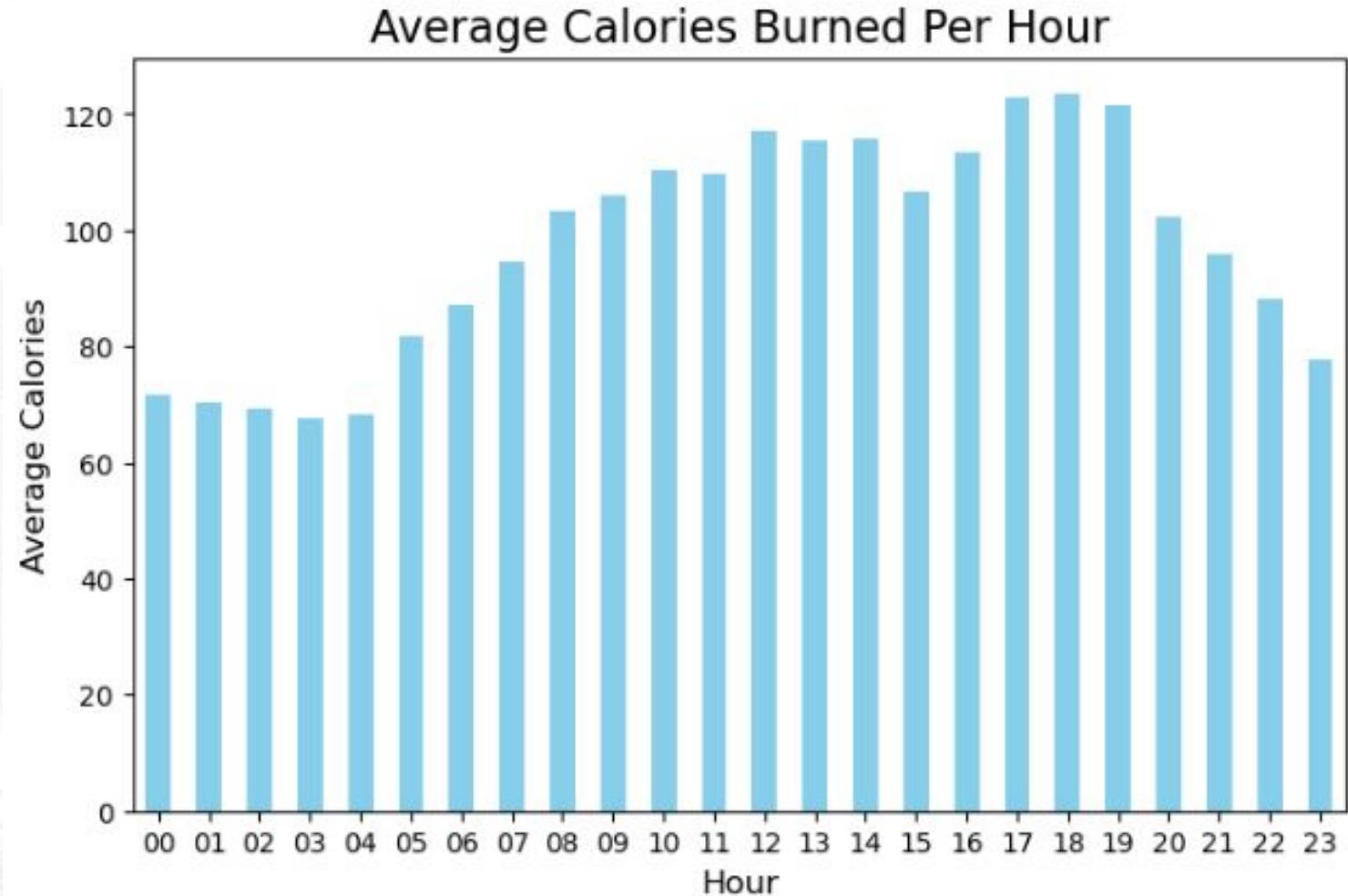
Distance paling berpengaruh terhadap Calories
(Yusof et al. 2018)

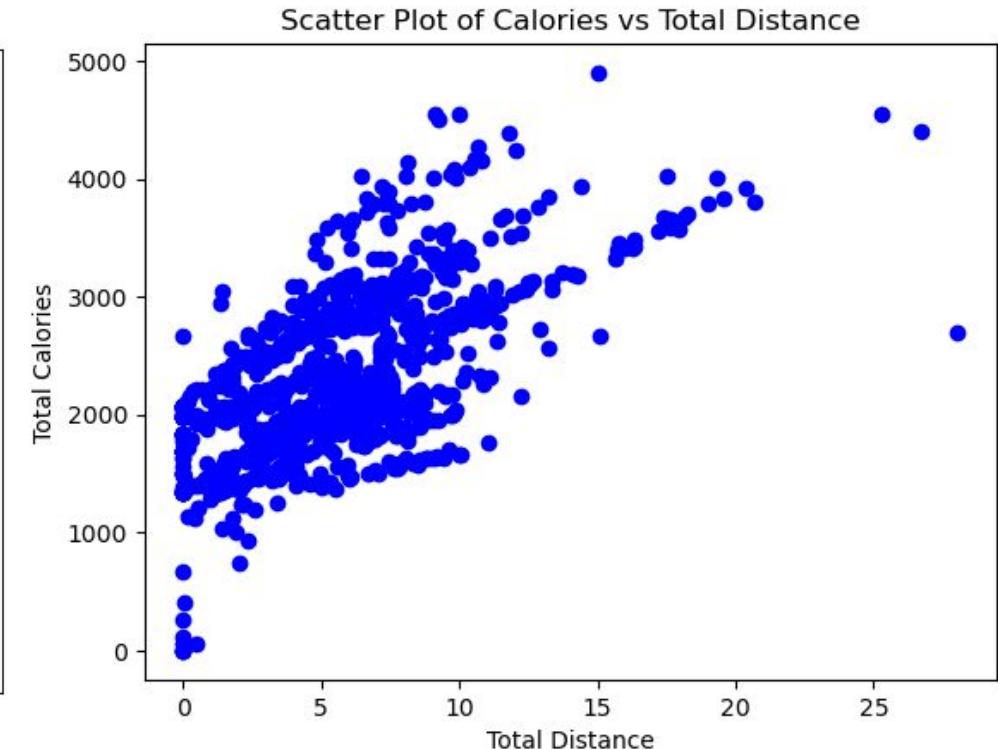
Steps dan Distance berpengaruh signifikan secara positif
(Karami et al. 2021)

Insight

BUSINESS INSIGHT

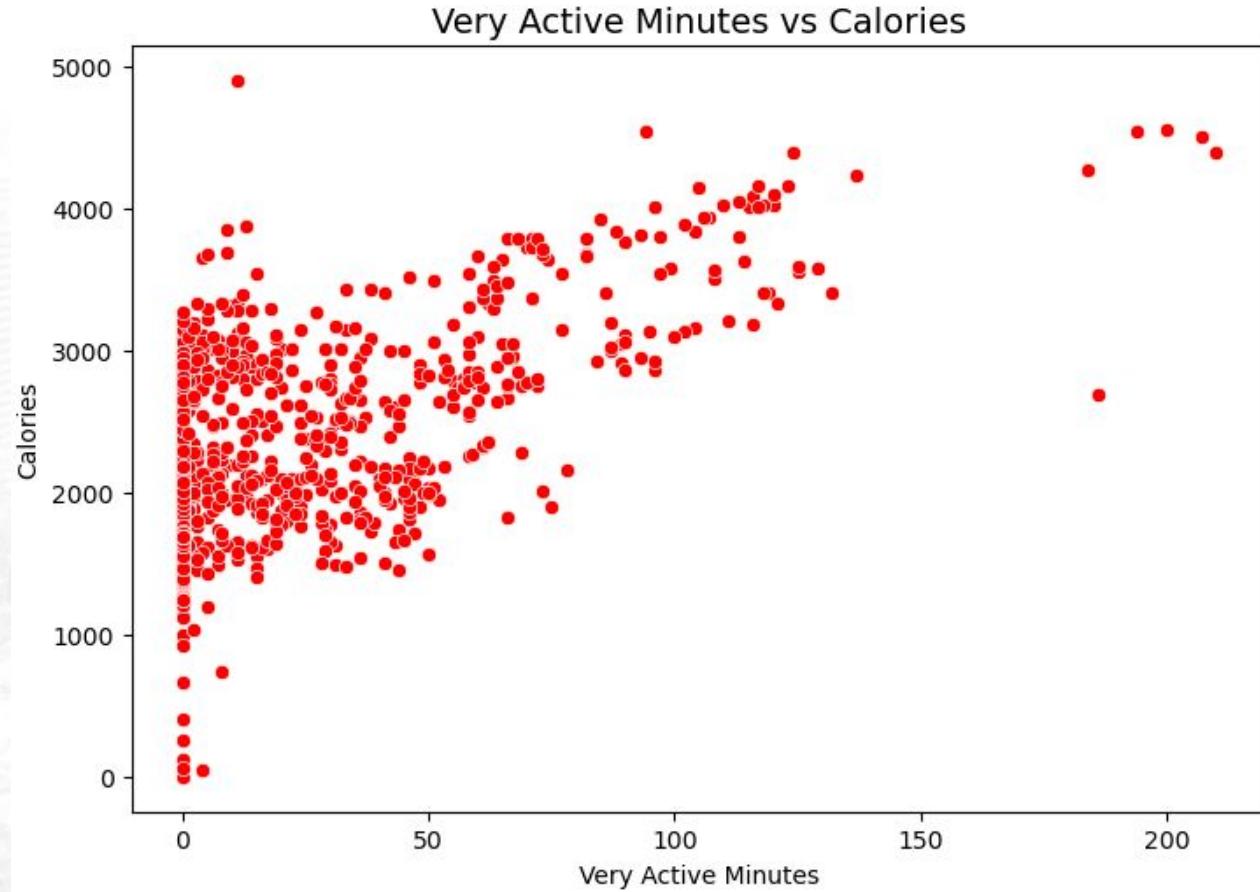
Waktu Kalori Terbakar: Peningkatan Kalori Terbakar dari jam 7 pagi hingga 7 malam.





Semakin jauh jarak dan lama menit aktif, semakin banyak kalori yang terbakar [Yusof et al. 2018, Karami et al. 2021]

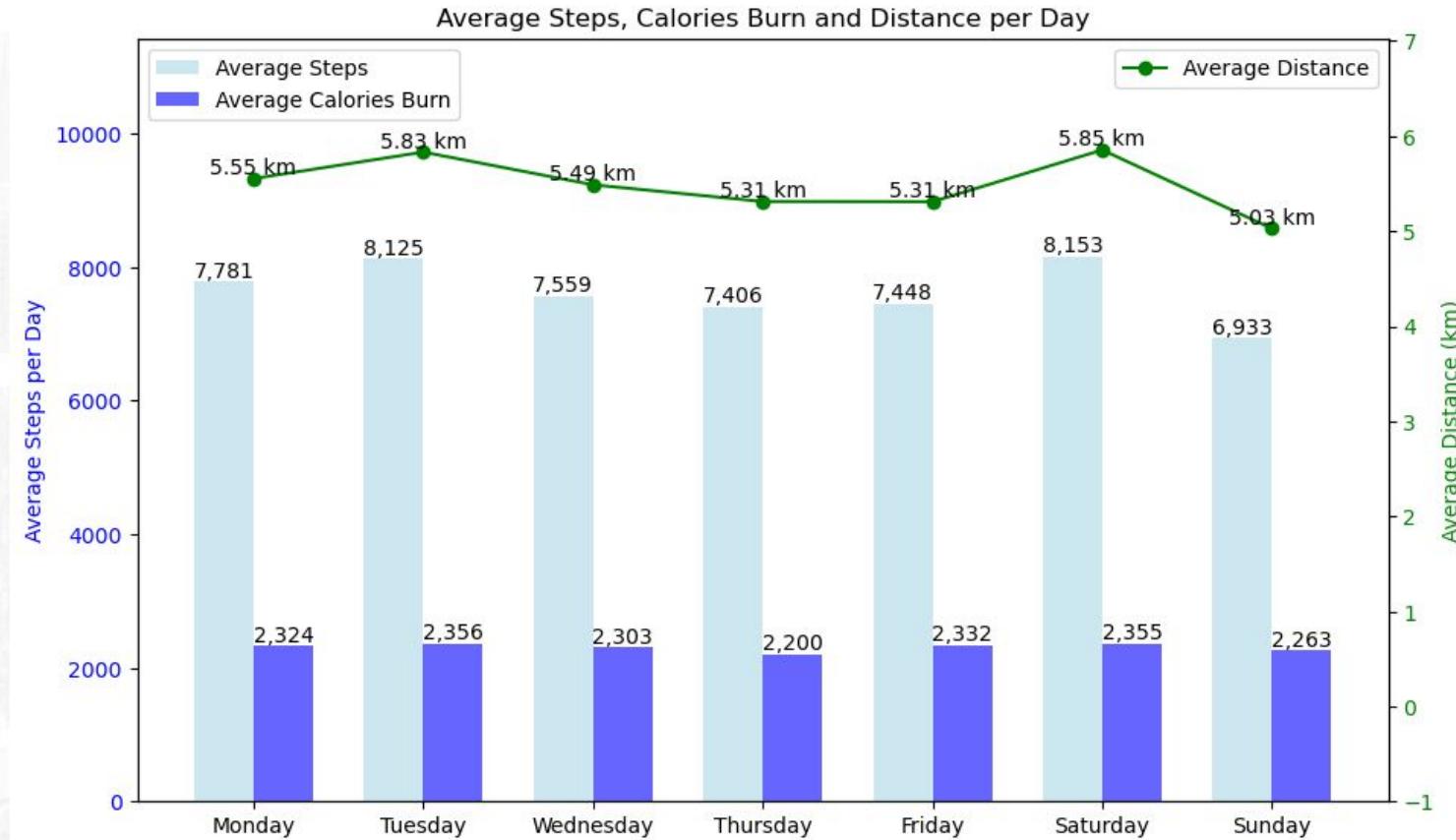
Lebih Aktif, Lebih Banyak Kalori: Hubungan Langsung Antara Aktivitas dan Pembakaran Energi



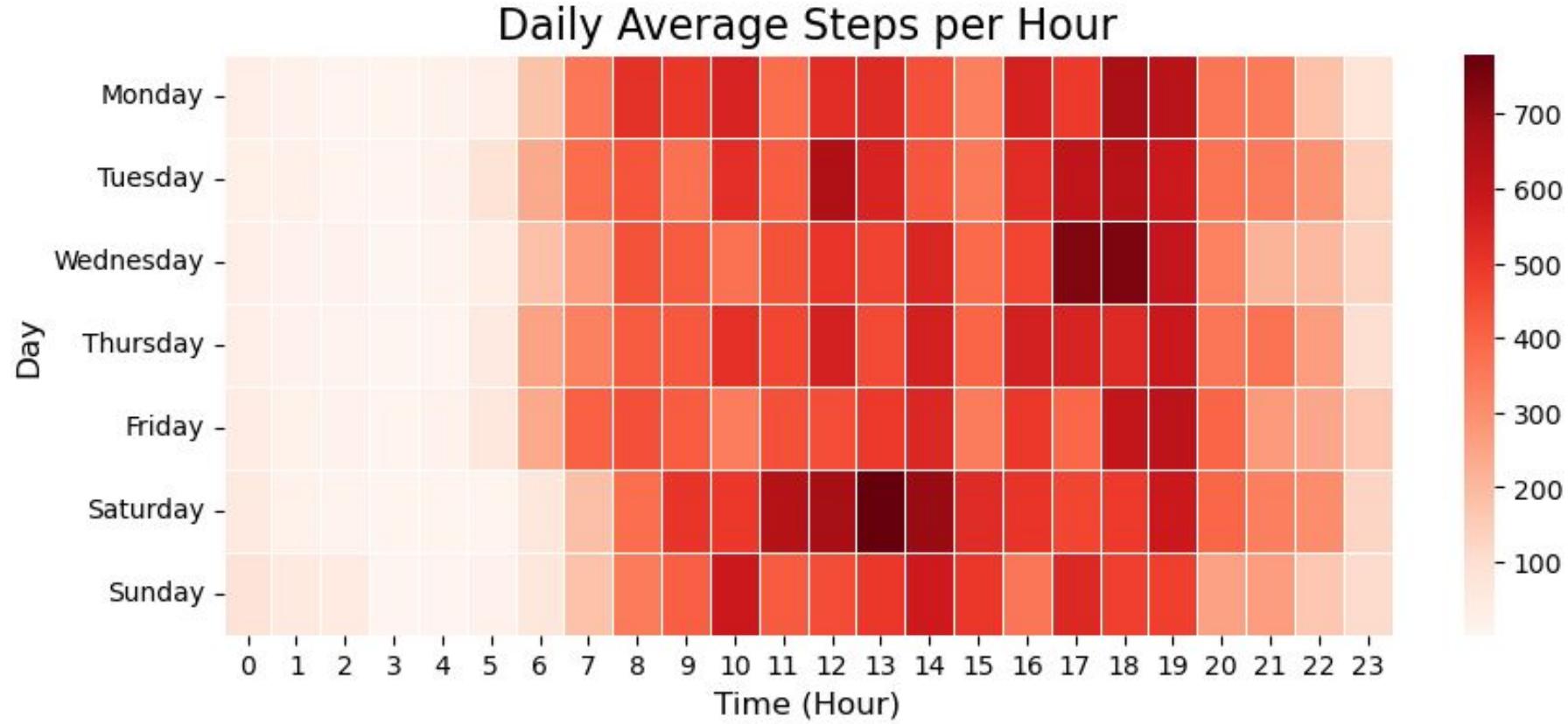
Insight

BUSINESS INSIGHT

**Memahami
Kebiasaan
Aktivitas Harian:
Bagaimana Pola
Aktivitas User
Dalam Setiap
Harinya?**



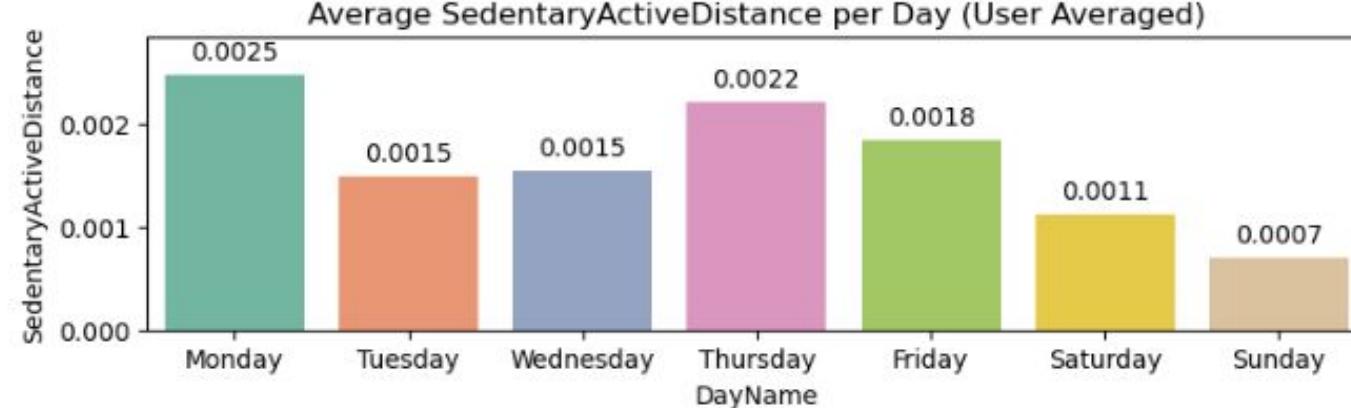
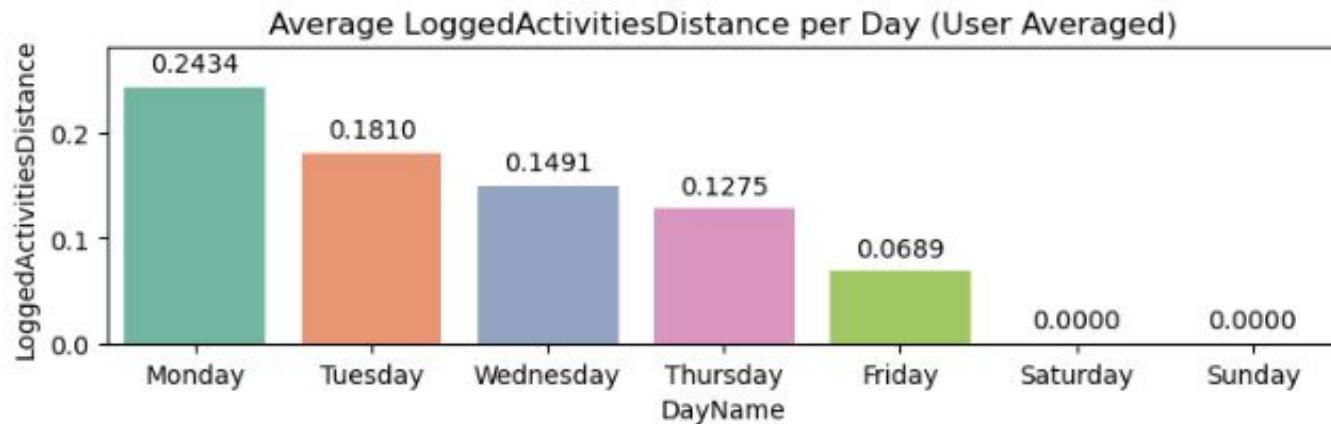
Pola Aktivitas Mingguan Pengguna



Insight

BUSINESS INSIGHT

Rata-Rata LoggedActivityDistance Harian



- **User tidak menginput aktivitas yang mereka lakukan pada Sabtu dan Minggu (weekend).**
- **Pada SedentaryActiveDistance, rata-rata terendah terdapat pada hari Minggu**

Poin-poin penting dari EDA:

- **Aktivitas sedentary mendominasi**, 81% waktu pengguna dihabiskan dalam kondisi minim pergerakan.
- Pola aktivitas harian menunjukkan **puncak aktivitas** terjadi pada hari kerja antara **pukul 07:00 hingga 19:00 dan Sabtu siang**.
- **Jarak tempuh** dan **waktu aktif** memiliki hubungan yang kuat dengan jumlah **kalori** yang terbakar.
- Sebagian besar pengguna masuk dalam kategori **frequent users** yang menunjukkan keterlibatan rutin.
- Sebagian besar waktu pengguna lebih banyak dihabiskan untuk aktivitas dengan **jumlah langkah yang rendah** atau minim pergerakan.
- Karena pada data kami ditemukan **pola tertentu**, kami dapat memanfaatkan **machine learning** untuk mengenali **hubungan kompleks** di antara fitur-fitur tersebut.

Data Preprocessing



Data Cleansing



Handle Missing Value

Tidak terdapat missing value



Handle Duplicate

Drop 1 data duplikat



Handle Outlier

Terdapat outlier pada hampir seluruh kolom, dan tidak diberikan perlakuan apapun



Feature Encoding

Feature encoding dilakukan bersamaan dengan feature extraction



Feature Transformation

Mengubah tipe data yang tidak sesuai

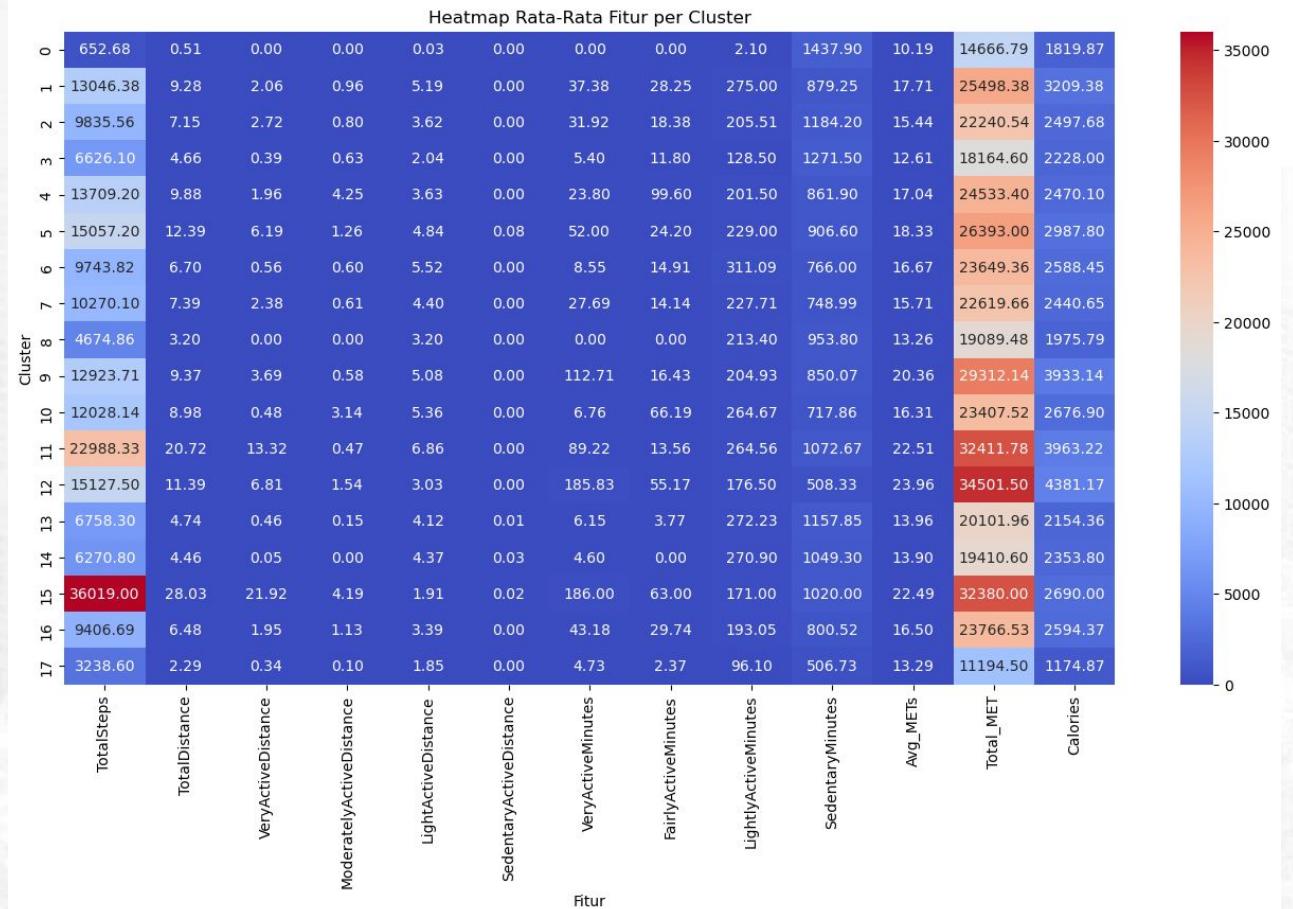
Feature selection

Menggunakan **SelectKBest** dari `sklearn.feature_selection` dengan rumus ***mutual information score*** didapat hasil berikut. Fitur yang dipilih adalah **fitur yang memiliki skor ≥ 0.1**

	Mutual Information Score (MIS) fitur terhadap target (Calories)				
MIS Range	< 0.1	0.1 \geq MIS \leq 0.3	0.3 > MIS < 0.5	≥ 0.5	
Nama Fitur	<ul style="list-style-type: none"> LoggedActivitiesDistance Month DayofWeek SedentaryActiveDistance Year Day is_weekend 	<ul style="list-style-type: none"> VeryActiveRatio VeryActiveDistance FairlyActiveRatio ActiveDistanceRatio FairlyActiveMinutes TotalUsageMinutes ModeratelyActiveDistance 	<ul style="list-style-type: none"> ActivityRank SedentaryRatio Cluster LightlyActiveMinutes VeryActiveMinutes LightActiveDistance LightlyActiveRatio AveragePace ActiveGroup 	<ul style="list-style-type: none"> Total_MET Avg_METs TotalDistance DistancePerStep StepsbyDistance TotalActiveDistance TotalSteps ActiveRatio AverageActiveMinutes DistanceIntensity InactiveRatio SedentaryMinutes 	
Total	7	7	9		12

Teknis Feature extraction dengan GMM Clustering

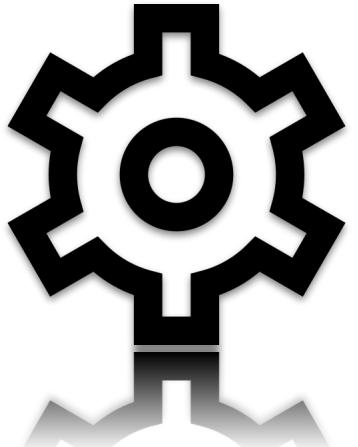
Langkah	Hasil
Terdapat 18 cluster optimal	kolom Cluster
Sorting cluster berdasarkan TotalSteps dan Total_MET secara descending	kolom ActivityRank
ActivityRank memiliki mutual information score lebih baik	Drop kolom Cluster, pertahankan ActivityRank



Feature selection

Menggunakan SelectKBest dari `sklearn.feature_selection` dengan rumus *mutual information score* didapat hasil berikut. Fitur yang dipilih adalah fitur yang memiliki skor di atas 0.1

	Feature	Mutual Information Score		Feature	Mutual Information Score		Feature	Mutual Information Score
0	Total_MET	1.163372	12	InactiveRatio	0.522456	25	FairlyActiveRatio	0.192835
1	Avg_METs	1.025490	13	SedentaryMinutes	0.519378	26	ActiveDistanceRatio	0.191887
2	TotalDistance	0.691249	14	ActivityRank	0.477328	27	FairlyActiveMinutes	0.173981
3	DistancePerStep	0.688869	15	SedentaryRatio	0.461425	28	TotalUsageMinutes	0.171106
4	TrackerDistance	0.686942	16	Cluster	0.446658	29	ModeratelyActiveDistance	0.147146
5	StepsbyDistance	0.675805	17	LightlyActiveMinutes	0.431148	30	LoggedActivitiesDistance	0.068171
6	TotalActiveDistance	0.640279	18	VeryActiveMinutes	0.404431	31	Month	0.047556
7	TotalSteps	0.638304	19	LightActiveDistance	0.401484	32	DayOfWeek	0.019579
8	ActiveRatio	0.562693	20	LightlyActiveRatio	0.382616	33	SedentaryActiveDistance	0.006427
9	AverageActiveMinutes	0.561732	21	AveragePace	0.364498	34	Year	0.002191
10	TotalActiveMinutes	0.560278	22	ActiveGroup	0.352107	35	Day	0.000000
11	DistanceIntensity	0.534578	23	VeryActiveRatio	0.285461	36	is_weekend	0.000000
			24	VeryActiveDistance	0.254702			



Machine Learning Modelling & Evaluation

Berbagai Model yang telah dicoba untuk memprediksi Calories

Nipas *et al.*
2022

Random Forest,
Linear Regression, dan
Ridge Regression

Calories Burnt Prediction

Aziz *et al.*
2023

XGBoost, linear
regression, SVM
dan Random forest

Reddy *et al.*
2023

Linear Regression, Ridge
Regression dan **Random
Forest Regression**

Modelling



Supervised
Learning:
Regression



Target:
Calories



Total:
**13 Algoritma
Model**

STAGE 3 - Modelling Experiments

Pada project ini dilakukan supervised learning jenis regression dengan target **Calories**. Semua experiment pada project ini menggunakan 13 algoritma yaitu:

- Random Forest
- Linear Regression
- Ridge
- Lasso
- Elastic Net
- Decision Tree
- XGBoost
- Gradient Boosting
- Support Vector Regressor (SVR)
- Bayesian Ridge
- CatBoost Regressor
- LightGBM Regressor
- Extra Trees Regressor

Selanjutnya dilakukan Hyperparameter Tuning dengan **GridSearch** pada algoritma yang memiliki hasil evaluasi terbaik.

Skenario Modelling Experiments

1

Experimen
Scaling

2

Experimen
Feature
Selection

3

Hyperparameter
Tuning

4

Evaluasi Model

5

Kesimpulan
Model
Terbaik

STAGE 3 - Modelling Experiments

1. Eksperimen Feature Selection

Metode: eliminasi fitur satu per satu

Teknik seleksi fitur:

- Mutual Information Regression
- Correlation analysis
- Feature Importance
- Particle Swarm Optimization (PSO)

Hasil: Feature Importance memberikan MAE lebih rendah dan lebih balanced antara score train dan test.

2. Eksperimen Scaling

- Membandingkan StandardScaler vs RobustScaler

Hasil: StandardScaler menghasilkan MAE lebih kecil

3. Hyperparameter Tuning

- Melakukan hyperparamter tuning pada 3 model terbaik

Metrics Evaluasi Menggunakan MAE Score

4. Evaluasi Metrics

- MAE Score

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

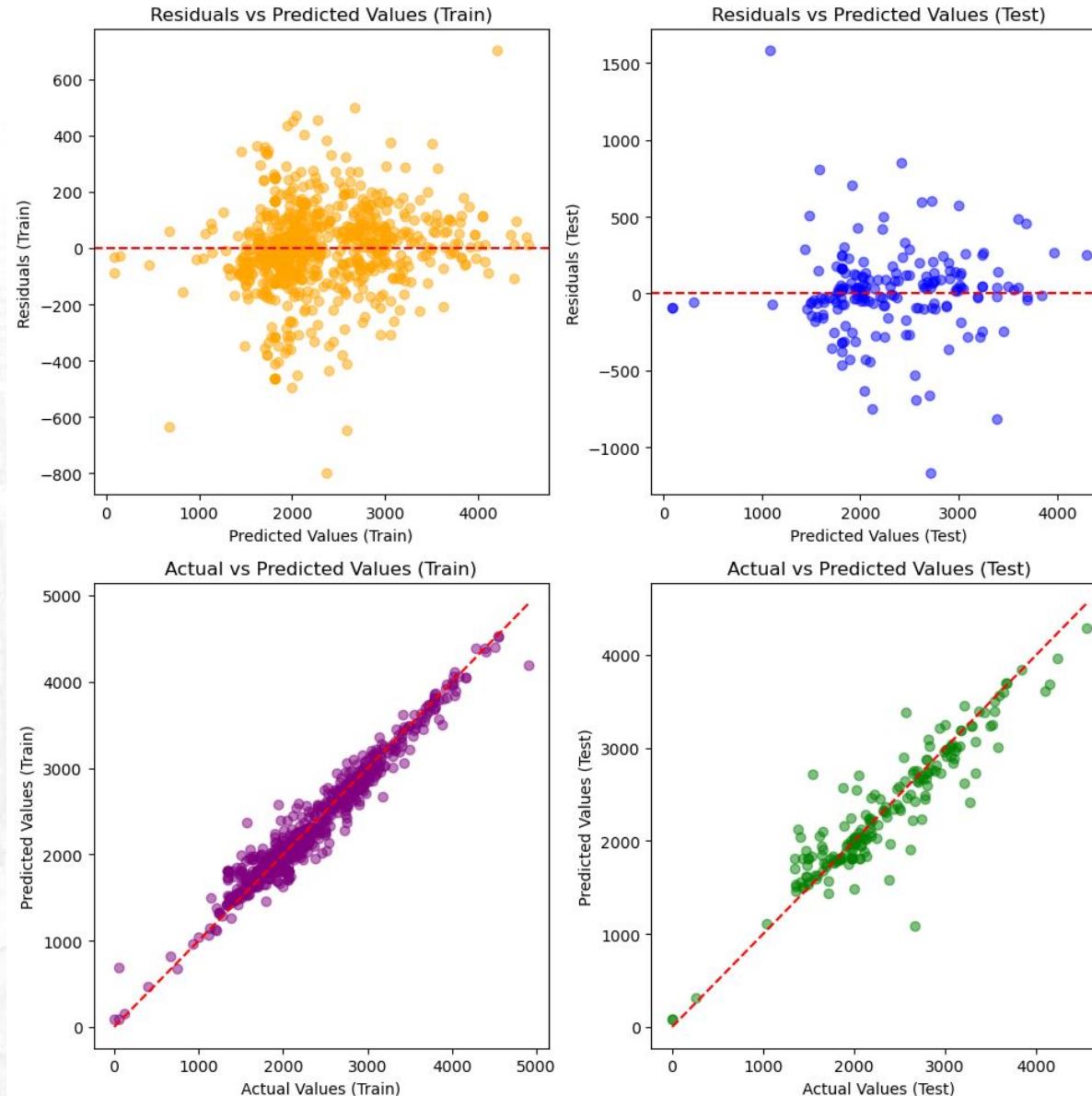
- MAE Score cocok untuk case ini karena data memiliki banyak outlier
- Semakin kecil MAE score, semakin baik hasil prediksi

5. Kesimpulan Model Terbaik



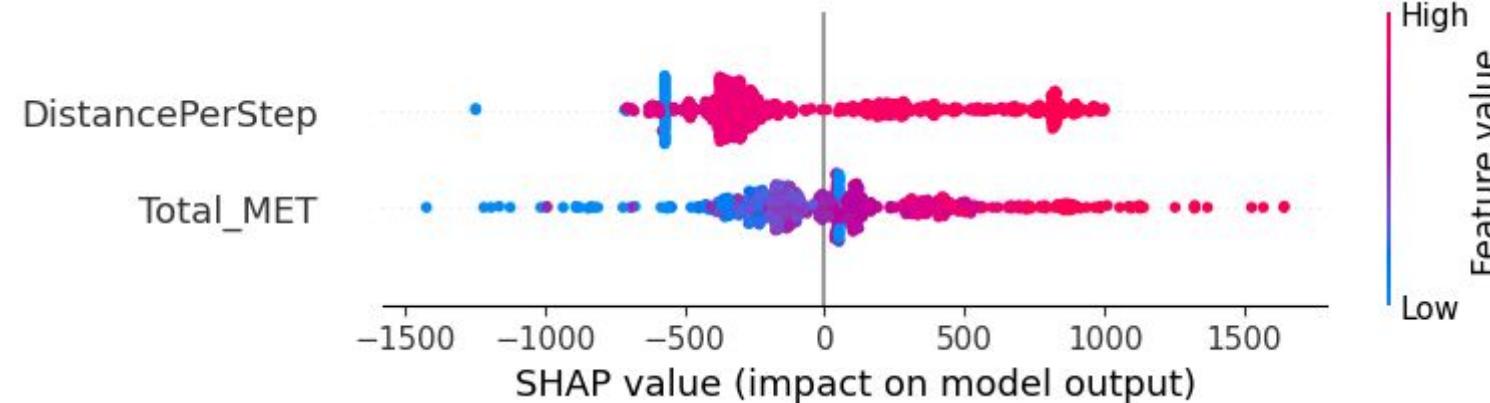
Model	MAE Train	MAE Test	MAE Diff
CatBoost	104.84	184.07	79.23
LightGBM	148.94	212.87	63.93
Gradien Boosting	146.01	204.92	58.91

STAGE 3 - Modelling Experiments



STAGE 3 - Modelling Experiments (1-3 slide)

Feature Importance CatBoost Regressor



Distance Per Step (Jarak Per Langkah):

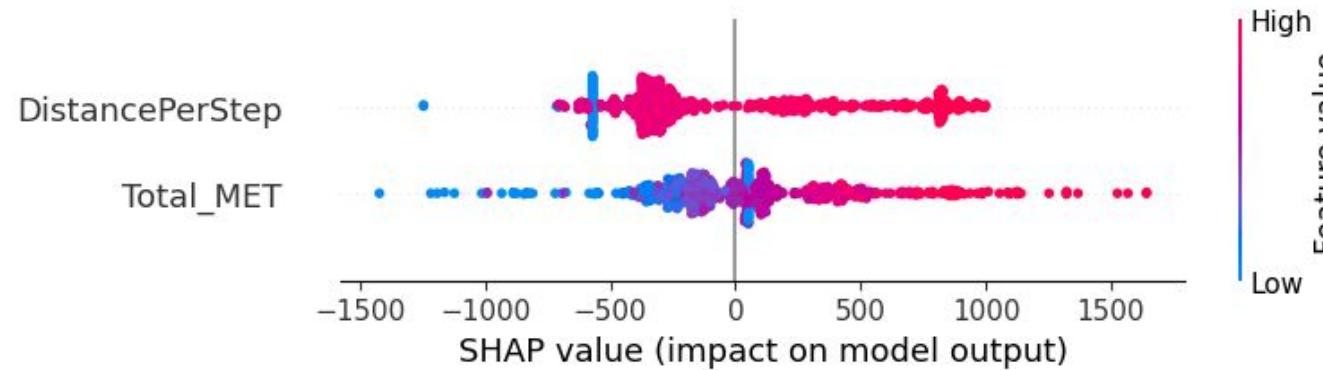
- Grafik ini menunjukkan kalau jarak per langkah punya pengaruh besar terhadap hasil prediksi model. Kalau jarak per langkahnya tinggi (ditunjukkan dengan warna merah di sebelah kanan), pengaruhnya sangat positif terhadap output model. Sebaliknya, kalau jaraknya kecil (warna biru di sebelah kiri), pengaruhnya malah jadi negatif.

Total MET (Metabolic Equivalent of Task):

- Di grafik ini, *Total_MET* punya pengaruh yang lebih tersebar, tapi tidak terlalu besar seperti *DistancePerStep*. Nilai tinggi dan rendah dari Total MET memang berdampak pada output, tapi efeknya tidak terlalu besar — lebih stabil dan mendekati nol.

STAGE 3 - Modelling Experiments (1-3 slide)

Feature Importance CatBoost Regressor

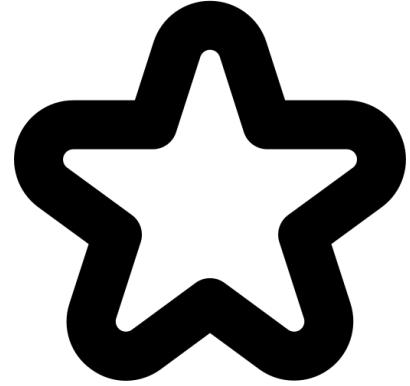


DistancePerStep (Jarak Per Langkah)

Semakin tinggi jarak per langkahnya, pengaruhnya sangat positif terhadap output model, begitupun sebaliknya

total_MET (Metabolic Equivalent of Task)

Nilai Total MET memang berdampak lebih stabil dan mendekati nol pada model



Business Recommendation

Business Recommendation

Fitur Personalisasi



Personalisasi Target Kalori Harian/Mingguan

Untuk meningkatkan fleksibilitas user dalam mencapai target kalori harian/mingguan dan meningkatkan motivasi user dalam pencapaian tujuan. Sesuai dengan teori pengaturan tujuan (goal-setting theory) (Locke dan Latham, 1990, 2022)



Deploy CatBoost dengan 2 Fitur Terbaik (MAE Terendah) untuk prediksi kalori

Dengan model machine learning, kami memprediksi kalori yang terbakar dari aktivitas pengguna, seperti langkah dan intensitas olahraga secara real time. Jika hasil prediksi pembakaran kalori masih kurang dari target harian, aplikasi akan memberikan rekomendasi penyesuaian aktivitas untuk memenuhi target pembakaran kalori secara mingguan.

Business Recommendation

Gamifikasi



Fitur Challenge

'Daily Challenge', seperti '10,000 Steps Challenge', dan 'Leaderboard' untuk memotivasi pengguna, karena mayoritas masih tergolong sedentary.



Fitur Streak

Mendorong pengguna untuk terus aktif dengan memberikan penghargaan atas setiap log harian atau pencapaian target kalori harian.

Notifikasi yang Memotivasi



Fitur Notifikasi/Pengingat

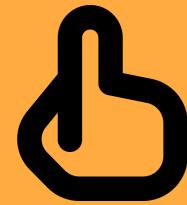
Fitur notifikasi untuk mendorong pengguna lebih sering menggunakan Fitbit dan mengingatkan waktu terbaik untuk berolahraga, seperti pukul 7 pagi hingga 7 malam.

Recommendation



Hubungkan Jarak Tempuh dengan Tujuan Kalori

seperti: “Ayo jalan 2 km lagi untuk mencapai target harian Anda.” Sesuai dengan teori pengaturan tujuan (goal-setting theory) (Locke dan Latham, 1990, 2022)



Fokus pada DistancePerStep dan Total_MET

DistancePerStep (diambil dari Total Distance dan Total Steps) dan Total_MET sebagai fitur utama untuk memberikan saran personal kepada pengguna.



Deploy CatBoost untuk prediksi kalori real-time

sehingga dapat diakses oleh pengguna di aplikasi



Fitur Streak

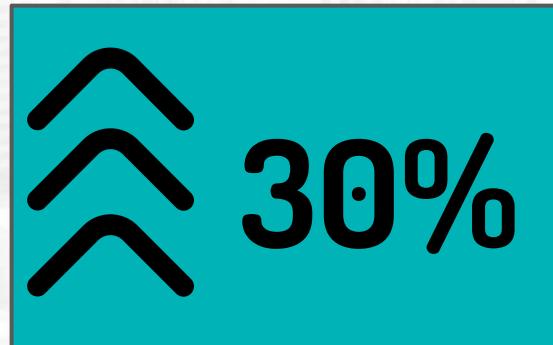
Untuk setiap daily Log ataupun setiap achieve goal calorie harian

Simulasi Sebelum dan Sesudah ML

Kami mengasumsikan ada **1000 user** baru pada bulan X, maka dengan rumus

$$Churn Rate = \frac{\text{Pelanggan yang Berhenti}}{\text{Total Pelanggan Awal}} \times 100\%$$

- **Sebelum ML, tersisa** 38% pengguna aktif, sehingga totalnya = **380 pengguna aktif** pada bulan X+5



- **Personalisasi pada aplikasi kesehatan dapat meningkatkan partisipasi jangka panjang** pengguna hingga 30% (Vorecol, 2024).

- **Setelah dilakukan pengaplikasian business recommendation,**

$(30\% + 38\%) \times \text{Total pelanggan awal}$

= $68\% \times 1000$

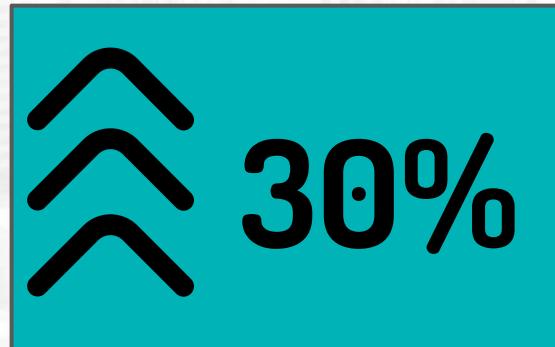
= **680 pengguna aktif** pada bulan X+5

Business Impact Simulation

We assume there are **1000 new users** in month X, then by the formula:

$$\text{Churn Rate} = \frac{\text{Number of Customers Lost During the Period}}{\text{Total Number of Customers at the Start of the Period}} \times 100$$

- **Before Model**, 38% of active users remained, so the total = **380 active users in month X+5**



- Personalization in health apps can **increase long-term user participation by up to 30%** (Vorecol, 2024).

- **After model + apply business recommendation**

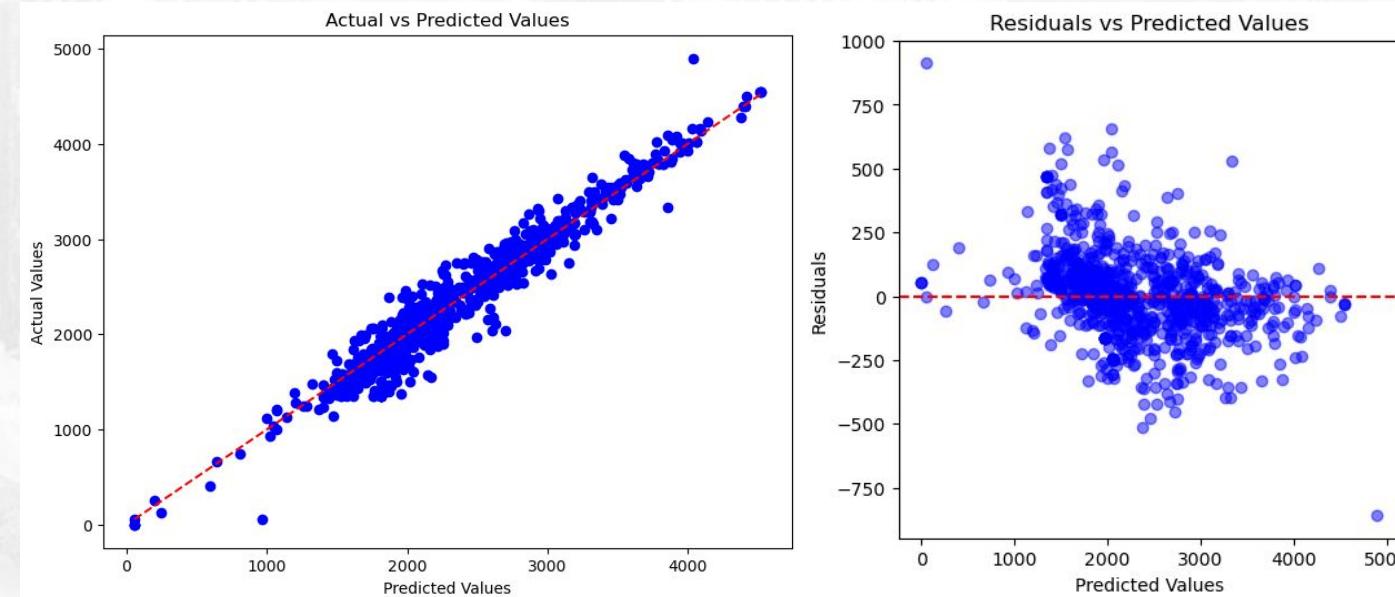
$(30\% + 38\%) \times \text{Total number of user at the start of the period}$

= $68\% \times 1000 = \mathbf{680 \text{ active user in month X+5}}$

Actual vs Predicted Calories

Kami membandingkan Actual Calories dengan Predicted Calories untuk mengetahui **kenaikan calories burn rate**:

Jumlah Data	Jumlah Data yang mengalami kenaikan calories burnt	Percentase kenaikan calories burnt
939	445	47.4%



Thank You

Source

American Health Association (2023)

<https://www.cdc.gov/physical-activity-basics/guidelines/adults.html#:~:text=Recommendations%20for%20adults&text=According%20to%20the%20current%20Physical,an%20and%20vigorous%2Dintensity%20physical%20activity.>

Gartner (2022).

<https://www.bitdefender.com/en-gb/blog/hotforsecurity/a-third-of-wearable-device-owners-quit-using-them-gartner-says>

Locke, et al. (2002)

<https://www-2.rotman.utoronto.ca/facbios/file/09%20-%20Locke%20&%20Latham%202002%20AP.pdf>

Statista (2024).

<https://www.statista.com/statistics/472600/fitbit-active-users/>

Tudor, et al. (2008)

<https://pubmed.ncbi.nlm.nih.gov/18562971/>

Varecol, et al. (2024).

<https://vorecol.com/blogs/blog-gamification-in-health-apps-boosting-engagement-and-wellbeing-among-employees-165832>

Zia, et al (2023)

<https://pubsonline.informs.org/doi/10.1287/mnsc.2022.4581>

Pembagian Tugas

Bintang (**Project Manager**):

- Memantau perkembangan project dan menjaga agar tetap berjalan tepat waktu dan sesuai dengan tujuan awal
- Mengarahkan dan mengatur strategi pembagian tugas kepada setiap anggota
- Membackup tugas anggota-anggota lain yang berhalangan

Asty (**Machine Learning**):

- Melakukan feature engineering
- Melakukan berbagai experiment dan modelling dalam machine learning

Tofan (**EDA**):

- Melakukan feature engineering
- Melakukan EDA dan mengambil insight dari data

David (**Data Preprocessing**):

- Melakukan feature engineering
- Melakukan berbagai experiment dan modelling dalam machine learning

Putri(**EDA**):

- Melakukan feature engineering
- Melakukan EDA dan mengambil insight dari data

Ade (**Preparation**):

- Menyusun latar belakang masalah dan problem statement
- Melakukan berbagai modelling dalam machine learning

CatBoost

- CatBoost adalah salah satu algoritma gradient boosting yang dikembangkan oleh perusahaan Yandex.
- CatBoost dirancang untuk menangani berbagai jenis data, terutama data kategorikal, dengan lebih efisien dibandingkan algoritma boosting lainnya seperti XGBoost atau LightGBM.
- Kelebihan: Bisa melakukan regularisasi untuk overfitting, terbukti less-overfit pada case kami dibandingkan dengan algoritma yang lainnya.

Kriteria Gradient Descent:

- Model Ensemble
Gradient boosting menggabungkan banyak model sederhana (weak learners), biasanya decision trees, menjadi model yang lebih kuat.
- Iteratif
Model dibuat secara berurutan. Setiap model baru mencoba memperbaiki kesalahan prediksi (residual) yang dibuat oleh model sebelumnya.
- Untuk memperbaiki kesalahan prediksi, algoritma meminimalkan suatu fungsi loss (seperti MSE atau log-loss) menggunakan pendekatan gradient descent.