DiaCare *- A Smart Caloric Expenditure Prediction App Using Machine Learning*

-: Team Member's :-

PRANAV CHOUHAN

UTKARSH RAJ

DEEKSHA REDDY

PRASHANT KUMAR DATTATREY

RUCHA BHIDE

Abstract:

The Comprehensive Caloric Expenditure Prediction System leverages machine learning algorithms and a diverse dataset encompassing physiological and contextual variables to develop a robust model for accurately estimating caloric burn during physical activities. Incorporating accelerometer data from wearable devices, heart rate measurements, demographic information, and contextual features like weather conditions, the system provides real-time feedback to users. By employing feature engineering and data preprocessing techniques, it achieves high accuracy in caloric expenditure prediction, catering to individuals with varying activity levels and health profiles. The system's short-term feasibility and long-term viability are emphasized, with monetization potential through subscription models, device sales, data insights, and partnerships. This solution supports individuals in maintaining healthy energy balance, aligning with the evolving landscape of health and fitness technology.

Introduction :

Problem Statement: Diabetes monitoring and support by Caloric Expenditure Prediction App Using Machine Learning

The relationship between calorie burn (caloric expenditure or energy expenditure) and diabetes is complex and multifaceted. Diabetes is a chronic medical condition characterized by high blood sugar levels, and it can be influenced by various factors, including diet, physical activity, and body weight. Calorie burn, which is the number of calories your body uses for various functions and activities, plays a role in managing diabetes in several ways:

Weight Management: For people with type 2 diabetes, achieving and maintaining a healthy weight can improve blood sugar control.

Insulin Sensitivity: Insulin is a hormone that helps regulate blood sugar levels, and insulin resistance is a common feature of type 2 diabetes. Exercise can make your cells more responsive to insulin, which can lead to better blood sugar control.

Blood Sugar Management: Engaging in physical activity, your muscles use glucose (sugar) for energy. This can help lower blood sugar levels in people with diabetes, particularly in those with type 2 diabetes.

It is evident that accurately monitoring and managing caloric expenditure is vital for individuals striving to balance their energy intake and physical activity effectively. While recent advancements in machine learning have shown promise in predicting calorie burn based on various inputs, there remains a need for a comprehensive solution that integrates diverse physiological and contextual variables. This project aims to develop a machine learning-based system capable of accurately estimating caloric expenditure by leveraging a wide range of data, including accelerometer readings from wearable devices, heart rate measurements, demographic information, and contextual features like weather conditions and time of day. The system should address the challenge of accurately predicting calorie burn across different activity levels and health profiles while providing real-time feedback to users. The primary objective is to create a practical and reliable tool that assists individuals in monitoring and managing their energy balance, promoting healthier lifestyles and aiding weight management efforts.

Prototype Selection:

1. Feasibility :

While building a calorie burn prediction app is feasible within a 2-3 year timeframe, the specific timeline will depend on the app's complexity, available resources, and the development team's expertise. Careful planning, a clear roadmap, and efficient development practices can help ensure a successful and timely app launch.

some key considerations that can impact the development timeline:

A. Scope and Features:

B. Data Availability

C. Algorithm development

D. UI Interface and design

E. Regulatory implementation

F. Integration, Testing and Quality assurance

G. Monetization strategy and Data security

2. Viability :

Diabetes is a chronic disease with no permanent solution currently for Type-2 and 3.

A diabetes prediction app will be relevant for more than 20 years atleast as it requires both preventive and management strategy to battle Diabetes. Ensuring the relevance of a diabetes monitoring app for 20 years is a significant challenge due to rapidly evolving technology, healthcare practices, and user expectations. However, with careful planning and adaptability, it is possible to create an app that remains valuable over the long term. Here are some strategies to consider:

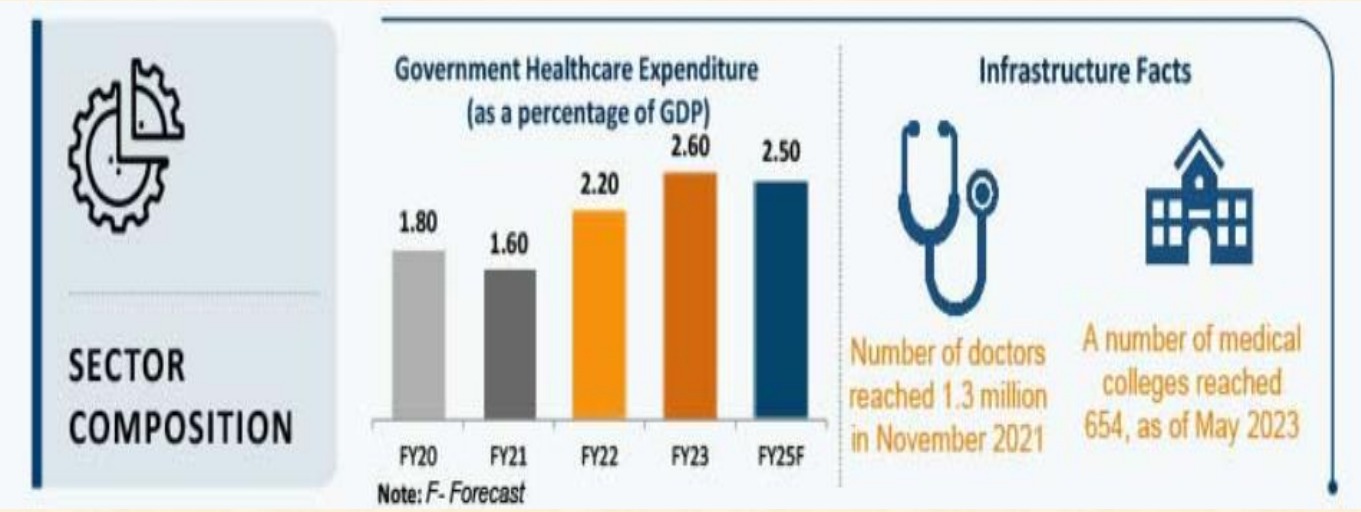
A. Continuous Updates and Improvements

B. Interoperability on wide range of devices available and on growing technologies

C. Telemedicine Integration which allows users to consult with healthcare professionals remotely, receive personalized advice, and track their progress over time.

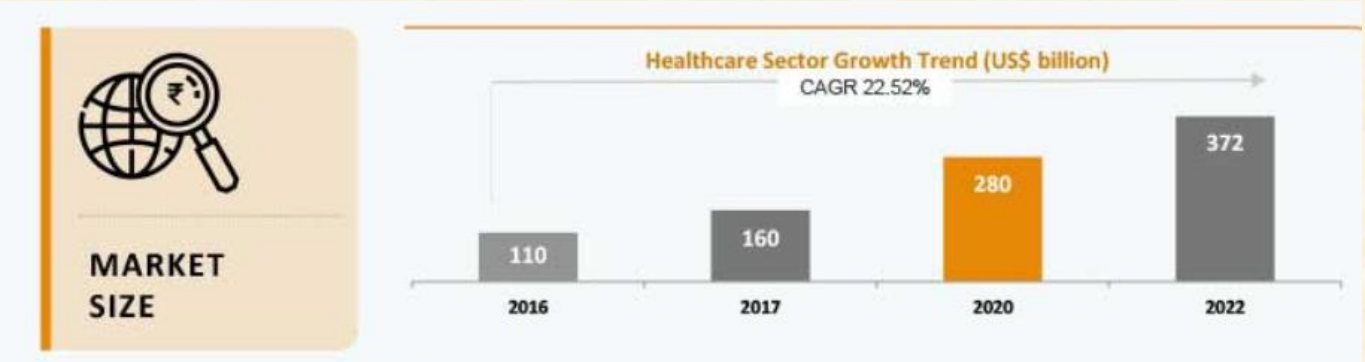
D. Conduct Longitudinal analysis and research on data collected over time to generate insights on diabetes management demonstrate the app's value to users.

E. Adaptability and Feedback loop to pivot and adapt as technology and healthcare landscapes change and improve features and services by listening to user inputs.



3. Monetization :

This App can have multiple mode of monetization. It can be directly monetized as subscription from users and commissions from partners.



Why is this the right time?

- Healthcare market in India is expected to reach US$ 450 billion by 2022, driven by rising income, better health awareness, lifestyle diseases and increasing access to insurance.
- India's public expenditure on healthcare touched 2.1 % of GDP in FY23 and 2.2% in FY22, against 1.6% in FY21, as per the Economic Survey 2022-23 and is forecasted to be 2.5 % in 2024.
- Availability of a large pool of well-trained medical professionals in the country.
- The Indian government is planning to introduce a credit incentive programme worth Rs. 500 billion (US$ 6.8 billion) to boost the country's healthcare infrastructure.
- Healthcare Startups like HealthifyMe, with a total user base of 30 million people, is adding half a million new users every month and crossed US$ 40 million ARR in January 2022.
- As of November 18, 2021, 638 e-Hospitals are established across India as part of the central government's 'Digital India' initiative.

Business Needs:

1. **Accurate Caloric Monitoring:** Individuals seeking to manage their energy balance require a reliable and accurate method for monitoring their caloric expenditure during physical activities. Current solutions often lack precision and fail to consider the wide range of physiological and contextual factors that influence calorie burn.
2. **Comprehensive Data Integration:** Existing fitness tracking devices offer limited insight into caloric expenditure, often relying on basic metrics. There is a need for a system that integrates diverse data sources, including wearable devices, heart rate measurements, demographic information, and contextual features such as weather conditions and time of day, to provide a more holistic view of caloric burn.
3. **Personalized Feedback:** Users benefit from receiving real-time feedback and insights tailored to their unique profiles and activities. Generic recommendations fall short in promoting effective energy balance management.
4. **Long-Term Viability:** The system should not only cater to current trends but also remain relevant and valuable over the next 20-30 years. It must be adaptable to evolving technologies and consumer preferences, ensuring its longevity in the dynamic health and wellness market.
5. **Monetization Opportunities:** To sustain the development and maintenance of the system, it needs direct monetization potential. The business model should encompass strategies such as subscription-based services, proprietary device sales, data insights for health organizations, and partnerships with fitness brands.
6. **Support for Health Professionals:** Health organizations, fitness trainers, and researchers require access to comprehensive and anonymized datasets to analyse trends and develop evidence-based strategies. The system should cater to this need, offering monetization avenues through data insights.
7. **User Engagement and Retention:** To ensure user satisfaction and long-term adoption, the system should provide engaging interfaces, continuous updates, and features that incentivize users to adhere to healthier lifestyles.

Target Specifications and Characterizations:

1. Accuracy and Precision:

- The system should achieve a mean absolute error (MAE) of less than 10% in predicting caloric expenditure across a diverse range of physical activities and user profiles.
- Precision should be maintained in varying environmental conditions, including different weather conditions and altitudes.

2. Data Integration:

- The system must seamlessly integrate multiple data sources, including accelerometer data from wearable devices, heart rate measurements, demographic information, and contextual features such as weather and time of day.
- The data integration process should be efficient and provide real-time updates to users.

3. Personalization:

- The system should offer personalized feedback and recommendations based on individual profiles, activity levels, and goals.
- Recommendations should adapt to changes in the user's routine and preferences, promoting sustained engagement.

4. Long-Term Viability:

- The architecture and design of the system should be adaptable to evolving technologies and consumer preferences over the next 20-30 years.
- Updates and improvements should be feasible without significant disruptions to user experience.

5. Monetization Strategies:

- The system should support various monetization avenues, including subscription-based services, proprietary wearable device sales, data insights for health organizations, and partnerships with fitness brands.

- Revenue generation models should be clear, well-structured, and provide sustainable income streams.

6. User Engagement and Retention:

- The system's user interface should be intuitive, engaging, and user-friendly.
- Gamification elements, rewards, and challenges can be integrated to enhance user engagement and motivation.

7. Security and Privacy:

- User data, especially health-related information, should be securely stored and processed in compliance with data protection regulations.
- Privacy settings and data-sharing options should be transparent and customizable for users.

8. Scalability:

- The system architecture should be designed to handle a large and growing user base without compromising performance.
- Scalability should apply to both data processing and server infrastructure.

9. Integration with Ecosystem:

- The system should integrate with other health and fitness applications, enabling users to synchronize data and access a comprehensive health management ecosystem.

10. Research and Collaboration:

- The system should facilitate data anonymization and aggregation for research purposes, enabling health organizations, researchers, and fitness professionals to derive insights.
- Collaboration features can be developed to foster engagement with professionals in the health and wellness industry.

11. User Support:

- The system should provide responsive and accessible customer support to address user queries, technical issues, and customization needs.

12. Accessibility:

- The user interface should be accessible to individuals with varying abilities, considering factors such as visual impairment and cognitive disabilities.

Benchmarking alternate Products:

Certainly, here are a few benchmarking alternate products that offer features related to health tracking, fitness monitoring, and caloric expenditure estimation:

1. **Fitbit:** Fitbit offers a range of wearable devices that track various health metrics, including steps taken, heart rate, sleep patterns, and estimated caloric expenditure. The devices sync with a mobile app that provides users with insights into their activity levels and health trends.
2. **Apple Watch:** Apple Watch includes features like heart rate monitoring, built-in accelerometer, and various fitness apps. It provides users with real-time data on their physical activity, heart health, and estimated calories burned during workouts.
3. **Garmin Fitness Watches:** Garmin offers fitness-oriented smartwatches with GPS tracking, heart rate monitoring, and specialized sport profiles. These watches provide detailed activity data, including calorie burn estimation, for various types of exercises.
4. **Samsung Health:** Samsung Health is a health and wellness app available on Samsung devices. It tracks steps, heart rate, sleep patterns, and offers a variety of workout modes with calorie burn estimation.
5. **Google Fit:** Google Fit is a fitness tracking app that integrates with various wearable devices and mobile apps. It tracks activity levels, heart points, and estimated calorie expenditure.

Applicable Patents/Regulations/Constraints:

Certainly, here are some applicable patents, regulations, and constraints that could impact the development of the Comprehensive Caloric Expenditure Prediction System:

Patents:

1. **Wearable Device Innovations:** Patents related to wearable devices, sensor integration, and data processing for health and fitness tracking.
2. **Algorithmic Methods:** Patents covering proprietary algorithms and machine learning techniques for accurate caloric expenditure prediction.
3. **Biometric Data Protection:** Patents related to secure storage and transmission of users' biometric data collected by wearable devices.
4. **Contextual Data Integration:** Patents involving the integration of contextual factors like weather conditions and time of day to enhance accuracy.

Regulations:

1. **Personal Data Protection Bill (Draft):** The forthcoming data protection law aims to regulate the collection, storage, and usage of personal data, including health-related information collected by wearable devices.
2. **Medical Device Regulation**: Wearable devices with medical functionalities might need to adhere to medical device regulations and obtain necessary certifications from relevant authorities.
3. **Telemedicine Guidelines**: If the system involves remote health monitoring, compliance with telemedicine guidelines and data privacy regulations is important.

Constraints:

1. **Data Privacy:** Ensuring user data privacy and obtaining informed consent for data collection and usage is crucial.
2. **Accuracy and Reliability:** The system's accuracy in caloric expenditure prediction needs to be well-established to avoid misleading users.
3. **Interoperability:** Compatibility with existing health and fitness platforms and devices to provide a seamless user experience**.**
4. **Regulatory Compliance:** Adhering to relevant regulations, especially if the system provides medical or health-related information.

5. **Ethical Considerations:** Ensuring transparency in data usage, providing meaningful insights, and avoiding potential harm to users.

Prototype Development:

GithubLink:https://github.com/PranavChouhan10/Feynn_Labs/blob/main/Calories_Burnt_Prediction.ipynb

- **Dependencies :**

Importing the Dependencies

```
In [ ]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        from sklearn.model_selection import train_test_split
        from xgboost import XGBRegressor
        from sklearn import metrics
```

- **Data Preparation**

Combining the two Dataframes

```
In [ ]: calories_data = pd.concat([exercise_data, calories['Calories']], axis=1)
```

```
In [ ]: calories_data.head()
```

Out[7]:

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
0	14733363	male	68	190.0	94.0	29.0	105.0	40.8	231.0
1	14861698	female	20	166.0	60.0	14.0	94.0	40.3	66.0
2	11179863	male	69	179.0	79.0	5.0	88.0	38.7	26.0
3	16180408	female	34	179.0	71.0	13.0	100.0	40.5	71.0
4	17771927	female	27	154.0	58.0	10.0	81.0	39.8	35.0

```
In [ ]: # checking the number of rows and columns
        calories_data.shape
```

```
Out[8]: (15000, 9)
```

- **Describing data**

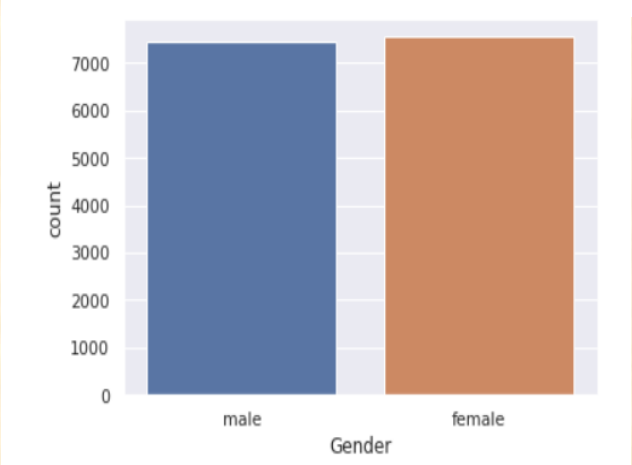
```
In [ ]: # get some statistical measures about the data
        calories_data.describe()
```

Out[11]:

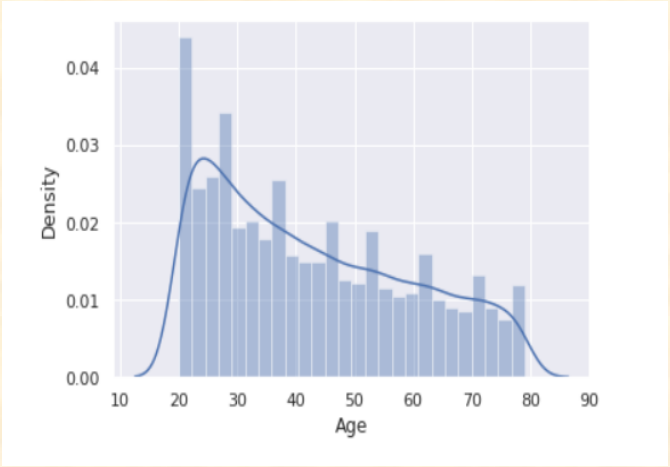
	User_ID	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
count	1.500000e+04	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000
mean	1.497736e+07	42.789800	174.465133	74.966867	15.530600	95.518533	40.025453	89.539533
std	2.872851e+06	16.980264	14.258114	15.035657	8.319203	9.583328	0.779230	62.456978
min	1.000116e+07	20.000000	123.000000	36.000000	1.000000	67.000000	37.100000	1.000000
25%	1.247419e+07	28.000000	164.000000	63.000000	8.000000	88.000000	39.600000	35.000000
50%	1.499728e+07	39.000000	175.000000	74.000000	16.000000	96.000000	40.200000	79.000000
75%	1.744928e+07	56.000000	185.000000	87.000000	23.000000	103.000000	40.600000	138.000000
max	1.999965e+07	79.000000	222.000000	132.000000	30.000000	128.000000	41.500000	314.000000

➢ Visualizing data

```
In [ ]: # plotting the gender column in count plot
        sns.countplot(calories_data['Gender'])
```




```
In [ ]: # finding the distribution of "Age" column
        sns.distplot(calories_data['Age'])
```



➢ Correlation in data

```
In [ ]: correlation = calories_data.corr()
```

```
In [ ]: # constructing a heatmap to understand the correlation

        plt.figure(figsize=(10,10))
        sns.heatmap(correlation, cbar=True, square=True, fmt='.1f', annot=True, annot_kws={'size':8}, cmap='Blu
```

➢ Splitting in Training and test sets

Splitting the data into training data and Test data

```
In [ ]: X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
```

```
In [ ]: print(X.shape, X_train.shape, X_test.shape)
```

```
(15000, 7) (12000, 7) (3000, 7)
```

➢ Model training

XGBoost Regressor

```
In [ ]: # loading the model
        model = XGBRegressor()
```

```
In [ ]: # training the model with X_train
        model.fit(X_train, Y_train)
```

```
[10:06:32] WARNING: /workspace/src/objective/regression_obj.cu:152: reg:linear is now deprecated in f
avor of reg:squarederror.
```

```
Out[27]: XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
                      colsample_bynode=1, colsample_bytree=1, gamma=0,
                      importance_type='gain', learning_rate=0.1, max_delta_step=0,
                      max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
                      n_jobs=1, nthread=None, objective='reg:linear', random_state=0,
                      reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None,
                      silent=None, subsample=1, verbosity=1)
```

➢ Model Evaluation and error calculation

Evaluation

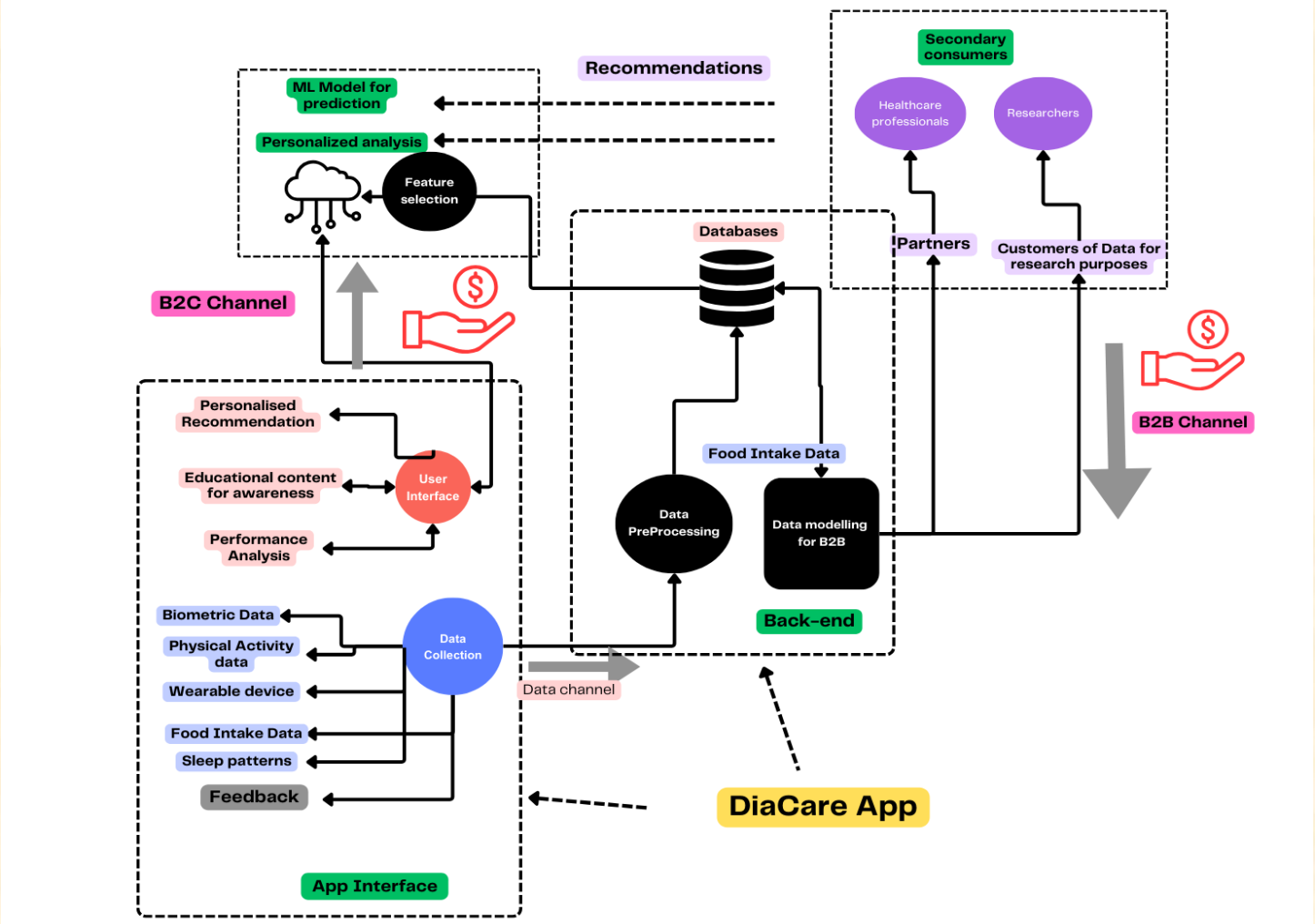
Prediction on Test Data

```
In [ ]: test_data_prediction = model.predict(X_test)

In [ ]: print(test_data_prediction)
        [129.06204   223.79721    39.181965 ... 145.59767    22.53474    92.29064 ]
```

Mean Absolute Error

```
In [ ]: mae = metrics.mean_absolute_error(Y_test, test_data_prediction)

In [ ]: print("Mean Absolute Error = ", mae)
        Mean Absolute Error =  2.7159012502233186
```

Final Prototype:



Business model:

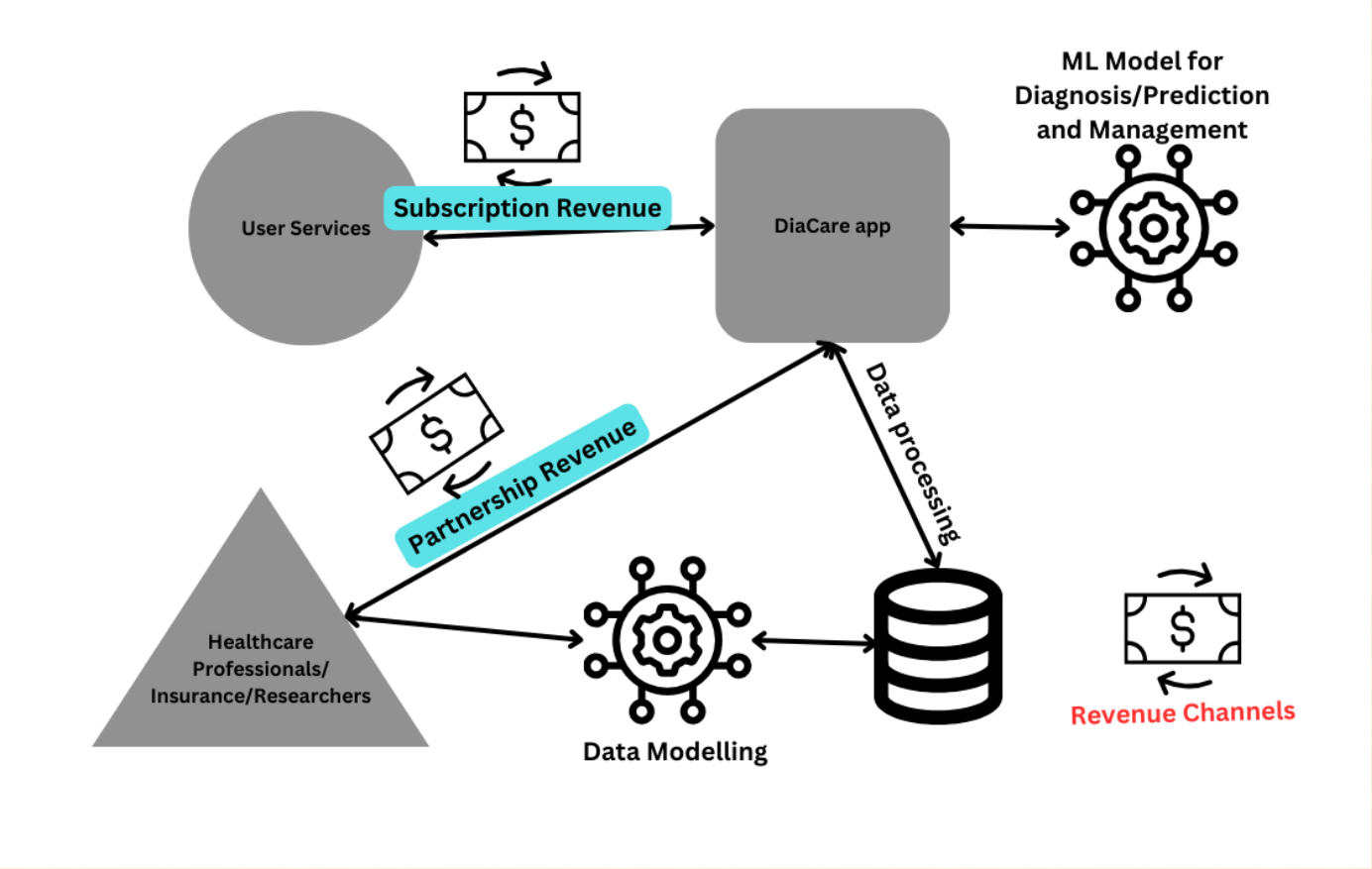
Our business model combines the Freemium and Partnership approaches to create a sustainable revenue stream while offering valuable services to users. The system's purpose is to predict caloric expenditure during physical activities accurately, promoting healthy energy balance management. We address both preventive and management strategies for diabetes prevention and management.

Freemium Model (Preventive Strategy):

1. **Free Version:** The basic version of the app offers core features like tracking, general educational content, and general diabetes risk assessments at no cost.
2. **Premium Version:** We offer a subscription-based premium version at INR 299/month or INR 2499/year, providing users with advanced features.
3. **Features:** The premium version includes personalized risk assessments, predictive analytics, integration with wearable devices, and free consultations with healthcare professionals.
4. **Conversion Rate:** We anticipate a conversion rate of 15% from free users to premium users.

Partnership Model (Management Strategy):

1. **Diabetes Management:** If a user is diagnosed with diabetes, we collaborate with healthcare providers, insurance companies, pharmaceutical firms, and dieticians.
2. **Recommendations:** Based on our predictive model, we recommend diagnosed users to these partners for personalized health plans.
3. **Commission:** In return, we receive a commission from these partners for customer referrals, ensuring monetization while helping users manage their condition.



Financial Model:

1. **User Acquisition and Growth:**
    - Initial Users (U(0)): 1,000 users
    - Growth Rate: 10% monthly
2. **Freemium and Premium Users:**
    - Conversion Rate: 15%
    - Premium Users (P(t)): Converted from free users
    - Freemium Users (F(t)): Total users - Premium users
3. **Subscription Pricing:**
    - Premium Version: INR 299/month or INR 2999/year
4. **Partnership Revenue:**
    - Partnership Revenue (Partner rev): Variable based on user opt-ins

5. **Costs:**
    - Development and Maintenance Costs (DMC)
    - Marketing and User Acquisition Costs (MAC)
    - Server and Hosting Costs (SHC)
    - Customer Support Costs (CSC)
6. **Customer Lifetime Value (CLV):**
    - CLV = (ARPU / CR)
    - ARPU (Average Revenue Per User): INR 2999 (monthly subscription)
    - CR (Churn Rate): 10%
7. **Monthly Revenue:**
    - Total Monthly Revenue (Total_rev) = Premium Subscription Revenue (Sub_rev) + Partnership Revenue (Partnr_rev)
8. **Monthly Costs:**
    - Total Monthly Costs (Total_cost) = DMC + MAC + SHC + CSC + (Costs of Partnership Agreements)
9. **Monthly Profit:**
    - Gross Profit (GP) = Total Revenue - Total Costs
    - Net Profit (NP) = Gross Profit - Taxes
10. **Break-Even Point:**
    - Determine when Total Monthly Revenue equals Total Monthly Costs.

Sensitivity Analysis: Conduct sensitivity analysis by adjusting variables like conversion rates, growth rates, and CLV to assess their impact on financial performance.

Funding and Investment: If initial expenses exceed available funds, consider seeking external funding through investors, loans, or grants.

Financial Equations :

1. **Total Users (U):**

$$U(t) = U(0) + (\text{Growth Rate}) * t$$

Where:
U(t) = Total users at time 't' (in months)

U(0) = Initial user base
Growth Rate = Assumed monthly user growth rate

2. **Freemium Users (F):**

F(t) = U(t) * (1 - Conversion Rate)

Where:
F(t) = Freemium users at time 't' (in months)

3. **Premium Users (P):**

P(t) = U(t) * Conversion Rate

Where:
P(t) = Premium users at time 't'

4. **Monthly Subscription Revenue (Sub_rev):**

Sub_rev(t) = P(t) * Subscription Price

Where:
P(t) = Premium users at time 't'
Sub_rev(t) = Monthly subscription revenue at time 't'

5. **Partnership Revenue (Partnr_rev):**

Partnr_rev(t) = Partnership Revenue (constant)

Where:

Partnr_rev(t) = Monthly partnership revenue at time 't' (constant as it depends on users taking opting partnership service and we getting commission so it can not be modeled)

6. **Total Monthly Revenue (Total_rev)**:

Total_rev(t) = Sub_rev(t) + Partnr_rev(t)

Where:
Total_rev(t) = Total monthly revenue at time 't'

7. **Total Monthly Costs (Total_cost) :**

Total_cost(t) = DMC + MAC + SHC + CSC + (Costs of Partnership Agreements)

Where:
Total_cost(t) = Total monthly cost at time 't'
DMC = Development and Maintenance Costs
MAC = Marketing and User Acquisition Costs
SHC = Server and Hosting Costs
CSC = Customer Support Costs

8. **Monthly Profits** :

Gross Profit (GP): The difference between total revenue and total costs.

GP = TR - TC

Net Profit (NP): Gross profit minus taxes and other expenses.

NP = GP - Taxes

9. **Break-Even Point:**

Determine when your monthly revenue equals your monthly costs, taking into account all revenue sources.

BEP = Total Costs / Total Revenue

7. **Customer Lifetime Value (CLV):**

CLV = (ARPU / CR)

Where
ARPU is Average revenue per user
CR is monthly churn rate

Revenue Model:

Revenue(t) = X * (1 - e^(-r * t)) * CLV

This formula takes into account the compounding effect of subscription growth and the value each customer brings over their lifetime.

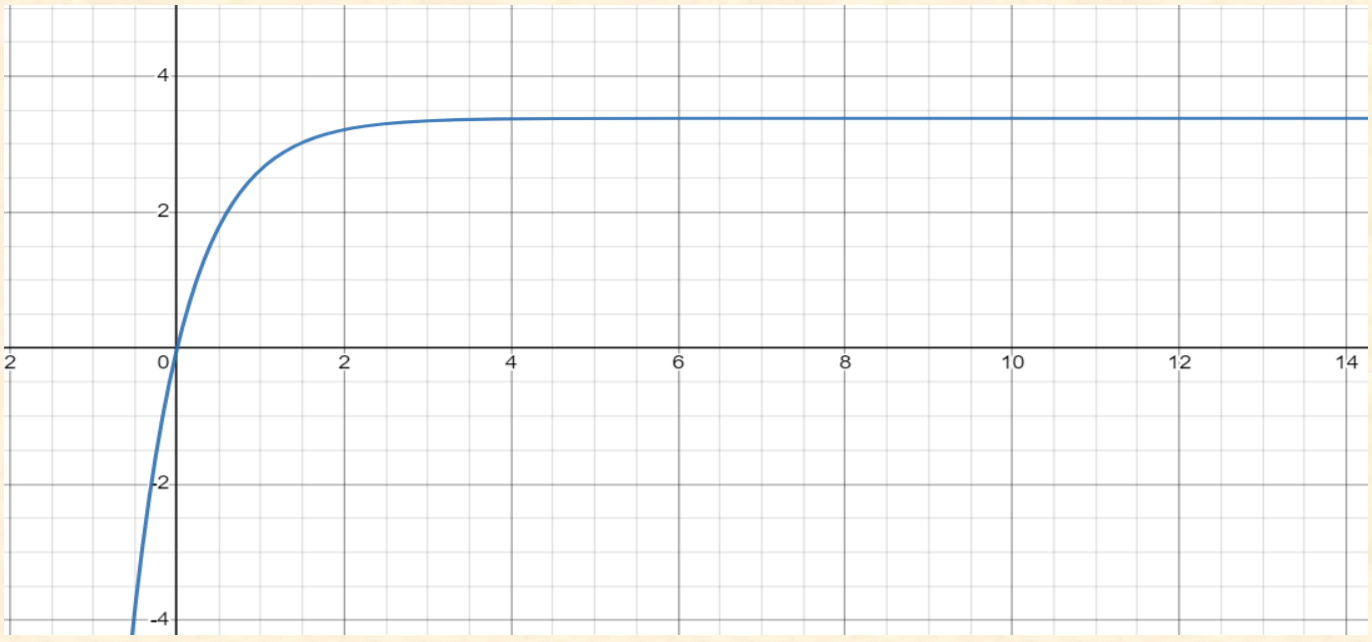
X is the initial subscriber count or revenue at the start (t = 0).

r is the conversion rate, which represents the rate at which new subscribers are added

t is the time period you're interested in (measured in months, years, etc.).

CLV stands for Customer Lifetime Value.

e^(-r * t) : This is the exponential decay function. The negative exponent models the decrease in growth rate over time. As **t** increases, the negative exponent decreases, indicating that growth slows down as your subscriber base gets larger.



Let's say you start with:

- Initial Subscribers (X) = 1000
- Monthly Growth Rate (r) = 0.05 (5%)
- Customer Lifetime Value (CLV) =(ARPU/Churn rate) =2999/0.02 = 149,950
- Subscription Period (t) = 24 months and 36 months

Using the formula: $R = X * (1 - e^{(-r * t)}) * CLV$

Let's calculate the revenue for 24 months and 36 months:

For t = 24 months: $R = 1000 * (1 - e^{(-0.1 * 24)}) * 149,950$

Calculations: $R \approx 1000 * (1 - e^{(-2.4)}) * 149,950$

$R \approx 1000 * (1 - 0.0907179) * 149,950$

$R \approx 1000 * 0.909282 * 149,950$

$R \approx$ INR 13,63,46,836

So, over 24 months, your projected revenue would be approximately INR 13,63,46,836.

For t = 36 months: $R = 1000 * (1 - e^{(-0.1 * 36)}) * 149,950$

Calculations: $R \approx 1000 * (1 - e^{(-3.6)}) * 149,950$

$R \approx 1000 * (1 - 0.027324) * 149,950$

$R \approx 1000 * 0.972646 * 149,950$

$R \approx$ INR 14,58,52,766

So, over 36 months, your projected revenue would be approximately INR 14,58,52,766.