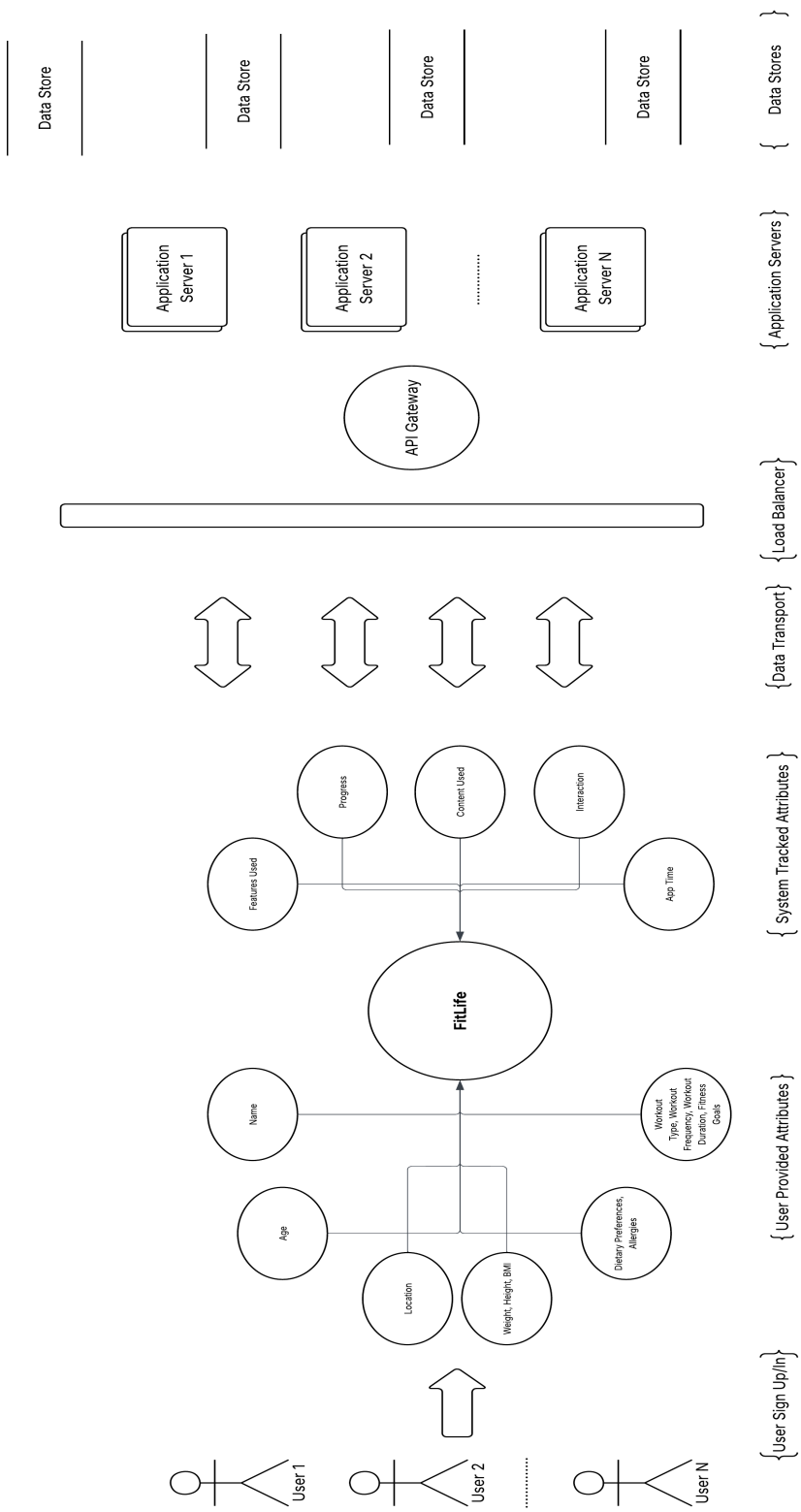
FitLife - Personalized nutritionist

FitLife collects user data, including health statistics, dietary habits, and workout routines, to provide personalized fitness and nutrition plans. It also offers community features, allowing users to connect with each other, share their progress, and even engage in friendly competition.

Section 1: Dataset

Column Name	Description	Data Type	identifier
Name	User's Full Name	String	**Direct Identifier**
Age	User's Age in years	Number	**Indirect Identifier**
Location	User's City and Country of Residence	String	**Indirect Identifier**
Gender	User's Gender	Enumeration	**Indirect Identifier**
Weight	User's Body weight in Kilograms	Number	**Indirect Identifier**
Height	User's Height in centimeters	Number	**Indirect Identifier**
BMI	User's Body Mass Index	Number	**Indirect Identifier**
Dietary Preference	User's Dietary Preference	String	**Indirect Identifier**
Allergies	User's Food Allergies	String	**Indirect Identifier**
Workout Frequency	Number of Workout Sessions of the User in a Week	Number	**Indirect Identifier**
Workout Type	Preffered Workout Type of the User	String	**Indirect Identifier**
Workout Duration	Average Workout Duration of each Session in minutes	Number	**Indirect Identifier**
Fitness Goals	Fitness objective of the User	Enumeration	**Indirect Identifier**
Progress	System tracked Fitness Progress of the User	Enumeration	**Indirect Identifier**
Features Used	App functionalities used by the User	Enumeration	**Indirect Identifier**
App Time	Average time spend on the Application in minutes	Number	**Indirect Identifier**
Interaction	Form of User Interaction on the Application	Enumeration	**Indirect Identifier**
Content Preference	Content Preffered by the User on the Application	Enumeration	**Indirect Identifier**

Section 2: System Reference Diagram (link)



Section 3: LINDDUN Threat Categories

1. **LINKABILITY**
    a. **Hotspot - Data collection for account registration**
        i. **Threat -** Same user data (Name, Age, Location) can be linked across entries related through friendship(connection) link.
    b. **Hotspot - Data Storage (Data Stores)**
        i. **Threat -** Repeated workout or usage of app can link the user's activity over the time.
2. **IDENTIFIABILITY**
    a. **Hotspot - Data Collection for account registration**
        i. **Threat -** Using real names and detailed location makes user easily identifiable.
    b. **Hotspot - Data Storage (Data Stores)**
        i. **Threat -** Combination of data points (Height, Weight, Location could be used to identify a user, if the data is leaked).
        ii. **Threat -** Misuse of Dangerous Data (Dietary Preferences) can be used to target a user.
3. **NON-REPUDIATION**
    a. **Hotspot - Application Servers**
        i. **Threat -** All activities are recorded by the user (Workout Duration, Workout Type, Workout Frequency).
4. **DETECTABILITY**
    a. **Hotspot - Data Transmission (Application -> Servers)**
        i. **Threat -** Man in the Middle attack, can possibly know when the user is online and working out.
    b. **Hotspot - Data Transmission (Servers -> Application)**
        i. **Threat -** Notifications can reveal sensitive details about the user.
5. **UNAWARENESS**
    a. **Hotspot - Continuous Sync**
        i. **Threat -** User's don't know that their fitness data keeps syncing automatically.
    b. **Hotspot - Usage Tracking**
        i. **Threat -** User might not be aware that the app tracks what features or other things they interact with.
6. **NON-COMPLIANCE**
    a. **Hotspot - Data Storage (Data Stores)**

i. **Threat -** Old user's data might not be deleted on time.
ii. **Threat -** Linking one user's data with other user's data without clearly asking for consent from the users.

b. **Hotspot - Data Collection**
   i. **Threat -** User's might not clearly agree before sharing their health data.

7. DISCLOURSE OF INFORMATION

a. **Hotspot - Data Storage (Data Stores)**
   i. **Threat -** Poor access control can lead to exposure of other users' personal health data.

Section 4: ARX Certificate - (ATTACHED CERTIFICATE FILE)

Section 5: Explanation of ARX Configuration Decisions

Input Specifications

The dataset includes **500 individual records** with **18 attributes** covering identification, demographic, and health or fitness information.

Classification	Attributes	Rationale
Identifying	Name	Directly identifies individuals. This attribute is removed from the anonymized dataset to prevent instant identification.
Quasi-Identifying	Age, Location, Gender	These demographic details can indirectly reveal identities when linked with external sources such as public records or social media.
Sensitive	Weight, Height, BMI, Workout Frequency, Workout Type, Workout Duration, Content Preference, Interaction, Allergies	These attributes reflect health status, fitness behavior, and medical conditions.
Insensitive	Progress, Fitness Goals, Features Used, Dietary Preference, App Time(minutes)	These are general usage and preference data points. They pose minimal privacy risk and help maintain the analytical utility of the dataset.

Privacy Models Applied:

1. k-Anonymity (k=10)

Each combination of quasi-identifiers (Age, Location, Gender) appears in at least **10 records**.This ensures that no individual can be uniquely identified based on demographic combinations alone.

2. ℓ-Diversity Models (for variety protection)

Ensures variety of values within groups, preventing homogeneity attacks where all members share similar behaviors/conditions

Attribute	ℓ Value	Justification
Workout Duration	10	High diversity required due to its strong behavioral uniqueness. Prevents homogeneity attacks.
Content Preference	3	Guarantees that all three preference categories are represented in each group.
Workout Type, Interaction, Allergies	3	Provides standard diversity for categorical sensitive attributes.

3. t-Closeness Models (for distribution protection)

Prevents attackers from inferring sensitive information by ensuring that attribute distributions within each group are similar to the overall dataset.

Attribute	t Value	Justification
Weight	0.3	Strong protection for stigmatized or sensitive health metrics. Prevents skewed clusters.
BMI, Height, Workout Duration	0.4	Moderate protection for correlated attributes to balance privacy with data usability.

Output Properties

- **Data Preservation:** 500 records retained (100%)
- **Attributes maintained:** All 18 columns kept after anonymization

Applied Transformations:

- **Quasi-identifiers:** (Age, Location, Gender) were generalized to their maximum levels.
- **Sensitive attributes:** retained their original values but were protected by the chosen privacy models.

Privacy Guarantees:

- **k-anonymity (k=10):** Each record is indistinguishable from at least 9 others.
- **ℓ-diversity:** All equivalence classes meet the required diversity levels for sensitive attributes.
- **t-closeness:** Attribute distributions remain within the defined similarity thresholds.
- **Re-identification risk:** Below 10% across all modeled attack scenarios.

Model Performance

- **Transformations Evaluated:** 938 possible combinations explored
- **Final Solution:** Achieved optimal balance between privacy and data utility (Anonymity = Yes)
- **Information Loss:** Normalized score of 1.0, considered acceptable given the high privacy protection levels.