

PRÁCTICA PROFESIONALIZANTE 1

Cohorte 2022 - 2do. Año

PROFESORES:

Ing. Narciso Héctor Perez

Lic. Héctor Prado

Lic. Carlos Ignacio Charletti

“Proyecto Practica Profesionalizante 1 - 2023 - Grupo 18”

ANÁLISIS PREDICTIVO DE LA COMPOSICIÓN CORPORAL

INTEGRANTES GRUPO 18:

Gutierrez	Diego	dgutierrez.m79@gmail.com
Gutierrez	Laura	laurasgneco@gmail.com
Lara	Mariela	marzular@gmail.com
Ruani	Julio	julio.ruani@gmail.com
Sánchez	Jorge Ignacio	nacho.sanchezsolano@gmail.com
Valdiviezo	Melisa	melisa.valdiviezo@gmail.com
Pineda	Mariano	pinedamariano.ds@gmail.com

ANÁLISIS PREDICTIVO DE LA COMPOSICIÓN CORPORAL

1. Introducción

El estilo de vida de la población mundial ha sufrido grandes cambios en los últimos años, en especial las prácticas de alimentación, asociadas a las condiciones económicas, el acceso a los alimentos, las costumbres de cada región y los horarios destinados a los tiempos de comida. Estas prácticas conducen a problemas de sobrepeso, obesidad incluso desnutrición, y son prácticas que actúan como factores de riesgo de padecer enfermedades cardiovasculares y metabólicas y algunos tipos de cáncer.

Se advierte sobre la importancia de tener una dieta equilibrada que contrarreste el efecto de la industrialización, movimiento que ha promovido un exceso de calorías en los alimentos y productos de consumo. Por ello, el estudio de la composición corporal es un tema de interés creciente y utilizado tanto para la investigación como para fines clínicos, ya que guardan relación directa con la salud y el riesgo de enfermedades.

1.1. Justificación

El control de peso es una estrategia para mantenernos sanos y prevenir futuras enfermedades. Se aplica de manera habitual en personas que ya sufren algún tipo de enfermedad crónica. Sin embargo, sobre todo se realiza en personas que necesitan adelgazar o engordar, en etapas de crecimiento y en casos de trastornos alimenticios.

Los controles de peso y el peso ideal van asociados casi siempre a la tendencia a adelgazar que impone la sociedad actual. Antiguamente se utilizaba el Índice de Masa Corporal (IMC) como indicador del peso de la persona. Para calcularlo se realiza aplicando la fórmula del Índice de Masa Corporal (IMC): $\text{peso (kg)} / \text{altura (m}^2\text{)}$. Dicho cálculo indica el rango de peso en el que la persona se encuentra.

Sin embargo, en la actualidad sabemos que no es un índice muy fiable a tener en cuenta, ya que también debemos conocer el porcentaje graso, el porcentaje de masa libre de grasa y otros datos para determinar si el paciente se encuentra en un peso correcto.

En este sentido los controles de peso presentan un desafío al momento de indagar en las razones de aumento/descenso de peso donde predominan multiplicidad de variables a tener en cuenta al momento de la medición.

1.2. Objetivo

Desarrollar un Modelo de Aprendizaje Automático, que permita evaluar y predecir la composición corporal de una persona a través de un seguimiento de los parámetros más importantes de lo que está compuesto el cuerpo (grasa, músculo, tejido óseo y agua).

1.2.1. Objetivos específicos

O.E 1 Evaluar el patrón del consumo de alimentos, permitiéndonos predecir los valores en un tiempo futuro con información precisa y valiosa, a fin de ajustar o regular ejercicios y alimentación adecuada para llegar a la meta personal propuesta.

O.E 1 Unificar los datos tomados de diversas fuentes y con diferentes metodologías un solo set de datos.

O.E 2 Construir un modelo que permite predecir y cuantificar las reservas corporales del organismo y en su caso, detectar y corregir problemas nutricionales.

2. Marco teórico

2.1. Recopilación de datos

Los datos que fundamentan la ejecución del modelo tienen su origen en el relevamiento de forma diaria por un integrante del grupo. Por un lado el pesaje se lleva a cabo utilizando una balanza digital que mediante conexión Bluetooth sincroniza los datos en una aplicación, que además permite el almacenamiento de datos del usuario relacionada con la edad, altura, entre otros valores, que completan el análisis corporal. Esta aplicación no tiene una función predeterminada para exportar los datos al formato necesario, por lo que debieron ser cargados a mano para conformar el set de datos.

Por otro lado se revela la ingesta diaria de alimentos en otra aplicación que determina la composición de los parámetros nutricionales más relevantes.

Estructura de datos:

→ Registros de peso en un período de 3 meses.

Fecha	Día	Peso	%Grasa
-------	-----	------	--------

→ Registro de alimentos consumidos con su respectiva información nutricional; y rutina de ejercicio durante el periodo de entrenamiento.

Fecha	Cals (kcal)	Grasa (g)	Sat (g)	Carb h(g)	Fibra (g)	Azúc ar(g)	Prot(g)	Sod (mg)	Coles (mg)	Potasio (mg)
-------	----------------	--------------	------------	--------------	--------------	---------------	-------------	-------------	---------------	-----------------

Variables:

Días transcurridos:

Esta variable representa el tiempo, es decir, el número de días transcurridos a lo largo de un período específico. Se usa como el predictor o variable independiente en el modelo para predecir la evolución de las variables de composición corporal a lo largo del tiempo.

Variables de la composición corporal:

Para predecir diferentes aspectos de la composición corporal, se emplean múltiples variables dependientes que se espera sean influenciadas por el tiempo. Estas variables incluyen, pero no se limitan a:

- Peso: El peso corporal.
- Grasa corporal (%): El porcentaje de grasa en el cuerpo.
- Músculo corporal (%): El porcentaje de masa muscular en el cuerpo.
- Metabolismo: El ritmo metabólico.
- Grasa visceral: La cantidad de grasa visceral.
- Agua corporal: El porcentaje de agua en el cuerpo.
- BMI (Índice de Masa Corporal): Una medida que relaciona el peso con la estatura.

Estas variables de la composición corporal se consideran como las variables dependientes o a predecir en el modelo de regresión lineal, con los días transcurridos como la variable independiente. El modelo se ajusta para encontrar la relación entre el tiempo y la evolución de estas diferentes métricas de la composición corporal.

Cada una de estas variables es utilizada para generar modelos de regresión lineal separados, permitiendo así predecir y proyectar la tendencia de cada componente de la composición corporal en función del tiempo transcurrido.

2.2. Limpieza y preprocesamiento de datos:

El proceso de limpieza y adecuación de los datos interactúa con la generación y consolidación del set de datos definitivo cargando los parámetros necesarios de las diferentes fuentes.

A continuación se describen los pasos necesarios.

Dataset de composición corporal:

- Unificar formato de fecha: Ambos datasets deben utilizar el mismo formato de fecha.
- Datos erróneos, vacíos, fuera de rango, etc.
- Generación de columna calculada "Dia": Esta columna se genera en base al valor de la fecha. La primera fecha del dataset será el día inicial 1, y luego el valor de día representará cuantos días pasaron desde el día inicial hasta la fecha de cada registro. Esta columna se utilizará para unificar ambos datasets, por lo cual, los valores deben estar "sincronizados".

Dataset de alimentación:

- Filtrado de datos: Deben tenerse en cuenta sólo las filas que corresponden a los totales de cada día (no tener en cuenta los datos detallados de desayuno, almuerzo, cena, etc).
- Datos erróneos, vacíos, fuera de rango, etc.
- Unificar formato de fecha: Ambos datasets deben utilizar el mismo formato de fecha.
- Generación de columna calculada "Dia".

3. Exploración de Datos:

El proceso de exploración de datos incluye la observación de la estructura del dataset midiendo su extensión y chequeando que los formatos sean los adecuados.

Esto es útil para una rápida visualización de los datos, mostrando las columnas y sus valores iniciales.

También se muestra la cantidad de entradas no nulas por columna, el tipo de datos, etc. Las estadísticas descriptivas son un resumen para las columnas numéricas. Esto incluye recuentos, media, desviación estándar, valores mínimos y máximos, así como los respectivos cuartiles. Proporciona información útil sobre la distribución de los datos numéricos, lo que puede ayudar a identificar valores atípicos, la dispersión de los datos y la tendencia central.

3.1. Exploración y visualización de las variables

Se presenta una exploración exhaustiva y visualización de los datos contenidos en el set de datos. La serie de gráficos y visualizaciones busca comprender las relaciones, distribuciones y anomalías presentes.

Relaciones entre variables:

Utiliza gráficos de dispersión para explorar la relación entre diferentes variables como peso y porcentaje de grasa, peso y calorías, porcentaje de grasa y músculo, entre otros.

La matriz de diagramas de dispersión muestra las relaciones entre varios pares de variables numéricas al mismo tiempo.

Análisis de tendencias:

Representa la tendencia del peso a lo largo del tiempo usando un gráfico de líneas. Visualiza la relación entre porcentajes de grasa y músculo en función del índice de masa corporal (BMI) tanto en gráficos 2D como 3D.

Correlaciones:

Calcula y visualiza la matriz de correlación entre todas las variables numéricas. En el cual se aprecia la correlación más fuerte y más baja entre las variables.

Análisis de datos atípicos:

Identifica valores atípicos (outliers) utilizando diagramas de caja (boxplot) para variables como peso, porcentaje de grasa, BMI y carbohidratos, así como un conteo de valores atípicos para carbohidratos.

Distribuciones de probabilidad:

Muestra la distribución de diversas variables numéricas mediante histogramas y gráficos de densidad (KDE), como la distribución de peso, grasa, metabolismo, agua, proteína, colesterol, entre otros.

4. Ingeniería de Características

En primera instancia, se procedió a generar los modelos empleando los datos en su estado original, es decir, sin modificaciones posteriores a la limpieza del dataset. Posteriormente, se llevó a cabo la evaluación de las métricas correspondientes a dichos modelos.

En un segundo momento, se realizó un ajuste en el conjunto de datos. Este ajuste consistió en el recorte de ciertos períodos específicos. Por ejemplo, se excluyó el intervalo cercano al día 36, donde se evidenció una disminución de peso considerable en un lapso de 3 días, seguida de una recuperación de peso en los días subsiguientes. Asimismo, se efectuó un

recorte a partir del mes de septiembre, debido a la interrupción del seguimiento del plan de entrenamiento, que resultó en un aumento de peso en lugar de la disminución esperada. El propósito de estos recortes radica en la evaluación más precisa de los resultados derivados del plan de entrenamiento, descartando aquellas instancias que no reflejen fielmente el impacto del proceso de entrenamiento en la pérdida de peso.

Una vez realizados estos ajustes en el dataset, se procedió a generar nuevamente los modelos pertinentes y se llevó a cabo una nueva evaluación de las métricas. Los resultados obtenidos revelaron mejoras significativas en las métricas luego de estos recortes.

Este procedimiento permitió obtener una visión más precisa y relevante del impacto del plan de entrenamiento en la pérdida de peso, al considerar únicamente los datos pertinentes y representativos, excluyendo aquellos que podrían distorsionar la evaluación del modelo de machine learning.

5. Selección del modelo

Optamos por emplear un modelo de regresión lineal, ya que nuestro propósito consiste en proyectar las tendencias actuales de los distintos componentes de la composición corporal hacia el futuro.

La regresión lineal asume una relación lineal entre las variables predictoras (días transcurridos) y las variables objetivo (metabolismo, grasa visceral, agua, BMI). Esta suposición puede ser limitante si la relación es no lineal.

La regresión lineal asume independencia entre las observaciones, lo cual es importante para la validez del modelo.

Tras aplicar el modelo, procedimos a evaluar las métricas resultantes, identificando algunas imprecisiones. Para abordar esta situación, realizamos ajustes en el conjunto de datos.

6. Evaluación del modelo

Durante la revisión, detectamos mediciones incorrectas que afectaban la precisión del modelo. Además, identificamos un período en el conjunto de datos en el cual el individuo desvió su dieta planificada. Este intervalo impactaba negativamente en el modelo, contradiciendo la evaluación de los resultados del plan de entrenamiento y dieta.

Luego de corregir estas inconsistencias, reconstruimos los modelos. Esta vez, al evaluar las métricas nuevamente, observamos mejoras significativas, alcanzando valores métricos aceptables. Estos ajustes y correcciones no solo optimizan la precisión del modelo, sino que

también garantizan una evaluación más fiel y precisa de los resultados del plan de entrenamiento y dieta en el estudio de la composición corporal.

Predecir la composición corporal a futuro, proyectando la tendencia actual. Predecir a 3, 6, 9 meses y un año el peso, porcentajes de grasa y músculo de una persona.

Especificando un porcentaje de grasa que se quiera lograr, el modelo calculará el tiempo necesario para llegar a ese objetivo.

7. Entrenamiento del Modelo

El entrenamiento del modelo se llevó a cabo dividiendo el conjunto de datos en dos segmentos: un 80% se utilizó para el entrenamiento del modelo, mientras que el 20% restante se reservó para probar y evaluar la eficacia del modelo.

Este enfoque, conocido como técnica de "train-test split" o división de entrenamiento y prueba, es fundamental en el aprendizaje automático. La porción más grande (80%) se empleó para ajustar y entrenar el modelo, permitiendo que el algoritmo aprendiera las relaciones entre las variables independientes (como los días transcurridos) y las variables dependientes (componentes de la composición corporal). Posteriormente, el 20% restante se utilizó para evaluar la capacidad predictiva del modelo en datos no vistos previamente durante el entrenamiento.

Esta separación facilita la verificación de la capacidad del modelo para generalizar y predecir con precisión datos nuevos, lo que es esencial para garantizar que el modelo no esté simplemente memorizando los datos de entrenamiento, sino que realmente esté aprendiendo patrones que puedan aplicarse a datos futuros.

8. Optimización del Modelo

Corrección de datos:

Identificar y corregir mediciones erróneas en el conjunto de datos ayudó a mejorar la precisión del modelo. Además, excluir el período con datos inexactos produjo predicciones más precisas al alinear los resultados con el plan de dieta y entrenamiento.

Reentrenamiento del modelo:

Después de realizar correcciones en el dataset, se volvieron a generar los modelos. Esto resultó en métricas mejoradas y predicciones más precisas para las diferentes variables de la composición corporal.

9. Software y Herramientas

- ❖ App FatSecret
- ❖ Herramientas colaborativas: Trello, Colab, Github
- ❖ Python y sus librerías Pandas Matplotlib, scikit-learn

10. Tendencias en la Composición Corporal

Peso Corporal:

Se observa una disminución gradual del peso a lo largo de los meses proyectados, lo que sugiere una tendencia a la baja en el peso del individuo.

Grasa Corporal (%):

El porcentaje de grasa corporal también presenta una disminución constante, lo que indica una tendencia a reducir el contenido de grasa en el cuerpo.

Músculo Corporal (%):

A la par de la disminución de la grasa, se aprecia una tendencia a mantener un porcentaje estable de músculo en el cuerpo a lo largo del tiempo proyectado.

Metabolismo:

El metabolismo parece disminuir de manera constante, lo que podría sugerir un cambio en el ritmo metabólico del individuo.

Grasa Visceral:

La grasa visceral muestra una disminución gradual, lo cual es positivo, ya que menores niveles de grasa visceral suelen asociarse con una mejor salud.

Agua Corporal:

Se prevé un aumento gradual en el porcentaje de agua en el cuerpo, lo que puede ser un indicador de hidratación adecuada o un cambio en la retención de líquidos.

BMI (Índice de Masa Corporal):

El BMI refleja una disminución constante, lo que sugiere una reducción del índice de masa corporal a lo largo del período proyectado.

11. Conclusiones generales

Las proyecciones indican una tendencia general hacia una composición corporal más saludable, con reducción en peso, porcentaje de grasa y BMI, acompañada de una preservación del porcentaje muscular.

La balanza digital permite el análisis de la composición corporal, a través de una técnica no invasiva de gran precisión que en un corto periodo de tiempo permite obtener datos de manera fiable para la evaluación del estado de hidratación y nutrición.

Es un método seguro, económico, preciso y no invasivo que proporciona datos sobre la composición corporal de una persona.

Importante que es de fácil mantenimiento, uso y mínima instalación. Y muy importante de destacar que el objetivo final es poder predecir la composición corporal a futuro de una persona, teniendo en cuenta los datos actuales y de esa manera lograr con anticipación encontrar el ejercicio y alimentación adecuada para obtener el objetivo en el tiempo propuesto.

Los resultados pueden verse afectados por factores como la temperatura y la hidratación de la piel, así como la ingesta reciente de alimentos o líquidos. Esto puede introducir errores en las mediciones y limitar la precisión de los resultados. Para obtener mediciones precisas, es necesario seguir ciertas pautas, como evitar el ejercicio físico intenso antes de la prueba y abstenerse de consumir ciertos alimentos.

Estas proyecciones proporcionan una visión general de la posible evolución de la composición corporal del individuo en un periodo futuro, pero es esencial considerarlas como estimaciones que podrían variar en función de múltiples factores.

12. Ética y Consideraciones de Privacidad

En este sentido es un punto a tener en cuenta el desarrollo de lineamientos éticos y de privacidad de los datos de estudios que no se tuvieron en cuenta al ser este un proyecto con fines educativos. Sin embargo debería revisarse en profundidad este punto.

13. Limitaciones de la Metodología

Las limitaciones de la metodología están dadas por las dificultades en la exportación y carga de datos que se espera puedan resolverse en una siguiente fase de desarrollo.

Una cuestión a considerar es la cantidad de datos que alimentaron el modelo; deberían ampliarse para obtener resultados más relevantes.

14. Recomendaciones y consideraciones futuras

A pesar de las mejoras logradas, se sugiere explorar modelos más avanzados que puedan capturar relaciones no lineales y considerar factores adicionales para enriquecer el análisis de la composición corporal. La validación continua y la interpretación clínica siguen siendo fundamentales para asegurar la utilidad y precisión del modelo en el contexto de la salud y la composición corporal.

Finalmente, cabe destacar que la información recopilada para este estudio no incluyó datos detallados sobre las comidas, específicamente en cuanto al consumo de macronutrientes como grasas, carbohidratos y proteínas. La ausencia de esta información limita la capacidad de realizar un análisis más profundo sobre cómo el consumo de estos macronutrientes podría influir en las variaciones de peso y, por ende, en la composición corporal del individuo.

La incorporación de datos alimenticios detallados permitiría una comprensión más holística de los factores que influyen en la evolución de la composición corporal. El análisis de la relación entre la ingesta de macronutrientes y las variaciones de peso ofrecería una perspectiva más completa sobre cómo la dieta impacta en la salud y en la modificación de la composición corporal a lo largo del tiempo.

15. Enlaces del Proyecto



[GitHub - julio-ruani-isp/Grupo18-PracticaProfesionalizante](https://github.com/julio-ruani-isp/Grupo18-PracticaProfesionalizante)



<https://trello.com/b/FYwfPsFE/pr%C3%A1ctica-profesionalizante-1grupo-18>