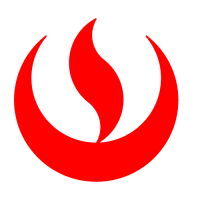
**UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS**

**FACULTAD DE INGENIERÍA**

**CARRERA PROFESIONAL**

**DE CIENCIAS DE LA COMPUTACIÓN**



**Asignatura:**

**CC72- Programación Concurrente y Distribuida**

**Sección: CC81**

**TRABAJO PARCIAL**

**Autores**

– u202013066 Christian Joel Cutipa Cañapataña

– u202021508 Jorge Omar Tarapa Peña

– u20201c579 Henry Josue Diaz Huarcaya

– u201713107 Juan Sebastian Osorio Quiroz

**Profesor**

Carlos Alberto Jara Garcia

Lima, septiembre de 2024

1. Introducción
2. Explicación del dataset
3. Planteamiento del problema
4. Explicación del algoritmo y el uso de los mecanismos de sincronización utilizado
5. Presentar las pantallas del menú
6. Explicación de las pruebas realizadas (Pegar capturas)
7. Bibliografía
8. Anexos

**1.** **DATASET**

El Diabetes Prediction Dataset es un conjunto de datos que incluye información médica y demográfica de pacientes, junto con su estado de diabetes (positivo o negativo). Las características del dataset incluyen variables como edad, género, índice de masa corporal (BMI), hipertensión, enfermedades cardíacas, historial de tabaquismo, nivel de HbA1c (hemoglobina glicosilada) y nivel de glucosa en sangre. Estas variables son utilizadas para entrenar modelos de aprendizaje automático con el objetivo de predecir la probabilidad de que una persona desarrolle diabetes.

Las principales columnas del dataset son:

* Género: Incluye tres categorías: masculino, femenino y otro. El 59% de los datos corresponde a mujeres, y el 41% a hombres.
* Edad: Rango de 0 a 80 años, con una media de 41.9 años.
* Hipertensión: Indica si el paciente tiene hipertensión (1) o no (0).
* Enfermedad cardíaca: Variable binaria que indica la presencia (1) o ausencia (0) de esta condición.
* Historial de tabaquismo: Seis categorías, donde "No Info" es la más frecuente.
* BMI (Índice de Masa Corporal): Valores que oscilan entre 10.16 y 71.55, siendo la media 27.3, lo cual indica un rango normal a obeso.
* Nivel de HbA1c: Proporciona el promedio de los niveles de azúcar en sangre en los últimos 2-3 meses. El promedio de HbA1c en la muestra es 5.53.
* Nivel de glucosa en sangre: Indicador clave de la presencia de diabetes, con un promedio de 138 mg/dL.
* Diabetes: Variable objetivo con valores de 0 (ausencia de diabetes) o 1 (presencia de diabetes).

**2.** **PLANTEAMIENTO DEL PROBLEMA**

La diabetes es una enfermedad crónica que afecta a millones de personas en todo el mundo. Un diagnóstico temprano es crucial para prevenir complicaciones graves y reducir los costos asociados con el tratamiento a largo plazo. Actualmente, muchos centros de salud enfrentan dificultades para detectar a tiempo a los pacientes que presentan un alto riesgo de desarrollar diabetes debido a la falta de herramientas predictivas eficientes y automatizadas.

- **Propuesta de Solución:**

El caso de negocio propuesto consiste en desarrollar un sistema de predicción de diabetes para instituciones de salud que permita identificar pacientes en riesgo de manera rápida y precisa utilizando algoritmos de Random Forests implementados en GO, sin librerías externas, lo que garantiza la flexibilidad y control sobre el proceso de predicción.

- **Objetivos:**

· Predicción temprana de diabetes: Utilizando datos médicos de pacientes (edad, género, nivel de glucosa, índice de masa corporal, antecedentes médicos, entre otros), el sistema será capaz de predecir el riesgo de desarrollar diabetes con un alto grado de precisión.

· Optimización del rendimiento: La implementación del algoritmo de Random Forests en GO permitirá aprovechar la paralelización y sincronización para mejorar la eficiencia, reduciendo los tiempos de procesamiento y adaptándose a grandes volúmenes de datos.

· Integración en sistemas de salud: El sistema podrá integrarse en las plataformas digitales de centros médicos, facilitando el acceso a las predicciones por parte del personal sanitario.

· Reducción de costos médicos: Al identificar de manera temprana a los pacientes en riesgo, los centros de salud podrán intervenir antes de que se desarrollen complicaciones, reduciendo significativamente los costos asociados al tratamiento de la diabetes avanzada.

**3.** **ALGORITMO**

El código utiliza concurrencia para entrenar varios árboles de decisión de manera simultánea, aprovechando las capacidades de paralelización y sincronización que ofrece el lenguaje GO. Esto se logra mediante la función buildForest, donde se entrenan múltiples árboles de decisión en paralelo, cada uno en su propia goroutine.

- **Goroutines para Entrenamiento Concurrente**

El entrenamiento de cada árbol de decisión se ejecuta dentro de una goroutine. Esto permite que varios árboles se construyan en paralelo, lo cual es fundamental para mejorar la eficiencia del algoritmo de Random Forest, ya que este requiere la construcción de múltiples árboles.

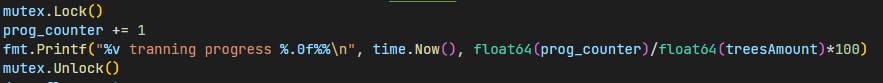
Figura 1:



- **Evitar la condición de carrera con Mutex:**

El código usa sync.Mutex para evitar la condición de carrera ya que tenemos una variable contador que será compartida.

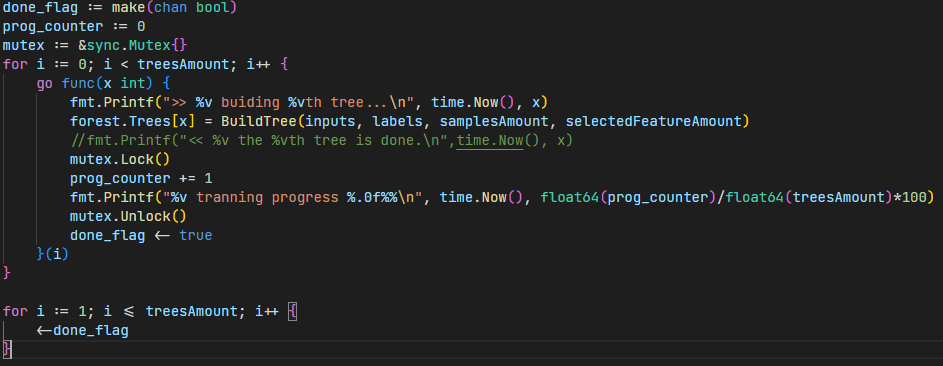
Figura 2:



- **BuildForest**

El código utiliza goroutines para entrenar múltiples árboles en paralelo, lanzando una goroutine por cada árbol en lugar de hacerlo de forma secuencial. Para gestionar la concurrencia, se emplea Canales, al final de cada goroutine enviamos un True al Channel. Finalmente, utilizamos un bucle que simula un Wait para garantizar que el programa no avance hasta que todas las goroutines hayan completado su trabajo.

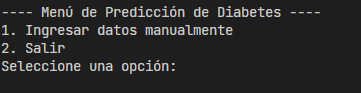
Figura 3



**4.** **INTERFAZ**

En la Figura 4 se ve un menú realizado en consola para que cualquiera usuario pueda hacer uso del algoritmo RandomForest y poder ingresar sus datos de manera manual.

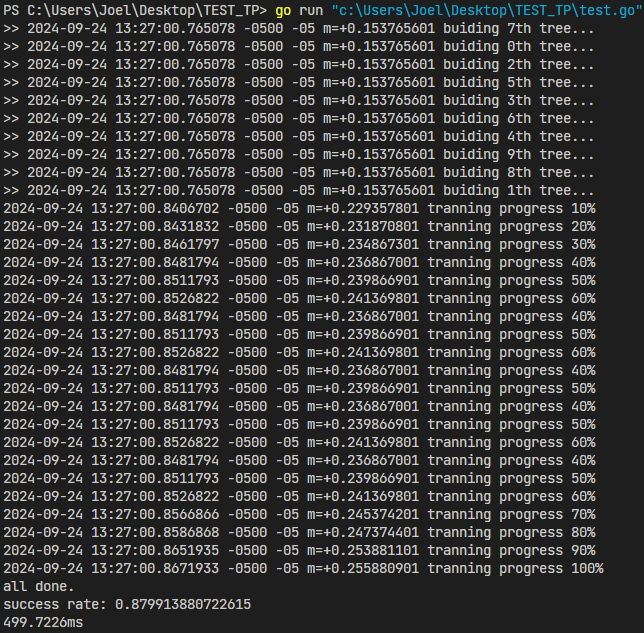
Figura 4



**5.** **PRUEBAS**

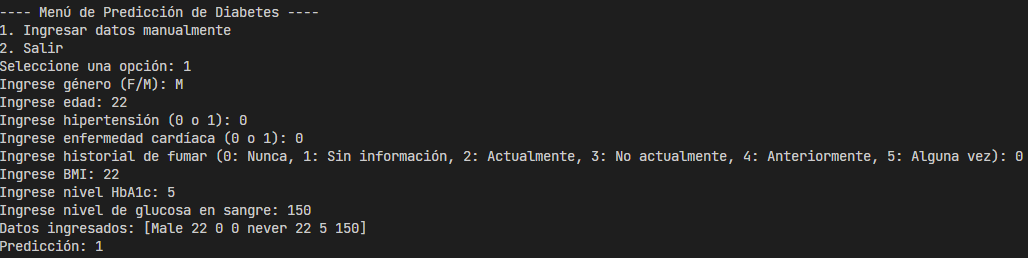
Durante el entrenamiento del modelo **Random Forest**, se construyeron y entrenaron 10 árboles de decisión. El progreso se reportó en intervalos del 10%, hasta completar el entrenamiento al 100%. Al finalizar, el sistema registró una **tasa de éxito del 87%** y un tiempo total de procesamiento de **409.72 segundos**. Como se aprecia en la Figura 5.

Figura 5



Y gracias a que hicimos un menú pudimos hacer pruebas con otros datos. El usuario ingresa datos como sexo, edad, hipertensión, antecedentes cardíacos, historia de fumar, IMC y nivel de glucosa. El algoritmo procesa esta información y devuelve una predicción, representada por un valor (en este caso "1", indicando posible riesgo de diabetes).

Figura 6



**6.** **BIBLIOGRAFÍA**

Mustafa, T. (2023). *Diabetes prediction dataset* [Data set]. Kaggle.<https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset/discussion?sort=undefined>

**7.** **GITHUB**

<https://github.com/joyel124/RF_GO>

**8.** **VIDEO**

[**https://youtu.be/e5tZvk36pWA**](https://youtu.be/e5tZvk36pWA)