

# MODELOS PREDICTIVOS DE RIESGO DE DIABETES BASADOS EN CONSUMO ALIMENTARIO Y NUTRIENTES (ID 1891)

**Jeison Suescum Holguín (1), Juliana Chaura Cortes (1),  
Delia Ortega Lenis (2), Andrés Jaramillo Botero (1).**

1-Instituto de Investigación iÓMICAS, Pontificia Universidad Javeriana Cali, Cali,  
Colombia

2-Departamento de Salud Pública y Epidemiología, Pontificia Universidad  
Javeriana Cali, Cali, Colombia





≈ 422 millones de enfermos en **el mundo**

- Diabetes tipo 2 (90-95%)
- Diabetes tipo 1
- Diabetes gestacional

**Es un desafío para la salud pública.**

≈ 1,6 millones de casos **en Colombia**

(Aschner, 2010; Ministerio de Salud y Protección Social, 2023).



## El foco: Diabetes tipo 2

Los factores de riesgo son variados, destacan la obesidad, la poca actividad física y **la alimentación**



En **Scopus** no se encontraron estudios que empleen ciencia de datos para relacionar/predecir diabetes usando información de alimentación y/o nutrientes

# Introducción

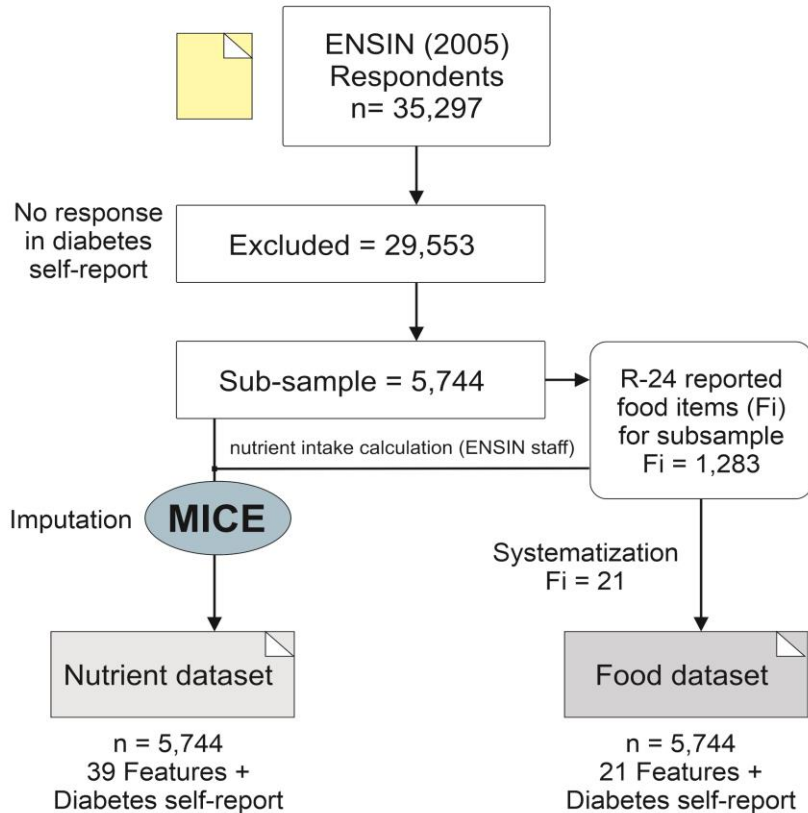
**Diabetes mellitus: Es una enfermedad metabólica crónica caracterizada por el incremento de azúcar en la sangre y la incapacidad de metabolizarla.**

(Organización Panamericana de la Salud, 2023)

# Materiales y métodos

Datos: Encuesta Nacional de Situación Nutricional de Colombia (ENSIN 2005)

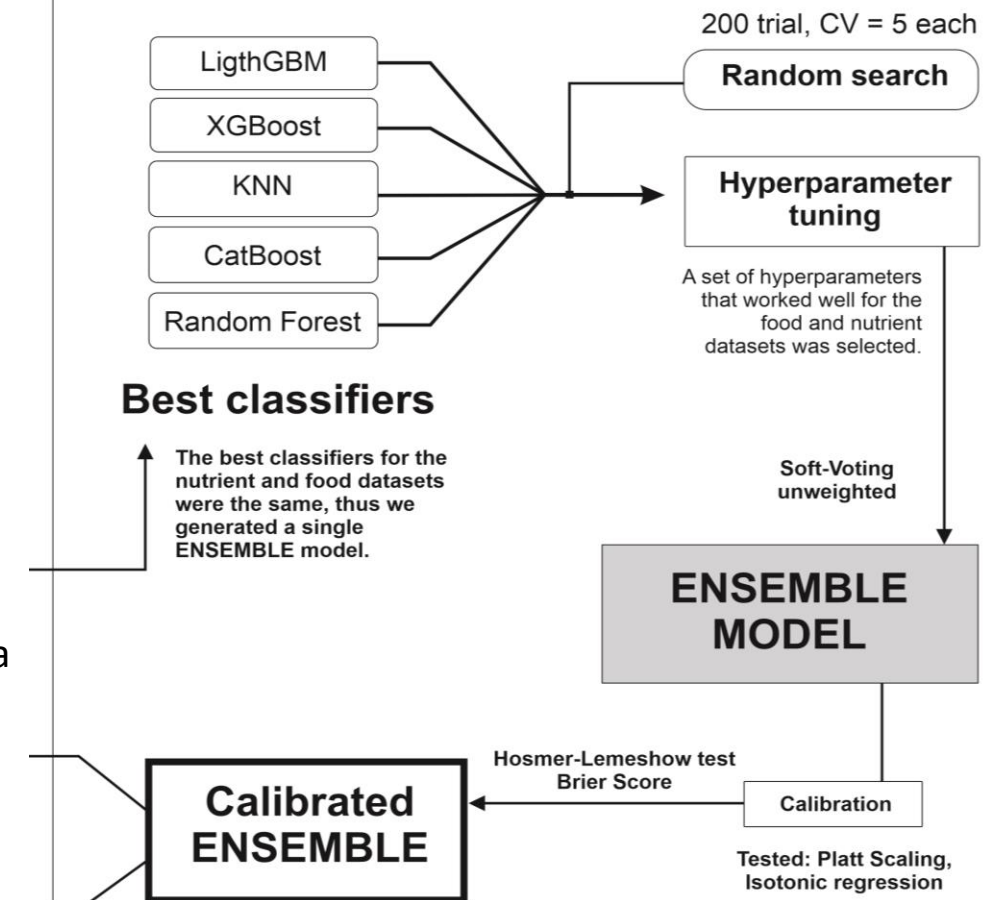
## 1 Data source & Sample



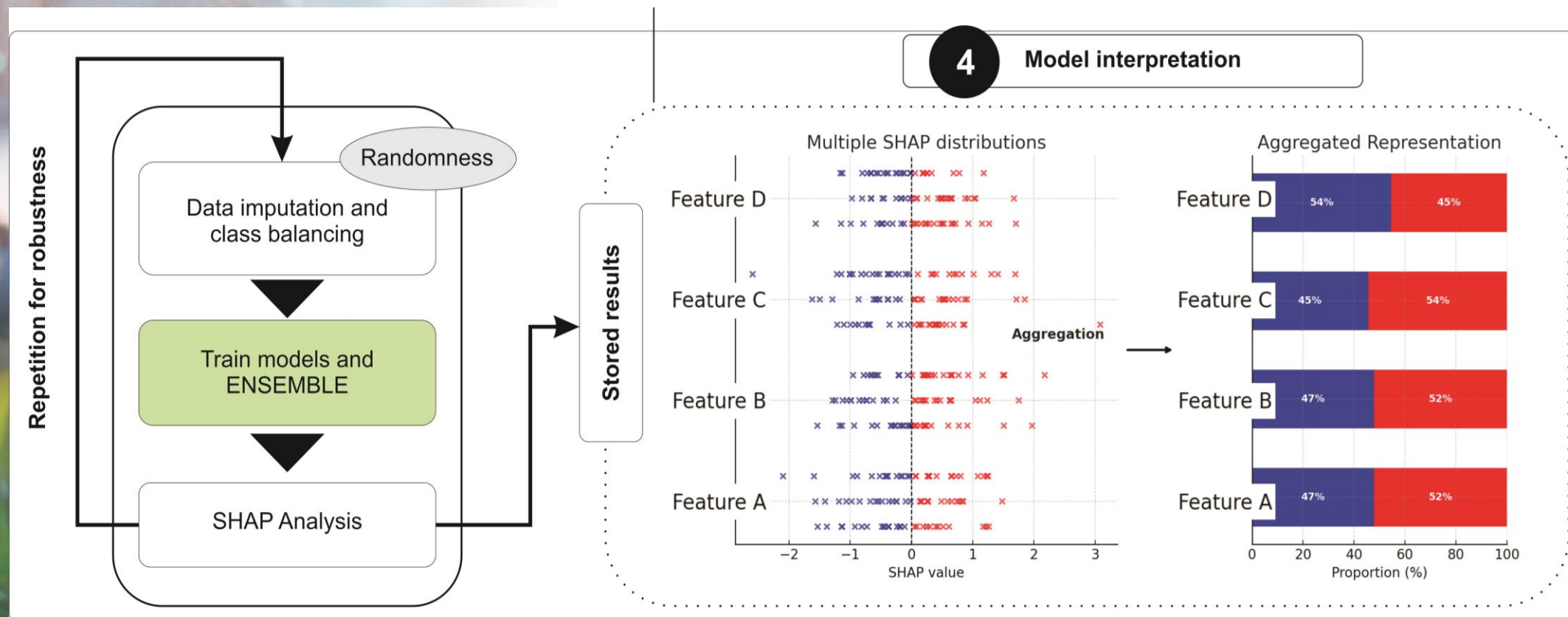
- Análisis exploratorio
- Selección de variables
- Balanceo de los datasets (SMOTE & ADASYN)
- Pruebas de clasificadores

- Datasets
  - Alimentación por persona VS **Diabetes**
  - Ingesta de nutrientes por persona VS **Diabetes**

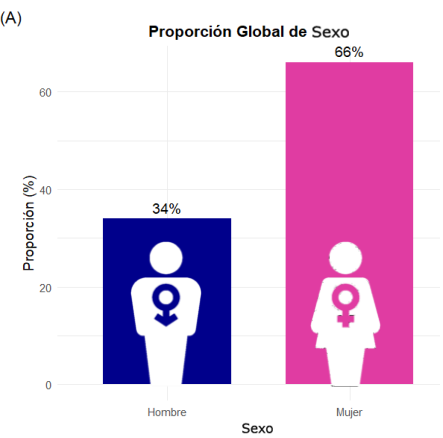
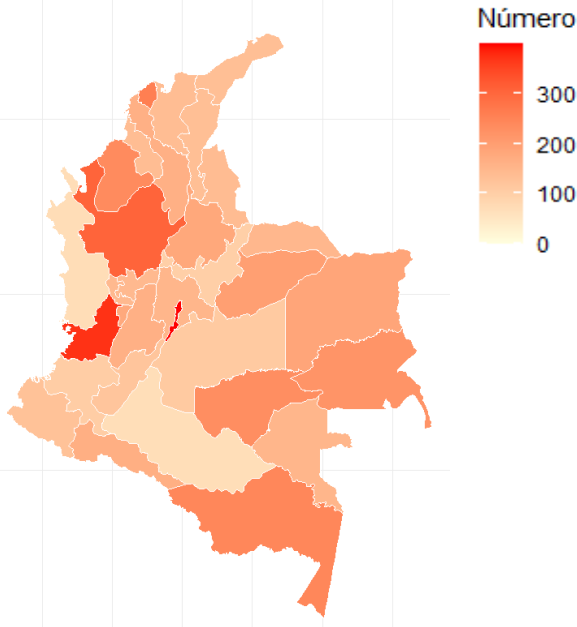
## 3 Ensemble & calibration



# Materiales y métodos





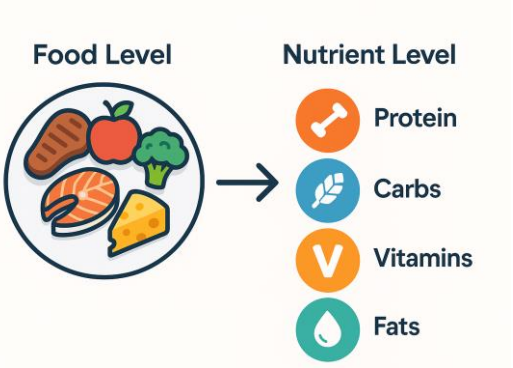


## Alimentos (21 Variables)

Grupo de Alimentos	Descripción
Lácteos	Productos derivados de la leche, como queso y yogur.
Grano y Cereales	Cereales y sus derivados, como arroz, pan y pasta.
Aceites o Grasas	Grasas y aceites utilizados en la alimentación.
Verduras	Hortalizas y vegetales consumidos en la dieta.
Proteínas	Fuentes de proteínas de origen animal y vegetal.
Otro	Alimentos no clasificados en otras categorías.
Bebidas Calientes	Bebidas como café, té y chocolate caliente.
Dulce	Productos azucarados y postres.
Frutas	Frutas frescas y naturales.
Carne Blanca	Carnes como pollo, pavo y pescado.
Tubérculos	Raíces y tubérculos como papa, yuca y ñame.
Alimento Procesado	Alimentos industrializados con procesamiento.
Bebida Azucarada	Jugos artificiales, gaseosas y otras bebidas dulces.
Carne Roja	Carnes como res, cerdo y cordero.
Snacks	Bocadillos y aperitivos de consumo rápido.
Frutas Jugo	Jugos naturales y procesados de frutas.

•  
•  
•

•  
•  
•



## Nutrientes (40 Variables)

Nombre de la Variable	Descripción
Kilocalorías (kcal)	Energía total proporcionada por los alimentos consumidos.
Proteína (g)	Contenido total de proteínas en la dieta.
Grasa (g)	Cantidad total de grasas consumidas.
Grasa Polisaturada (g)	Porción de grasas poliinsaturadas en la dieta.
Grasa Monosaturada (g)	Porción de grasas monoinsaturadas en la dieta.
Grasa Saturada (g)	Porción de grasas saturadas en la dieta.
Carbohidratos (g)	Cantidad total de carbohidratos consumidos.
Calcio (mg)	Contenido de calcio en la dieta.
Colesterol (mg)	Cantidad de colesterol presente en los alimentos.
Fibra Cruda (g)	Porción de fibra no soluble presente en los alimentos.
Fibra Dietética (g)	Contenido total de fibra dietética en la dieta.
Cenizas (g)	Minerales residuales presentes en los alimentos.
Fósforo (mg)	Cantidad de fósforo consumido.
Hierro (mg)	Contenido de hierro en la dieta.
Sodio (mg)	Cantidad total de sodio consumido.
Potasio (mg)	Contenido de potasio en los alimentos.
Magnesio (mg)	Cantidad de magnesio presente en la dieta.
Zinc (mg)	Contenido de zinc en la dieta.

•  
•  
•

# Resultados

XGBoost

LightGBM

CatBoost

Random  
Forest

KNN

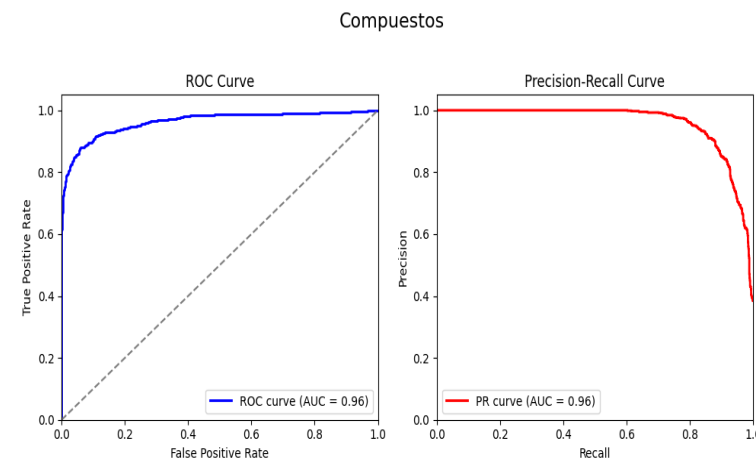
Modelo Ensemble

Compuestos / nutrientes					
Fold	Accuracy	Precision	Recall	F1	
1	0,913	0,912	0,858	0,884	
2	0,932	0,930	0,892	0,911	
3	0,923	0,886	0,916	0,901	
4	0,942	0,932	0,916	0,924	
5	0,932	0,915	0,908	0,911	
6	0,945	0,955	0,899	0,926	
7	0,939	0,939	0,899	0,919	
8	0,906	0,909	0,840	0,873	
9	0,955	0,965	0,916	0,940	
10	0,916	0,904	0,874	0,889	
Promedio	0,930	0,925	0,892	0,908	

89,2% Diabéticos identificados

Alimentos					
Fold	Accuracy	Precision	Recall	F1	
1	0,849	0,814	0,793	0,804	
2	0,869	0,808	0,871	0,838	
3	0,896	0,921	0,802	0,857	
4	0,822	0,794	0,733	0,762	
5	0,842	0,822	0,759	0,789	
6	0,852	0,827	0,785	0,805	
7	0,838	0,813	0,757	0,784	
8	0,805	0,752	0,739	0,746	
9	0,852	0,809	0,809	0,809	
10	0,872	0,829	0,844	0,836	
Promedio	0,850	0,819	0,789	0,803	

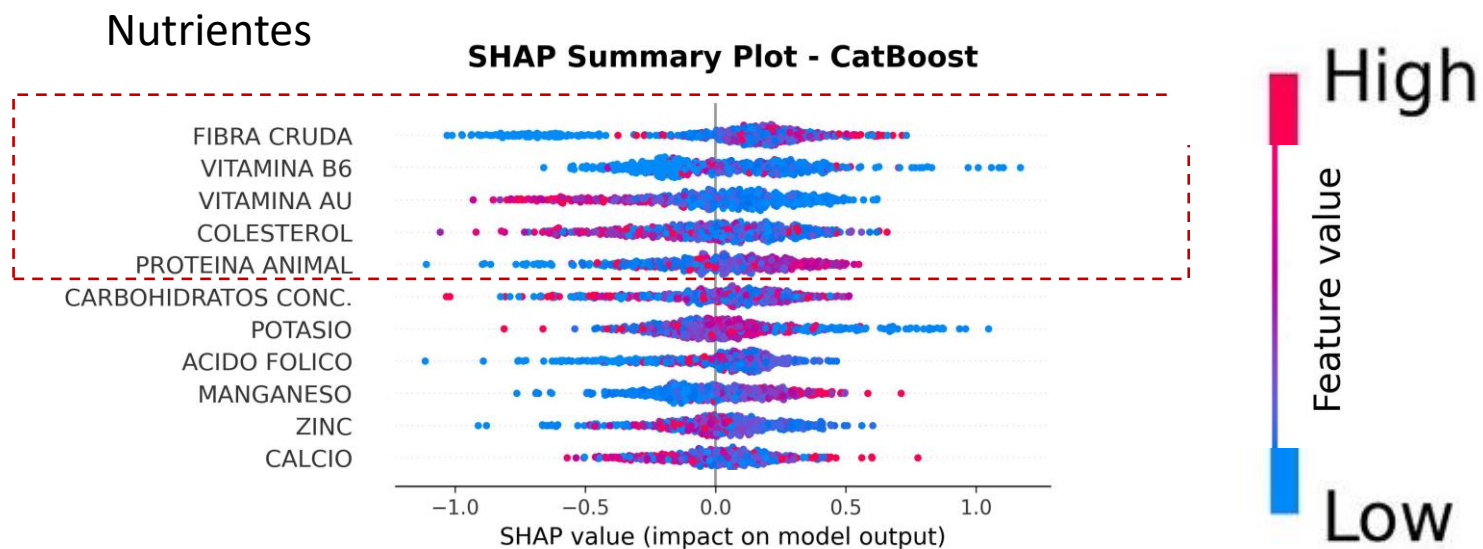
78,9% Diabéticos identificados



ROC curve (AUC = 0.96)

- ↑ Frutas y bebidas calientes
- ↑ Dulces y golosinas
- ↑ Vitamina A
- ↑ Fibra y proteína animal

Resultados



XGBoost

LightGBM

CatBoost

Random  
Forest

KNN

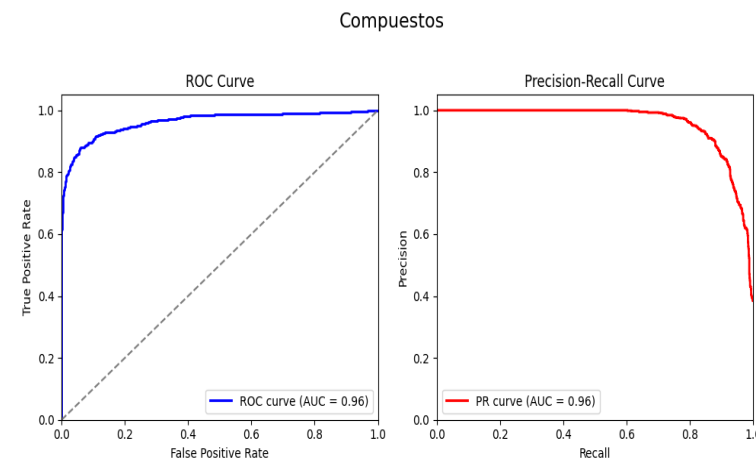
Modelo Ensemble

Compuestos / nutrientes					
Fold	Accuracy	Precision	Recall	F1	
1	0,913	0,912	0,858	0,884	
2	0,932	0,930	0,892	0,911	
3	0,923	0,886	0,916	0,901	
4	0,942	0,932	0,916	0,924	
5	0,932	0,915	0,908	0,911	
6	0,945	0,955	0,899	0,926	
7	0,939	0,939	0,899	0,919	
8	0,906	0,909	0,840	0,873	
9	0,955	0,965	0,916	0,940	
10	0,916	0,904	0,874	0,889	
Promedio	0,930	0,925	0,892	0,908	

89,2% Diabéticos identificados

Alimentos					
Fold	Accuracy	Precision	Recall	F1	
1	0,849	0,814	0,793	0,804	
2	0,869	0,808	0,871	0,838	
3	0,896	0,921	0,802	0,857	
4	0,822	0,794	0,733	0,762	
5	0,842	0,822	0,759	0,789	
6	0,852	0,827	0,785	0,805	
7	0,838	0,813	0,757	0,784	
8	0,805	0,752	0,739	0,746	
9	0,852	0,809	0,809	0,809	
10	0,872	0,829	0,844	0,836	
Promedio	0,850	0,819	0,789	0,803	

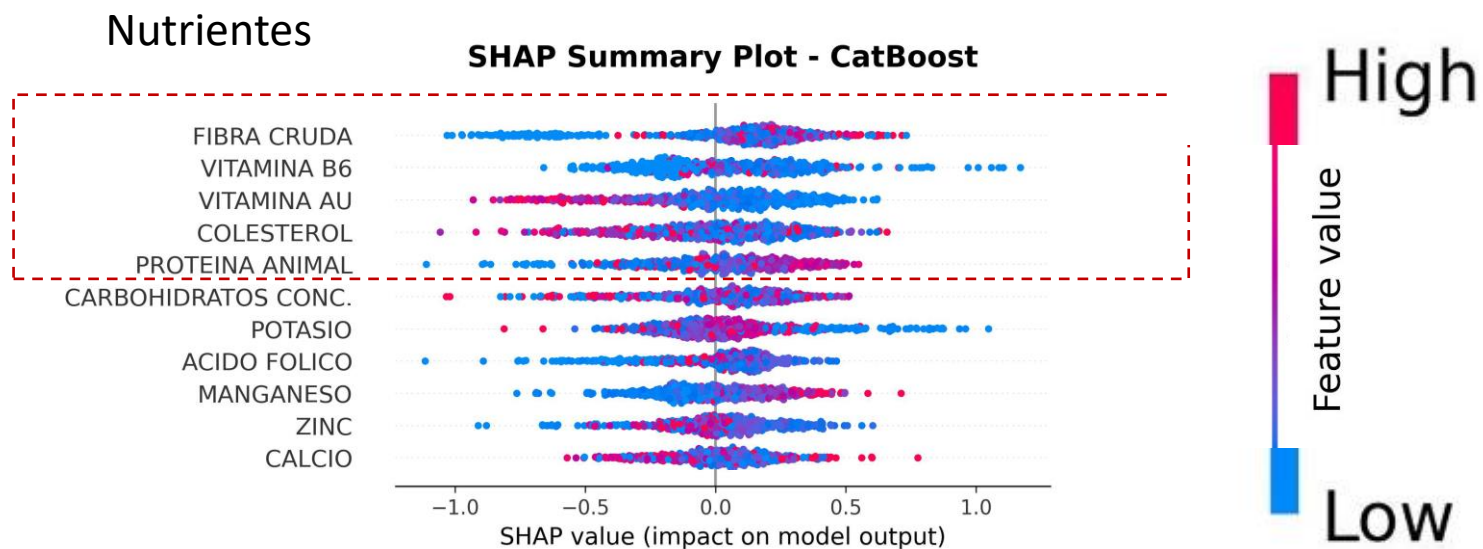
78,9% Diabéticos identificados



ROC curve (AUC = 0.96)

- ↑ Frutas y bebidas calientes
- ↑ Dulces y golosinas
- ↑ Vitamina A
- ↑ Fibra y proteína animal

Resultados





# Conclusión

- **Primer modelo** que combina datos de consumo alimentario y composición nutricional para predecir diabetes mellitus en el contexto colombiano
- **Alto poder discriminativo:**  $AUC = 0,96$  y  $F1 = 0,90$  al diferenciar personas con y sin diagnóstico.
- **Interpretabilidad:** el marco SHAP revela cómo alimentos y nutrientes podrían estar asociados al riesgo de diabetes en Colombia. Generación de hipótesis para estudios clínicos futuros.





# Referencias

- Organización Panamericana de la Salud. (2023). *Diabetes*. Recuperado el 19 de mayo de 2024 de <https://www.paho.org/es/temas/diabetes>
- Ministerio de Salud y Protección Social. (2023, 13 de noviembre). *En el Día Mundial de la Diabetes, Minsalud promueve acciones prioritarias para evitar el crecimiento de esta enfermedad*. Recuperado de <https://www.minsalud.gov.co>
- Aschner, P. (2010). Epidemiología de la diabetes en Colombia. *Avances en Diabetología*, 26(2), 95-100. [https://doi.org/10.1016/S1134-3230\(10\)62005-4](https://doi.org/10.1016/S1134-3230(10)62005-4)



# VII Congreso Chileno de Salud Pública X Congreso Chileno de Epidemiología



## ORGANIZAN



## AUSPICIAN

