

PROPUESTA DE MONOGRAFÍA

Título del proyecto	Predicción de probabilidad de diabetes en pacientes	
Estudiante 1		
Nombres completos	Ana Estefanía Henao Restrepo	e-mail: aestefania.henao@udea.edu.co
		GitHub: https://github.com/AnaEstefaniaHenaoRestrepo/ Monograf-a-Especializaci-n-En-Anal-tica-Y-Ciencia-De-Datos-UdeA.git
Estudiante 2		
Nombres completos	Juan José Gil Hoyos	e-mail: jjose.gil@udea.edu.co
		GitHub: https://github.com/JuanJoseGilH/Seminario-Especializaci-n-en-Anal-tica-y-Ciencia-de-Datos-UdeA.git

1. Descripción del problema

La diabetes se está convirtiendo en una enfermedad cada vez más común. Existe evidencia científica abundante que establece que los casos de diabetes están incrementando dramáticamente en el mundo. Tanto así, que se la ha catalogado como una epidemia para la humanidad por parte del foro económico mundial durante el año 2020 [1] [2].

Factores como el consumo excesivo de alcohol, una dieta desbalanceada, poca actividad física y fumar, tienen una relación estrecha con la probabilidad de contraer diabetes tipo 2 [1].

¿Qué implicaciones tiene para la salud en general la diabetes?

- Si tiene diabetes es mucho más probable que desarrolle una enfermedad cardíaca o derrame cerebral que aquellas que no la padecen [3].
- La asociación americana de enfermedades cardiovasculares, (AHA por sus siglas en inglés), establece que la diabetes disminuye los niveles de colesterol HDL (también conocido como colesterol bueno), aumenta los niveles de triglicéridos y los niveles de colesterol LDL (también conocido como colesterol malo). Estos últimos incrementan el riesgo de padecer enfermedades cardiovasculares y derrames cerebrales [4].

Pero la diabetes no solo ataca la salud física sino también tiene grandes consecuencias sobre la salud mental:

- Hay medicinas que pueden alterar las células beta o alterar el funcionamiento de la insulina, como los medicamentos para enfermedades psiquiátricas [5].

- Los enfermos con esquizofrenia tienen de 2 a 4 veces más riesgo de presentar diabetes tipo 2. Los antecedentes familiares de diabetes son más comunes en pacientes con esquizofrenia [6].
- Las personas con diabetes tienen entre 2 y 3 veces más probabilidades de presentar depresión que las personas sin diabetes [6].

Se pretende hacer uso del dataset extraído de la plataforma Kaggle para predecir si un paciente padece o no diabetes, basados en atributos relacionados con hábitos de alimentación y condición de salud.

2. Descripción del dataset

Esta fuente de datos proviene de la limpieza de los resultados de una encuesta realizada por *Centers for Disease Control and Prevention (CDC)* por sus siglas en inglés, cuyo nombre fue *Behavioral Risk Factor Surveillance System*, la cual inicialmente contenía 441.456 registros y fue llevada a cabo durante el año 2015 [7]. Esta información se adaptó para utilizarse en una de las competiciones que se generan dentro de la herramienta Kaggle. Su lanzamiento dentro de esta plataforma se produjo en noviembre de 2022.

El dataset consta de tres archivos en formato csv: **el primero** de estos corresponde a los datos que servirán como insumo para entrenar el modelo predictivo que se proponga, **el segundo** servirá para poner a prueba el modelo y **finalmente**, un tercer archivo de prueba para llevar a cabo el cargue de las respuestas del reto (submission file). Dentro del dataset, existen 18 columnas, las cuales se dividen en 17 variables características y una sola variable de respuesta. El archivo que permitirá entrenar el modelo tiene un total de 80.692 registros, el archivo de puesta en marcha tiene 20.000 registros, al igual que el archivo de cargue de respuestas.

El objetivo principal de nuestra variable de respuesta es determinar si un paciente tiene o no diabetes a partir de los datos dados. El link para consultar la fuente de datos es el siguiente: <https://acortar.link/c7DNzC> [8].

3. Métricas de desempeño

Como **métricas de desempeño**, para este caso de estudio la mejor opción es el F1 score, sin embargo, hay que recordar que es necesario también analizar la matriz de confusión. Se selecciona un F1 score aproximado de 75% como mínimo. Finalmente, se propone utilizar como métrica de validación la pérdida logarítmica, la cual se calculará una vez se tenga alguna predicción del modelo propuesto.

4. Criterio de desempeño

Como **métrica de negocio**, se propone hacer uso del modelo predictivo que se desarrolle, para que, a partir de un conjunto de datos recolectados de un paciente sea posible ofrecerle al médico tratante una herramienta de apoyo al diagnóstico temprano de diabetes.

5. Referencias

[1] O. Folorunso y O. Oguntibeju. "The Role of Nutrition in the Management of Diabetes Mellitus". IntechOpen - Open Science Open Minds | IntechOpen. <https://www.intechopen.com/chapters/42086> (accedido el 15 de marzo de 2023).

[2] Centro Nacional para la Prevención de Enfermedades Crónicas y Promoción de la Salud, División de Diabetes Aplicada. "La diabetes y su corazón". Centers for Disease Control and Prevention. <https://www.cdc.gov/diabetes/spanish/resources/features/diabetes-and-heart.html#:~:text=Modo%20en%20la%20diabetes%20afecta%20el%20corazón&text=La%20presión%20arterial%20alta%20aumenta,el%20riesgo%20de%20enfermedad%20cardiaca> (accedido el 15 de marzo de 2023).

[3] C. M. Story. "A Guide to Living with Diabetes and High Cholesterol". Healthline. [https://www.healthline.com/health/high-cholesterol/treating-with-statins/guide-to-diabetes-and-high-cholesterol#:~:text=Diabetes%20and%20high%20cholesterol%20often%20occur%20together&text=The%20American%20Heart%20Association%20\(AHA,for%20heart%20disease%20and%20stroke](https://www.healthline.com/health/high-cholesterol/treating-with-statins/guide-to-diabetes-and-high-cholesterol#:~:text=Diabetes%20and%20high%20cholesterol%20often%20occur%20together&text=The%20American%20Heart%20Association%20(AHA,for%20heart%20disease%20and%20stroke) (accedido el 15 de marzo de 2023).

[4] Instituto Nacional de la Diabetes y las Enfermedades Digestivas y Renales. "Síntomas y causas de la diabetes - NIDDK". National Institute of Diabetes and Digestive and Kidney Diseases. <https://www.niddk.nih.gov/health-information/informacion-de-la-salud/diabetes/informacion-general/sintomas-causas> (accedido el 15 de marzo de 2023).

[5] S. B. YP. "DIABETES, TRASTORNOS PSIQUIÁTRICOS E INTERACCIONES ENTRE AMBAS ENTIDADES". Bienvenidos a siicsalud. <https://www.siicsalud.com/dato/resiiccompleto.php/130807#:~:text=Los%20enfermos%20con%20esquizofrenia%20tienen,en%20los%20pacientes%20con%20esquizofrenia> (accedido el 15 de marzo de 2023).

[6] Centro Nacional para la Prevención de Enfermedades Crónicas y Promoción de la Salud, División de Diabetes Aplicada. "La diabetes y la salud mental". Centers for Disease Control and Prevention. <https://www.cdc.gov/diabetes/spanish/living/mental-health.html> (accedido el 15 de marzo de 2023).

[7] National Center for Chronic Disease Prevention and Health Promotion, Division of Population Health. "CDC - 2015 BRFSS Survey Data and Documentation". Centers for Disease Control and Prevention. https://www.cdc.gov/brfss/annual_data/annual_2015.html (accedido el 15 de marzo de 2023).

[8] TFUG Chandigarh. "Diabetes Prediction Competition(TFUG Chd Nov 2022) | Kaggle". Kaggle: Your Machine Learning and Data Science Community. <https://www.kaggle.com/competitions/diabetes-prediction-competitiontfug-chd-nov-2022/data> (accedido el 15 de marzo de 2023).