

Aplicación web para la predicción de diabetes tipo I y II utilizando aprendizaje automático

Trabajo Terminal No.

Alumnos: * Jaimes Mackay Carlos Alejandro, Robles Sosa Paula Guadalupe

Directores: Hernández Rubio Erika, Cruz Cortes Nareli

e-mail: alexjaimes38@gmail.com, probleml500@alumno.ipn.mx

Resumen – En el presente trabajo se tiene como objetivo desarrollar una aplicación web en .NET utilizando algoritmos de aprendizaje automático con base a datos específicos proporcionados por el INEGI. Conforme a datos arrojados del INEGI sobre las principales defunciones en México, se ha detectado que la diabetes es la tercera causa principal de muerte y se propone crear una herramienta que sea de apoyo en la predicción de diabetes tipo I y tipo II para beneficiar al paciente sin que tenga que acudir físicamente a un hospital o laboratorio.

Palabras clave – Aprendizaje automático, Diabetes, Diagnóstico, Predicción

1. Introducción

La diabetes es una enfermedad crónica, la cual se caracteriza por niveles elevados de glucosa en sangre y por muy poca producción de insulina o cuando esta no puede utilizarse de manera eficaz. Existen cuatro tipos de diabetes; Diabetes mellitus tipo 1, diabetes mellitus tipo 2, diabetes tipo LADA y la diabetes gestacional, sin embargo, la diabetes mellitus tipo 2 es la más común pues según la fundación de diabetes abarca entre el 85% y el 90% del total de los casos. Cualquier tipo de diabetes de no ser tratada ni controlada puede producir complicaciones en muchas partes del cuerpo y elevar el riesgo de una muerte prematura [1].

El número de personas que desarrolla la diabetes tipo 1 va en aumento cada año, esta suele aparecer en la infancia o juventud, y puede desarrollarse durante el desarrollo del feto en el útero, la alimentación durante los primeros años de vida, etc. En cuanto a la diabetes tipo 2 se presenta en su mayoría a partir de los cuarenta años y a las personas que padecen obesidad u otras enfermedades crónicas [2].

Los algoritmos de aprendizaje automático son fragmentos de código, que ayudan a analizar datos y a buscar un significado. El objetivo es detectar patrones para que puedan ser utilizados para hacer predicciones o clasificar información [3].

Basándonos en las cifras del año 2020 que se han arrojado en los últimos años por el INEGI, en México, la diabetes es la tercera causa principal de muerte. Si la diabetes es diagnosticada a tiempo, se puede ayudar a las personas a gestionar su enfermedad [4].

En este sentido, el aprendizaje automático actúa de forma en que no necesita explícitamente ser programada y en realidad trabaja con los datos que se le exponen, utilizando dos tipos de algoritmos: aprendizaje supervisado y aprendizaje no supervisado. El modelo de aprendizaje supervisado es aquel en los que se aprenden funciones, relaciones que asocian entradas con salidas, por lo que se ajustan a un conjunto de ejemplos de los que conocemos la relación entre la entrada y la salida deseada. En el caso del modelo de aprendizaje no supervisado, son aquellos en los que no estamos interesados en ajustar pares (entrada, salida), sino en aumentar el conocimiento estructural de los datos disponibles (y posibles datos futuros que provengan del mismo fenómeno), por ejemplo, dando una agrupación de los datos según su similitud (clustering), simplificando la estructura de los mismos manteniendo sus características fundamentales (como en los procesos de reducción de la dimensionalidad), o extrayendo la estructura interna con la que se distribuyen los datos en su espacio original (aprendizaje topológico) [5].

A continuación, se mencionan tres algoritmos relacionados con el aprendizaje automático. Las Máquinas de Vector Soporte se fundamentan en el Maximal Margin Classifier, que, a su vez, se basa en el concepto de hiperplano. Consiste en seleccionar como clasificador óptimo al que se conoce como maximal margin hyperplane o hiperplano óptimo de separación, que se corresponde con el hiperplano que se encuentra más alejado de todas las observaciones de entrenamiento. Para obtenerlo, se tiene que calcular la distancia perpendicular de cada observación a un determinado hiperplano. La menor de estas distancias (conocida como margen) determina como de alejado está el hiperplano de las observaciones de entrenamiento. El maximal margin hyperplane se define como el hiperplano que consigue un mayor margen, es decir, que la distancia mínima entre el hiperplano y las observaciones es lo más grande posible [11].

Algoritmos Naive Bayes, en un sentido amplio, los modelos de Naive Bayes son una clase especial de algoritmos de clasificación de Aprendizaje Automático, o Machine Learning. Se basan en una técnica de clasificación estadística llamada “teorema de Bayes”.

En ellos se asume que las variables predictoras son independientes entre sí. En otras palabras, que la presencia de una cierta característica en un conjunto de datos no está en absoluto relacionada con la presencia de cualquier otra característica [12].

Un árbol de decisión en Machine Learning es una estructura de árbol similar a un diagrama de flujo donde un nodo interno representa una característica (o atributo), la rama representa una regla de decisión y cada nodo hoja representa el resultado. Es una visualización como un diagrama de flujo que imita fácilmente el pensamiento a nivel humano. Es por eso por lo que los árboles de decisión son fáciles de entender e interpretar [13].

1.1.Estado del arte

En la tabla 1 se muestran diferentes tipos de trabajos relacionados con el pre-diagnóstico de diabetes.

Título del proyecto	Tipo	Características	Autor	País	Publicación	Año	Costo
Sistema de Pre-diagnóstico de diabetes mellitus mediante memorias asociativas [6].	Trabajo terminal	Sistema capaz de proporcionar pre-diagnóstico de diabetes mellitus tipo 2, mediante un módulo que implemente una memoria asociativa CHAT. Obtención de pre-diagnósticos a través de una aplicación móvil Módulo de seguimiento de enfermedad en la web	López González David García Maya Edgar	México	Trabajo terminal No: 2012-A042	2014	NA
Sistema Bayesiano para la Predicción de la Diabetes [7].	Artículo científico	Propuesta de un sistema de clasificación Bayesiano para la identificación temprana de la diabetes Mellitus con base en el análisis de algunas variables tales como número de embarazos, presión arterial diastólica, espesor cutáneo del tríceps, índice de masa corporal, herencia y edad	Omar D. Castrillón, William Sarache y Eduardo Castaño		NA	2017	
mySugr [8].	Aplicación móvil	<ul style="list-style-type: none"> • Aplicación dirigida a pacientes tipo 1 y tipo 2 • Monitoriza la dieta • Gestiona la dosis de insulina • Mantiene un control de la enfermedad Disponible en IOS y Android	mySugr GmbH			2018	MX\$11.00 - MX\$629.00
Diabetes:M Management & Blood Sugar Tracker App [8].		Permite hacer un seguimiento del tratamiento de la diabetes, ya que puede analizar los datos importados de distintos glucómetros y bombas de insulina	Sirma Medical Systems			2013	MX\$55.00 - MX\$999.00

Tabla 1. Trabajos relacionados con el pre-diagnóstico de diabetes.

2. Objetivo

Desarrollar una aplicación web para la predicción de diabetes tipo I y II mediante el uso de algoritmos de aprendizaje automático de clasificación.

2.1. Objetivos específicos

- Preparar y limpiar el banco de datos.
- Diseñar la interfaz web.
- Implementar algoritmos de aprendizaje automático para el desarrollo de la aplicación web.
- Realizar pruebas de rendimiento.

3. Justificación

Llevar a cabo el proceso de diagnóstico de diabetes tipo I y tipo II suele realizarse en consultorios médicos u hospitales en los que el seguimiento es laborioso y tardado. Con esta propuesta, el proceso de detección y diagnóstico será facilitado con el uso de algoritmos de aprendizaje automático como son naive bayes, árbol de decisiones y máquina de vectores reduciendo así el margen de error, estos algoritmos facilitan la predicción y evaluación de datos con base en la comparación y selección de información en un banco de datos.

Se desarrollará un marco utilizando información basada en bases de datos de Salud para anticipar si el paciente presenta síntomas de diabetes o no, proporcionando un diagnóstico básico para anticipar el nivel de amenaza con una mayor exactitud disminuyendo costos de personal, facilitando el direccionamiento del paciente hacia el médico y agilizando el proceso de tratamiento.

4. Productos o Resultados esperados

Gestión de información: Recolección de datos del paciente y organización de fórmulas comparativas.

Procesamiento de información: Comparativo con bases de datos de salud, para determinar estado de salud del paciente.

Diagnostico tentativo: Otorgar un pre-diagnóstico para su posterior revisión con un especialista.

Procedimiento propuesto esta sintetizado en el siguiente diagrama. La figura 1 nos muestra el flujo de investigación que se realizará para la construcción del modelo.

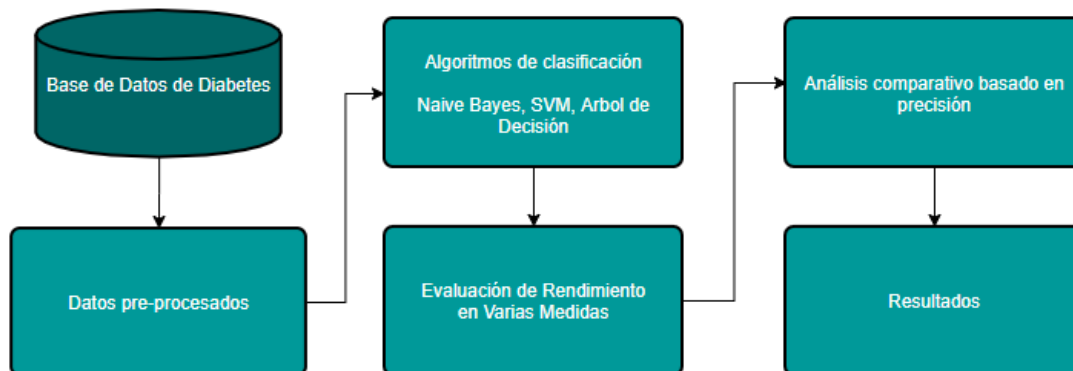


Figura 1. Diagrama de modelo propuesto. Fuente: Creación propia

Los productos esperados al finalizar el Trabajo Terminal son:

- La aplicación web de prediagnóstico.
- Manual de usuario.
- Manual Técnico.
- Documentación de todo el sistema.

5. Metodología

Se propone combinar dos metodologías debido a la limpieza y procesamiento de datos que se pueda requerir; se tratará de incorporar ciertas fases del modelo CRISP-DM en los niveles del modelo V.

CRISP-DM

La metodología Cross Industry Standard Process for Data Mining (CRISP-DM), es una metodología para el proceso de minería de datos, en la que permite movimiento hacia delante y hacia atrás entre diferentes fases, de las cuales nos interesa la fase de preparación de datos que es donde se seleccionan, limpian e integran los datos [9].

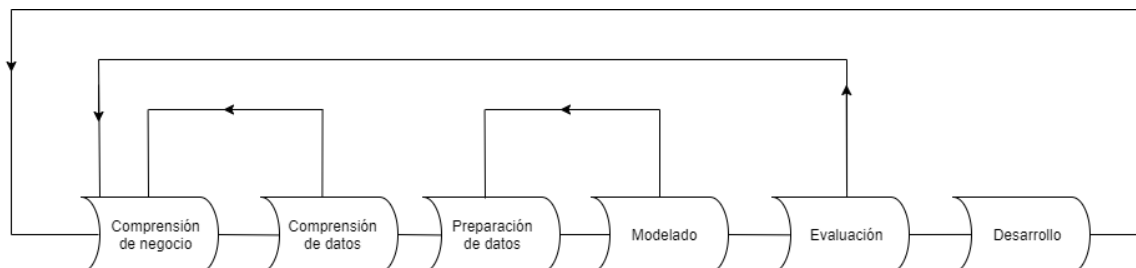


Figura 2. Metodología CRISP-DM [14].

El Modelo en V

El Modelo V, brinda las características que se necesitan para trabajar el proyecto, donde se aprecia la relación entre las acciones para el aseguramiento de la calidad y aquellas asociadas con la comunicación, modelado y construcción temprana [10].

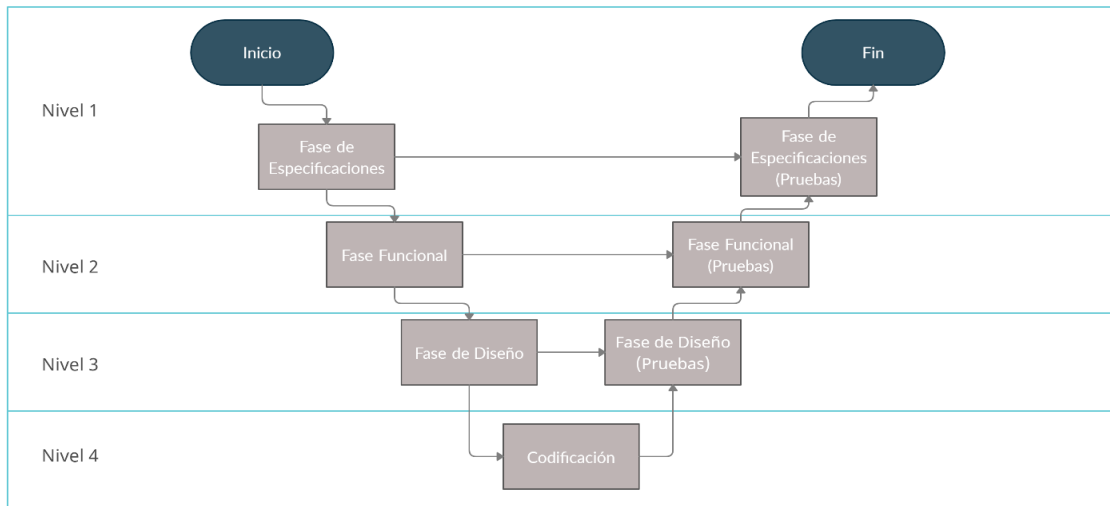


Figura 2.1 Modelo en V

Por lo cual, para el desarrollo de la plataforma web se propone combinar la metodología CRISP-DM y el modelo en V, agregando la fase “Preparación de datos”, donde se implementará la limpieza y proceso de datos, de la metodología CRISP-DM en el modelo en V.

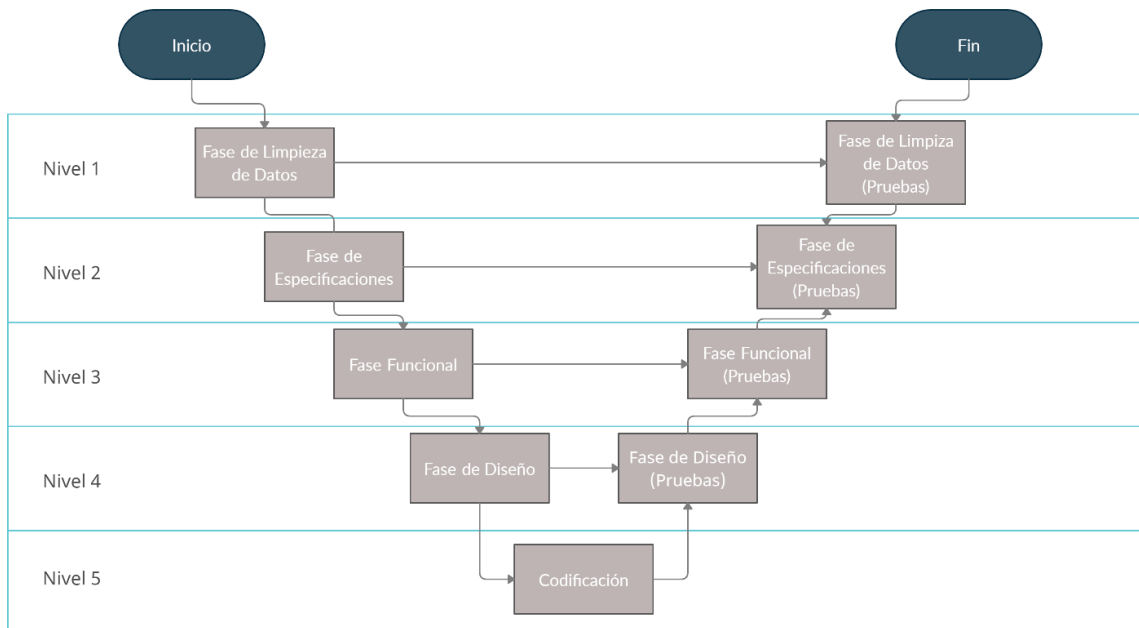


Figura 2.2 Modelo en V. Fuente: Creación propia

6. Cronograma

Nombre del alumno: Jaimes Mackay Carlos Alejandro

Título de TT: Aplicación web de aprendizaje automático para la predicción de diabetes tipo I y II

[illegible]

Nombre del alumno: Robles Sosa Paula Guadalupe

Título de TT: Aplicación web de aprendizaje automático para la predicción de diabetes tipo I y II

[illegible]

7. Referencias

- [1] Diabetes, uno de los principales problemas de salud en México. (2019, 14 noviembre). Gobierno de la Ciudad de México. Recuperado 31 de marzo de 2021, <https://www.gob.mx/issste/es/articulos/diabetes-uno-de-los-principales-problemas-de-salud-en-mexico?idiom=es>
- [2] Maset, D. M. J. D. C., & Maset, D. M. J. D. C. (2020, 5 mayo). ¿Qué tipos de Diabetes existen? - CinfaSalud. Cinfasalud. Recuperado 31 de marzo de 2021, <https://cinfasalud.cinfa.com/p/diabetes/>
- [3] Algoritmos de aprendizaje automático. (s. f.). Microsoft Azure. Recuperado 31 de marzo de 2021, <https://azure.microsoft.com/es-mx/overview/machine-learning-algorithms/>
- [4] Geografía, E. D. N. I. Y. (s. f.). Prevalencia de Obesidad, Hipertensión y Diabetes para los Municipios de México 2018. INEGI. Recuperado 31 de marzo de 2021, de <https://www.inegi.org.mx/investigacion/pohd/2018/>
- [5] Caparrini, F. S. (2020, 14 diciembre). Aprendizaje Supervisado y No Supervisado - Fernando Sancho Caparrini. CS. Recuperado 15 de abril de 2021, <http://www.cs.us.es/%7Efsancho/?e=77>
- [6] López González, D (2014). *Sistema de Pre-diagnóstico de diabetes mellitus mediante memorias asociativas (Licenciatura)*. Recuperado 02 de abril de 2021, Escuela Superior de Cómputo del Instituto Politécnico Nacional, Ciudad de México.
- [7] Castrillón, O., Sarache, W., & Castaño, E. (2017, diciembre). *Sistema Bayesiano para la Predicción de la Diabetes*. Scielo. https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0718-07642017000600017
- [8] Cuáles son las mejores apps de diabetes. (2020, 31 agosto). Campus Sanofi. <https://campussanofi.es/smart-care/noticias/cuales-son-las-mejores-apps-de-diabetes/>
- [9] CRISP-DM. (2021, 29 Marzo). Data Science Project Management. <https://www.datascience-pm.com/crisp-dm-2/>
- [10] Metodología de desarrollo de software - El Método en V: http://www.ingenieriadesoftware.mex.tl/61885_Modelo-V.html
- [11] Sisodia, D., Singh, L., Sisodia, S., 2014. Fast and Accurate Face Recognition Using SVM and DCT, in: Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012, Springer. pp. 1027–1038.

[12] Ray, S., 2017. 6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python).

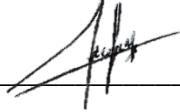
[13] Iyer, A., S, J., Sumbaly, R., 2015. Diagnosis of Diabetes Using Classification Mining Techniques. International Journal of Data Mining & Knowledge Management Process 5, 1–14. doi:10.5121/ijdkp.2015.5101, arXiv:1502.03774.

[14] Acotzi Santillán, F (2017). “*Sistema para la detección de innovación en empresas a través de minería de datos*” Recuperado 15 de abril 2021, Escuela Superior de Cómputo del Instituto Politécnico Nacional, Ciudad de México. pp. 21.

8. Alumnos y Directores

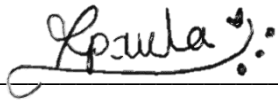
Carlos Alejandro Jaimes Mackay. -
Alumno de la carrera de Ingeniería en
Sistemas Computacionales en la
Escuela Superior de Cómputo del
Instituto Politécnico Nacional, Boleta:
2016630467, Tel. 55 4481 5866,
email: alexjaimes38@gmail.com

Firma:



Paula Guadalupe Robles Sosa. -
Alumna de la carrera de Ingeniería en
Sistemas Computacionales en la
Escuela Superior de Cómputo del
Instituto Politécnico Nacional, Boleta:
2016601461, Tel. 55 6006 9097,
email: proble1500@alumno.ipn.mx

Firma:



Hernández Rubio Erika. - Maestra en
Ciencias de la Computación Instituto
Politécnico Nacional - Centro de
Investigación en Computación.
Licenciada en Ciencias de la
Informática Instituto Politécnico
Nacional – Unidad Profesional
Interdisciplinaria de Ingeniería y
Ciencias Sociales y Administrativas
(UPIICSA)

Áreas de interés y líneas de
investigación:

Bases de Datos, Inteligencia Artificial,
Cómputo Móvil.

Teléfono: (55) 5729-6000 ext. 52038
Correo electrónico:
ehernandezru@ipn.mx.

Firma:

Cruz Cortés Nareli. - Doctora en
Ciencias de la Ingeniería Eléctrica,
Sección de Computación del
Departamento de Ingeniería Eléctrica
del CINVESTAV-IPN

Áreas de interés y líneas de
investigación:

Aprendizaje automático,
Ciberseguridad

Teléfono: (55) 5729-6000 ext: 56574
Correo electrónico: ncruzc@ipn.mx

Firma:

CARÁCTER: Confidencial
FUNDAMENTO LEGAL: Artículo 11 Fracc. V
y Artículos 108, 113 y 117 de la Ley Federal de
Transparencia y Acceso a la Información
Pública.
PARTES CONFIDENCIALES: Número de

Acuses

Re: Protocolo - Acuse de recibido



Erika Hernandez Rubio
To Paula Guadalupe Robles Sosa

[↩ Reply](#) [↩ Reply All](#) [→ Forward](#) [📧](#) [...](#)

martes 09/11/2021 08:09 a. m.

[🌐 Translate message to: English](#) | [Never translate from: Spanish](#) | [Translation preferences](#)

Estoy de acuerdo con el protocolo. Acuso de recibido.
Saludos.

M. en C. Erika Hernández Rubio.
Profesora de la SEPI-ESCOM
ehernandezru@ipn.mx
57296000 ext. 52038

De: Paula Guadalupe Robles Sosa <probless1500@alumno.ipn.mx>
Enviado: lunes, 8 de noviembre de 2021 9:48 p. m.
Para: Nareli Cruz Cortes <ncruz@ipn.mx>; Erika Hernandez Rubio <ehernandezru@ipn.mx>
CC: Carlos Alejandro Jaimes Mackay <cjaimesm2100tmp@alumnoguinda.mx>
Asunto: Protocolo - Acuse de recibido

Buenas noches,

Como parte del proceso de registro de protocolo, se solicita su firma o acuse por lo que les envío el documento el cual deberá ser firmado o en caso contrario responder este correo como un acuse de confirmación.

Desde ya muchas gracias.
Saludos,
Paula Robles

Re: Protocolo - Acuse de recibido



Nareli Cruz Cortes
To Paula Guadalupe Robles Sosa; Erika Hernandez Rubio
Cc Carlos Alejandro Jaimes Mackay

[↩ Reply](#) [↩ Reply All](#) [→ Forward](#) [📧](#) [...](#)

martes 09/11/2021 08:53 a. m.

[🌐 Translate message to: English](#) | [Never translate from: Spanish](#) | [Translation preferences](#)

Estimada Paula, confirmo que recibí el documento de protocolo.
Saludos cordiales
Nareli Cruz Cortés

From: Paula Guadalupe Robles Sosa <probless1500@alumno.ipn.mx>
Sent: 08 November 2021 21:48
To: Nareli Cruz Cortes <ncruz@ipn.mx>; Erika Hernandez Rubio <ehernandezru@ipn.mx>
Cc: Carlos Alejandro Jaimes Mackay <cjaimesm2100tmp@alumnoguinda.mx>
Subject: Protocolo - Acuse de recibido

Buenas noches,

Como parte del proceso de registro de protocolo, se solicita su firma o acuse por lo que les envío el documento el cual deberá ser firmado o en caso contrario responder este correo como un acuse de confirmación.

Desde ya muchas gracias.
Saludos,
Paula Robles