

Sistema gestor de pacientes con modelos de predicción de COVID-19 en México

Trabajo Terminal No. _____ - _____

*Alumnos: Bernal Leocadio Josué Eduardo, *Cruz Pérez Raúl Eduardo, Pérez Garduño José Emiliano.*

Directores: Dr. Rosas Trigueros Jorge Luis, M. en C. González Albarrán Gisela.

***email: rcruzp1401@alumno.ipn.mx**

Resumen – Crear una aplicación web responsiva para computadoras de escritorio y dispositivos móviles que ayude en la gestión de una clínica a través del registro completo de un paciente, registro de un diagnóstico, así como el registro del historial clínico del paciente para poder llevar un mejor control de posibles infectados y el tratamiento que se llevará a cabo.

Implementaremos un módulo de uso exclusivo para médicos, el cual implementará modelos de predicción de SARS-CoV-2 en México a través de técnicas de aprendizaje automático (Machine learning), esto como un medio de apoyo para el diagnóstico.

Palabras clave – COVID-19, SARS-CoV-2, modelos de predicción, Machine Learning, sistema gestor de pacientes, aplicación web.

1. Introducción

En la actualidad los sistemas informáticos tienen un papel muy relevante en la vida diaria de la sociedad en general, ya que, gracias a ellos es posible automatizar y mejorar procesos de una entidad en un tiempo menor, esto se logra a través del almacenamiento, procesamiento y transmisión de datos.

Por ello crear un sistema informático en el área de salud supone una mejora sustancial en la gestión de los recursos, que en la mayoría de las veces son reducidos, estos recursos son tales como materias primas y capital de trabajo.

Debemos de entender como sistema a un conjunto de elementos ya sean software o hardware que cooperan entre sí para llevar a cabo determinadas tareas en específico.

Las primeras entidades en utilizar los sistemas de información digitales como medios de producción fueron los bancos, ya que ellos vieron en dichos sistemas una forma de adecuar y automatizar sus procesos, esta gestión que anteriormente la hacía una persona fue delegada a un nuevo agente informático, básicamente una persona especializada en programar dicho agente se encargó de programar las reglas de operación de la entidad en cuestión.

El concepto de digitalización se ha inmiscuido demasiado en todos los ámbitos del individuo, en una primera instancia fue a través de las empresas que empezaron a ofrecer poco a poco sus servicios a través de canales de distribución digitalizados, ya que esto presentaba una nueva área de oportunidad en la cual un consumidor podría llevar a cabo transacciones sin acudir a un lugar físico.

Es por eso que planteamos aplicar estas nuevas tecnologías a un problema actual, la pandemia del COVID-19, la cual logró cambiar por completo la forma en la que se vivía el día a día en todo el mundo, afectando mayormente al sector médico debido al poco conocimiento que se tiene de la enfermedad y los mitos que la rodean actualmente, esto junto a las complicaciones que tiene el mantenerse a salvo de un posible contagio complicaron el conseguir resultados concretos de posibles infecciones y exposiciones al virus SARS-Cov-2.

Nuestro proyecto busca facilitar la forma en la que es detectada en etapas tempranas la enfermedad COVID-19 aplicando tecnología de Machine Learning para poder predecir de acuerdo a modelos de predicción la posibilidad de que un paciente posiblemente infectado pueda dar positivo a la hora de realizarle la prueba viral y la facilidad de gestionar a estos posibles infectados para así disminuir las posibles infecciones que se puedan llegar a dar por una persona a la que no se le han entregado los resultados de la prueba lo suficientemente rápido o desconozca la condición en la que se pueda encontrar.

Actualmente hay varios modelos sugeridos para realizar un análisis similar, como es la propuesta [1] que utiliza inteligencia artificial para identificar los procesos biológicos que suceden en un paciente diagnosticado de COVID-19, sin embargo se enfoca más en la manera que responde el virus dentro del cuerpo y no como detectarlo con anticipación. Finalmente encontramos un modelo de predicción enfocado a posibles pacientes positivos de SARS-Cov-2, el cual utilizando datos recopilados de diferentes bases de datos abiertas pertenecientes a hospitales con pacientes afectados lograron entrenar modelos de Machine Learning y lograr encontrar los infectados en base a predicciones matemáticas.

“Las características del conjunto de datos está estructurado como sigue: información básica con respecto a edad, género, residencia habitual, trabajo e historia de contacto Wuhan/Hubei; la información de trayectoria es el tiempo, lugar, transportación y eventos hasta el 20 de Febrero del 2020. Extrajimos características importantes de información básica (edad, género), síntomas (fiebre, tos, cansancio muscular), resultados de diagnóstico (infección de pulmones, imágenes radiográficas), historial de enfermedades previas (neumonía, diarrea, rinorrea) e información de trayectorias (estado del tratamiento aislado, historial de viajes) que son directamente o indirectamente relacionados a la enfermedad COVID-19 ” [2]

Software	Características	Precio en el mercado
Artificial intelligence and machine learning to fight COVID-19	Modelo de redes neuronales junto a Machine Learning para poder determinar un posible tratamiento, prevención de contagios y operaciones de hospitales para atacar la pandemia de COVID-19	N/A
Solución Propuesta	Modelo de Machine Learning para detectar de manera temprana posibles pacientes infectados de COVID-19 para evitar futuros contagios y ofrecer un tratamiento más eficaz y conciso.	N/A

Tabla 1. Resumen de productos similares

Basándonos en los resultados encontrados, podemos enfocarnos en las variables de importancia como los síntomas y el grupo perteneciente al posible infectado y así predecir con un alto grado de consistencia los casos en los que puedan resultar positivos en la prueba de infección viral.

2. Objetivo

Objetivo general

Desarrollar una aplicación web basada en machine learning y modelos de predicciones para ofrecer un servicio para gestionar pacientes posiblemente infectados de COVID-19 para fácil detección y temprano tratamiento de la enfermedad. La información de estos pacientes será extraída del sitio web Kaggle[3], el cual contiene miles de datos sobre personas infectadas y que estuvieron enfermas de Covid 19.

Objetivos Específicos

1. Desarrollar una interfaz donde el usuario ingrese los datos del paciente y un diagnóstico médico.
2. Implementar un modelo predictivo basado en Machine Learning para poder predecir el potencial número de infectados.
3. Generar un reporte fácil de interpretar para el usuario que explique la posibilidad de que alguno de sus pacientes se encuentre infectado.

3. Justificación

El reciente brote de SARS-CoV-2 creó una crisis sin precedentes, aunque lo más preocupante es que demostró que el sistema médico actual no tenía las herramientas necesarias para poder detectar con exactitud la cantidad de posibles infectados que pudiesen llegar a tener síntomas de COVID-19, implicando una complicación a la hora de detectar casos positivos debido al tiempo de espera de las pruebas comunes, los posibles falsos positivos que llegasen a ocurrir en ellas y los casos de pacientes asintomáticos.

Las pruebas actuales se dividen en dos categorías de acuerdo al centro para el control y la prevención de enfermedades [2], que son pruebas virales y pruebas de anticuerpos, sin embargo ambas son realizadas una vez el paciente presenta síntomas más definidos o considera que ha sido expuesto a otra persona infectada por lo que considera apropiado realizar la prueba para determinar su salud física, esto junto al tiempo necesario para obtener los resultados de la prueba ralentizan el proceso lo suficiente como para que haya más posibles infectados sin darse cuenta de ello.

Utilizando machine learning para entrenar modelos de predicción con previos diagnósticos de pacientes infectados con COVID-19 se planea realizar varias iteraciones de diagnósticos realizados a pacientes infectados previos a conseguir los resultados de su prueba para así poder determinar de manera temprana la posibilidad de que realmente se encuentre afectado por la enfermedad.

Este trabajo busca que se agilicen las detecciones de COVID-19 en pacientes sintomáticos y que las mismas tengan un respaldo más acertado al basarse en un modelo de machine learning previamente entrenado, disminuyendo de esta manera los falsos positivos que puedan llegar a existir.

Hoy en día existen modelos teóricos que buscan la probabilidad de que un posible paciente infectado de COVID-19 pase de un estado estable a un estado crítico, aunque ninguno de estos modelos permite saber con exactitud los casos de posibles infectados a partir de un diagnóstico médico realizado, ya que funcionan basándose en imágenes del paciente infectado “*Para la detección del COVID-19, cuatro conjuntos de datos fueron formados tomando 150 imágenes de CT con tamaños de 16x16, 32x32, 48x48 y 64x64.*” [4] y radiografías realizadas a la zona pulmonar, casos en los que es posible que el afectado ya haya sido ingresado a la sala de urgencias.

La audiencia que se toma en cuenta para el rango de este trabajo son médicos en el área de COVID-19 y pacientes que presenten síntomas relacionados a la enfermedad.

En contraste con modelos previamente planteados por otros proyectos de investigación [5], el producto de este trabajo propone establecer una detección temprana de la enfermedad, no una vez se tiene un paciente infectado, para así prevenir posibles infecciones y poder notificar al usuario de la posible detección para llevar a cabo medidas preventivas en caso de ser un grupo de alto riesgo.[6]

Para alimentar el modelo de Machine Learning se planteó utilizar el repositorio de GitHub dedicado a recopilación de datos para COVID-19[8] el cual fue elegido debido a su veracidad respaldada por worldometers.

“Para los datos de COVID-19, recolectamos datos de reportes oficiales, directamente de canales de comunicación del gobierno o indirectamente, a través de fuentes de medios locales cuando se podían confirmar como confiables. Proveemos la fuente de cada conjunto de datos en la sección de “Actualizaciones más recientes”. Se realizan actualizaciones diariamente gracias a la participación de usuarios alrededor del mundo y la dedicación de un equipo de analistas e investigadores que validan datos de un conjunto infinitamente creciente de más de 5000 fuentes ”.[7, 8]

4. Productos o resultados esperados

Productos finales del sistema:

- Sistema web
- Modelo predictivo
- Reporte de resultados
- Manual de usuario

El producto utilizará la arquitectura descrita en el siguiente diagrama:

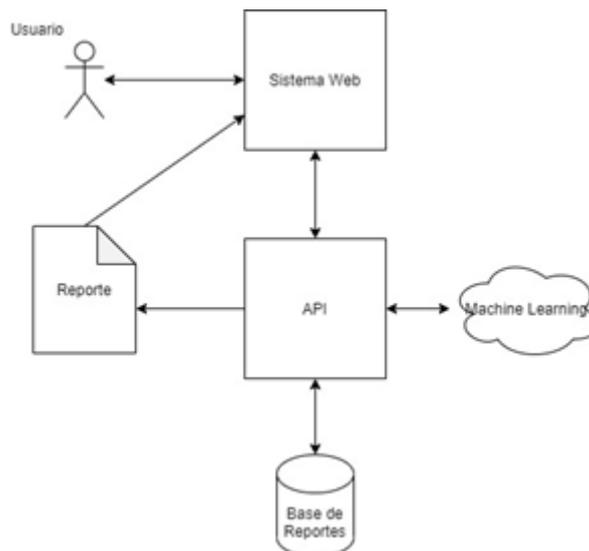


Imagen 4.1: Arquitectura del sistema web utilizando una API para alimentar al modelo de Machine Learning con los datos ingresados por el usuario y retornando los resultados para imprimirlos en un reporte desplegado en la interfaz de usuario.

5. Metodología

Se decidió utilizar la metodología ágil Scrum debido a que el tiempo de entrega es relativamente corto para el alcance del proyecto, por lo que dividiremos el desarrollo en Sprints de un mes aprovechando al máximo el tiempo disponible y adecuándose a posibles modificaciones que se llegasen a necesitar para así poder entrenar el modelo de predicción desde un principio.

De acuerdo con las lineaciones que ofrece Scrum en su manual de usuario, se especificarán tareas en un storyboard asignado por el Scrum Master al equipo de desarrollo y diariamente se realizarán reuniones de media hora para poder entablar conversación entre los desarrolladores e identificar los problemas que pudiesen llegar a existir en el desarrollo.

Los Sprints serán revisados al final de cada mes para poder obtener una retroalimentación efectiva y poder definir el siguiente Sprint atacando las debilidades encontradas durante el desarrollo previo y agilizar el progreso del proyecto, registrando los avances mediante la bitácora llamada Backlog y anotando las actividades concluidas para tener un registro de horas dispuestas en él.

Herramientas a utilizar para facilitar el desarrollo

Programación:

- Visual Studio Code: Editor de texto de código abierto desarrollado por Microsoft para Windows, Linux y MacOS.
- Django: Es un framework de código abierto escrito en python desarrollado para facilitar la creación de sitios web complejos.
- Node.js: Es un framework de código abierto, para la capa del servidor (pero no limitándose a ello) basado en el lenguaje de programación JavaScript.

Diseño:

- Bootstrap: Bootstrap es una biblioteca multiplataforma o conjunto de herramientas de código abierto para diseño de sitios y aplicaciones web.

Machine Learning:

- Tensor Flow: Tensor Flow es una biblioteca de código abierto para aprendizaje automático a través de un rango de tareas, redes neuronales, problemas análogos al aprendizaje y razonamiento usados por los humanos.

Repositorio:

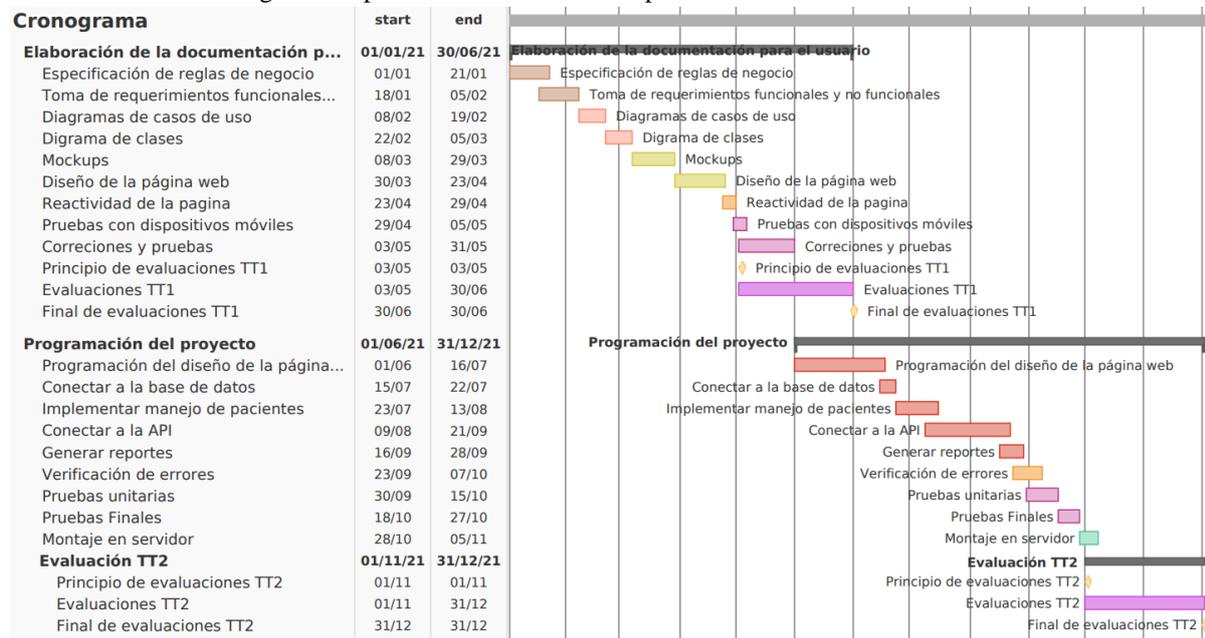
- GitLab: GitLab es un servicio web de control de versiones y desarrollo de software colaborativo basado en Git. Además de gestor de repositorios, el servicio ofrece también alojamiento de wikis y un sistema de seguimiento de errores.

6. Cronograma

Nombre del alumno(a): Pérez Garduño José Emiliano

TT No.:

Título del TT: Sistema gestor de pacientes con modelos de predicción de COVID-19 en México



Nombre del alumno(a): Cruz Pérez Raúl Eduardo

TT No.:

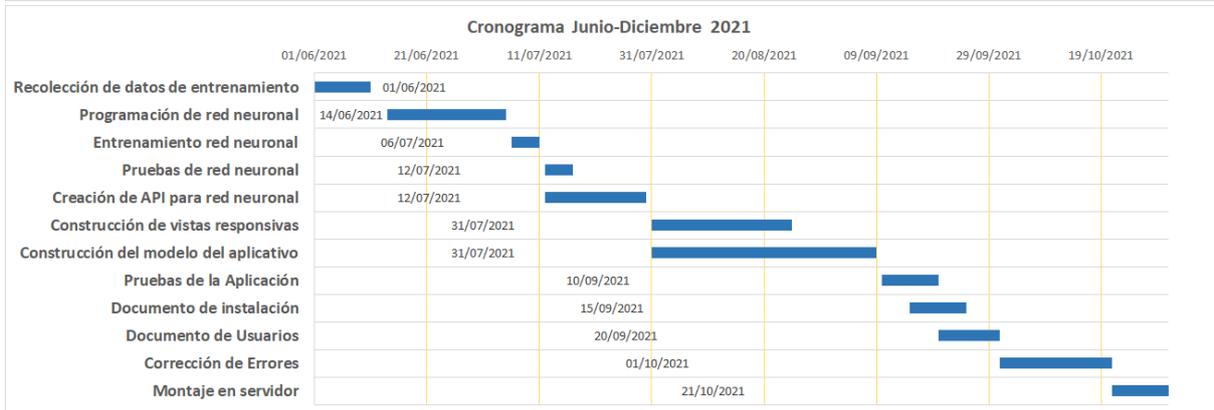
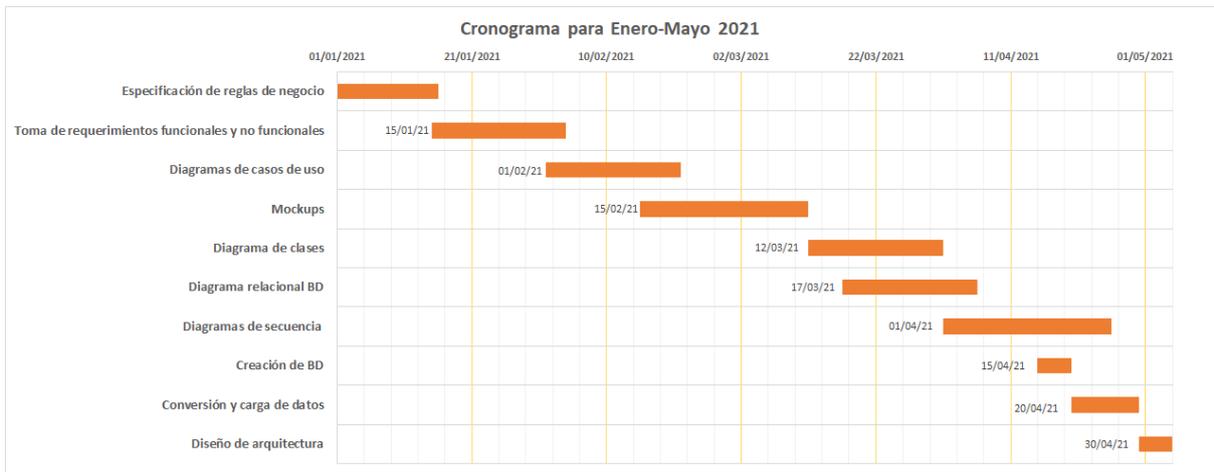
Título del TT: Sistema gestor de pacientes con modelos de predicción de COVID-19 en México

Cronograma	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Especificación de reglas de negocio	■											
Toma de requerimientos funcionales y no funcionales		■	■									
Diagramas de casos de uso				■								
Diagrama de clases					■							
Mockups						■						
Creación modelo relación BD							■					
Creación de BD y primeras pruebas de alta de información								■				
Programación del diseño de la página web									■			
Conectar a la base de datos										■		
Implementar manejo de pacientes											■	
Conectar a la API												■
Generar reportes												■
Verificación de errores												■
Pruebas unitarias												■
Pruebas Finales												■
Montaje en servidor												■

Nombre del alumno(a): Bernal Leocadio Josué Eduardo

TT No.:

Título del TT: Sistema gestor de pacientes con modelos de predicción de COVID-19 en México



7. Referencias

- [1] Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P., Joe, B., & Cheng, X. (2020). *Artificial intelligence and machine learning to fight COVID-19 | Physiological Genomics*. Journals.physiology.org. Retrieved 5 November 2020, from <https://journals.physiology.org/doi/full/10.1152/physiolgenomics.00029.2020>.
- [2] Enfermedad del coronavirus 2019 (COVID-19). Centers for Disease Control and Prevention. (2020). Retrieved 5 November 2020, from <https://espanol.cdc.gov/coronavirus/2019-ncov/testing/diagnostic-testing.html>.
- [3] *Kaggle: Your Machine Learning and Data Science Community*. (s. f.). Kaggle. <https://www.kaggle.com/>
- [4] Barstugan, M., Ozkaya, U. and Ozturk, S., 2020. *Coronavirus (COVID-19) Classification Using CT Images By Machine Learning Methods*. [online] arXiv.org. Available at: <<https://arxiv.org/abs/2003.09424>> [Accessed 5 November 2020].
- [5] BMJ 2020; 369 doi: <https://doi.org/10.1136/bmj.m1328> (Published 07 April 2020)
- [6] Zhao, Z., & Chen, A. (2020). Prediction model and risk scores of ICU admission and mortality in COVID-19. PLOS ONE. Retrieved 5 November 2020, from <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0236618>.
- [7] kp, D. (2020). *imdevskp/covid_19_jhu_data_web_scrap_and_cleaning*. GitHub. Retrieved 9 November 2020, from https://github.com/imdevskp/covid_19_jhu_data_web_scrap_and_cleaning.
- [8] *Worldometer - About us*. Worldometers.info. (2020). Retrieved 9 November 2020, from <https://www.worldometers.info/about/>.

8. Alumnos y Directores

Bernal Leocadio Josué Eduardo. - Alumno de la carrera de Ing. de Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta:2014360954 , Tel. 5548644190, email jbernal1000@alumno.ipn.mx

Cruz Pérez Raúl Eduardo. - Alumno de la carrera de Ing. de Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta:2015630105 , Tel.5529651840, email rcruzp1401@alumno.ipn.mx

Pérez Garduño José Emiliano. - Alumno de la carrera de Ing. de Sistemas Computacionales en ESCOM, Especialidad Sistema, Boleta:2016630300 , Tel.5580220614, email jperezg1510@alumno.ipn.mx

Rosas Trigueros Jorge Luis. - Dr. en Biotecnología por el IPN, (2012), M. en C. en Ing. Eléctrica por la Universidad de Texas A & M en College Station, Estados Unidos (2002), Ing. en Sistemas Computacionales por la Escuela Superior de Cómputo del IPN (1998). Actualmente es profesor Titular en ESCOM y sus áreas de interés son: *Sistemas Evolutivos, Bioinformática y Graficación.* Ext. 52073, correo electrónico: jlorosas@ipn.mx

González Albarran Gisela. - M. en C. de la Educación. Docente-Investigadora de la Academia de Ciencias Sociales. Del Departamento de Formación Integral e Institucional de la ESCOM. Temas de interés: Educación, Psicología, Ciencias Sociales, Género, Estrés, Educación y tecnología, Tutorías, Covid19; E-mail: gisgisela7@gmail.com, Tel. 57296000, ext. 52070, email: gisgisela7@gmail.com

 Eduardo bernal <leocadio624@gmail.com>
Lun 09/11/2020 04:36 PM
Para: Raul Eduardo Cruz Perez

Confirmando de recibido

 Raul Eduardo Cruz Perez
Lun 09/11/2020 07:07 PM
Para: Jorge Luis Rosas Trigueros
CC: Jose Emiliano Perez Garduno; leocadio624@gmail.com

Confirmando de recibido

Cruz Pérez Raúl Eduardo

 Jose Emiliano Perez Garduno
Lun 09/11/2020 04:39 PM
Para: Jorge Luis Rosas Trigueros; Raul Eduardo Cruz Perez
CC: leocadio624@gmail.com

Recibido

 **Jorge Luis Rosas Trigueros**
Lun 09/11/2020 06:46 PM
Para: Raul Eduardo Cruz Perez

Buenas noches.

Acuso de recibido.

Dr. Jorge Luis Rosas Trigueros

 **Gisela González Albarrán** <gisgisela7@gmail.com>
Lun 09/11/2020 06:48 PM
Para: Raul Eduardo Cruz Perez

Acuso de recibido.

M. en C. Gisela González Albarrán.

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 boleta y teléfono.