

**EVALUACION DE MODELO DE DATOS EN PACIENTES CON DIABETES
MELLITUS CONTAGIADOS DE COVID-19 EN COLOMBIA**

DIANA MILENA BELTRAN BONILLA



**MAESTRIA EN INGENIERIA Y ANALITICA DE DATOS (MIAD)
FACULTAD DE INGENIERIA
UNIVERSIDAD JORGE TADEO LOZANO
BOGOTA D.C.
2021**

**EVALUACION DE MODELO DE DATOS EN PACIENTES CON DIABETES
MELLITUS CONTAGIADOS DE COVID-19 EN COLOMBIA**

DIANA MILENA BELTRAN BONILLA

**Trabajo de grado presentado como requisito para optar al título de MASTER EN
INGENIERIA Y ANALITICA DE DATOS**



**MAESTRIA EN INGENIERIA Y ANALITICA DE DATOS (MIAD)
FACULTAD DE INGENIERIA
UNIVERSIDAD JORGE TADEO LOZANO
BOGOTA D.C.
2021**

Tabla de contenido

Tabla de contenido	3
1. INTRODUCCION.....	4
2. MARCO TEÓRICO	5
2.1. MODELADO DE DATOS.....	5
2.2. MACHINE LEARNING.....	5
2.2.1. APRENDIZAJE SUPERVISADO	5
2.2.2. APRENDIZAJE NO SUPERVISADO	5
2.2.3. APRENDIZAJE REFORZADO.....	6
2.3. ENFERMEDAD NO TRANSMISIBLE	6
2.4. DIABETES MELLITUS	6
2.5. COVID-19.....	6
2.6. INSULINA.....	7
2.7. HIPOGLICEMIA.....	7
2.8. ENFERMEDAD CRONICA	7
2.9. TASA DE MORTALIDAD	7
2.10. DANE.....	7
2.11. SISPRO.....	7
3. ESTADO DEL ARTE	8
4. PLANTEAMIENTO DEL PROBLEMA.....	11
5. OBJETIVOS	12
5.1. OBJETIVO GENERAL:.....	12
5.2. OBJETIVOS ESPECIFICOS:.....	12
6. METODOLOGIA	13
7. CRONOGRAMA	14
8. PRESUPUESTO.....	14
9. BIBLIOGRAFIA.....	15

1. INTRODUCCION

La estructuración del proyecto tiene como fin la evaluación del impacto que ha tenido la reciente pandemia del **SARS-CoV-2 (COVID-19)**, en los servicios hospitalarios y en la tasa de muertes para pacientes con enfermedad crónica como la **DIABETES MELLITUS**, para llegar a los resultados esperados los estudios se realizaran bajo los datos estadísticos de entidades como El Departamento Administrativo Nacional de Estadística (**DANE**), Sistema Integral de Información de la Protección Social (**SISPRO**), la **ORGANIZACIÓN MUNDIAL DE LA SALUD (OMS)**, entre otros.

Para contextualizar sobre el impacto de la **DIABETES MELLITUS** en el mundo, nos basamos en los datos de la **OMS**, que en su reporte del año **2014** indica para ese tiempo se alcanzaba los **422 millones** de personas adultas con Diabetes Mellitus a nivel mundial, siendo el **8,5%** de participación comparada con las otras enfermedades clasificadas como crónicas en la población adulta, adicional que es una de las comorbilidades con mayor crecimiento si es comparada con el año **1980** que tan solo alcanzaba los **108 millones** de personas con la enfermedad, y con una participación del **4,7%** sobre el total de las enfermedades existentes mundialmente para ese año[1].

Si retornamos a la actualidad, en donde nos enfrentamos a una pandemia como el **COVID-19**, y que con el tiempo está dejando retos al sector salud, mismo que a la fecha viene con un gran déficit de personal especialista y calificado, que a su vez viene siendo desmejorado y descuidado estatalmente por carencia de presupuesto y de condiciones laborales, además el cual en emergencias sanitarias como las que vivimos en la actualidad, debe tener el mismo nivel de respuesta para sus pacientes con comorbilidades es importante evaluarlo para tal vez bajo estos resultados evaluar mejoras y planes de acción preventivos.

En cuanto al servicio que se presta por las entidades de salud, puede que se evidencien algunas falencias como los son la falta unificación de Subredes y planeación de las organizaciones, sin embargo logran ser omitidas por altos directivos del sector y dar reportes de estabilidad e informes con servicios de calidad, tal como se indica en la **Boletín de Prensa No 848 de 2020**, donde el Ministro de Salud **Fernando Ruiz Gómez** donde enumera lo siguiente, *"Si no hubiéramos tenido un sistema de salud mixto público-privado integrado, un sistema con una capacidad de respuesta de aseguramiento y con una cobertura universal superior al 95% no hubiéramos podido responder"*, es decir , el sector salud esta con todo en orden[2].

Sin embargo se puede aclarar que los datos que se arrojan por estas entidades son tardíos, poco veraces y en muchas ocasiones se generan para ser más estudios correctivos y no preventivos. Debido a la falta de integración de las entidades públicas y privadas dueñas de la información y que complican los estudios bajo este sector primordial.

Por lo expuesto anteriormente, en este proyecto se realizara un estudio donde se estima cuantificar y calcular, las posibles alteraciones en el servicio prestado a pacientes diagnosticados con **DIABETES MELLITUS** e infectados con el virus del **COVID-19**, durante el primer año tras la llegada de la pandemia a nuestro país y pretende evaluar, si toda la crisis hospitalaria que ha generado la pandemia, ha ocasionado desmejoras en los procesos habituales para estos pacientes, en la asignación de medicamentos y si de forma directa ha afectado en los resultados de la letalidad, todo esto bajo modelos de aprendizaje No supervisados y supervisados de *Machine Learning (ML)*.

2. MARCO TEÓRICO

En esta sección se encuentran los diferentes componentes para el desarrollo de este proyecto tanto de las especificaciones médicas, como los conceptos técnicos y teóricos a utilizar para la realización del Estudio transversal.

2.1. MODELADO DE DATOS

Es la forma de estructura, organizar y clasificar los datos para que puedan ser usado fácilmente por las bases de datos. Lo que se busca en el modelamiento de datos es dejar disponible la información para la correcta lectura y que los diferentes procesos se faciliten tanto su integración como lectura. Existen tres tipos de modelamientos de datos, **Modelos de datos conceptuales**, que se encargan de un conocimiento previo de la información, basándose en las estructuras y finalmente precursores de los **Modelos de datos lógicos(MDL)**, este modelo representa los tipos de entidades lógicas, se denominan tipos de entidades, los atributos de datos son los que detallan esas entidades y las relaciones entre entidades. Finalmente, **Modelos de datos físicos (MDF)**, que son usados para el diseño interno de las bases de datos y la relación entre datos[3].

2.2. MACHINE LEARNING.

Es también conocido como APRENDIZAJE AUTOMATICO, es una rama científica de la inteligencia Artificial (IA), la cual facilita a los sistemas la adquisición de conocimiento progresivo, con mejoras a las tareas y análisis de los datos continuo, sin algún tipo de programación. El *machine learning (ML)*, toma como base la información conseguida en los datos analizados por los sistemas, de alguna manera imitando el comportamiento humano; Con esto ya se pueden crear modelos predictivos, para las posteriores tomas de decisiones, con gran nivel de eficiencia en los resultados[4].

Existen varios tipos de Machine Learning(ML), como son el aprendizaje supervisado, aprendizaje no supervisado, aprendizaje profundo.

2.2.1. APRENDIZAJE SUPERVISADO

El aprendizaje del tipo supervisado hace referencia a un Modelo Específico de Aprendizaje Automático, y se basa en la generación de conocimiento basada en ejemplos o data, esperando que se arrojen resultado esperados. Con los resultados obtenidos el modelo sugiere un tipo de ajuste en la parametrización interior para siempre tener adaptación al ingreso de nueva información. Con estos modelos se logra realizar predicciones pertinentes de las conductas de los datos, por ello es muy utilizado en aplicaciones tecnológicas como detectores de imágenes, de Spam en los correos o reconocimientos de voz[4].

2.2.2. APRENDIZAJE NO SUPERVISADO

Este tipo de aprendizaje, se incluyen los grupos de datos sin etiqueta, de los que se desconoce su estructura, debido a que en este tipo de aprendizaje se busca obtener información sin tener conocimiento del resultado, estimando conseguir variables claves

o relevantes para el modelo. En Aprendizaje no supervisado encontramos dos categorías que son: **Reducción Dimensional** y **Clustering**. La Reducción dimensional, abarca datos de alta complejidad, contemplando la correlación de estos mismos y eliminando la redundancia de información, analizando con mayor eficiencia los datos considerados relevantes para el modelo. En cambio el *Clustering*, es un proceso exploratorio, que clasifica los datos por grupos aun sin conocer su estructura, todo esto lo hace bajo características de datos similares. Estas tipologías son comúnmente usadas en proyectos de mercadeo, donde se busca segmentación de clientes, nichos y demás variables específicas[4].

2.2.3. APRENDIZAJE REFORZADO

Este modelo es diferenciado a los mencionados anteriormente, pues en este modelo se hace el uso de *Deep Learning*(DL), este tiene como finalidad el mejoramiento del rendimiento , tomando como base los resultados o las interacciones realizadas, es un sistema de recompensa, en conclusión es un modelo de entrenamiento constante que busca programar todas las combinaciones posibles, con el fin de tener mejores resultados y predicciones aún más acertadas[5].

2.3. ENFERMEDAD NO TRANSMISIBLE

Las Enfermedades No transmisibles (ENT), vienen siendo la mayor causal tanto de muertes como discapacidades en el mundo, se les da ese término de (**ENT**) porque no son adquiridas de manera transmisible y son susceptibles a duraciones extensas, evolutivas y a ser propensas de tratamiento, con cuidados a mediano y largo plazo. Entre ellas podemos encontrar enfermedades, cardiovasculares, pulmonares crónicas, cáncer, diabetes, trastornos mentales, entre otras[6].

2.4. DIABETES MELLITUS

La diabetes es una enfermedad que se desarrolla cuando no se produce suficiente insulina en el páncreas, uno de los efectos de este trastorno previos al diagnóstico de la diabetes es la hipoglicemia, que es el aumento de azúcar en la sangre, la diabetes se puede clasificar en dos tipos la tipo (1) que es la insulinodependiente o juvenil, se caracteriza por la ausencia de síntesis de insulina y la tipo (2), que no es insulinodependiente y se desarrolla en las personas adultas, se ocasiona generalmente por excesos de peso y sedentarismo. Finalmente encontramos la diabetes gestacional que se encuentra en la hiperglicemia que se detecta durante el embarazo[7].

2.5. COVID-19

El virus **COVID-19**, tiene como origen es en Wuhan (China) desde el 31 de diciembre de 2019[13], la enfermedad causada por el nuevo coronavirus conocido como **SARS-CoV-2** y la que se estimaba en sus inicios como una “neumonía vírica”[8].

2.6. INSULINA

La insulina es un componente natural que se libera desde el páncreas y es el encargado de mantener los niveles del azúcar en la sangre controlado. Durante la noche y entre comidas, los niveles normales, no diabéticos oscilan entre **60-100mg/dl** y **140 mg/dl** o menos después de las comidas y aperitivos[9].

2.7. HIPOGLICEMIA

La Hipoglucemia representa un nivel muy bajo de azúcar en la sangre (o glucosa). A su vez la glucosa, es uno de los principales componentes en el cuerpo para la formación de carbohidratos los cuales son vitales para el correcto funcionamiento del cerebro y cuerpo[10].

2.8. ENFERMEDAD CRONICA

Una enfermedad Crónica es aquella que su duración es prolongada y con progresión lenta y continua. Entre ella podemos encontrar el cáncer, enfermedades de tipo cardiaco, Artritis, asma, diabetes, etc. Este tipo de enfermedades son las causales del aproximadamente **63%** de todas las muertes en el mundo, sin distinción de género y teniendo para el año **2008**, un aproximado de 36 millones en personas menores de 60 años[11].

2.9. TASA DE MORTALIDAD

Básicamente la tasa de mortalidad o mortandad es la proporción de decesos registrados respecto al total de individuos pertenecientes a un país, región, en un determinado rango de tiempo, este tipo de dato estadístico, generalmente se representa porcentualmente, sin embargo también es correcto estimar la cifra numérica de los mismos. La importancia de este cálculo es estimar las causas de las muertes y tomar medidas correctivas o preventivas frente a los eventos [12].

2.10. DANE

El Departamento Administrativo Nacional de Estadística (DANE), es la entidad encargada del levantamiento, procesamiento, análisis y divulgación de las estadísticas oficiales de Colombia. El DANE ofrece al no solo a las personas del país y si no al mundo alrededor de 30 investigaciones anualmente de los sectores más relevantes del país como son la economía, industria, población, sector agropecuario y calidad de vida, entre otras[13].

2.11. SISPRO

El Sistema Integral de Información de la Protección Social – SISPRO, es una plataforma que integra la información varias instituciones, básicamente es usada para el monitoreo, para la validación de servicios prestados y también sirve de base para la toma de decisiones políticas. El SISPRO es un sistema que integra información de varias fuentes, para la protección social, que buscan garantizar el aseguramiento y la asistencia social[14].

3. ESTADO DEL ARTE

En el siguiente cuadro se expondrán algunos de los estudios relacionados previos a esta investigación, el ideal es indicar estudios su enfoque e identificar el valor diferencial de este proyecto.

TITULO	OBJETIVO Y METODOLOGIA	AÑO
Smart Healthcare for Diabetes During COVID-19[15]	Estudio del riesgo que poseen los pacientes con Diabetes en adquirir el Virus del SARS-CoV-2 , se analizar tanto los riesgos y las recomendaciones que se dan para manejar los perfiles glucémicos y el impacto que tendría en la reducción de contagios.	2020
Social determinates of health and COVID-19 mortality rates at the county level[16]	Este artículo analiza la importancia de la disponibilidad de resultados del sector salud de forma detallada y precisa, en el condado de evaluación. Explica cómo estos datos pueden ser significativos para los análisis, sin poner en peligro la privacidad del paciente, todo esto basado en modelos de regresión Lineal.	2020
Medicine Allotment for COVID-19 Patients by Statistical Data Analysis[17]	En este proyecto, se evalúa un método de asignación de medicamentos a pacientes diagnosticados con COVID-19, en los cuales se estimen factores de riesgo, como lo son la presión arterial, sufrir de diabetes, alcoholismo y cáncer, el estudio se ejecutó con algoritmos de agrupamiento.	2020
La Salud de las Personas Adultas Mayores durante la Pandemia de COVID-19[18]	Por medio de revisión al sistema de Salud en México, se evalúan los impactos que la enfermedad COVID-19, tiene sobre pacientes con enfermedades crónicas como lo son Hipertensión, diabetes, cáncer y en personas adultas.	2020
Ultrasound Imaging: A Silent Hero in COVID-19 and Lung Diagnostics[19]	Estudia el impacto y la relevancia que tiene un estudio temprano en la salud de pacientes con comorbilidades y el gran aporte que este estudio genera para reducir procesos más críticos y prevenir complicaciones, mediante desarrollo algorítmico orientados a los ultrasonidos.	2020
Predictions of COVID-19 Infection Severity Based on Co-associations between the SNPs of Co-morbid Diseases and COVID-19 through Machine Learning of Genetic Data[20]	Estudia por medio de <i>Machine Learning</i> , las susceptibilidades que presentan pacientes con comorbilidades o enfermedades crónicas, como lo son el Cáncer, Diabetes Mellitus, Obesidad, entre otras. Detectar de forma temprana las vulnerabilidades y estimar evaluar la efectividad de la predicción.	2020
Comorbidity Impact on COVID-19[21]	Realiza un análisis estadístico de los pacientes con comorbilidad y el impacto que tiene en la atención de para ellos, por requerir tratamiento diferenciado a los demás infectados por el virus en Brasil.	2020
COVID-19 y su relación con poblaciones vulnerables[22]	Encontró la relación entre el contagio con Covid-19 y población más vulnerable socioeconómicamente, mediante revisión bibliográfica de artículos científicos.	2020
Data Associated with Epigenetic Changes Brought by SARS-CoV-2[23]	Se efectúa un estudio del impacto que tiene exacerbar la comorbilidad a través de datos epigenéticos (cambio en los genes).	2020

Tabla 1. Estado del Arte - Fuente Propia

Smart Healthcare for Diabetes During COVID-19. Los resultados obtenidos en este Estudio revelan que las personas con diagnosticadas con diabetes tienen un riesgo mayor de 2 o 3 veces que las personas que no lo padecen, adicional que una sana alimentación mitiga la posibilidad de contagio[15], este estudio abarca tanto la mortalidad de los contagios y si realmente impacta el servicio médico prestado habitualmente en los pacientes que padecen Diabetes Mellitus en todo el territorio colombiano.

Social determinates of health and COVID-19 mortality rates at the county level. Se encontró que en el condado no aumento la tasa de mortalidad por causas como obesidad, tabaquismo o diabetes, al contrario si aumento 7.6% en personas con edad mayor y era directamente dependiente de los confinamientos[16]. Se realiza la evaluación en un segmento distinto en este caso Colombia, con una metodología diferenciada a la de este proyecto, pues hacemos uso de lenguaje no supervisado y supervisado que nos aporte mayor eficacia en los resultados a obtener.

Medicine Allotment for COVID-19 Patients by Statistical Data Analysis. La investigación es basada en algoritmos de Agrupamiento, este procedimiento presenta muchos factores positivos, entre los cuales pueden estar la segmentación de pacientes para la asignación de medicamentos, entre otras virtudes encontradas, es la actualización constante del algoritmo, que permite incluir variables de otras enfermedades y a si calcular la asignación de medicamentos, de acuerdo a su condición e información registrada[17], la diferencia entre los estudios es que este proyecto se realiza únicamente con pacientes diagnosticados con Diabetes Mellitus, busca evaluar entre muchos factores el servicio farmacológico, por ende este proyecto puede servir de base para encontrar mecanismos de servicio en los medicamentos suministrados fuera de la condición del COVID-19.

La Salud de las Personas Adultas Mayores durante la Pandemia de COVID-19. Se encuentra que las personas con mayor factor de riesgo en México son las personas adultas con edades superiores a los 56 años, seguidas por las personas con comorbilidades (Enfermedades Cardiacas, diabetes y Enfermedad respiratoria), adicionalmente se pudo encontrar un factor adicional que es su situación económica y social que posean los pacientes[18]. Este estudio se enfoca en la afectación del servicio y las consecuencias en la letalidad de las personas de todo el país, entre los 0 años en adelante, de todas las regiones del país y de todos los estratos sociales.

Ultrasound Imaging: A Silent Hero in COVID-19 and Lung Diagnostics. Es una investigación en curso que tiene como objetivo promover el uso de robótica y de elementos de inteligencia artificial (IA), para la creación de algoritmos de detección de enfermedades respiratorias crónicas, en exámenes como son el diagnóstico de pulmón y los ultrasonidos, los modelos de **Ultrasound Imaging**[19]; La investigación abarca líneas de prevención y de estudios minuciosos, que son viables para un diagnóstico detallado en pacientes con enfermedades respiratorias, sin embargo el estudio se enfoca en la afectación del servicio y en consecuencia en la letalidad de las personas que padecen esta enfermedad y están contagiadas por el virus, y así determinar cuál es el impacto que al mes de Abril 2021 ha tenido la emergencia sanitaria en nuestro país.

Predictions of COVID-19 Infection Serverity Based on Co-associations between the SNPs of Co-morbid Diseases and COVID-19 through Machine Learning of Genetic Data. Esta predicción se obtuvo como resultado de incluir las patologías y sintomatologías de los pacientes, los resultados logrados con las correlaciones son altamente confiables y podrían estimar el margen de contagio que se tiene según cuadro clínico en cada individuo[20]. Los modelos se realizan respecto a los cuadros clínicos o

las historias clínicas, de un grupo de pacientes, sin embargo este proyecto no estima abarcar ese nivel de detalle, si no estudiar de manera general por Región, edad, sexo, entre otros factores. Las posibles causas que el virus tuvo en pacientes con Diabetes Mellitus contagiados por el virus, si se vieron alteradas sus cifras históricas de muertes y que otros indicadores en la actualidad se ven perjudicados.

Comorbidity Impact on COVID-19. En este proyecto se evidencia en sus resultados, que los pacientes con comorbilidades representan un mayor riesgo de contagio para el COVID-19, se estudiaron múltiples factores, como los son algunos de los componentes en los medicamentos que se suministran en tratamientos de enfermedades crónicas y medicamentos de control en edades avanzadas, con el fin de determinar si estos se atribuyen un factor de riesgo adicional al ya existente por su prescripción o si por el contrario al ser consumidos reducen el riesgo de contagio[21]. Este estudio se realiza bajo condiciones médicas de pacientes con comorbilidad, es muy similar a este estudio, sin embargo este proyecto se enfoca en la tasa de mortalidad y en el impacto del servicio a pacientes con DIABETES MELLITUS infectados por el virus, no en las prescripciones de los medicamentos y sus secuelas.

COVID-19 y su relación con poblaciones vulnerables. El proyecto concluye que la pandemia es una problemática social, que refleja condiciones diferenciadas entre la población, que hace que las personas con bajos recursos y que padecen enfermedades crónicas tiendan a tener mayor vulnerabilidad que personas con condiciones económicas poco superiores, por lo cual propone plantear un sistema de salud más incluyente y menos selectivo a la hora de los tratamientos [22]. La mayor discrepancia entre los proyectos es que realmente se estima cuantificar el impacto por género, edad y ubicación, entre otros. Y a su vez realizar una evaluación más detallada del impacto social en la atención médica de pacientes con Diabetes Mellitus infectados por COVID-19.

Data Associated with Epigenetic Changes Brought by SARS-Cov-2, Tras toda la ejecución del proyecto se concluye que los pacientes con dos afecciones simultáneas, es decir, alguna comorbilidad y el Virus del COVID-19. Son más propensos a padecer complicaciones dentro del proceso de la incubación del virus, lo que hace que se proponga dentro como recomendación, incluir a pacientes con estas enfermedades crónicas, en plan de seguimiento preventivo y así reducir la cantidad de muertes por la adquisición del virus[23] Uno de los diferenciadores del estudio, es que no está orientado a un nivel genético, sin embargo este tipo de estudios aportan de forma médica, los estudios estadísticos y generar valor en la consecución de patologías que deban estimarse para la ejecución de este proyecto.

Podemos verificar que el problema de DIABETES MELLITUS es un problema muy directo con el virus del COVID-19, su aparición no solo afecta su atención regular, si no que según estudios agudiza e incrementa la vulnerabilidad a la adquisición del virus, y no solo eso sino que a su vez acelera en ocasiones la sintomatología y letalidad de dichos pacientes.

4. PLANTEAMIENTO DEL PROBLEMA

Una de las mayores problemáticas actuales en el Colombia, es el sector salud y no solo por el déficit que se presenta de personal de la salud, su alta demanda de usuarios, el poco tiempo de atención con el que cuentan los profesionales de la salud, para evaluar el estado real de los pacientes, la infraestructura deficiente, la alineación de subredes, si no que a todo lo anterior se suma la crisis sanitaria generada por la pandemia del **COVID-19**, situación que al 30 de Abril de 2021, lo posiciona en el ranking mundial de contagios de Número **12** y a nivel de Latinoamérica en el tercer lugar[24].

Si bien es un problema de carácter mundial, este ha centrado la atención tanto médicos generales, como de especialistas en la vigilancia de pacientes que resultan infectados por el virus. Se estima que de una manera directa, influye en la atención, programación de los profesionales y tener menor capacidad instalada, para la atención enfermedades crónicas como es la **DIABETES MELLITUS**, enfermedad que tiene una tasa de mortalidad significativa a nivel mundial y la cual viene por años aumentando de forma exponencial[25].

Por esta causa, es importante estudiar de manera detallas, cual es el impacto que estas personas han sufrido desde el inicio de la pandemia en el país, y evidenciar si es percepción de decadencia de los servicios o si en realidad la pandemia ha desmejorado las condiciones clínicas de mencionados pacientes aumentando su probabilidad de contagio y por ende acrecentando la letalidad[11].

Otra de las falencias detectadas son los efectos colaterales en la salud y en los tratamientos e incapacidades, que se requieren para mantener las enfermedades crónicas como una alerta latente, siendo las que ocasionan la mayor cantidad de muertes a nivel mundial y Colombia no es la excepción, y quien lo afirma es la **OPS** “*Los servicios de prevención y tratamiento de las enfermedades no transmisibles (ENT) se han visto gravemente afectados desde el comienzo de la pandemia de COVID-19 en la región de las Américas, según una encuesta de la Organización Panamericana de la Salud/ Organización Mundial de la Salud (OPS/OMS)*”[26].

Así que en el siguiente proyecto se realizara una evaluación de Modelos de datos, de la población que padece **DIABETES MELLITUS** contagiada con **COVID-19**, según edad, sexo, región, y demás factores que se encuentren en una data, difícilmente localizable y con complejidad de integración, debido a la falta de unificación en las bases de datos a nivel nacional.

5. OBJETIVOS

El proyecto de investigación cuenta con tres objetivos específicos, que permitirán la consecución del objetivo general y la medición de metas parciales basada en la metodología **CRISP-DM**, para así lograr la obtención de los resultados esperados. A continuación, se enuncian cada uno de ellos:

5.1. OBJETIVO GENERAL:

Generar un Modelo de datos de diferentes fuentes de Información para la evaluación de Pacientes con DIABETES MELLITUS CONTAGIADOS DE COVID-19 EN COLOMBIA

5.2. OBJETIVOS ESPECIFICOS:

- Identificar las bases de datos pertinentes para la creación de Modelos de datos.
- Probar diferentes modelos de Aprendizaje automático no supervisado y supervisado que permita encontrar patrones.
- Evaluar el impacto de la llegada del **COVID-19** en la salud de **DIABETES MELLITUS**.

6. METODOLOGIA

Dentro del progreso del proyecto es sustancial a la definición de actividades, a su vez se especifica el cómo se irán ejecutando cada una de ellas, para el este proyecto se estimaron 6 etapas con la Metodología CRISP-DM[27], detalladas a continuación:

ETAPA	DESCRIPCION	DETALLE
FASE DE ANALISIS	<ul style="list-style-type: none"> * Se Realizará un análisis detallado, que permita conocer las verdaderas problemáticas del proyecto y validar la información disponible para la ejecución. * Dentro de las fuentes conocidas tenemos el DANE, SISPRO, SALUDATA, entre otras. 	<ul style="list-style-type: none"> * Comprensión de los objetivos y requerimientos del proyecto. * Definición del problema de Minería de Datos
COMPRENSION DE LOS DATOS	<ul style="list-style-type: none"> * Con las fuentes encontradas en la etapa anterior, se realiza el análisis de la información y por ende la descripción, para verificar la utilidad y pertinencia de los datos y si son factibles para iniciar la siguiente fase. * Se analizará, si los datos son suficientes y/o se requiere algo adicional. 	<ul style="list-style-type: none"> * Obtención conjunto inicial de datos. * Exploración del conjunto de datos. * Identificar las características de calidad de los datos * Identificar los resultados iniciales.
PREPARACION DE DATOS	<ul style="list-style-type: none"> * En esta fase se inicia el proceso de calidad de los datos y se realiza la preparación las fuentes para dar inicio al procesamiento. 	<ul style="list-style-type: none"> * Selección de datos * Limpieza de datos * Integrar los datos * Formatear los datos
MODELAMIENTO	<ul style="list-style-type: none"> * Se inicia la construcción del Modelo de datos, para el desarrollo del proyecto 	<ul style="list-style-type: none"> * Seleccionar Técnica de Modelamiento * Verificar la pertinencia del modelo * Iniciar fase de desarrollo * Generar el Diseño de Prueba * Construir el Modelo incluyendo la Parametrización * Revisión de la configuración de los parámetros del modelo
EVALUACION	<p>Se inicia la fase de evaluación de los datos obtenidos y la precisión de la información.</p>	<ul style="list-style-type: none"> * Evaluación de los Resultados Obtenidos * Evaluar si cumplen los Criterios de los objetivos * Revisar el Proceso * Revisión del Modelo * Estimar si es factible para promoción a la siguiente fase
DESPLIEGUE	<p>Posterior a las fases mencionadas anteriormente y luego de concluir las satisfactoriamente, se inicia la fase del despliegue.</p>	<ul style="list-style-type: none"> * Desplegar el Plan * Desarrollar el reporte final * Revisión del Proyecto * Realizar la Documentación de las experiencias

Tabla 2. Metodología - CRISP-DM (Fuente Propia)

7. CRONOGRAMA

Se realiza un cronograma de las actividades estimadas por semana y la duración de estas, encontrando relacionadas las que se denotan a continuación:

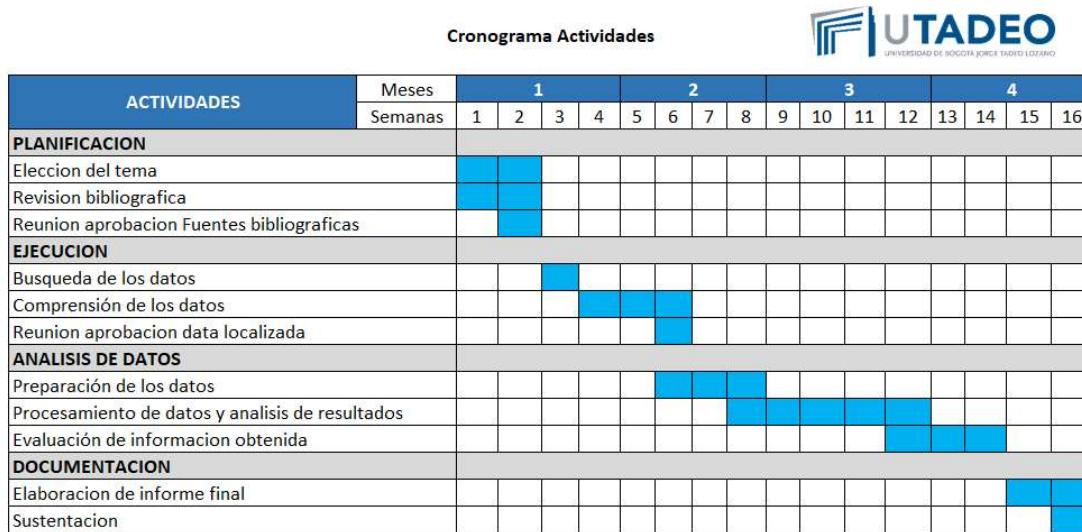


Ilustración 3. Propia

8. PRESUPUESTO

Para el desarrollo y ejecución del proyecto se tiene en cuenta los recursos tanto tangibles como recurso humano requeridos para la correcta elaboración del proyecto, se detallan de la siguiente manera:



Presupuesto

DESCRIPCION	CANTIDAD	VALOR UNITARIO	TOTAL
Hora Estudiante	50	80.000	4.000.000
Hora Director	20	160.000	3.200.000
Equipo de Computo (Depresiacion)	4	20.000	80.000
Internet	4	30.000	120.000
TOTAL			7.400.000

Ilustración 4. Propia

9. BIBLIOGRAFIA

- [1] Banco Mundial, “Evaluación Externa de la Calidad de la Atención en el Sector de la Salud en Colombia,” 2019.
<https://www.bancomundial.org/es/topic/health/publication/external-assessment-of-quality-of-care-in-the-health-sector-in-colombia> (accessed Apr. 04, 2021).
- [2] Ministerio de salud y protección Social, “Los retos del sistema de salud que dejó la pandemia por covid-19,” *Página Web*, 2020.
<https://www.minsalud.gov.co/Paginas/Los-retos-del-sistema-de-salud-que-dejo-la-pandemia-por-covid-19.aspx> (accessed Apr. 04, 2021).
- [3] M. Soto, “Modelado de Datos: Definición, Usos y Tipos.”
<https://www.tecnologias-informacion.com/modeladodatos.html> (accessed May 05, 2021).
- [4] grapheverywhere, “Machine Learning Qué es, tipos, ejemplos y cómo implementarlo,” 2020. <https://www.grapheverywhere.com/machine-learning-que-es-tipos-ejemplos-y-como-implementarlo/> (accessed May 04, 2021).
- [5] P. S. J. Barrios, “Comparación de técnicas de aprendizaje por refuerzo jugando a un videojuego de tenis.,” 2019.
- [6] OPS/OMS, “Enfermedades no transmisibles - OPS/OMS | Organización Panamericana de la Salud,” Ops, 2018.
<https://www.paho.org/es/temas/enfermedades-no-transmisibles> (accessed Apr. 25, 2021).
- [7] “OMS | Informe mundial sobre la diabetes,” *Who*, 2016.
<https://www.who.int/diabetes/global-report/es/#:~:text=Según las estimaciones%2C 422 millones,5%25 en la población adulta.> (accessed Apr. 23, 2021).
- [8] “Nuevo coronavirus 2019,” 2020.
https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019?gclid=Cj0KCQiA0fr_BRDaARIsAABw4EuFOV8nG27mhLr-MQ3FrYrxs9NKS3QqpMKScY_eMx6JBhDORdv7mdoaAle2EALw_wcB%0Ahttps://www.who.int/es/emergencies/diseases/novel-coronavirus-2019%0Ahttps://www.who.int/ (accessed Apr. 24, 2021).
- [9] “OMS Diabetes,” *Who*, 2017.
http://www.who.int/topics/diabetes_mellitus/es/%0Ahttp://www.who.int/mediacentre/factsheets/fs312/es/ (accessed Apr. 24, 2021).
- [10] S. N. Davis and G. Lastra-González, “Diabetes y Bajo Nivel de Glucosa (Hipoglicemia),” *J. Clin. Endocrinol. Metab.*, vol. 93, no. 8, pp. E1–E1, Aug. 2008, doi: 10.1210/jcem.93.8.9992.
- [11] Ministerio de Salud Pública, CORAPE, OPS, and OMS, “Enfermedad crónicas y COVID-19,” *Centro de Coordinación de Alertas y Emergencias Sanitarias*, 2020.
<https://www.paho.org/sites/default/files/enfermedades-cronicas-covid-19.pdf> (accessed Apr. 25, 2021).
- [12] “Tasa de mortalidad - Qué es, definición y concepto | 2021 | Economipedia.”
<https://economipedia.com/definiciones/tasa-de-mortalidad.html> (accessed Apr. 25, 2021).

- [13] DANE, "Que es el DANE," p. 1, 2008.
- [14] MINSALUD, "Sistema integral de la protección social." <http://www.sispro.gov.co/Pages/Observatorios/cancer.aspx> (accessed May 05, 2021).
- [15] A. M. Joshi, U. P. Shukla, and S. P. Mohanty, "Smart healthcare for diabetes: A COVID-19 perspective," *arXiv*, 2020.
- [16] S. Lam *et al.*, "Social determinates of health and COVID-19 mortality rates at the county level," *2020 4th Int. Conf. Multimed. Comput. Netw. Appl. MCNA 2020*, pp. 159–165, 2020, doi: 10.1109/MCNA50957.2020.9264276.
- [17] S. S. Prasad and S. N. Korra, "Medicine Allotment for COVID-19 Patients by Statistical Data Analysis," pp. 665–669, 2021.
- [18] J. A. Vega Rivero, J. C. Ruvalcaba Ledezma, I. Hernández Pacheco, M. del R. Acuña Gurrola, and L. López Pontigo, "La salud de las personas adultas mayores durante la pandemia de COVID-19," *J. negat. no posit. results*, pp. 726–739, 2020, doi: 10.19230/jonnpr.3772.
- [19] A. C. H. Yu, L. Demi, M. Muller, and Q. Zhou, "Ultrasound Imaging: A Silent Hero in COVID-19 and Lung Diagnostics," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 67, no. 11, pp. 2194–2196, 2020, doi: 10.1109/TUFFC.2020.3031444.
- [20] W. L. Y. Wanng Ryan Yixiang, Qinsong Guo Tim, Guanhua Li Leo, Jiao Julia Yutian, "Predictions of COVID-19 Infection Severity Basedon Co-associations between the SNPs of Co-morbidDiseases and COVID-19 through Machine Learningof Genetic Data," *2020 IEEE 8th Int. Conf. Comput. Sci. Netw. Technol.*, vol. 1, no. 0, pp. 1–5, 2020, doi: 10.1088/1751-8113/44/8/085201.
- [21] N. Darapaneni *et al.*, "Comorbidity Impact on COVID-19," *Proc. 2020 IEEE Int. Conf. Mach. Learn. Appl. Netw. Technol. ICMLANT 2020*, pp. 3–8, 2020, doi: 10.1109/ICMLANT50963.2020.9355994.
- [22] C. R. Aquino-Canchari, R. del C. Quispe-Arrieta, and K. M. Huaman Castillon, "COVID-19 y su relación con poblaciones vulnerables," *Rev. habanera cienc. méd*, vol. 19, pp. 1–18, 2020.
- [23] Tianze Qiu, "'dwd \$vvrlfdwhg zlwk (sljhqhwlf &kdqjhv %urxjkw e\ 6\$56 &ry," *Data Assoc. with Epigenetic Chang. Brought by SARS-CoV-2*, pp. 1–5, 2020.
- [24] A. Orús, "• Países con más casos de coronavirus | Statista," *5 De Enero*, 2021. <https://es.statista.com/estadisticas/1091192/paises-afectados-por-el-coronavirus-de-wuhan-segun-los-casos-confirmados/> (accessed May 04, 2021).
- [25] OPS, "Informe de la evaluación rápida de la prestación de servicios para enfermedades no transmisibles durante la pandemia de COVID-19 en las Américas," *Ops*, 2020.
- [26] L. Linn, S. Oliel, and A. Baldwin, "La COVID-19 afectó el funcionamiento de los servicios de salud para enfermedades no transmisibles en las Américas - OPS/OMS | Organización Panamericana de la Salud," *Organización Panamericana de la Salud/Organización Mundial de la Salud*, 2020. <https://www.paho.org/es/noticias/17-6-2020-covid-19-afecto-funcionamiento-servicios-salud-para-enfermedades-no> (accessed Apr. 05, 2021).

- [27] J. Gallardo, "Modelos de proceso para proyectos de Data Mining (DM) CRISP-DM (Cross Industry Standard Process for Data Mining)," 2010.