



Anexo 1. Formato de protocolo de investigación

1. Datos generales de la propuesta

- **Nombre de la persona proponente:** Dagoberto Armenta Medina
- **Título de la propuesta:** Enfoques computacionales para la detección de blancos moleculares de interés farmacológico y predicción de enfermedades.
- **Línea de Generación y Aplicación del Conocimiento (LGAC) de INFOTEC en la que incide la propuesta:** Ciencia de datos

Periodo de ejecución (Del 06/08/2021 al 06/11/2027):

2. Descripción de la propuesta:

- **Resumen** (ejecutivo):

En este proyecto mediante el uso de información derivada del gran cúmulo de datos de origen biológico y técnicas de procesamiento de lenguaje natural y minería de texto desarrollaremos enfoques computacionales capaces de detectar blancos de drogas asociados a enfermedades. En una primera etapa del proyecto seleccionaremos enfermedades de interés regional y nacional ubicando mediante enfoques de performance analysis y science mapping los actores principales, la estructura y dinámica del dominio permitiéndonos acceder al estado del arte del campo de estudio.

En una segunda etapa de proyecto desarrollaremos enfoques computacionales capaces de pronosticar enfermedades, así como detectar potenciales blancos moleculares asociados, por medio de diferentes fuentes de información biológica, basándose en la extracción de conocimiento. En una tercera etapa del proyecto se implementará en plataformas web los enfoques y modelos desarrollados en la etapa 2 del proyecto.



- **Antecedentes:**

Las tecnologías omicas (datos genómica, proteómica y metabolómica) actuales en conjunto con los expedientes electrónicos médicos han generado un enorme cúmulo de datos sin precedentes en la historia de la biología y el área biomédica. La acumulación astronómica de datos de origen biológico, así como la necesidad de su almacenamiento, anotación, análisis, organización e integración en redes biológicas y bases de datos han sido las principales fuerzas para la emergencia y desarrollo de la ciencia de datos aplicadas al área de las biociencias la cual puede ser a grandes rasgos nombrada como sinónimo de bioinformática.

La bioinformática, citando una de sus definiciones minimalistas, es la aplicación de tecnologías computacionales a la gestión y análisis de datos biológicos. El reto de manejar el gran volumen de datos omicos solo es posible gracias al poder de cómputo, algoritmos y métodos utilizados por la ciencia de datos y específicamente por bioinformáticos los cuales ayudan a revelar aspectos desconocidos de los sistemas biológicos. Sin estos recursos la simple tarea de leer un solo genoma humano a ojo requeriría alrededor de 26 años de trabajo.

De esta manera la bioinformática es utilizada para ahorrar tiempo y costo en los laboratorios de diferentes áreas relacionadas con la biología. Por ejemplo, en la anotación automática de genomas, predicción de la estructura y función de genes, diseño de drogas, detección molecular de organismos entre muchas otras. En general la bioinformática es considerada por varios países una de las áreas primordiales para la innovación de tecnologías relacionadas con las ciencias de la vida y de gran relevancia para el desarrollo de la bio-economía mundial.

Debido a las necesidades internacionales y que México no está exento de la explosión de los datos biológicos, el desarrollo de la bioinformática en México constituye un área prioritaria y de oportunidad. Lo anterior denota la importancia de contar en el país con líneas de investigación enfocadas en resolver las diferentes demandas que apoyen el desarrollo de las investigaciones en materia de bioinformática y las bioindustrias. Uno de los sectores con mayor potencialidad de aplicaciones bioinformáticas es el sector salud e industrias como la farmacéutica, específicamente en el descubrimiento de nuevas drogas o fármacos consiste en encontrar moléculas capaces de unirse e inhibir o activar eficientemente un blanco macromolecular específico, generalmente una proteína. A pesar de que la identificación de blancos de drogas por medio de



herramientas bioinformáticas es una realidad, estas se encuentran poco desarrolladas.

En el presente proyecto proponemos la generación de enfoques computacionales para la detección de elementos genéticos y factores de riesgos asociados a enfermedades mediante el análisis masivo de datos de origen biológico proveniente de diferentes fuentes que puedan favorecer el descubrimiento de soluciones terapéuticas y diagnóstico de enfermedades mediante técnicas de extracción de conocimiento y generación de modelos computacionales.

- **Justificación:**

El uso de herramientas computacionales en conjunto con las técnicas de ciencia de datos y machine learning nos permitirá desarrollar modelos y enfoques computacionales útiles en el entendimiento y predicción de fenómenos biológicos como las enfermedades.

- **Objetivo general:** Desarrollar líneas de investigación en bioinformática que permita la creación de enfoques computacionales para el análisis de grandes cúmulos de datos de origen biológico enfocados en el entendimiento y predicción de enfermedades, así como en la detección de potenciales blancos moleculares.

- **Objetivos específicos:**

Desarrollar líneas de investigación en análisis de datos biológicos.

Desarrollar enfoques computacionales para el análisis de grandes cúmulos de datos de origen biológico.

Desarrollar modelos computacionales que permitan el entendimiento y pronóstico de enfermedades de interés nacional e internacional.

Desarrollo de conocimiento en el entendimiento y predicción de fenómenos biológicos.



**GOBIERNO DE
MÉXICO**



CONACYT
Consejo Nacional de Ciencia y Tecnología



**Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental**

Desarrollar prototipos que permitan implementar los modelos desarrollados en los objetivos anterior mediante aplicaciones web y otros dispositivos útiles para la sociedad.

- **Metas:**

Generar líneas de investigación en ciencia de datos aplicados al análisis de datos biológicos.

Generación de divulgación por medio publicación en revistas jcr de corte internacional y de divulgación nacional sobre el conocimiento básico y aplicado obtenido de nuestros estudios.

Generar enfoques y modelos computacionales que logren un mejor entendimiento y predicción de los fenómenos biológicos como las enfermedades.

Implementar los enfoques y modelos computacionales en plataformas web y otros dispositivos.

- **Metodología:**

Se utilizarán herramientas y técnicas de la ciencia de datos para recuperar información de bases de datos publicas como PUBMED y de expediente medico facilitado por la red de hospitales nacionales. Se generarán modelos por medio de algoritmos de inteligencia computacional a través de librerías de los lenguajes de programación r y Python. Los modelos y enfoques computacionales desarrollados serán utilizados en el entendimiento y predicción de fenómenos biológicos. Los desarrollos anteriores se implementarán en aplicaciones web y dispositivos por medio lenguajes de programación como r.



**GOBIERNO DE
MÉXICO**



CONACYT
Consejo Nacional de Ciencia y Tecnología



**Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental**

- **Beneficios esperados:**

Mediante desarrollo del proyecto se generará una nueva línea de investigación en ciencia de datos en el área biológica, siendo pionera para INFOTEC.

Del trabajo anterior se obtendrán publicaciones internacionales que posicionarán a INFOTEC en el área de la ciencia de datos y bioinformática.

Se generarán enfoques computacionales con utilidad en el entendimiento y pronóstico de enfermedades en conjunto con la detección de potenciales blancos farmacológicos, los cuales servirán como herramienta a INFOTEC en la colaboración con sectores interesados en el área.

En las diferentes etapas del proyecto se colaborará con el área ciencia de datos fortaleciendo la integración de los recursos humanos de INFOTEC.

- **Resultados esperados:**

Generación de conocimiento y formación de profesionales en la ciencia de datos.

Generación de publicaciones internacionales en revistas JCR y artículos de divulgación.

Generación de prototipos de enfoques computacionales y modelos útiles en el entendimiento y pronóstico de fenómenos biológicos como las enfermedades.

Los enfoques y modelos computacionales obtenidos en presente proyecto serán implementados en aplicaciones web y dispositivos.



3. Plan de actividades

- **Descripción de las actividades**

Etapa núm.: 1		
#	Actividad	Descripción de la actividad
1	Recopilación de datos	Busca de información referente al tema de estudios
2	Integración y generación de base de datos	Colocar la información pertinente en bases de datos
3	Enfoques computacionales	Detectar y emplear enfoques computacionales útiles en el problema de estudio
4	Análisis	Análisis de los resultados obtenidos por los enfoques
5	Escritura de artículo	Escribir las metodologías y resultados derivado de los análisis previos
6	Publicaciones	Someter y Publicar en revistas nacionales e internacionales

- **Descripción de las metas**

Etapa núm.: 1		
#	Actividad	Meta
1	Recopilación de datos	
2	Integración y generación de base de datos	Base de datos sobre el área
3	Enfoques computacionales	
4	Análisis	Estrategia computacional
5	Escritura de artículo	
6	Publicaciones	Publicaciones en revistas

- **Productos (entregables)**

Etapa núm.:1		
#	Actividad	Producto
1	Recopilación de datos	





2	Integración y generación de base de datos	
3	Enfoques computacionales	Enfoque computacional
4	Análisis	
5	Escritura de artículo	
6	Publicaciones	Publicaciones en JCR al menos 2

• **Descripción de las actividades**

Etapa núm.: 2		
#	Actividad	Descripción de la actividad
1	Identificación y entendimiento del estado del arte de enfermedades de interés nacional e internacional	Ubicar las enfermedades de interés nacional e internacional con potencial de estudio.
2	Obtención y generación de bases de datos asociados en las enfermedades	Localizar las mejores fuentes de información e integrarlas en bases de datos útiles para ser procesadas
3	Desarrollo de enfoques para la extracción de conocimiento y pronóstico de enfermedades	Adaptar y generar enfoques computacionales por medio de técnicas de bigdata y machine learning útiles en la extracción de conocimiento y pronóstico de enfermedades.
4	Desarrollo de modelos útiles en el pronóstico de enfermedades	Generar modelos útiles en la extracción de conocimiento y pronóstico de enfermedades.
5	Evaluación de los modelos	Seleccionar los mejores modelos mediante su evaluación con diferentes métricas de desempeño.
6	Generación de manuscritos	Escribir los resultados mas relevantes encontrados en la investigación.
7	Publicación y divulgación	Someter a diferentes revistas y una vez publicado generar divulgación de los trabajos.

• **Descripción de las metas**

Etapa núm.: 2		
#	Actividad	Meta



1	Identificación y entendimiento del estado del arte de enfermedades de interés nacional e internacional	
2	Obtención y generación de bases de datos asociados en las enfermedades	
3	Desarrollo de enfoques para la extracción de conocimiento en enfermedades	Generación de enfoques útiles en la extracción de conocimiento
4	Desarrollo de modelos útiles en el pronóstico de enfermedades	Obtención de modelos útiles en la extracción de conocimiento y pronóstico de enfermedades
5	Evaluación de los modelos	
6	Generación de manuscritos	
7	Publicación y divulgación	Publicaciones en revistas

• **Productos (entregables)**

Etapas		
Etapas		
#	Actividad	Producto
1	Identificación y entendimiento del estado del arte de enfermedades de interés nacional e internacional	
2	Obtención y generación de bases de datos asociados en las enfermedades	
3	Desarrollo de enfoques para la extracción de conocimiento en enfermedades	
4	Desarrollo de modelos útiles en el pronóstico de enfermedades	



5	Evaluación de los modelos	Obtención de modelos útiles en la extracción de conocimiento y pronóstico de enfermedades
6	Generación de manuscritos	
7	Publicación y divulgación	Publicaciones en revistas

• **Descripción de las actividades**

Etapa núm.: 3		
#	Actividad	Descripción de la actividad
1	Generación de la interfaz web para extracción de conocimiento o pronóstico de enfermedades	Programar interfaz web para visualizar modelos
2	Integración de los modelos o enfoques computacionales en la interfaz web	Aplicación web que visualice los resultados derivados de los modelos y enfoques computacionales
3	Implementación de los servicios anteriores en un servidor o plataforma web de consulta útil para la sociedad	Ejecución remota del servicio de aplicación web
4	Divulgación de los servicios	Diversos eventos de divulgación

• **Descripción de las metas**

Etapa núm.: 3		
#	Actividad	Meta
1	Generación de la interfaz web para extracción de conocimiento o pronóstico de enfermedades	
2	Integración de los modelos o enfoques computacionales en la interfaz web	Aplicación web que visualice los resultados derivados de los modelos y enfoques computacionales





3	Implementación de los servicios anteriores en un servidor o plataforma web de consulta útil para la sociedad	Ejecución remota del servicio de aplicación web
4	Divulgación de los servicios	Diversos eventos de divulgación

• **Productos (entregables)**

Etapa núm.:3		
#	Actividad	Producto
1	Generación de la interfaz web para extracción de conocimiento o pronóstico de enfermedades	
2	Integración de los modelos o enfoques computacionales en la interfaz web	
3	Implementación de los servicios anteriores en un servidor o plataforma web de consulta útil para la sociedad	Aplicación web
4	Divulgación de los servicios	Diversos eventos de divulgación sobre las aplicaciones desarrolladas

• **Especificaciones de los productos(N/A)**

Etapa núm.:		
#	Actividad	Especificación del producto (entregable)
1		
2		

• **Mecanismos de transferencia(N/A)**

Etapa núm.:





#	Actividad	Mecanismo de transferencia
1		
2		

4. Cronograma de actividades

Etapa núm.:1				
#	Actividad	Productos (entregables)	Fecha de inicio (#de semanas)	Duración (núm. de semanas)
1	Recopilación de datos		1	10
2	Integración y generación de base de datos		4	12
3	Enfoques computacionales	Enfoque computacional	13	23
4	Análisis		30	6
5	Escritura de artículo		28	38
6	Publicaciones	Publicaciones en JCR al menos 2	36	16

Cronograma de actividades

Etapa núm.:2				
#	Actividad	Productos (entregables)	Fecha de inicio	Duración (núm. de semanas)
1	Identificación y entendimiento del estado del arte de enfermedad		30	26





**Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental**

	ades de interés nacional e internacional			
2	Obtención y generación de bases de datos asociados en las enfermedades		40	26
3	Desarrollo de enfoques para la extracción de conocimiento en enfermedades		50	30
4	Desarrollo de modelos útiles en el pronóstico de enfermedades		50	60
5	Evaluación de los modelos	Obtención de modelos útiles en la extracción de conocimiento y pronóstico de enfermedades	50	70





**Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental**

6	Generación de manuscritos		60	80
7	Publicación y divulgación	Publicaciones en revistas	120	25

Cronograma de actividades

Etapa núm.:3				
#	Actividad	Productos (entregables)	Fecha de inicio	Duración (núm. de semanas)
1	Generación de la interfaz web para extracción de conocimiento o pronóstico de enfermedades		50	80
2	Integración de los modelos o enfoques computacionales en la interfaz web		120	110
3	Implementación de los servicios	Ejecución remota del servicio de aplicación web	230	70





	anteriores en un servidor o plataforma web de consulta útil para la sociedad			
4	Divulgación de los servicios	Diversos eventos de divulgación sobre las aplicaciones desarrolladas	230	75

1. Referencias

1. Gawande, K.; Rane, D. Exploring the Applications and Potential of Bioinformatics. IOSR J. Comput. Eng. 2016, 8, 2278–2661.
2. Hogeweg, P. The Roots of Bioinformatics in Theoretical Biology. PLoS Comput. Biol. 2011, 7, e1002021. PubMed]
3. Hesper, B.; Hogeweg, P. Bioinformatica: Een werkconcept. Kameleon 1 (6): 28–29. Dutch.) Leiden Leidse Biol. Club. arXiv 1970, arXiv:2111.11832.
4. Schneider, M.V.; Orchard, S. Omics Technologies, Data and Bioinformatics Principles; Humana Press: Totowa, NJ, USA, 2011; pp. 3–30.
5. Kumar, A.; Chordia, N. Role of Bioinformatics in Biotechnology. Res. Rev. Biosci. 2017, 12, 116.
6. Salter, B.; Zhou, Y.; Datta, S.; Salter, C. Bioinformatics and the Politics of Innovation in the Life Sciences: Science and the State in the United Kingdom, China, and India. Sci. Technol. Hum. Values 2016, 41, 793–826.
7. Wei, L.; Yu, J. Bioinformatics in China: A Personal Perspective. PLoS Comput. Biol. 2008, 4, e1000020.
8. Baillie Gerritsen, V.; Palagi, P.M.; Durinx, C. Bioinformatics on a national scale: An example from Switzerland. Brief. Bioinform. 2019, 20, 361–369.
9. Cobo, M.J.; López-Herrera, A.G.; Herrera-Viedma, E.; Herrera, F. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. J. Informetr. 2011, 5, 146–166.





**GOBIERNO DE
MÉXICO**



CONACYT
Consejo Nacional de Ciencia y Tecnología



**Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental**

10. Manoharan, A.; Kanagavel, B.; Muthuchidambaram, A.; Kumaravel, J.P.S. Bioinformatics Research: An Informetric View. In Proceedings of the 2011 International Conference on Information, Communication and Management, Singapore, 14 October 2011; Volume 6, pp. 199–204.
11. Song, M.; Kim, S.; Zhang, G.; Ding, Y.; Chambers, T. Productivity and influence in bioinformatics: A bibliometric analysis using PubMed central. J. Assoc. Inf. Sci. Technol. 2014, 65, 352–371.
12. De Las Rivas, J.; Bonavides-Martínez, C.; Campos-Laborie, F.J. Bioinformatics in Latin America and SolBio impact, a tale of spin-off and expansion around genomes and protein structures. Brief. Bioinform. 2017, 20, 390–397.
13. Armenta-Medina, D.; Díaz de Leon, C.; Valderrama-Blanco, B. Bioinformatics in Mexico: A diagnostic from the academic perspective and recommendations for a public policy. PLoS ONE 2020, 15, e0243531.
14. Roberts, R.J. PubMed Central: The GenBank of the published literature. Proc. Natl. Acad. Sci. USA 2001, 98, 381–382.

ATENTAMENTE

[Dr. Dagoberto Armenta Medina]
[CONACYT-INFOTEC investigador]

C.c.p. **Mtro. Carlos Josué Lavandeira Portillo**, Director Adjunto de Innovación y Conocimiento. Presente.
Dr. Juan Antonio Vega Garfias, Subgerente de Innovación Gubernamental. Presente.



**GOBIERNO DE
MÉXICO**



CONACYT
Consejo Nacional de Ciencia y Tecnología



**Dirección Adjunta de Innovación y Conocimiento
Gerencia de Innovación
Subgerencia de Innovación Gubernamental**

Avenida San Fernando No. 37, Col. Toriello Guerra, CP. 14050, CDMX, México.
Tel: 55 5624 2800 www.infotec.mx



2022 Flores
Año de Magón
PRECURSOR DE LA REVOLUCIÓN MEXICANA