

6 de junio de 2025.
Dictamen C.I. 12/2025

DICTAMEN
QUE PRESENTA LA COMISIÓN DE INVESTIGACIÓN DE LA DIVISIÓN DE CIENCIAS DE LA
COMUNICACIÓN Y DISEÑO

ANTECEDENTES

- I. El Consejo Divisional de Ciencias de la Comunicación y Diseño, en la sesión 09.25, celebrada el 30 de abril de 2025, integró esta Comisión en los términos señalados en el artículo 56 de Reglamento Interno de los Órganos Colegiados Académicos.

- II. El Consejo Divisional designó para esta Comisión a las siguientes personas integrantes:
 - a) Órganos personales:
 - ✓ Mtro. José Alfredo Andrade García, Jefe del Departamento de Ciencias de la Comunicación;
 - ✓ Mtra. Brenda García Parra, Jefa del Departamento de Teoría y Procesos del Diseño;
 - ✓ Dr. Carlos Roberto Jaimez González, Jefe del Departamento de Tecnologías de la Información.

 - b) Representantes propietarios:
 - Personal académico:
 - ✓ Dra. María Alejandra Osorio Olave, Departamento de Ciencias de la Comunicación;
 - ✓ Dr. Christopher Heard Wade, Departamento de Teoría y Procesos del Diseño;
 - ✓ Dr. Carlos Rodríguez Lucatero, Departamento de Tecnologías de la Información.

CONSIDERACIONES

- I. La Comisión recibió, para análisis y discusión, el proyecto de investigación denominado **“Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa”**, que presenta el Dr. Alfredo Piero Mateos Papis.

II. La Comisión de Investigación sesionó el 6 de junio de 2025, fecha en la que concluyó su trabajo de análisis y evaluación de la propuesta, con el presente Dictamen.

III. La Comisión tomó en consideración los siguientes elementos:

- *"Lineamientos para la creación de grupos de investigación y para el registro, seguimiento y evaluación de proyectos de investigación. División de Ciencias de la Comunicación y Diseño"* aprobados en la Sesión 20.24 del Consejo Divisional, celebrada el 18 de noviembre de 2024, mediante Acuerdo DCCD.CD.23.20.24.
- Coherencia interna del proyecto, la relación entre el planteamiento de problema, sus objetivos y metas, y su metodología.

IV. **Objetivo general:**

Aplicar el método PCA con orientación novedosa planteada a problemas diversos, principalmente de salud pública, para mostrar la eficacia de dicha aplicación en el hallazgo de relaciones entre atributos.

V. **Objetivos específicos:**

- Realizar búsqueda de "problemáticas" (conjunto de problemas) con los cuales se cuente con datos suficientes y apropiados para su análisis con métodos PCA y de agrupamiento no supervisado. Principalmente problemas del tipo de salud pública.
- Obtener y registrar los datos asociados a los problemas a estudiar, tal que queden indicados los atributos asociados, especialmente al menos un atributo de interés.
- Realizar la aplicación de los métodos PCA y de agrupamiento a los datos para tratar de encontrar relaciones entre atributos.
- Analizar las soluciones, realizar comparaciones, hacer registros, obtener conclusiones.
- Generar reportes. Si es posible, publicar los reportes en forma de artículos académicos / científicos, de tal forma que se "trate" de tener al menos una publicación en cada año de la duración de este proyecto.

VI. Productos de investigación o metas:

Generación de un reporte de investigación anual (dos anuales) y uno final.
Se pretende realizar al menos una publicación anual durante el tiempo que dure el proyecto, ya sea en revista o en conferencia con memorias.

VII. Participantes del proyecto:

- Dr. Alfredo Piero Mateos Papis, Responsable del proyecto.

Participantes internos del DTI:

- Dr. Carlos Joel Rivero Moreno.
- Dr. Christian Sánchez Sánchez

VIII. La evaluación de los resultados de investigación se llevará a cabo de acuerdo con los lineamientos vigentes.

DICTAMEN

ÚNICO:

Se recomienda al Consejo Divisional de Ciencias de la Comunicación y Diseño, aprobar el proyecto de investigación **“Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa”**, que presenta el Dr. Alfredo Piero Mateos Papis.

La **duración** del proyecto será del 24 de junio de 2025 al 23 de junio de 2027.

Los departamentos de adscripción del personal académico participante, proporcionarán un financiamiento básico, sujeto a disponibilidad presupuestal, para la realización de los proyectos.

Se recomienda a las personas titulares de las jefaturas de departamento informar oportunamente del monto anual del que disponen los profesores para la realización del proyecto.

Se recomienda a los profesores, la búsqueda de fuentes adicionales de financiamiento, externas a la Universidad.

VOTOS:

Integrantes	Sentido de los votos
Mtro. José Alfredo Andrade García	A favor
Mtra. Brenda García Parra	A favor
Dr. Carlos Roberto Jaimez González	A favor
Dra. María Alejandra Osorio Olave	A favor
Dr. Christopher Heard Wade	A favor
Dr. Carlos Rodríguez Lucatero	A favor
Total de los votos	6 votos a favor

Coordinadora



Martínez



Secretaria del Consejo Divisional de Ciencias de la Comunicación y Diseño



Ciudad de México, 24 de mayo de 2025

Dra. Angélica Martínez de la Peña

Directora de la División de Ciencias de la Comunicación y Diseño

Universidad Autónoma Metropolitana, Unidad Cuajimalpa

P R E S E N T E

ASUNTO: Registro de Proyecto de Investigación

Por medio de la presente, me permito enviarle el Proyecto de Investigación **Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa**, del cual es responsable el Dr. Alfredo Piero Mateos Papis. Se solicita el registro de este Proyecto de Investigación ante el Consejo Divisional de la División de Ciencias de la Comunicación y Diseño, con una duración de dos años.

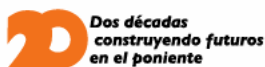
El objetivo general del proyecto es el siguiente: *Aplicar el método PCA con orientación novedosa a problemas diversos, principalmente de salud pública, para mostrar la eficacia de dicha aplicación en el hallazgo de relaciones entre atributos.*

Entre los objetivos específicos del proyecto se encuentran los siguientes:

- Realizar búsqueda de “problemáticas” (conjunto de problemas) con los cuales se cuente con datos suficientes y apropiados para su análisis con métodos PCA y de agrupamiento no supervisado. Principalmente problemas del tipo de salud pública.
- Obtener y registrar los datos asociados a los problemas a estudiar, tal que queden indicados los atributos asociados, especialmente al menos un atributo de interés.
- Realizar la aplicación de los métodos PCA y de agrupamiento a los datos para tratar de encontrar relaciones entre atributos.



Departamento
de Tecnologías
de la Información



Unidad Cuajimalpa
DCCD | División de Ciencias de la Comunicación y Diseño
Jefatura del Departamento de Tecnologías de la Información



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Cuajimalpa

- Analizar las soluciones, realizar comparaciones, hacer registros, obtener conclusiones.
- Generar reportes. Si es posible, publicar los reportes en forma de artículos académicos / científicos, de tal forma que se trate de tener al menos una publicación en cada año de la duración de este proyecto.

Los entregables que se tendrán al concluir el Proyecto de Investigación, son los siguientes:

- Dos reportes técnicos con los resultados obtenidos (uno por año). Se pretende que se conviertan en publicaciones de artículos de revista o de congreso.

Le agradecería se turne esta solicitud de registro del proyecto de investigación ante la Comisión de Investigación para su revisión, y en su caso, al Consejo Divisional de la División de Ciencias de la Comunicación y Diseño que usted preside, para su aprobación.

Sin más por el momento, aprovecho la oportunidad para enviarle un cordial saludo.

Atentamente

Casa abierta al tiempo

Dr. Carlos Roberto Jaimez González

Jefe del Departamento de Tecnologías de la Información



DTI

Departamento
de Tecnologías
de la Información

20 **Dos décadas
construyendo futuros
en el poniente**

Unidad Cuajimalpa
DCCD | División de Ciencias de la Comunicación y Diseño
Jefatura del Departamento de Tecnologías de la Información

Ciudad de México a 22 de abril de 2025

Dr. Carlos Jaimez González. Jefe.
Departamento de Tecnologías de Información
División de Ciencias de la Comunicación y Diseño
Universidad Autónoma Metropolitana, Unidad Cuajimalpa

Estimado Dr. Jaimez González

Por este medio presentamos a usted nuestro documento descriptivo de proyecto de investigación, ya con los cambios realizados motivados por la revisión y observaciones de usted a nuestra versión anterior del proyecto, presentado a usted el 4 de abril de 2025. Agradecemos por la revisión tan minuciosa realizada.

Le solicitamos que si usted considera ya pertinente, este documento se canalice al Consejo Divisional de la División de Ciencias de la Comunicación y Diseño, para su revisión, con la expectativa de su aprobación.

Este documento se realiza con atención a la versión vigente de los: “Lineamientos para la creación de grupos de investigación y para el registro, seguimiento y evaluación de proyectos de investigación”, de la División de Ciencias de la Comunicación y Diseño.

El nombre del proyecto es: “Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa”.

Agradecemos su atención.

Atentamente,

Dr. Alfredo Piero Mateos Papis. Coordinador del proyecto
Departamento de Tecnologías de Información
División de Ciencias de la Comunicación y Diseño
Universidad Autónoma Metropolitana. Unidad Cuajimalpa.

Ciudad de México, a 22 de abril de 2025

Dra. Gloria Angélica Martínez de la Peña. Presidente
Consejo Divisional de la División de Ciencias de la Comunicación y Diseño
UAM. Unidad Cuajimalpa

Estimado Dra. Martínez de la Peña

Por este medio presentamos a usted nuestro documento descriptivo de proyecto de investigación, con el fin de que se envíe al Consejo Divisional de la División de Ciencias de la Comunicación y Diseño para su revisión, con la expectativa de su aprobación.

Este documento se presentó el día 04 del presente a la jefatura del Departamento de Tecnologías de la Información (DTI). El Dr. Carlos Jaimez González, jefe de DTI, hizo una revisión del documento con detalle, y sus observaciones ayudaron a mejorar significativamente el contenido y forma de presentación del documento.

Este documento se realizó con atención a la versión vigente de los: “Lineamientos para la creación de grupos de investigación y para el registro, seguimiento y evaluación de proyectos de investigación”, de la División de Ciencias de la Comunicación y Diseño.

El nombre del proyecto es: “Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa”.

Agradecemos su atención, y quedamos al pendiente de sus comentarios.

Atentamente,

Dr. Alfredo Piero Mateos Papis. Coordinador del proyecto
Departamento de Tecnologías de Información
División de Ciencias de la Comunicación y Diseño
Universidad Autónoma Metropolitana. Unidad Cuajimalpa.

Proyecto de Investigación

Contenido

Proyecto de Investigación.....	1
1. Datos generales	3
1.1. Título del proyecto	3
1.2. Departamento en el que se inscribe el proyecto de investigación	3
1.3. Colectividad.....	3
1.4. Tipo de investigación	3
1.5. Datos académicos de las personas participantes	3
1.5.1. Persona responsable del proyecto.....	3
1.5.2. Participantes internos.....	3
1.5.3. Tiempo estimado de dedicación a las actividades del proyecto	3
1.5.4. Participantes externos al proyecto.....	4
2. Protocolo de Investigación	4
2.1. Título del proyecto	4
2.2. Resumen del Proyecto.....	4
2.3. Planteamiento del problema.....	5
2.4. Justificación.....	6
2.5. Planteamiento general del objeto de estudio. Explicar brevemente el contexto social / político / científico / cultural en el que se inscribe el proyecto	7
2.6. Hipótesis o supuesto del proyecto	8
2.7. Preguntas de investigación	8
2.8. Objetivo general	8
2.9. Objetivos específicos	9
2.10. Antecedentes históricos, teóricos y conceptuales	9
2.10.1. Antecedentes históricos	9
2.10.2. Antecedentes Conceptuales.....	9
2.10.3. Explicación del método PCA planteado	15
2.11. Estado del arte del problema de investigación: Estudios académicos en torno al objeto de estudio	19
2.12. Alcances y limitaciones del proyecto (establecidos de forma preliminar) con relación directa al problema que se aborda.....	20

Proyecto de Investigación: *Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa*

2.13.	Metodología. (estrategia de investigación y descripción del diseño de investigación, así como la selección de técnicas e instrumentos de investigación).....	21
2.14.	Referencias y fuentes de consulta	22
	Bibliografía.....	22
3.	Planeación y requerimientos	23
3.1.	Productos de Investigación o metas	23
3.2.	Cronograma de actividades	24
3.3.	Requerimientos y justificación de los recursos necesarios.....	24
4.	Vinculación con los planes y programas de estudio de la División y la Unidad	25
4.1.	Planes y programas de estudio de la División de Ciencias de la Comunicación y Diseño (DCCD).....	25
4.2.	Planes y programas de estudio de la División de Ciencias Naturales e Ingeniería (CNI) 26	
4.3.	Vinculación Institucional	26
5.	Incidencia social.....	27
5.1.	Con relación a los Programas Nacionales Estratégicos	27
5.2.	Con relación a los Objetivos de Desarrollo Sostenible de la ONU	28

1. Datos generales

1.1. Título del proyecto

Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa.

1.2. Departamento en el que se inscribe el proyecto de investigación

Departamento de Tecnologías de Información.

1.3. Colectividad

Investigación colectiva, no perteneciente a algún grupo de investigación.

1.4. Tipo de investigación

Proyecto de Investigación Científica

1.5. Datos académicos de las personas participantes

Dr. Alfredo Piero Mateos Papis. Profesor Asociado D de tiempo indeterminado, del Departamento de Tecnologías de Información. Número económico 33729.

Dr. Carlos Joel Rivero Moreno. Profesor Titular C de tiempo indeterminado, del Departamento de Tecnologías de Información. Número económico 39201.

Dr. Christian Sánchez Sánchez. Profesor Titular C de tiempo indeterminado, del Departamento de Tecnologías de Información. Número económico 34585.

1.5.1. Persona responsable del proyecto.

Dr. Alfredo Piero Mateos Papis.

1.5.2. Participantes internos.

Dr. Alfredo Piero Mateos Papis.

Dr. Carlos Joel Rivero Moreno.

Dr. Christian Sánchez Sánchez.

1.5.3. Tiempo estimado de dedicación a las actividades del proyecto

En horas semanales:

Dr. Alfredo Piero Mateos Papis. 3 horas semanales al menos.

Dr. Carlos Joel Rivero Moreno. 2 horas semanales.

Dr. Christian Sánchez Sánchez. 2 horas semanales.

Se adjunta carta compromiso de dedicación, por cada participante, dirigida al Consejo Divisional de la División de Ciencias de la Comunicación y Diseño.

La duración del proyecto se propone para DOS años.

1.5.4. Participantes externos al proyecto

Pudiera haber participantes externos, pero no se plantea compromiso actualmente. Se indicaría al respecto al Consejo Divisional de la División de Ciencias de la Comunicación y Diseño si fuera el caso y en su momento, si hubiera adición de participantes externos, solicitando, consecuentemente, la modificación del proyecto. Se incluiría, en su caso, una carta de aceptación de participación por parte de cada uno de los participantes externos, así como una semblanza curricular de una cuartilla de extensión, por cada uno.

2. Protocolo de Investigación

2.1. Título del proyecto

Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa.

2.2. Resumen del Proyecto

El objeto de este proyecto es el conjunto de datos disponibles, que están expresados con un formato donde hay diversos individuos e iguales atributos por cada individuo. Cada individuo se describe o clasifica mediante sus atributos, por ejemplo atributos físicos como el género, el peso, y la altura de una persona; o atributos de costumbres como los horarios de alimentación.

El tipo de análisis de este proyecto es de tipo “exploratorio”¹, en el cual se analizan datos de los que se tiene poco conocimiento para identificar patrones o relaciones: específicamente en este proyecto la búsqueda tiene un propósito inicial de encontrar cómo un atributo seleccionado se relaciona con los demás atributos.

Hay trabajos que se apoyan en la metodología de análisis de datos llamada Análisis de Componentes Principales (PCA), en situaciones donde por la cantidad de atributos (varias decenas al menos) hace que la obtención de dependencias entre ellos sea difícil. Esta metodología es conocida en el medio científico y es aplicable para estas situaciones.

¹ <https://worldcampus.saintleo.edu/blog/cuantos-tipos-de-analisis-de-datos-hay-cuales-son-los-tipos-de-analisis-de-datos#:~:text=Existen%20cinco%20tipos%20de%20an%C3%A1lisis,Vea-mos%20cada%20uno%20de%20ellos>

Los autores de este proyecto han encontrado una orientación novedosa de PCA que es prometedora para tener resultados exitosos cuando la metodología “tradicional” PCA falla para obtener información² adicional oculta entre los datos.

Ya el proyecto ha tenido impacto en estudios en conjunto con la Facultad de Medicina de la UNAM, pero el impacto puede ser básicamente en cualquier disciplina, con tal de que se tengan datos con el formato planteado.

Para el caso de la Facultad de Medicina de la UNAM, la aplicación que se ha obtenido con la metodología novedosa es encontrar la relación entre el atributo “obesidad” de las personas, con diversos atributos asociados con costumbres alimenticias y con la disrupción a su ciclo circadiano [1].

Por otro lado, los participantes de este proyecto han trabajado en métodos de agrupamiento no supervisado (automáticos) [2], y buscan aplicar estos métodos como alternativa a PCA, y confrontar, si es posible, los resultados obtenidos sobre la búsqueda de relaciones entre atributos.

Los participantes de este proyecto buscan también encontrar relaciones entre atributos con el uso de ANOVA (análisis de variancias), como ya lo han realizado los participantes de este proyecto [3].

Se busca comparar los resultados de los diversos métodos aquí planteados, pero principalmente se busca probar la efectividad del método PCA con la orientación novedosa que le han dado los participantes de este proyecto [1].

La motivación de este trabajo es la oportunidad que ofrecen las computadoras modernas para automatizar labores de análisis de datos, de tal forma que se puedan descubrir relaciones entre atributos en situaciones o fenómenos físicos o sociales de interés

2.3. Planteamiento del problema

En fenómenos físicos o sociales de los cuales se dispusiera de datos, los cuales tuvieran “muchos” (decenas o centenas) atributos (características), en donde interesara conocer o “descubrir” la relación entre un atributo de interés, con los demás atributos, es común que este descubrimiento sea difícil. Mediante métodos convencionales de análisis de datos, como el graficado, o el análisis estadístico, suele ser difícil, o básicamente inviable, poder descubrir la relación [4] [5] [6] entre atributos.

Este proyecto busca analizar la efectividad del método PCA (*Principal Component Analysis*) con orientación novedosa dada por los participantes de este proyecto [1], en problemas con datos suficientes y apropiados, para encontrar relaciones entre atributos

Con PCA, se transforma la forma en que se observan los datos, normalmente observados gráficamente en un espacio *euclidiano* “original”, ahora ubicándolos en un espacio *euclidiano*

² Los teóricos de la información **no la definen** (para evitar controversias sobre su significado), solamente la miden. Aquí nos damos libertad de manejar el término “información” de manera informal.

“alternativo”, donde es más fácil poder observar relaciones entre atributos [6]. Especialmente el método PCA, con una orientación novedosa planteada por los participantes de este proyecto, se vislumbra prometedor para obtener buenos resultados (que se puedan identificar las relaciones buscadas).

Problemas de interés en este proyecto, para la aplicación del método PCA son los de salud pública. Los participantes de este proyecto han trabajado [1] en problemas sobre obesidad, relacionados con hábitos alimenticios y hábitos de disrupción circadiana. La obesidad es el atributo de interés, y los hábitos alimenticios, y los de disrupción circadiana se consideran como los demás atributos.

Este proyecto está previsto para estudiar también problemas de otro tipo.

2.4. Justificación

La sociedad está continuamente en búsqueda de alternativas que provean asistencia para solución de problemas de interés público. Se pueden encontrar soluciones mediante métodos científicos, y también mediante métodos empíricos que usan herramientas tecnológicas emanadas de conocimiento científico y de la experiencia. El análisis de datos en este proyecto es una mezcla de ambos tipos de métodos³.

Cuando hay datos disponibles sobre algún fenómeno, esos datos pueden guardar información oculta importante. Esta es la razón por la que, por ejemplo, los datos de astrofísica [7], aunque sean de miles de millones de bytes diarios, nunca se borran, se almacenan “para siempre”. Esto se debe a que esos datos pueden analizarse posteriormente, con nuevos métodos, con objetivos nuevos, para encontrar información nueva “oculta”, no fácil de observar, o simplemente para extraer un tipo de información que no se había buscado.

La búsqueda de información en datos disponibles es una forma de participar en la solución de problemas. Actualmente, con el poder del cómputo, es posible llevar a realidad cálculos matemáticos elaborados o muy grandes, que no es posible realizar “manualmente”, para encontrar información “oculta” en datos. Esta posibilidad se quiere aprovechar en este proyecto.

El análisis de datos implica no solamente la disponibilidad de equipo de cómputo, sino la aplicación de métodos de solución que incluyen matemáticas, algoritmos e imaginación, y eso se puede realizar en países que no tienen tantos recursos como los países más avanzados y ricos del mundo. Eso es aprovechable y benéfico para un país como México.

³ Los estudios con base en herramientas de métodos empíricos ofrecen mayor posibilidad de generación de resultados, por su posibilidad “casi infinita” de caminos de búsqueda, lo que no es tan fácil con los estudios puramente científicos. Eso puede causar una disparidad de generación de resultados en el trabajo de investigación, para investigadores con diversas orientaciones, lo cual puede generar injusticia cuando se mide el esfuerzo investigativo con tan solo la llamada “productividad”, inclusive generando un daño al país al sopesar de forma “errónea” el esfuerzo y trabajo de los diversos investigadores.

2.5. Planteamiento general del objeto de estudio. Explicar brevemente el contexto social / político / científico / cultural en el que se inscribe el proyecto

El objeto de estudio de este proyecto es en sí mismo la efectividad del método PCA con una orientación novedosa dada por los autores de este proyecto, en [1]. Es probar esta efectividad.

Un objeto de estudio paralelamente importante en este proyecto son los problemas que cuentan con datos suficientes y apropiados para su estudio mediante el método planteado en este proyecto, siendo estos problemas fundamentales de salud pública, como el problema de obesidad, aunque también es deseable incidir en solución a problemas graves como el cáncer, pero eso depende en gran medida de la disponibilidad de datos apropiados, recordando que la obtención de datos con relación a los padecimientos humanos ha sido generalmente una labor que requiere mucho esfuerzo y es tardada, y frecuentemente escasa (se pueden obtener datos para análisis de cáncer en sitios como el del *National Cancer Institute* [8]).

Este tipo de análisis de datos de interés de este proyecto se puede clasificar como “de análisis exploratorio”, para investigar datos de los cuales identificar patrones o relaciones [9]. Este tipo de análisis puede orientarse para la asistencia sanitaria en la “Detección precoz de enfermedades”, que anticipa el deterioro de la salud, o en la “Medicina Personaliza”, que personaliza el tratamiento, ambas realizando análisis de datos que indiquen, entre otras cosas, el estilo de vida y los factores ambientales [10].

Es importante observar que la obesidad y el cáncer están entre el número 1 de enfermedades crónicas no transmisibles (ECNT) en México, según un análisis de los principales retos que enfrenta la salud pública en México para 2030 [11].

Con relación a problemas de salud en que los participantes ya han trabajado y en que piensan continuar en este proyecto es sobre obesidad de personas, y la relación de este “atributo” con las costumbres de alimentación y con la interrupción del ciclo circadiano [1]. Los datos para estos trabajos se han obtenido con colegas de la Facultad de Medicina de la UNAM, en Ciudad Universitaria, Ciudad de México. Está previsto en este proyecto continuar trabajando con estos colegas para profundizar y ampliar los resultados.

En este estudio, sin embargo, se pueden aceptar problemas de otro tipo, de interés general en donde haya datos disponibles suficientes y apropiados para su análisis mediante los métodos aquí planteados. Por ejemplo, colegas del Instituto de Ciencias del Mar y Limnología de la UNAM, en Ciudad Universitaria, han mostrado interés en el tipo de análisis de datos de este proyecto para estudiar fenómenos como la disponibilidad de oxígeno en el mar, en diversas profundidades, con relación a sustancias u organismos encontrados.

Los participantes de este proyecto también tienen interés en interactuar con los miembros de la División de Ciencias Naturales e Ingeniería, especialmente las especialidades de Biología

Molecular, e Ingeniería Biológica, y aunque ha habido pláticas que nos indican algunas cosas de lo que ellos realizan, como búsqueda de fármacos contra el cáncer, mediante aplicación de productos propuestos, a células cancerosas, no ha habido aún intercambio que indique alguna forma en que este proyecto podría incidir con ellos, pero el interés está planteado y hay una búsqueda.

La participación de este proyecto con datos relacionados con Ciencias Sociales o del análisis de soluciones a problemas sociales, humanos o físicos basados en el Diseño, podría dar la oportunidad de interacción con las especialidades de los departamentos de Teoría y Procesos del Diseño, y de Ciencias de la Comunicación. Aún no ha habido acercamiento para conocer de las posibilidades de interacción.

2.6. Hipótesis o supuesto del proyecto

En situaciones de datos en donde es difícil encontrar relaciones entre atributos es posible, mediante el PCA (y el Biplot), encontrar relaciones entre atributos que no se encontrarían, al menos no fácilmente, con otros métodos, especialmente usando PCA con el procedimiento de orientación novedosa que participantes de este proyecto ya han planteado con éxito sobre obesidad [1].

Puede ser que algunos métodos adicionales, tales como el agrupamiento no supervisado (en donde no hay una supervisión humana), en lo que han trabajado los participantes de este proyecto [2], encuentren también relaciones entre atributos.

Debe quedar claro que sobre la aplicación de estos métodos, como aquí se plantea, no es claro que se pueda tener éxito.

Debe quedar claro que la aplicación de los métodos de agrupamiento y de clasificación aquí mencionados no forman parte del objeto de estudio de este proyecto, pero se incluye su uso como una forma de completar el análisis de los datos.

2.7. Preguntas de investigación

¿Es posible encontrar relaciones entre atributos en donde métodos como el graficado o el análisis estadístico no lo encuentran?

¿Es posible encontrar relaciones entre atributos en donde el método de PCA (convencional) no lo encuentra?, usando el método de PCA con orientación novedosa, planteada en [1]

Además, es de interés también tratar de encontrar relaciones entre atributos usando métodos como el agrupamiento no supervisado? (sin que esto forma una pregunta de investigación).

2.8. Objetivo general

Aplicar el método PCA con orientación novedosa planteada en [1] a problemas diversos, principalmente de salud pública, para mostrar la eficacia de dicha aplicación en el hallazgo de relaciones entre atributos.

Se suma el interés de aplicar métodos como el agrupamiento no supervisado para “tratar” de encontrar soluciones para hallar relaciones entre atributos y, en caso positivo, comparar la solución, en su similitud, con aquella encontrada con el método PCA con orientación novedosa. Debe quedar claro que este paréntesis no forma parte del objetivo general de este proyecto.

2.9. Objetivos específicos

- Realizar búsqueda de “problemáticas” (conjunto de problemas) con los cuales se cuente con datos suficientes y apropiados para su análisis con métodos PCA y de agrupamiento no supervisado. Principalmente problemas del tipo de salud pública.
- Obtener y registrar los datos asociados a los problemas a estudiar, tal que queden indicados los atributos asociados, especialmente al menos un atributo de interés.
- Realizar la aplicación de los métodos PCA y de agrupamiento a los datos para tratar de encontrar relaciones entre atributos.
- Analizar las soluciones, realizar comparaciones, hacer registros, obtener conclusiones.
- Generar reportes. Si es posible, publicar los reportes en forma de artículos académicos / científicos, de tal forma que se “trate” de tener al menos una publicación en cada año de la duración de este proyecto.

Debe quedar claro que las publicaciones no se pueden asegurar pues eso depende de los tiempos y de la actuación y evaluación de personas externas al proyecto, pero en caso de no poderse dar las publicaciones esperadas, al menos sí se asegura que se presentará un reporte técnico y una memoria de las actividades y datos de envío a publicación.

2.10. Antecedentes históricos, teóricos y conceptuales

2.10.1. Antecedentes históricos

La técnica PCA proviene del año aproximadamente 1900 [12], la cual no pudo desarrollarse sino hasta recientemente, contando ya con un significativo poder de cómputo por la disponibilidad de las computadoras digitales. Con PCA se realiza una secuencia de transformaciones lineales de los datos “originales”, hacia un espacio “alternativo” donde se pueden graficar y observar de forma simplificada los datos. Los ejes coordenados alternativos son los componentes principales, dejando claro que se tiene en cuenta que en trabajos como [4] Pág. 3 se alerta sobre la confusión del uso de la palabra “componente”.

Hay bibliografía que dedica en forma exclusiva, o en una parte importante, su publicación a la explicación del concepto de PCA, por ejemplo las siguientes referencias generadas entre 2014 hasta 2021: [4], [5], [6], y [12].

2.10.2. Antecedentes Conceptuales

Imaginemos que buscamos comprender algún fenómeno⁴, y que del mismo hemos identificado dos características que lo pueden representar (al menos parcialmente). Dichas

⁴ Hecho o evento cuya causa y / o explicación no es clara.

características deben ser perceptibles de tal forma que éstas se puedan “leer” y registrar⁵. La percepción se hace en intervalos, ya sea de tiempo o de espacio (más comúnmente de tiempo), para finalmente formar una tabla de registro de dos atributos. La tabla tiene:

- 1) Dos columnas, cada una correspondiente a cada característica.
- 2) n renglones, cada uno conteniendo dos lecturas, una de cada característica.
- 3) Un encabezado que indica el nombre de cada característica.
- 4) Puede contener una columna adicional indicando los intervalos en que se hizo cada lectura.

Es común, en análisis de datos, que a cada característica se le denomine “atributo”. Considere que las lecturas corresponden a números.

Como ejemplo, la Tabla 1 muestra una tabla de registro de dos atributos, por lo que tiene dos columnas. La tabla tiene un encabezado que tiene los nombres de ambos atributos, que para este ejemplo se llaman: *Atributo 1* y *Atributo 2*. Cada uno de los renglones de la tabla corresponde a cada “lectura” realizada. Se puede decir que cada renglón es “un dato”, formado por valores de los dos atributos.

<i>Atributo 1</i>	<i>Atributo 2</i>
Dato 1 (atributo 1)	Dato 1 (atributo 2)
Dato 2 (atributo 1)	Dato 2 (atributo 2)
⋮	⋮
Dato n (atributo 1)	Dato n (atributo 2)

Tabla 1. Tabla de registro de n datos, con 2 atributos por cada dato.

Para facilitar la observación del fenómeno relacionado con la tabla de registro de dos atributos, se pueden expresar los datos de la tabla como puntos en una gráfica de un espacio euclidiano con coordenadas cartesianas⁶, teniendo la gráfica dos ejes coordenados perpendiculares, cada eje representando a uno de los dos atributos. La Ilustración 1:

- Muestra un ejemplo de cómo podría ser una gráfica así, donde los ejes coordenados parten de un origen, tal que entre más lejos del origen el valor representado es mayor.
- Incluye un solo dato, que se muestra como un pequeño círculo situado según los valores los dos atributos, valores que son: “ a_1 ” y “ a_2 ”.
- Incluye una línea pintada entre el origen y el punto que representa el dato. La magnitud de esta línea es proporcional a la magnitud del dato.
- La línea entre el origen y el dato se puede representar como:

$$(a_1, a_2)$$

⁵ Si la lectura implica una conversión en valores numéricos, la lectura se vuelve una medición, y los atributos entonces tienen valores numéricos, y a los atributos se les llama “variables”.

⁶ El espacio euclidiano representa originalmente el espacio de tres dimensiones en el que nos movemos, donde se satisfacen los axiomas de Euclides. Posteriormente, en 1637, René Descartes agrega a este espacio unos ejes coordenados con cotas, lo que permite enlazar a la geometría y el álgebra.

Esta representación de la línea se conoce como vector. Un vector es un arreglo de números, en este caso los números son a_1 y a_2 . Se sabe que la magnitud del vector es⁷:

$$d_1^2 = a_1^2 + a_2^2$$

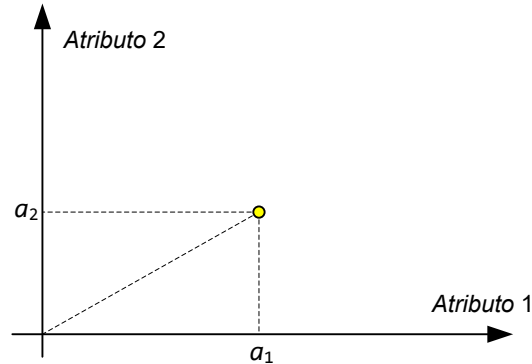


Ilustración 1

La Ilustración 2 es similar a la Ilustración 1, pero presenta más datos, situados a forma de ejemplo, donde a partir de estos datos parece haber una relación “positiva lineal” entre el *Atributo 1* y el *Atributo 2*, es decir, si un dato tiene un valor grande del *Atributo 1*, también lo tiene del *Atributo 2*, tal que la relación se puede aproximar a una relación lineal.

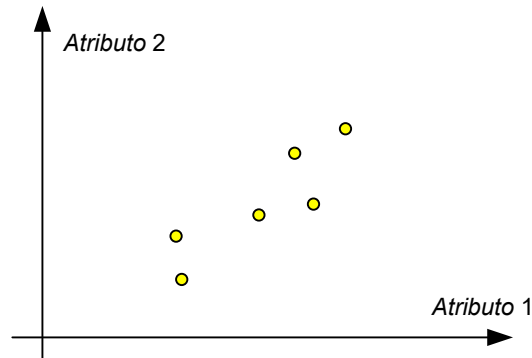


Ilustración 2

La Ilustración 3 es similar a la Ilustración 2, pero en donde se sitúan los datos a forma de ejemplo, de tal forma que no se observa una relación clara de crecimiento entre el *Atributo 1* y el *Atributo 2*.

⁷ Por el Teorema de Pitágoras.

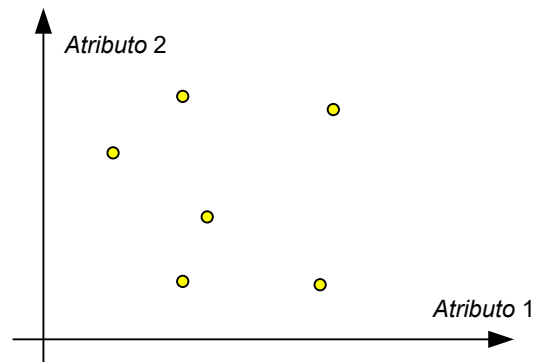


Ilustración 3

La Ilustración 4 es similar a la Ilustración 1, pero un solo dato que en este caso tiene tres atributos. En la ilustración se incluye una línea entre el origen y el punto del dato, que representa el vector (a_1, a_2, a_3) , de magnitud $d_1^2 = a_1^2 + a_2^2 + a_3^2$.

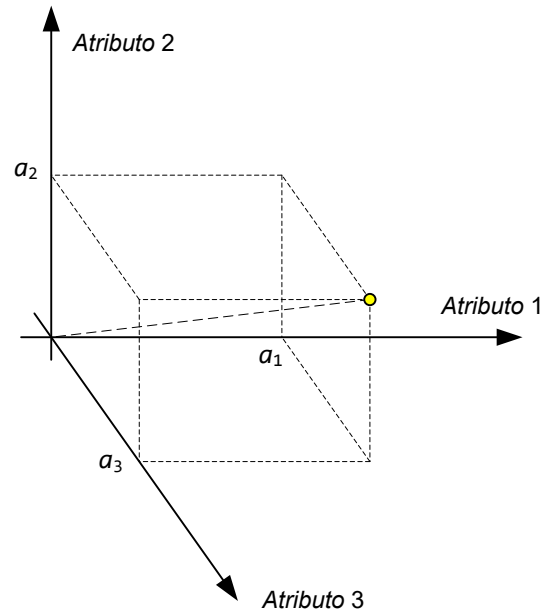


Ilustración 4

La Ilustración 5 es similar a la Ilustración 4, pero muestra también las proyecciones del vector del dato: (a_1, a_2, a_3) sobre cada uno de los tres ejes coordenados.

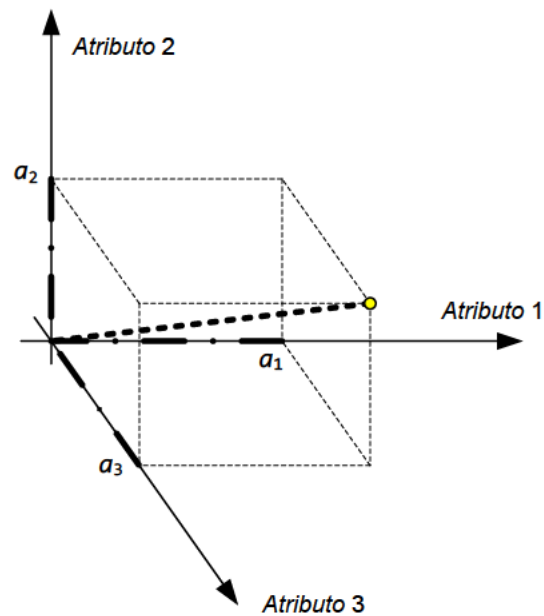


Ilustración 5

La Ilustración 6 es similar a la Ilustración 4, pero muestra las proyecciones del vector del dato: (a_1, a_2, a_3) , en cada uno de los tres planos: 1) Plano *Atributo 1-Atributo 2*; 2) Plano *Atributo 1-Atributo 3*, y 3) Plano *Atributo 2-Atributo 3*. Por simplicidad de la ilustración, solamente se muestra el nombre *Atributo 1-Atributo 3*.

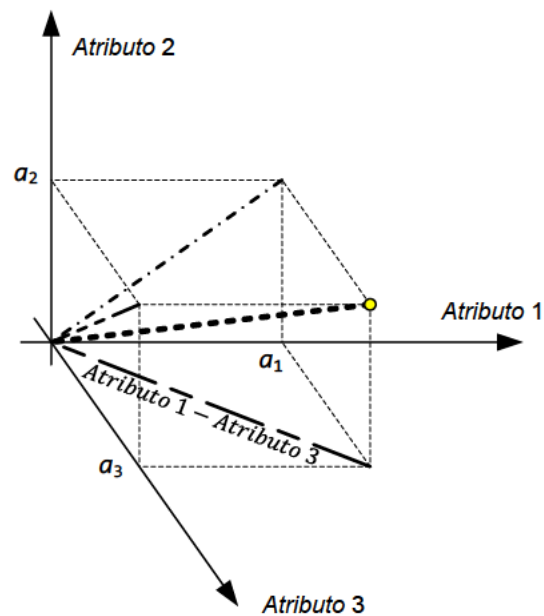


Ilustración 6

La Ilustración 7 es similar a la Ilustración 6, pero donde se muestra un caso de un dato cuyo vector (a_1, a_2, a_3) tiene una proyección en el plano *Atributo 2-Atributo 3* con magnitud significativamente menor que la magnitud de su proyección en los otros dos planos, como lo sería en el plano *Atributo 1-Atributo 3*. Esto indica que si se observase solamente la proyección de

algún vector en un plano específico, esto podría no reflejar correctamente la verdadera magnitud del vector.

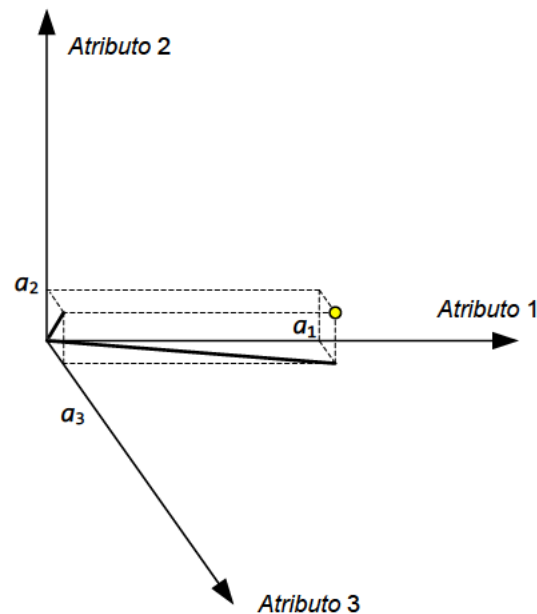


Ilustración 7

Una reflexión importante emanada de las ilustraciones anteriores es que conforme mayor es el número de atributos, la relación entre éstos puede ser más difícil de observar.

En casos de fenómenos a estudiar, pudiese haber decenas o centenas de atributos, número que indicamos con la letra m , tal que los datos observados se expresaran en una tabla como la Tabla 2, con m columnas, habiendo una columna para cada atributo. El número de datos observados, número que expresamos con la letra n , pudiese ser de centenas o millardos.

Estos datos se podrían expresar en una gráfica que tuviese m dimensiones (una dimensión por cada atributo). Una gráfica así no es representable en 3 dimensiones, como normalmente se representa con un dibujo en un papel, pero sí es manejable matemáticamente. Al espacio euclidiano en donde se sitúa una gráfica así se le suele llamar “hiper espacio”. El poder observar la relación entre los m atributos a partir de una gráfica así, aunque sea mediante el apoyo de una computadora, es generalmente difícil.

Atributo 1	Atributo 2	...	Atributo m
Dato 1 (atributo 1)	Dato 1 (atributo 2)	...	Dato 1 (atributo m)
Dato 2 (atributo 1)	Dato 2 (atributo 2)	...	Dato 2 (atributo m)
⋮	⋮	⋮	⋮
Dato n (atributo 1)	Dato n (atributo 2)	...	Dato n (atributo m)

Tabla 2. Tabla de registro de n datos, con m atributos por cada dato.

2.10.3. Explicación del método PCA planteado

Observación de un fenómeno en un espacio euclidiano original

Considerando un fenómeno a estudiar, que tiene m atributos y del cual se registran n datos (como lo que se indica en la Tabla 2), se pueden representar los datos y atributos en una matriz⁸ que llamamos Y , tal como se representa en la Ilustración 8. Cada renglón de la matriz aloja a uno de los n datos. Cada uno y todos los datos están compuestos de m atributos, tal que la matriz tiene m columnas. Los valores de cada atributo se alojan en una sola columna de la matriz, provenientes de los datos que se tienen en la matriz.

El fundamento del problema es el interés en descubrir la relación que puede haber entre los atributos, con base en los n datos obtenidos. Muchas veces hay un atributo de interés, e interesa conocer la relación que puede tener con los demás atributos.

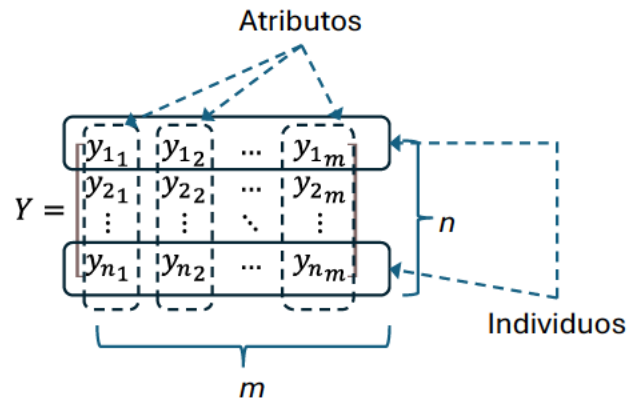


Ilustración 8

En la Ilustración 8 se puede observar que cada elemento de la matriz se representa como y_{ij} , $i = 1, \dots, n$; $j = 1, \dots, m$. El primer subíndice de y_{ij} representa el número de dato (de 1 a n), y el segundo subíndice (más pequeño) representa el número de atributo (de 1 a m).

Los n datos se pueden graficar como " n vectores" en una gráfica que represente un espacio euclidiano m -dimensional con coordenadas cartesianas, donde cada uno de los m ejes coordenados es cada uno de los m atributos. En la Ilustración 9 se observa una versión de dicha gráfica, en donde se ilustra, por simplicidad, solamente un vector que representa a uno de los n datos. También se presentan, por simplicidad, solamente los ejes coordenados 1, 2 y m (referentes consecutivamente a los: *Atributo 1*, *Atributo 2* y *Atributo m*).

⁸ Una matriz en matemáticas es un arreglo de números.

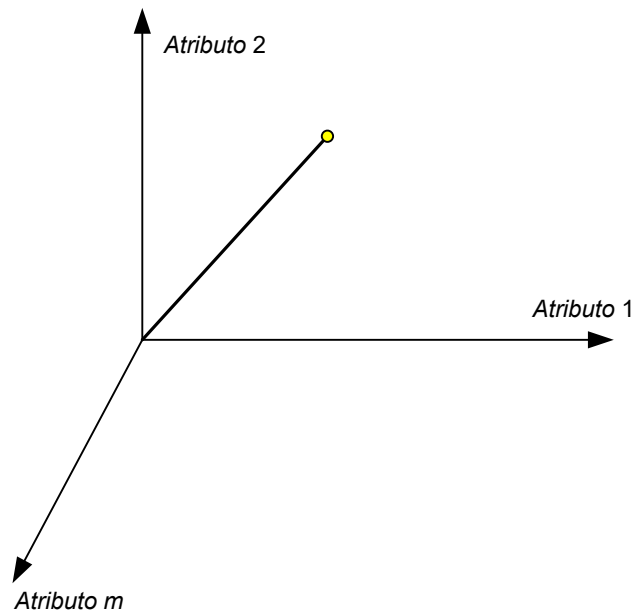


Ilustración 9

En el espacio euclidiano de la Ilustración 9, que llamamos “espacio original”, es difícil poder descubrir, visual o analíticamente, cómo se relacionen los atributos.

Observación de un fenómeno en un espacio euclidiano “alternativo”

Se puede usar el *Análisis de Componentes Principales* (PCA), que es una técnica matemática que permite ver los datos en otro espacio euclidiano “alternativo”, que llamamos, por facilidad, simplemente “espacio alternativo”, con ejes coordenados que llamamos “alternativos”, que están girados con relación a los ejes coordenados “originales”, manteniendo el mismo origen. El primer eje coordenado alternativo, que llamamos “EjeAlt_1”, recibe la mayor proyección global en promedio de los vectores que representan a los datos. Por eso se considera que este eje tiene “la mayor cantidad de información”.

El segundo eje coordenado alternativo, que llamamos “EjeAlt_2”, recibe la segunda mayor proyección global en promedio de los vectores que representan los datos. Por eso se considera que este eje tiene “la segunda mayor cantidad de información”. El eje EjeAlt_2 es perpendicular al eje EjeAlt_1.

De forma similar, se generan los restantes $m - 2$ ejes coordenados alternativos.

Si se desechan la mayoría de los ejes coordenados alternativos, para crear una gráfica de un espacio alternativo formado solamente con los dos o tres primeros ejes coordenados alternativos, y dibujando ahí las proyecciones de los vectores que representan a los datos, puede ser mucho más fácil observar la relación entre los atributos, sabiendo que los primeros ejes coordenados alternativos mantienen una cantidad “significativa” de toda la información (hay que calcularlo).

Ya que los ejes coordenados alternativos no representan a atributo alguno, es muy importante observar a los ejes coordenados originales (que sí representan a los atributos)

proyectados en la gráfica del espacio alternativo formado con los dos o tres primeros ejes coordenados alternativos, ya que con esto se puede descubrir de forma “relativamente” fácil las relaciones entre atributos, si las hubiese. Las proyecciones de los ejes coordenados originales no tienen por qué ser perpendiculares, es más, si las proyecciones de algunos ejes coordenados originales tienden a ser paralelas, los atributos a los cuales representan pueden tener una relación importante pues si hay crecimiento con relación a uno, lo hay aproximadamente también con relación a otro. Una gráfica así se llama “Biplot”.

La Ilustración 10 representa una gráfica del espacio alternativo considerando tan solo los ejes coordenados alternativos que tienen más información. Ahí se dibuja también las proyecciones de los ejes coordenados originales (por facilidad se dibuja solamente las proyecciones de dos ejes coordenados originales), y las proyecciones de los vectores de los datos (por facilidad se dibujan solamente la proyección de un eje coordenado original).

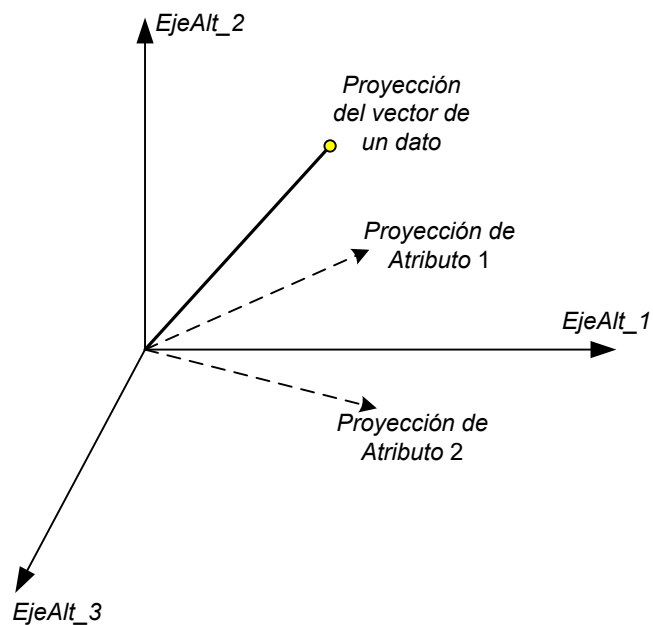


Ilustración 10

La Ilustración 11 muestra un Biplot bidimensional, tomado de [1] (Fig. 4). Ahí se muestra la proyección de 34 ejes coordenados originales, que representan a 34 atributos siendo C4 el eje de mayor importancia, y se muestra la proyección de 2485 los vectores correspondientes a 2,485 datos. No se observa relación clara entre los ejes coordenados (es decir entre los atributos).

El Biplot y el análisis PCA se realiza mediante análisis computacional de datos, apoyado en algoritmos y recursos obtenidos con bibliotecas de análisis, usando Python, un lenguaje de programación de alto nivel, interpretado y de código abierto.

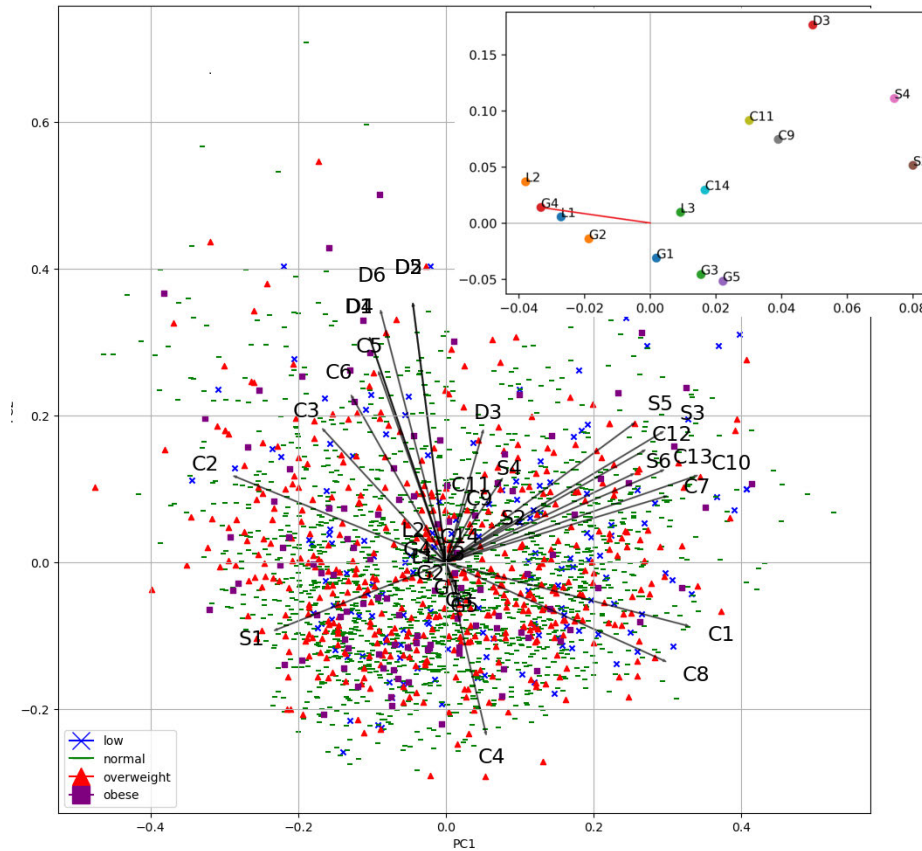


Ilustración 11. Biplot tomado de [1], Fig. 4.

Observación del fenómeno en un espacio euclidiano “alternativo” con orientación novedosa

Mateos et al [1] muestran que es posible buscar una solución diversa al Biplot convencionalmente propuesto, que usa los ejes coordenados alternativos que tienen más información. La propuesta novedosa es usar los ejes coordenados alternativos en donde tenga mayor proyección el eje coordenado original correspondiente al atributo de interés. Eso mismo se hizo en [1], obteniendo un Biplot ([1] Fig. 7), que se muestra en la Ilustración 12, y que contrasta con el Biplot mostrado en la Ilustración 11 (con orientación *digamos* “tradicional”). Este Biplot se presenta en la Ilustración 12, donde se observa claramente que los atributos más influyentes en G4 son G2, D3, C3, los no relacionados son L2, S5, C6, y los más contrapuestos son C7, S2 y S1.

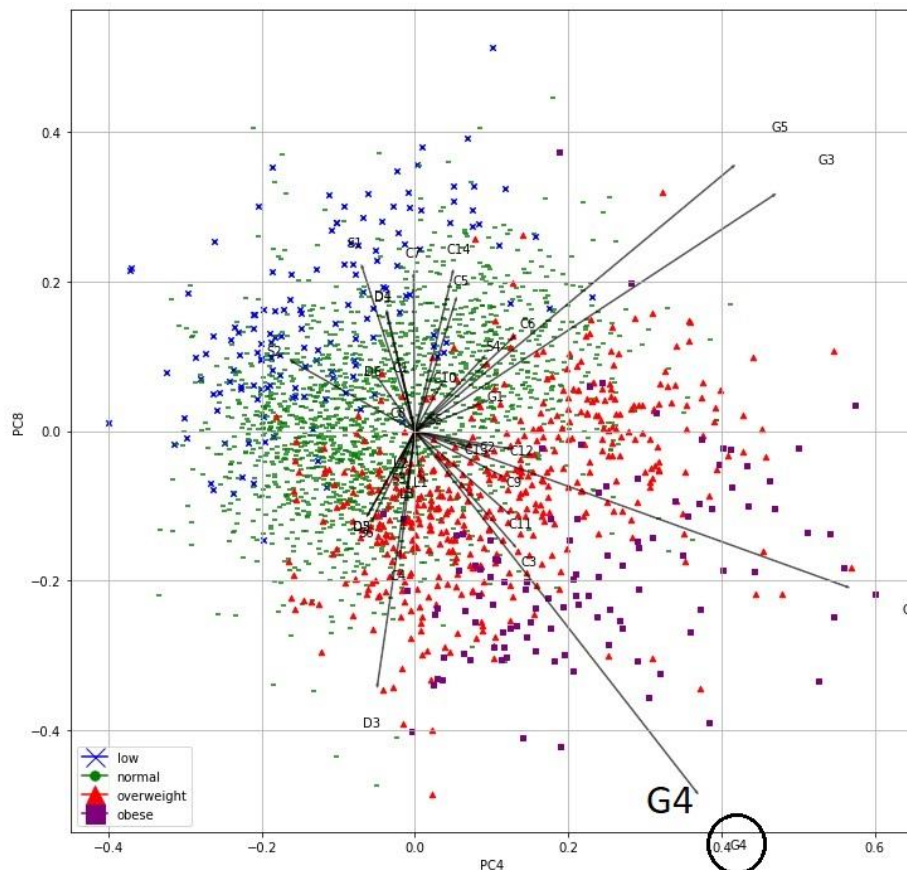


Ilustración 12. Biplot tomado de [1], Fig. 7.

2.11. Estado del arte del problema de investigación: Estudios académicos en torno al objeto de estudio

Para encontrar las relaciones entre los atributos de un conjunto de datos, se realiza lo que a veces se llama *análisis multivariado*, lo que permite descubrir patrones de interrelación entre atributos, aunque no sean fáciles de obtener (por lo que se dice que “están ocultos”).

Para esta búsqueda de patrones, se pueden usar herramientas como [13]:

- 1- Diagramas de dispersión. Éstas son herramientas de visualización gráfica que pueden favorecer la identificación visual de relaciones entre dos variables numéricas.
- 2- Diagramas de pares. Son múltiples diagramas de dispersión, cada uno relativo a dos variables numéricas.
- 3- Tabla cruzada. Esta tabla muestra mismas variables situados sus nombres de forma ordenada tanto en la primera columna como en el primer renglón (el encabezado), estando los valores de interacción entre las variables en el cuerpo de la tabla. Esta forma proporciona una forma sencilla de observar la interacción entre variables.
- 4- Mapa de calor: los mapas de calor presentan visualmente valores en las tablas cruzadas mediante cuadritos coloreados, donde la magnitud de los valores se asocia con

colores. Esta forma puede ayudar visualmente a observar patrones o correlaciones entre variables.

- 5- Mapa de agrupamientos. Estos mapas se aplican sobre los mapas de calor, aplicando algoritmos de agrupamiento a los cuadritos, revelando estructuras y agrupaciones más profundas.
- 6- Gráfica de barras (con intervalos de confianza). Estos gráficos resumen la tendencia del valor, por separado para cada variable, y pueden incluir líneas (sobre la barra) que indican intervalos de confianza sobre la variabilidad.
- 7- Gráfico de caja. Estos muestran la distribución de cada variable en diferentes categorías, destacando la mediana, los cuartiles y los valores atípicos.
- 8- Diagramas de distribución. Con éstos se dibuja la distribución de valores (o diagramas de densidad) para diversas variables, y pueden ayudar a comparar las distribuciones de diversas variables.
- 9- Análisis de Componentes Principales (*Principal Component Analysis –PCA*). Es el método que usamos en el presente proyecto de investigación.

PCA es una alternativa al análisis del tipo indicado en los puntos anteriores. El PCA Se usa actualmente de forma extensiva para el análisis de diversos tipos de datos, como por ejemplo los datos biológicos o sociales. Con PCA se transforman los datos originales, representándolos en espacios euclidianos con ejes coordenados alternativos, donde el primer eje alternativo recaba la mayor información de los datos [14], el segundo eje alternativo la segunda mayor información. Se desechan los ejes coordenados que no sean los dos o tres primeros con el objeto de simplificar la manera en que se observan los datos y con esto simplificar el análisis. En la literatura sobre PCA se recomienda seleccionar los PC con más información [4] [5] [6].

Los participantes de este proyecto, al usar PCA han planteado, como algo novedoso [1], no necesariamente usar los ejes coordenados con mayor información para la transformación de los datos, sino los ejes coordenados donde se proyecta mayormente (donde refleja su mayor información) el atributo de interés, y este tipo de propuesta de uso de PCA no se ha encontrado en otros sitios, al menos los participantes de este proyecto no lo han encontrado. Con esta orientación novedosa de transformación PCA, los participantes de este proyecto han tenido buenos resultados [1].

Se espera que con este proyecto se continúen obteniendo resultados exitosos, aplicado a problemas primordialmente de salud pública.

2.12. Alcances y limitaciones del proyecto (establecidos de forma preliminar) con relación directa al problema que se aborda

Todo el trabajo de este proyecto depende de la disponibilidad de datos apropiados relacionados con los problemas que se quieran estudiar.

La cantidad de relaciones a estudiar está limitada por el poder computacional y el tiempo del que se dispone, previendo que no sean más de una centena de atributos que estén

considerados entre los cuales se estudie su relación. El número de datos obtenidos en un experimento hasta ahora fue de “unos” 2,500 (dos mil quinientos) [1]. Se prevé que el número de datos pudiese crecer no más de en una orden de magnitud (10 veces), pero no es claro si pudiese por este proyecto manejarse una cantidad mayor.

La posibilidad de encontrar relaciones entre atributos no se garantiza, pues se prevé trabajar con problemas en donde es difícil obtener relaciones. En toda investigación seria, real, la obtención de éxito en la búsqueda de soluciones no está asegurada, aunque se tienen expectativas reales de que sí se pueda lograr.

2.13. Metodología. (estrategia de investigación y descripción del diseño de investigación, así como la selección de técnicas e instrumentos de investigación)

La metodología empleada en este proyecto incluye una combinación de técnicas cualitativas y cuantitativas. A continuación, se detallan los pasos seguidos.

- Investigación. Recolección de datos a través de encuestas, entrevistas y análisis de documentos relevantes.
- Generación de tablas de datos apropiadas, reduciendo datos incompletos, incorrectos y repetidos. Selección de datos prioritarios y deshecho de datos de interés secundario. Identificación de atributos de interés o de aquellos de mayor significancia en el proyecto.
- Exploración. Uso de herramientas especializadas, como métodos de análisis (PCA y agrupamiento), y aplicación de programas en el lenguaje interpretado Python, con el aprovechamiento de su vasta cantidad de bibliotecas afines al proyecto, para analizar los datos recopilados.
- Análisis. Revisión de los resultados obtenidos. En caso de PCA revisión de los datos puestos en Biplot, mediante métodos visuales y analíticos, para encontrar las relaciones entre atributos buscadas.
- Prueba de la efectividad del método. La efectividad del método PCA con la orientación novedosa aquí planteada puede evaluarse mediante la participación de expertos (como ya se ha hecho en [1]), aunque esto solamente opera para datos que ya se han estudiado con otros métodos y hay resultados evaluados.
- Obtención de conclusiones.
- Reporte. Incorporación de los conocimientos relativos y explicativos del proyecto, en todas sus etapas, incluyendo conclusiones.
- Publicación (socialización). Se procede a “tratar” de socializar los reportes del proyecto.

2.14. Referencias y fuentes de consulta

Bibliografía

- [1 A. P. Mateos-Papis, C. Sánchez-Sánchez, H. Jiménez-Salazar, N. Guerrero-Vargas, A. M. Ángeles-Castellanos y C. Escobar, «Exploración de representaciones para identificar relaciones entre atributos,» *RESEARCH IN COMPUTING SCIENCE*, vol. 151, nº 10, 2022.
- [2 C. Sánchez-Sánchez y A. P. Mateos-Papis, «A method for detecting recurring types of projects using clustering,» de *Encuentro Nacional de Computación Eureka ENC 2024*, Ciudad Madero, Tamaulipas. México, 2024.
- [3 C. Sánchez-Sánchez, A. P. Mateos-Papis, N. N. Guerrero-Vargas, A. M. Ángeles-Castellanos y C. Escobar, «Feature Selection and Classification for Searching Light at Night Exposure and Students' Weight Relationship,» de *Advances in Soft Computing*, Cham, 2024.
- [4 I. T. Jolliffe y J. Cadima, «Principal component analysis: a review and recent developments,» *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, p. 20150202, 2016.
- [5 J. Shlens, «A Tutorial on Principal Component Analysis,» *CoRR*, vol. abs/1404.1100, 2014.
- [6 C. R. Shalizi, «Advanced Data Analysis from an Elementary Point of View (Principal Component Analysis chapter),» Carnegie Mellon University, 22 03 2021. [En línea]. Available: <https://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/>. [Último acceso: 18 10 2022].
- [7 National Aeronautics and Space Administration, «Imagine the Universe. Data Archives.,» High Energy Astrophysics Science Archive Research Center (HEASARC), 11 12 2018. [En línea]. Available: https://imagine.gsfc.nasa.gov/observatories/data/data_archives.html. [Último acceso: 11 05 2025].
- [8 National Cancer Institute, «Exploring and Analysing Data: The Basics,» National Cancer Institute. Center for Biomedical Informatics & Information Technology, 20 12 2023. [En línea]. Available: <https://datascience.cancer.gov/training/learn-data-science/explore-analyze-data-basics>. [Último acceso: 21 04 2025].
- [9 K. Ortega, «¿Cuántos tipos de análisis de datos existen?,» Saint Leo University, 11 02 2025. [En línea]. Available: <https://worldcampus.saintleo.edu/blog/cuantos-tipos-de-analisis-de-datos-hay-cuales-son-los-tipos-de-analisis-de-datos>. [Último acceso: 12 04 2025].

- [1 Organización Internacional de Normalización, «Guía sencilla para entender el análisis de 0] datos sanitarios,» [En línea]. Available: <https://www.iso.org/es/asistencia-sanitaria/analisis-datos-sanitarios#:~:text=El%20an%C3%A1lisis%20de%20datos%20cl%C3%ADnicos,los%20flujos%20de%20trabajo%20cl%C3%ADnicos..>
- [1 LaSalud.mx, «Principales retos que enfrenta la salud pública en México para 2030,» 1] Mundodehoy.com, 27 1 2025. [En línea]. Available: <https://lasalud.mx/2025/01/27/principales-retos-que-enfrenta-la-salud-publica-en-mexico-para-2025/>. [Último acceso: 12 04 2025].
- [1 G. W. Stewart, «On the Early History of the Singular Value Decomposition,» *SIAM Review*, 2] vol. 35, p. 551–566, 1993.
- [1 N. Fatima, «Multivariate Analysis: Exploring Relationships Between Variables,» Medium, 3] 25 06 2024. [En línea]. Available: <https://medium.com/@noorfatimaafzalbutt/multivariate-analysis-exploring-relationships-between-variables-87a3363bd320>. [Último acceso: 14 02 2025].
- [1 I. Borg and P. J. F. Groenen, *Modern Multidimensional Scaling. Theory and Applications*, 4] Ney York: Springer, 2005.
- [1 Sea Bird Scientific, «Sea Bird University,» Sea Bird Scientific, [En línea]. Available: 5] <https://www.seabird.com/about-sea-bird>; <https://www.seabird.com/sbu>. [Último acceso: 22 05 2025].
- [1 Consejo Nacional de Humanidades, Ciencia y Tecnología, «Programas Nacionales 6] Estratégicos,» Secretaría de Ciencia, Humanidades, Tecnología e Innovación, 2025. [En línea]. Available: <https://secihti.mx/pronaces/>. [Último acceso: 22 05 2025].
- [1 O. d. N. Unidas, «Objetivos de Desarrollo Sostenible,» Organización de Naciones Unidas, 7] [En línea]. Available: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>. [Último acceso: 22 04 2025].

3. Planeación y requerimientos

3.1. Productos de Investigación o metas

Generación de un reporte de investigación anual (dos anuales) y uno final.

Se pretende realizar al menos una publicación anual durante el tiempo que dure el proyecto, ya sea en revista o en conferencia con memorias. Para cada pretensión de publicación se

genera un reporte técnico de publicación, en donde se incluye la copia del trabajo enviado a revisión para publicación. Debe quedar claro que las publicaciones no se pueden asegurar pues eso depende de los tiempos y de la actuación y evaluación de personas externas al proyecto.

Un Proyecto Terminal relativo a este tema de un alumno de licenciatura, ya sea de LIC o de LTSI. Alfredo Mateos actualmente tiene dos alumnos de PT trabajando en PCA, uno de LTSI y otro de LIC. También debe quedar claro que la generación del PT exitoso depende no solamente de los participantes de este proyecto, sino de los alumnos. De cualquier forma se presentan los reportes de PT como parte de los productos de esta proyecto.

3.2. Cronograma de actividades

El cronograma de actividades de los productos de investigación a desarrollar anualmente en este proyecto se da en la Tabla 3.

Tabla 3

Año	Trimestre	Productos de Investigación
1	1	Conclusión de la integración por escrito (ya está avanzada) de la metodología y antecedentes matemáticos y de programación que se requiera comunicar (socializar) para dar entender este proyecto a nuevos participantes, o para dejar un reporte que ayude a un mejor control y revisión. Obtención de datos en diversas fuentes (como en la Facultad de Medicina de la UNAM), o datos oceanográficos, o de salud. Realización de experimentación con datos obtenidos. Generación de reporte técnico con los resultados obtenidos (se pretende que se convierta en publicación de artículo de revista o de congreso). Generación de reporte anual de trabajo.
	2	
	3	
2	1	Continuidad de obtención de datos en diversas fuentes, para análisis. Realización de experimentación con datos obtenidos. Generación de reporte técnico con los resultados obtenidos (se pretende que se convierta en publicación de artículo de revista o de congreso). Generación de reporte anual de trabajo. Generación de reporte final de trabajo.
	2	
	3	

3.3. Requerimientos y justificación de los recursos necesarios

No se plantean requerimientos financieros, otros que los que habitualmente otorga el Departamento de Tecnologías de Información a sus miembros, que han sido, aproximadamente \$32,000 pesos anuales por persona. Los gastos adicionales que se puedan requerir se aportarían de forma personal por los participantes de este proyecto; o se buscaría apoyo externo (que por ahora no se prevé).

4. Vinculación con los planes y programas de estudio de la División y la Unidad

4.1. Planes y programas de estudio de la División de Ciencias de la Comunicación y Diseño (DCCD)

Para la licenciatura en Tecnologías y Sistemas de la Información (LTSI) el material de este proyecto puede completar los programas de estudio de las UEA optativas siguientes (varía según el profesor)⁹:

- 450228, Inteligencia Artificial II. Horas Teoría: 3; Horas Práctica: 2.
- 450232, Seminario de Sistemas Inteligentes I. Horas Teoría: 3; Horas Práctica: 2.
- 450233, Seminario de Sistemas Inteligentes II. Horas Teoría: 3; Horas Práctica: 2.

También en LTSI están las UEA obligatorias de Proyecto Terminal (PT). Actualmente hay dos alumnos de PT a cargo del Dr. Alfredo Piero Mateos Papis, tanto de LTSI como de LIC (Licenciatura en Ingeniería en Computación) que están cursando PT con el tema de este mismo proyecto de investigación:

- 2173033864 AMBROSIO LUCAS JUANA de LTSI
- 2203066382 JACOBO RAMIREZ NOE de LIC.

Las UEA de PT son:

- 450218 Proyecto Terminal I. Horas Teoría: 3; Horas Práctica: 6.
- 450219 Proyecto Terminal II. Horas Teoría: 2; Horas Práctica: 8.
- 450220 Proyecto Terminal III Horas Teoría: 2; Horas Práctica: 8.

Dentro de la disciplina de Ciencias de la Comunicación, cultivada en DCCD tanto a nivel licenciatura como en una ruta de estudio en la Maestría en Diseño, Información y Comunicación (MADIC), se puede dar un ejemplo de análisis de datos cuantitativos, como es el análisis relativo a emisiones de audio, imagen, o documento escrito, de contenido informativo diverso, donde cada emisión o documento se considera un individuo, en el análisis, y cada individuo tiene los siguientes atributos (se quiere conocer la relación entre los atributos):

- 1- Frecuencia de una o más palabras clave.
- 2- Categorías temáticas.
- 3- Categoría en lenguaje utilizado.
- 4- Número de usuarios o audiencia.
- 5- Número de "Clicks", o de llamadas, o de eventos de participación.
- 6- Alcance geográfico.
- 7- Categoría de edad de usuarios o audiencia.
- 8- Categorías de ubicación geográfica de usuarios o audiencia.

⁹ [http://dccd.cua.uam.mx/Licenciatura en Tecnologías y Sistemas de Información](http://dccd.cua.uam.mx/Licenciatura_en_Tecnologias_y_Sistemas_de_Informacion)

9- Categorías de la satisfacción o de la percepción de parte de los usuarios o audiencia.

4.2. Planes y programas de estudio de la División de Ciencias Naturales e Ingeniería (CNI)

Para la Licenciatura en Computación (LIC) el material de este proyecto puede completar los programas de estudio de las UEA optativas siguientes:

- 460043, Minería de Datos. Créditos: 8.

También en LIC están las UEA obligatorias de Proyecto Terminal (PT):

- 4604094 Proyecto Terminal I. Horas Teoría: 2; Horas Práctica: 6.
- 4604095 Proyecto Terminal II. Horas Teoría: 2; Horas Práctica: 6.
- 4604096 Proyecto Terminal III. Horas Teoría: 2; Horas Práctica: 6.

Para el posgrado en Ciencias Naturales e Ingeniería (PCNI), este proyecto podría apoyar en la las siguientes UEA optativas:

- 4607027 Computación Científica. Horas Teoría: 4; Horas Práctica: 2.
- 4607049 Temas Selectos I. Horas Teoría: 4; Horas Práctica: 2.
- 4607050 Temas Selectos II. Horas Teoría: 4; Horas Práctica: 2.
- 4607051 Temas Selectos III. Horas Teoría: 4; Horas Práctica: 2.
- 4607052 Temas Selectos IV. Horas Teoría: 4; Horas Práctica: 2.
- 4607053 Temas Selectos V. Horas Teoría: 4; Horas Práctica: 2.

4.3. Vinculación Institucional

Con la Facultad de Medicina de la UNAM, en Ciudad Universitaria, Ciudad de México se ha desarrollado ya trabajo conjunto con el mismo tema de este proyecto [1]. Está previsto en el desarrollo de este proyecto continuar trabajando con los colegas de la Facultad de Medicina para profundizar y ampliar los resultados. El interés de la Facultad de Medicina es encontrar la relación entre el Índice de Masa Corporal (IMC)¹⁰ de las personas, con las costumbres de alimentación y de interrupción del ciclo circadiano de las personas.

Con miembros del Instituto de Ciencias del Mar y Limnología de la UNAM, se ha hablado de su interés en este proyecto para ayudar a encontrar la relación entre (el atributo de) la cantidad de oxígeno en el mar, con los atributos de temperatura y de profundidad, y con otros atributos asociados con la existencia de organismos o de sustancias diversas. Este interés tiene relación con el proyecto *Sea-Bird University* de la empresa *Sea-Bird Scientific*¹¹ que, como esta empresa lo indica, “es el mayor desarrollador y fabricante mundial de productos para la medición de salinidad, temperatura, presión, oxígeno disuelto, fluorescencia, nutrientes y parámetros oceanográficos relacionados en aguas marinas” [15].

¹⁰ Se calcula como el peso entre la altura de la persona.

¹¹ Empresa con sede en Bellevue, WA 98005, Estados Unidos.

Los participantes de este proyecto de investigación tienen interés en explorar, también, la relación que pueda existir entre categorías de malestares o enfermedades, con relación a exposición a situaciones ambientales diversas.

En este proyecto de investigación se pueden aceptar problemas de otro tipo, de interés general en donde haya datos disponibles, suficientes y con formato apropiado para su análisis mediante los métodos aquí planteados.

5. Incidencia social

5.1. Con relación a los Programas Nacionales Estratégicos

Este proyecto de investigación puede tener incidencia en prácticamente todos los Programas Nacionales Estratégicos (PRONACES) planteados por el CONAHCYT [16]. Los PRONACES son 10:

- Agentes Tóxicos y Procesos Contaminantes
- Agua
- Cultura
- Educación
- Energía y Cambio Climático
- Salud
- Seguridad Humana
- Sistemas Socio-Ecológicos
- Soberanía Alimentaria
- Vivienda

Por ejemplo, el programa: “Agentes Tóxicos y Procesos Contaminantes” impulsa los derechos colectivos relativos a la salud, al medio ambiente, y al territorio, mediante el mejoramiento de las condiciones “salubres” y “de calidad de vida y bienestar” de las comunidades. Esto incluye el esfuerzo y el trabajo por restaurar los ecosistemas ya dañados.

Parte de la problemática que aborda este programa se indica como sigue: “Las numerosas fuentes de contaminación como minas, cementeras, agroindustrias, empresas manufactureras, corredores urbano-industriales y aguas residuales sin tratamiento afectan regiones en las que se pierden ecosistemas y recursos básicos para el sustento de la salud humana y para una vida digna”. Justamente el análisis de los datos que se hace en este proyecto de investigación puede analizar datos que por persona (individuo) indiquen su lugar de residencia, de trabajo, sus datos personales como edad, peso, y otros, e indique sus condiciones de salud o enfermedades existentes, y asocie como atributos las fuentes de contaminación existentes en los lugares de residencia o trabajo, con datos de temporalidad en la existencia de todos estos datos.

5.2. Con relación a los Objetivos de Desarrollo Sostenible de la ONU

La Organización de las Naciones Unidas (ONU –UNO) plantea 17 objetivos de desarrollo sostenible [17]. Cada uno de estos objetivos ya tiene metas planteadas, que parecen (al menos a entendimiento de los participantes de este proyecto de investigación) estar basadas en el cumplimiento de acciones de trabajo físico, como por ejemplo, las metas del objetivo #7 “Energía Asequible y No Contaminante” incluyen:

1. Garantizar acceso a servicios energéticos.
2. Aumentar la proporción de energía renovable.
3. Duplicar la tasa de mejora de eficiencia energética.
4. Aumentar la cooperación internacional.
5. Ampliar infraestructura.

Siempre que los objetivos de desarrollo planteados se acompañen de búsqueda exploratoria informativa con relación a datos disponibles, este proyecto es de ayuda. Nosotros, los participantes de este proyecto de investigación encontramos que sí existe esta búsqueda ya que la ONU produce datos y cifras como por ejemplo, la ONU patrocina al “Intergovernmental Panel on Climate Change” que produce estudios diversos, como estudios de emisiones de CO_2 . Se da como ejemplo el estudio relacionado con emisiones utilizada en Edificios (<https://www.ipcc.ch/report/ar6/wg3/chapter/chapter-9/>).



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Unidad Cuajimalpa

Ciudad de México, a 22 de mayo de 2025

Comisión de Investigación
División de Ciencias de la Comunicación y Diseño
Universidad Autónoma Metropolitana, Unidad Cuajimalpa

ASUNTO: Carta Compromiso

Por medio de la presente manifiesto mi acuerdo y compromiso de trabajo con el proyecto de investigación titulado "Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa" en el cual trabajo como coordinador.

Me comprometo a dedicar cuatro horas a la semana a actividades de este proyecto durante el trimestre 25-P.

A partir del trimestre 25-O tengo planeado el goce de un periodo sabático, que he solicitado. En este periodo me comprometo a dedicar al menos diez horas a la semana a actividades de este proyecto, hasta su conclusión.

Agradezco su atención a la presente.

ATENTAMENTE

Dr. Alfredo Piero Mateos Papis
Departamento de Tecnologías de Información
División de Ciencias de la Comunicación y Diseño
Unidad Cuajimalpa

Ciudad de México, 14 de mayo de 2025

Comisión de Investigación
División de Ciencias de la Comunicación y Diseño
Universidad Autónoma Metropolitana, Unidad Cuajimalpa
PRESENTE

ASUNTO: Carta Compromiso

Por medio de la presente, el que suscribe manifiesta su acuerdo y compromiso de trabajo con el proyecto de investigación titulado "Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa" en el cual seré colaborador. Participaré particularmente en las actividades relacionadas con la programación de algunos experimentos, y en la comparación y análisis de los resultados. Dedicaré dos horas a la semana a esta actividad.

Sin más por el momento, les envío un cordial saludo.

ATENTAMENTE

Dr. Christian Sánchez Sánchez
Departamento de Tecnologías de la Información
DCCD
Unidad Cuajimalpa



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Cuajimalpa

Ciudad de México, 16 mayo 2025

Comisión de Investigación
División de Ciencias de la Comunicación y Diseño
Universidad Autónoma Metropolitana, Unidad Cuajimalpa
PRESENTE
P r e s e n t e

ASUNTO: Carta compromiso del Dr. Carlos Joel RIVERO MORENO en el proyecto de investigación.

Por medio de la presente, quien suscribe manifiesta su acuerdo y compromiso de participar en el proyecto de investigación titulado "**Búsqueda de relaciones entre atributos mediante Análisis de Componentes Principales, con una orientación novedosa**".

Dedicaré dos horas a la semana a esta actividad.

Sin más por el momento, les envío un cordial saludo.

Atentamente,

Casa abierta al tiempo,

Dr. Carlos Joel Rivero Moreno

Profesor Investigador del Departamento de Tecnologías de la Información

División de Ciencias de la Comunicación y Diseño

Universidad Autónoma Metropolitana. Unidad Cuajimalpa

Número económico:

Email: crivero@cua.uam.mx

DTI

Departamento
de Tecnologías
de la Información

Unidad Cuajimalpa

DCCD | División de Ciencias de la Comunicación y Diseño

Departamento de Tecnologías de la Información

Torre III, 5to. piso. Av. Vasco de Quiroga 4871,

Colonia Santa Fe Cuajimalpa, Alcaldía Cuajimalpa de Morelos.

C.P. 05348, Ciudad de México.

Tel.: (+52) 55.5814.6557

<http://dccd.cua.uam.mx>