

Jhon Wilber Ajata Ascarrunz

Análisis documental y predictivo de las causas de defunción en Región de Puno utilizando técnicas de

 Universidad Nacional del Altiplano

Detalles del documento

Identificador de la entrega

trn:oid::8254:412968766

Fecha de entrega

5 dic 2024, 1:12 p.m. GMT-5

Fecha de descarga

5 dic 2024, 1:15 p.m. GMT-5

Nombre de archivo

Formato de proyecto de tesis -b Finesi.docx

Tamaño de archivo

3.3 MB

9 Páginas

3,724 Palabras

22,985 Caracteres




27% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 10 palabras)

Fuentes principales

- 22%  Fuentes de Internet
- 7%  Publicaciones
- 20%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Fuentes principales

- 22% Fuentes de Internet
- 7% Publicaciones
- 20% Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	Internet	
vriunap.pe		4%
2	Trabajos entregados	
Ho Chi Minh University of Technology and Education on 2023-10-16		4%
3	Trabajos entregados	
unap on 2023-12-21		3%
4	Internet	
www.researchgate.net		3%
5	Internet	
www.coursehero.com		2%
6	Trabajos entregados	
Charles University on 2024-07-30		1%
7	Trabajos entregados	
unap on 2022-11-29		1%
8	Trabajos entregados	
		0%
9	Trabajos entregados	
National University College - Online on 2024-11-23		0%
10	Internet	
acikbilim.yok.gov.tr		0%
11	Internet	
es.slideshare.net		0%

12	Trabajos entregados	Universidad Europea de Madrid on 2024-09-29	0%
13	Trabajos entregados	dyci on 2024-12-02	0%
14	Trabajos entregados	Universidad San Francisco de Quito on 2023-11-06	0%
15	Internet	www.bandes.com.br	0%
16	Trabajos entregados	ITESM: Instituto Tecnológico y de Estudios Superiores de Monterrey on 2024-10-06	0%
17	Trabajos entregados	Universidad Internacional de la Rioja on 2024-06-26	0%
18	Trabajos entregados	Universidad Tecnológica Centroamericana UNITEC on 2024-09-09	0%
19	Internet	repositorio.espe.edu.ec	0%
20	Internet	upc.aws.openrepository.com	0%
21	Trabajos entregados	Centro Europeo de Postgrado - CEUPE on 2024-07-25	0%
22	Trabajos entregados	National University College - Online on 2024-03-22	0%
23	Trabajos entregados	Universidad Anahuac México Sur on 2024-10-13	0%
24	Internet	www.drsl.info	0%
25	Internet	www.lareferencia.info	0%

26	Internet	1library.co	0%
27	Trabajos entregados	Corporación Universitaria Minuto de Dios, UNIMINUTO on 2023-10-13	0%
28	Trabajos entregados	Corporación Universitaria Minuto de Dios,UNIMINUTO on 2024-11-04	0%
29	Trabajos entregados	Universidad Internacional de la Rioja on 2024-05-20	0%
30	Trabajos entregados	Universidad Tecnológica Centroamericana UNITEC on 2024-06-24	0%
31	Internet	core.ac.uk	0%
32	Internet	domstat.net	0%
33	Internet	idoc.pub	0%
34	Internet	repositorio.minsa.gob.pe	0%
35	Internet	repositorio.unap.edu.pe	0%
36	Internet	repository.javeriana.edu.co	0%
37	Internet	senegal.luxdev.lu	0%
38	Trabajos entregados	unap on 2022-02-04	0%
39	Internet	www.archbronconeumol.org	0%

40 Internet

www.dykinson.com 0%

41 Internet

www.scribd.com 0%

PROYECTO DE INVESTIGACIÓN

(EL PROYECTO DEBE CONTENER UN MÁXIMO DE 15 HOJAS)
(Retire todas las marcas guía en paréntesis y la descripción de los ítems)

- I. **Título** (El proyecto de tesis debe llevar un título que exprese en forma sucinta su contenido, haciendo referencia en lo posible, al resultado final que se pretende lograr. Máx. palabras 25)

Análisis Documental y Predictivo de las Causas de Defunción en la Región de Puno utilizando Técnicas de Machine Learning y Estadística Multivariada

- II. **Resumen del Proyecto de Tesis** (Debe ser suficientemente informativo, presentando -igual que un trabajo científico- una descripción de los principales puntos que se abordarán, objetivos, metodología y resultados que se esperan)

El presente proyecto de tesis tiene como finalidad realizar un análisis exhaustivo de las causas de defunción en la región de Puno, utilizando la base de datos del Sistema Informático Nacional de Defunciones (SINADEF). Como estudiante de ingeniería estadística e informática, reconozco el gran potencial que tienen los análisis estadísticos para generar información valiosa a partir de datos masivos, así como el poder de la informática y el *machine learning* para crear modelos predictivos precisos.

El enfoque metodológico de esta investigación es de naturaleza documental, estadística y predictiva. Se emplearán técnicas de estadística multivariada y algoritmos de *machine learning* para identificar patrones, factores asociados y realizar proyecciones sobre la mortalidad en Puno. El análisis abarcará variables sociodemográficas (edad, sexo, estado civil, nivel de instrucción, etnia), geográficas (departamento, provincia, distrito), socioeconómicas (ocupación, tipo de seguro) y clínicas (tipo de lugar de defunción, muerte violenta, necropsia).

Los objetivos específicos incluyen. **Describir** las características sociodemográficas y geográficas de los fallecidos en Puno registrados en SINADEF. **Identificar** las principales causas de defunción en la región y su distribución. **Analizar** las relaciones entre las variables independientes y las causas de muerte para determinar factores de riesgo significativos. **Desarrollar** modelos predictivos que permitan anticipar patrones de mortalidad y causas de defunción, contribuyendo a la toma de decisiones en salud pública.

La metodología contempla un análisis estadístico descriptivo para caracterizar la población de estudio, seguido de análisis multivariado y la aplicación de técnicas de *machine learning* como árboles de decisión, bosques aleatorios y redes neuronales, para identificar patrones y construir modelos predictivos robustos. Se utilizarán herramientas como Python y R para el procesamiento y análisis de datos.

Se espera que los resultados proporcionen un modelo predictivo capaz de predecir las causas de defunción en función de las características sociodemográficas y geográficas. Este modelo permitirá a las autoridades sanitarias diseñar estrategias preventivas y políticas de salud adaptadas al contexto local de Puno. Además, el estudio contribuirá al avance del conocimiento científico en el área de epidemiología y análisis de datos masivos aplicados a la salud pública, sirviendo como referencia para investigaciones futuras en otras regiones o contextos similares.

- III. **Palabras claves (Keywords)** (Colocadas en orden de importancia. Máx. palabras: cinco)

- Causas de defunción
- Estadística multivariada
- Machine Learning
- Salud pública
- SINADEF

- IV. **Justificación del proyecto** (Describa el problema y su relevancia como objeto de investigación. Es relevante una clara definición y delimitación del problema que abordará la investigación, ya que temas cuya definición es difusa o amplísima son difíciles de evaluar y desarrollar)

La mortalidad es un indicador esencial para comprender el estado de salud de una población. En la región de Puno, las características demográficas, socioeconómicas y geográficas específicas afectan las dinámicas de mortalidad, pero existen pocos estudios que analicen a profundidad estas

variables. El uso de la base de datos SINADEF permite realizar un análisis detallado y predictivo de las causas de defunción. Este estudio es crucial para diseñar políticas de salud pública adaptadas a la región, mejorar la calidad de vida y reducir la mortalidad evitable mediante estrategias basadas en datos.

V. Antecedentes del proyecto (Incluya el estado actual del conocimiento en el ámbito nacional e internacional. La revisión bibliográfica debe incluir en lo posible artículos científicos actuales, para dejar en claro el conocimiento existente y lo que agregará la Tesis propuesta. Esto es importante para el futuro artículo que resultará como producto de este trabajo)

A nivel internacional, el uso de bases de datos de mortalidad ha demostrado ser efectivo para el análisis de patrones de defunción, utilizando técnicas de machine learning y estadística avanzada. En Perú, estudios previos han explorado causas generales de mortalidad, pero pocos han integrado enfoques predictivos en regiones específicas como Puno, donde las condiciones de vida, altitud y acceso a servicios de salud presentan particularidades significativas. Este proyecto pretende llenar este vacío mediante un análisis innovador y contextualizado.

1. A nivel internacional

Cerda y Valdivia (2021) sostienen que encontrar causalidad en medicina es de gran interés en la investigación, con el fin de generar intervenciones que traten o curen enfermedades. La mayoría de los modelos estadísticos clásicos permiten inferir asociaciones, y solo unos pocos diseños son capaces de demostrar causa y efecto con una metodología adecuada y evidencia sólida. La medicina basada en la evidencia respalda sus hallazgos en modelos que van desde una hipótesis hasta la búsqueda de datos para probarla o descartarla. Esto también se aplica al desarrollo de modelos predictivos que deben ser confiables y producir impacto en la práctica clínica. La gran cantidad de datos almacenados en registros electrónicos de salud y el mayor poder computacional han permitido que las técnicas de machine learning jueguen un papel preponderante en el desarrollo de nuevos análisis predictivos y el reconocimiento de patrones desconocidos con estos modelos computacionales modernos. Estos modelos, junto con el cambio de la perspectiva de datos a información, están siendo cada vez más incorporados en la práctica clínica diaria, proporcionando mayor precisión y rapidez para apoyar la toma de decisiones.

Smith (2022) menciona que la intención de esta revisión es proporcionar bases teóricas y evidencia de cómo estas modernas técnicas computacionales de machine learning han permitido lograr mejores resultados y están siendo ampliamente utilizadas. Este artículo revisará los aspectos más relevantes de la ciencia de datos en salud en América Latina.

Mora y Pineda (2022) en su artículo titulado "Modelos predictivos en salud basados en aprendizaje de máquina (machine learning)", publicado en la *Revista Médica Clínica Las Condes*, analiza el impacto de las técnicas de aprendizaje automático en el desarrollo de modelos predictivos dentro del ámbito de la salud. El autor destaca que el incremento en la disponibilidad de grandes volúmenes de datos clínicos y el avance en la capacidad computacional han permitido la aplicación efectiva de algoritmos de *machine learning* para identificar patrones complejos y realizar predicciones precisas sobre diversos indicadores de salud. El estudio enfatiza la transición de datos brutos a información útil mediante el uso de modelos computacionales modernos, los cuales están siendo cada vez más incorporados en la práctica clínica diaria. Esta integración no solo mejora la precisión y velocidad en la toma de decisiones médicas, sino que también proporciona un soporte significativo para la gestión de riesgos y la implementación de intervenciones preventivas. Mora Pineda subraya que, aunque estos modelos son altamente efectivos en la identificación de patrones, es fundamental garantizar una interpretación adecuada de los resultados para evitar malentendidos y asegurar su correcta aplicación en contextos clínicos.

Además, el autor aborda las implicaciones éticas y prácticas de la incorporación de *machine learning* en la salud, destacando la necesidad de un enfoque equilibrado que combine la experiencia clínica con la precisión de los algoritmos predictivos. Este enfoque es esencial para maximizar el impacto positivo de los modelos predictivos en la mejora de la atención médica y en la reducción de la mortalidad a través de intervenciones basadas en evidencia.

2. A nivel nacional

El Ministerio de Salud del Perú (MINSA, 2018) publicó el "Análisis de las causas de mortalidad en el Perú, 1986-2015", que examina las causas de muerte en el país durante tres décadas. Este informe considera variables como ámbito urbano-rural, regiones naturales, departamentos y condición de pobreza, proporcionando una visión integral de las tendencias de mortalidad y las desigualdades existentes. Los resultados indican disparidades significativas en las causas de muerte entre diferentes regiones y grupos socioeconómicos, resaltando la necesidad de estudios focalizados en áreas específicas como Puno.

Álvarez-Calderón, Ortiz y Rivas (2018) describen en su artículo "Resultados preliminares del fortalecimiento del sistema informático nacional de defunciones en Perú" la implementación y mejoras del SINADEF. Este sistema ha sido clave para mejorar la calidad y cobertura de los datos de mortalidad en el país, permitiendo un registro más preciso y oportuno de las defunciones. La disponibilidad de datos más confiables ha facilitado análisis más detallados y confiables sobre mortalidad, lo cual es esencial para investigaciones como la presente.

3. A nivel local

Huata Apaza (2024) realizó un estudio titulado "Clústeres de causas de fallecimiento en la Región Puno 2020-2023", presentado en la Universidad Nacional del Altiplano. Este trabajo aplicó algoritmos de agrupamiento para detectar patrones inusuales de mortalidad en Puno, analizando datos de 31,880 pacientes fallecidos registrados en el Sistema Informático Nacional de Defunciones (SINADEF). Los resultados revelaron que la insuficiencia respiratoria aguda (J96.0) fue responsable del 14% de las defunciones, seguida de septicemia no especificada (A41.9) y COVID-19 (U07.1), cada una representando el 5% de las causas secundarias de fallecimiento. Estos hallazgos evidencian la relevancia de enfermedades respiratorias e infecciosas en la mortalidad regional.

Estos antecedentes evidencian el uso creciente de técnicas de análisis documental, estadística multivariada y machine learning en el estudio de las causas de defunción en la región de Puno y en el Perú en general. Sin embargo, aún existe un vacío en la aplicación de modelos predictivos que integren múltiples variables sociodemográficas y geográficas para anticipar patrones de mortalidad específicos de Puno. El presente proyecto busca llenar este vacío, aportando un enfoque innovador que combine análisis estadísticos avanzados y técnicas de machine learning para proporcionar información valiosa en la formulación de políticas de salud más efectivas y adaptadas al contexto regional.

VI. Hipótesis del trabajo (Es el aporte proyectado de la investigación en la solución del problema)

Las características sociodemográficas y geográficas de los fallecidos en Puno, registradas en la

base de datos de SINAEF, están significativamente asociadas con las causas de defunción. Además, estas variables permiten desarrollar modelos predictivos fiables que anticipen tendencias futuras de mortalidad en la región mediante la aplicación de técnicas de estadística multivariada y aprendizaje automático (machine learning).

VII. Objetivo general

Analizar y predecir las causas de defunción en Puno utilizando datos de SINAEF, aplicando técnicas de estadística multivariada y machine learning para identificar factores asociados y patrones de mortalidad.

VIII. Objetivos específicos

1. Describir las características sociodemográficas y geográficas de los fallecidos en Puno registrados en SINAEF.
2. Identificar las principales causas de defunción en la región.
3. Analizar las relaciones entre variables sociodemográficas y geográficas con las causas de muerte.
4. Desarrollar modelos predictivos para anticipar patrones de mortalidad en Puno.

IX. Metodología de investigación (Describir el(los) método(s) científico(s) que se empleará(n) para alcanzar los objetivos específicos, en forma coherente a la hipótesis de la investigación. Sustentar, con base bibliográfica, la pertinencia del(los) método(s) en términos de la representatividad de la muestra y de los resultados que se esperan alcanzar. Incluir los análisis estadísticos a utilizar)

Diseño de Investigación

El presente estudio seguirá un enfoque **cuantitativo**, con un diseño **no experimental, transversal, descriptivo, correlacional y predictivo**. Este diseño es adecuado para analizar datos existentes sin manipular variables, permitiendo describir las características de la población, explorar relaciones entre variables y desarrollar modelos predictivos sobre las causas de defunción en la región de Puno.

Población y Muestra

- **Población:** Todos los registros de defunciones en la región de Puno registrados en el Sistema Informático Nacional de Defunciones (SINAEF) durante el período **2019-2024**. Este rango temporal permite captar tendencias recientes y contar con un volumen significativo de datos para el análisis.
- **Muestra:** Se utilizará la **totalidad de la población** disponible (muestreo censal), lo que garantiza una alta representatividad y evita sesgos de muestreo. Esto es posible gracias al acceso completo a la base de datos de SINAEF para la región de Puno.

Variables del Estudio

- **Variable Dependiente:**
 - **Causa de Defunción Principal (CAUSA A (CIE-X)):** Clasificada según la **Clasificación Internacional de Enfermedades, 10.ª Revisión (CIE-10)**.
- **Variables Independientes:**
 - **Sociodemográficas:** Edad, sexo, estado civil, nivel de instrucción, etnia, ocupación, tipo de seguro.
 - **Geográficas:** Departamento, provincia, distrito, UBIGEO del domicilio.
 - **Clínicas:** Tipo de lugar de defunción, muerte violenta, necropsia, tiempo de hospitalización, muerte durante (embarazo, parto, puerperio).
 - **Temporales:** Fecha, año, mes y hora de defunción.

Procedimiento

1. Recolección de Datos:

- **Acceso a la Base de Datos:** Se obtendrá la base de datos de defunciones del SINAEF para la región de Puno, correspondiente al período 2019-2024, siguiendo los protocolos y normativas legales y éticas para el uso de datos personales y sensibles.

2. Preparación y Limpieza de Datos:

- **Depuración de Datos:** Se realizará una revisión exhaustiva para identificar y corregir datos faltantes, inconsistencias o errores. Se estandarizarán formatos y unidades de medida (por ejemplo, convertir todas las edades a años).
- **Codificación de Variables:** Las variables categóricas serán codificadas numéricamente para facilitar el análisis estadístico. Se elaborará un diccionario de datos que documente la codificación utilizada.

3. Análisis Descriptivo:

- **Estadística Descriptiva:** Cálculo de medidas de tendencia central (media, mediana) y dispersión (desviación estándar) para variables numéricas. Frecuencias y porcentajes para variables categóricas.
- **Visualización de Datos:** Creación de gráficos (histogramas, gráficos de barras, diagramas de caja) para visualizar la distribución de las variables y detectar patrones iniciales.

4. Análisis Bivariado:

- **Pruebas de Asociación:** Uso de pruebas **chi-cuadrado** para evaluar la asociación entre variables categóricas (e.g., sexo y causa de defunción).
- **Comparación de Medias:** Aplicación de pruebas **t de Student** o **ANOVA** para comparar medias de variables numéricas entre diferentes grupos.

5. Análisis Multivariado:

- **Análisis de Componentes Principales (ACP):** Utilizado para reducir la dimensionalidad de los datos y explorar relaciones entre múltiples variables.
- **Regresión Logística Multinomial:** Modela la relación entre múltiples variables independientes y una variable dependiente categórica con más de dos categorías (causa de defunción).
- **Justificación:** La estadística multivariada permite identificar factores que influyen simultáneamente en las causas de defunción, proporcionando una comprensión más profunda de los patrones de mortalidad (Hair et al., 2010).

6. Modelos Predictivos de Machine Learning:

- **Selección de Algoritmos:** Implementación de **árboles de decisión**, **Random Forest** y **redes neuronales** para construir modelos predictivos.
 - **Árboles de Decisión:** Facilitan la interpretación de las decisiones y variables más influyentes.
 - **Random Forest:** Mejoran la precisión y evitan el sobreajuste al combinar múltiples árboles de decisión (Breiman, 2001).
 - **Redes Neuronales Artificiales:** Capturan relaciones complejas y no lineales entre variables (Goodfellow et al., 2016).

○ Preparación de Datos para Modelado:

- **División de Datos:** Separación del conjunto de datos en **entrenamiento** (70%) y **prueba** (30%).
- **Balanceo de Clases:** Si existe desbalance en las categorías de la variable dependiente, se aplicarán técnicas como **SMOTE** para equilibrar las clases (Chawla et al., 2002).

○ Entrenamiento y Validación de Modelos:

- **Validación Cruzada k-Fold:** Se utilizará una validación cruzada con k=10 para evaluar la consistencia de los modelos y prevenir el sobreajuste.
- **Métricas de Evaluación:** Se analizarán medidas como **precisión**, **recall**, **F1-score** y **AUC-ROC** para seleccionar el modelo más adecuado.

7. Interpretación de Resultados:

- **Importancia de Variables:** Identificación de las variables que más contribuyen a la predicción, lo que ayuda a entender los factores de riesgo asociados a diferentes causas de defunción.
- **Análisis de Patrones:** Interpretación de los patrones identificados en el contexto socioeconómico y cultural de Puno.

8. Ética y Confidencialidad:

- **Protección de Datos Personales:** Se garantizará la anonimización de los datos y el cumplimiento de las normativas legales vigentes, como la **Ley N° 29733 - Ley de Protección de Datos Personales** en Perú.
- **Uso Ético de la Información:** Los resultados serán utilizados exclusivamente con fines académicos y para contribuir a la mejora de la salud pública.

Justificación de la Metodología

- **Representatividad de la Muestra:** Al trabajar con todos los registros disponibles en SINADEF para Puno, se asegura una alta representatividad y la posibilidad de generalizar los hallazgos a la población estudiada.
- **Pertinencia de los Métodos:**
 - **Estadística Multivariada:** Es esencial para analizar múltiples variables simultáneamente y comprender cómo interactúan entre sí (Tabachnick & Fidell, 2013).
 - **Machine Learning:** Las técnicas de aprendizaje automático son adecuadas para manejar grandes volúmenes de datos y pueden descubrir patrones complejos no detectables con métodos tradicionales (Esteve et al., 2017).
- **Resultados Esperados:** Se espera obtener modelos predictivos precisos que identifiquen factores de riesgo y patrones de mortalidad, contribuyendo a estrategias de prevención y políticas de salud más efectivas.

Análisis Estadísticos a Utilizar

1. Estadística Descriptiva:

- **Medidas de Tendencia Central:** Media, mediana, moda.
- **Medidas de Dispersión:** Desviación estándar, rango intercuartílico.
- **Distribuciones de Frecuencia:** Tablas y gráficos para variables categóricas.

2. Estadística Inferencial:

- **Pruebas Chi-Cuadrado:** Para evaluar asociaciones entre variables categóricas.
- **Pruebas t y ANOVA:** Para comparar medias de variables numéricas entre grupos.

3. Estadística Multivariada:

- **Análisis de Componentes Principales (ACP):** Para identificar patrones y reducir dimensionalidad.
- **Regresión Logística Multinomial:** Para modelar relaciones entre variables independientes y una variable dependiente categórica multinomial.

4. Modelos de Machine Learning:

- **Árboles de Decisión:** Para identificar las variables más influyentes en la predicción de causas de defunción.
- **Random Forest:** Para mejorar la precisión y estabilidad del modelo.
- **Redes Neuronales Artificiales:** Para capturar relaciones no lineales y complejas.

5. Validación y Evaluación de Modelos:

- **Validación Cruzada k-Fold:** Para asegurar la generalización de los modelos.
- **Métricas de Rendimiento:** Precisión, recall, F1-score, AUC-ROC.

Herramientas y Software

● Python:

- **Pandas:** Para manipulación y limpieza de datos.
- **Scikit-learn:** Para implementación de algoritmos de machine learning y validación de modelos.
- **NumPy:** Para operaciones numéricas avanzadas.
- **Matplotlib y Seaborn:** Para visualización de datos.

● R:

- **Caret:** Para modelado predictivo y validación cruzada.
- **Ggplot2:** Para visualización gráfica avanzada.
- **Dplyr:** Para manipulación eficiente de datos.

Base Bibliográfica

- **Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010).** Multivariate Data Analysis.

- Pearson.
- **Breiman, L. (2001).** Random Forests. *Machine Learning*, 45(1), 5-32.
 - **Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002).** SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
 - **Esteva, A., et al. (2017).** Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.
 - **Goodfellow, I., Bengio, Y., & Courville, A. (2016).** Deep Learning. MIT Press.
 - **Tabachnick, B. G., & Fidell, L. S. (2013).** Using Multivariate Statistics. Pearson.

Limitaciones del Estudio

- **Calidad de los Datos:** Posibles errores o inconsistencias en los registros pueden afectar la precisión de los resultados. Se implementarán procedimientos de limpieza y validación para minimizar este impacto.
- **Variables No Consideradas:** La falta de ciertas variables potencialmente relevantes (e.g., hábitos de vida, antecedentes médicos) puede limitar la explicación de algunos resultados.
- **Generalización Temporal:** Los resultados reflejan el período 2015-2020 y pueden no ser aplicables a años posteriores si las condiciones cambian significativamente.

Consideraciones Éticas y Legales

- **Confidencialidad y Anonimato:** Se garantizará la anonimización de los datos personales para proteger la privacidad de los individuos.
- **Consentimiento y Uso de Datos:** Se respetarán todas las regulaciones y normativas relacionadas con el manejo de datos personales y se utilizarán únicamente con fines académicos y de investigación.
- **Ética en la Investigación:** Se seguirán los principios éticos establecidos para investigaciones en salud y manejo de datos, asegurando la integridad y transparencia del estudio.

X. Referencias (Listar las citas bibliográficas con el estilo adecuado a su especialidad)

- Revista Médica Clínica Las Condes. (2022a). Modelos predictivos en salud basados en aprendizaje de máquina. Revista Médica Clínica Las Condes. Recuperado de <https://www.elsevier.es/es-revista-revista-medica-clinica-las-condes-202-articulo-modelos-predictivos-salud-basados-aprendizaje-S0716864022001213>
- Revista Médica. (2023). Aplicación de Big Data en la predicción de enfermedades. Revista Médica. Recuperado de <https://revistamedica.com/aplicacion-big-data-prediccion-enfermedades>
- Revista Médica Clínica Las Condes. (2022b). Aplicaciones de aprendizaje automático en salud. Revista Médica Clínica Las Condes. Recuperado de <https://www.elsevier.es/es-revista-revista-medica-clinica-las-condes-202-articulo-aplicaciones-aprendizaje-automatico-salud-S0716864022001195>
- Revista Médica de Chile. (2021). Generalidades del Machine Learning y su aplicación en la gestión de salud. Revista Médica de Chile. Recuperado de https://www.scielo.cl/scielo.php?pid=S0034-98872021000200248&script=sci_arttext
- Medicina Intensiva. (2019). Big Data Analysis y Machine Learning en medicina intensiva. Medicina Intensiva. Recuperado de <https://medintensiva.org/es-big-data-analysis-machine-learning-articulo-S0210569118303139>
- Huata Apaza, J. M. (2024). Clústeres de causas de fallecimiento en la Región Puno 2020-2023 [Tesis de pregrado, Universidad Nacional del Altiplano]. Alicia Concytec. Recuperado de https://alicia.concytec.gob.pe/vufind/Record/RNAP_3aaff4d7b24de60b77f7b1fdfe0f224
- Ministerio de Salud del Perú (MINSA). (2018). Análisis de las causas de mortalidad en el Perú,

1986-2015. Recuperado de <https://www.gob.pe/institucion/minsa/informes-publicaciones/279665-analisis-de-las-causas-de-mortalidad-en-el-peru-1986-2015>

Álvares-Calderón, M., Ortiz, P., & Rivas, J. (2018). Resultados preliminares del fortalecimiento del sistema informático nacional de defunciones en Perú. Revista Peruana de Medicina Experimental y Salud Pública, 35(3), 484–492. Recuperado de https://www.scielo.org.pe/scielo.php?pid=S1726-46342018000300019&script=sci_arttext

Smith, J. (2022). "Data Analytics for Mortality Patterns." International Journal of Epidemiology. Pérez, R., & Gómez, M. (2020). "Machine Learning Applications in Public Health." Journal of Health Analytics.

INEI (2021). "Estadísticas Vitales del Perú." Lima, Perú.

OMS (2020). "CIE-X Manual." Organización Mundial de la Salud.

XI. Uso de los resultados y contribuciones del proyecto (Señalar el posible uso de los resultados y la contribución de los mismos)

Resultados Esperados:

- Modelos predictivos para causas de defunción.
- Identificación de factores de riesgo prioritarios.

Contribuciones:

- Apoyo en la formulación de políticas públicas.
- Desarrollo de estrategias preventivas de salud.

XII. Impactos esperados

i. Impactos en Ciencia y Tecnología

Avance en la integración de machine learning y salud pública.

ii. Impactos económicos

Optimización de recursos en intervenciones sanitarias.

iii. Impactos sociales

Reducción de mortalidad evitable y mejora en la calidad de vida.

iv. Impactos ambientales

Sin impacto directo.

XIII. Recursos necesarios (Infraestructura, equipos y principales tecnologías en uso relacionadas con la temática del proyecto, señale medios y recursos para realizar el proyecto)

- Infraestructura:** Computadora con alto rendimiento.
- Software:** Python.
- Base de Datos:** Acceso a SINADEF.

XIV. Localización del proyecto (indicar donde se llevará a cabo el proyecto)

El proyecto se llevará a cabo en la región de Puno, Perú, utilizando datos recopilados por SINADEF.

XV. Cronograma de actividades

Actividad	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
Revisión de literatura	X			
Recolección y preparación de datos	X	X		
Análisis exploratorio de datos		X		
Desarrollo de modelos estadísticos		X	X	
Interpretación y redacción de resultados			X	X
Revisión final y presentación				X

XVI. Presupuesto

Descripción	Unidad de medida	Costo Unitario (S/.)	Cantidad	Costo Total (S/.)
Software Python/R	Licencia	0	1	0
Hardware (PC/Servidor)	Unidad	4,000	1	4,000
Acceso a SINADEF	Licencia	0	1	0
Material de oficina	Unidad	100	1	100