

Macroproceso

Proceso

Título: Formato de registro de trabajo de grado



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Programa Curricular

Administración de Sistemas
Informáticos

Fecha

06

06

2025

Datos del estudiante

Nombres: Juan Manuel						Apellidos: Garcia Cifuentes							
Tipo de identificación:		T.I.		C.C.	X	C.E.		Número: 1.053.867.032					
Correo institucional: jugarciaci@unal.edu.co						Teléfono: +57 3053509190							

Formulación Trabajo de Grado

1. MODALIDAD/SUBMODALIDAD DE TRABAJO DE GRADO

Indique en cuál de las siguientes submodalidades se desarrollará el trabajo de grado. Según Artículo 1 del Acuerdo 006 de 2022 del Consejo de Facultad de Administración

- Trabajos investigativos/Trabajo Monográfico
Por trabajo monográfico se entiende un trabajo escrito en el cual se exponen, de manera sistemática y organizada, los resultados de un estudio sobre un tema o problema específico en alguna de las áreas de conocimiento del programa curricular.
- Trabajos investigativos/Proyecto Final
Por proyecto final se entiende la aplicación de conocimientos en la solución de problemas específicos en el campo de acción del perfil profesional.

Este trabajo corresponde a un Trabajo investigativo/Proyecto Final, ya que se centra en la aplicación de conocimientos en ciencia de datos y Machine Learning para abordar un problema real relacionado con salud pública y calidad del aire en la ciudad de Manizales. A través del desarrollo de una aplicación funcional, se propone una solución tecnológica concreta alineada con el perfil profesional.

2. TÍTULO DEL TRABAJO DE GRADO

El título deberá expresar el contenido de la propuesta con concisión y claridad. Hay que poner un gran cuidado en la elección de las palabras empleadas ya que el título no debe sugerir ni más ni menos que lo que se pretende desarrollar en el trabajo de grado.

Machine Learning para predecir el impacto en la salud según la calidad del aire en la ciudad de Manizales.

3. INTRODUCCIÓN Y JUSTIFICACIÓN

Esta sección es una exposición clara, concisa y bien fundamentada del problema de investigación o de desarrollo/aplicación que se plantea. Se usan algunas referencias generales (soportes documentales, estadísticas, datos, bibliografía, etc.) para establecer la justificación (es decir, la importancia y la pertinencia) del trabajo de grado propuesto.

La creciente preocupación por los efectos de la contaminación atmosférica en la salud humana ha generado la necesidad de herramientas efectivas que permitan prever y mitigar sus impactos. En ciudades como Manizales, donde la geografía, las fuentes de emisión antropogénicas y condiciones meteorológicas complejas influyen significativamente en la calidad del aire, el desarrollo de modelos predictivos se vuelve esencial.

La evidencia científica ha mostrado asociaciones entre la exposición prolongada a contaminantes como PM10, PM2.5, NO₂, SO₂ y O₃ con enfermedades respiratorias y cardiovasculares. A pesar de los avances en monitoreo ambiental, persiste una necesidad urgente de soluciones tecnológicas que traduzcan estos datos en conocimiento útil y accesible para la toma de decisiones.

Este proyecto propone el uso de técnicas de Machine Learning para predecir el impacto en la salud de la población manizaleña basado en datos históricos de calidad del aire, salud y variables contextuales. Como producto final, se desarrollará una aplicación interactiva que permitirá visualizar riesgos y fomentar la toma de decisiones informadas por parte de autoridades y ciudadanos.

**4. OBJETIVO GENERAL**

El objetivo general resume la idea completa de la propuesta. Casi siempre está muy relacionado con el título. La redacción del objetivo general debe comenzar con un verbo.

Aplicar y evaluar modelos de Machine Learning capaz de predecir el riesgo de impacto en la salud de la población de Manizales, basado en datos históricos de calidad del aire y variables contextuales relevantes, e implementarlo en una aplicación interactiva para la visualización y la concienciación.

5. OBJETIVOS ESPECÍFICOS

Cada objetivo específico debe comenzar con un verbo. La redacción de los objetivos específicos es muy importante porque establece los parámetros sobre los cuales se realiza la evaluación de la propuesta. Consecuentemente deben cumplir con los siguientes criterios:

- *Específicos. ¿Responden a las preguntas qué, cómo, cuándo, por qué?*
 - *Medibles. ¿Cuánto?*
 - *Alcanzables. Habilidades, esfuerzos y limitaciones.*
 - *Realistas. Se pueden lograr.*
 - *Tangibles. Se puede apreciar.*
-
- Recopilar y consolidar datos históricos de calidad del aire de las estaciones de monitoreo ubicadas en Manizales, abarcando un periodo de tiempo significativo.
 - Obtener y preprocesar datos de salud anonimizados relevantes para el estudio, identificando indicadores clave de impacto en la salud relacionados con la calidad del aire (casos de enfermedades respiratorias y cardiovasculares, ingresos hospitalarios).
 - Integrar y analizar las bases de datos de calidad del aire y salud, junto con otras variables contextuales como datos meteorológicos, para identificar posibles correlaciones y patrones.
 - Implementar y comparar diversos algoritmos de Machine Learning (KNN, Regresión Logística, XGBoots, Redes Neuronales) para la tarea de predicción del impacto en la salud.
 - Evaluar el rendimiento de los modelos predictivos utilizando métricas apropiadas y técnicas.
 - Desarrollar una aplicación interactiva para los usuarios donde pueden obtener una predicción del nivel de impacto en la salud.

6. ANTECEDENTES

Presentación de los trabajos previos relacionados con la propuesta, con base en la recopilación y análisis de la información bibliográfica. Para la redacción de los antecedentes se hace uso principalmente de bases de datos, libros y artículos científicos o divulgativos. En esta sección de la propuesta son de vital importancia las citaciones y las referencias bibliográficas para soportar las afirmaciones hechas en el documento.

La contaminación atmosférica representa un problema crítico de salud pública a nivel global. Según la Organización Mundial de la Salud (2021), la exposición a contaminantes como PM_{2.5}, PM₁₀ y NO₂ está asociada con 7 millones de muertes prematuras anuales, principalmente por enfermedades respiratorias y cardiovasculares.

En Colombia, el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) ha desarrollado el Sistema de Información sobre Calidad del Aire (SISAIRE), que constituye la principal fuente de datos históricos de monitoreo ambiental del país (IDEAM, 2020). Este sistema proporciona información de las estaciones de monitoreo a nivel nacional, incluyendo datos de contaminantes criterio como PM₁₀, PM_{2.5}, NO₂, SO₂ y O₃.

Específicamente para Manizales, diversos estudios han abordado la problemática de calidad del aire desde diferentes perspectivas. Saffon et al. (2022) analizaron la relación entre la calidad del aire y las enfermedades respiratorias en la ciudad, documentando correlaciones significativas entre los niveles de contaminantes atmosféricos y la incidencia de patologías respiratorias en la población local. Complementariamente, Li-Ramírez et al. (2016) desarrollaron metodologías para la generación y representación de indicadores de calidad del aire aplicados específicamente a Manizales, estableciendo parámetros de referencia y técnicas de visualización que facilitan la interpretación de datos ambientales. Estos antecedentes locales proporcionan una base empírica sólida sobre los efectos de la contaminación atmosférica en la salud de los manizaleños y los métodos de análisis de datos ambientales en la región.

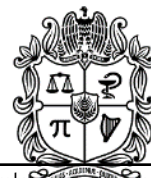
A nivel metodológico, Vásquez (2022) destaca los beneficios de los modelos basados en algoritmos de Machine Learning para la predicción, evidenciando la superioridad de estas técnicas sobre métodos estadísticos tradicionales en términos de precisión y capacidad de manejo de grandes volúmenes de datos.

Para la región de Caldas, la Universidad Nacional de Colombia en conjunto con CORPOCALDAS han desarrollado el Centro de Datos e Indicadores Ambientales de Caldas (CDIAC), que documenta las condiciones ambientales locales y proporciona datos históricos de calidad del aire específicos para Manizales (Universidad Nacional de Colombia & CORPOCALDAS, 2015).

Macroproceso

Proceso

Título: Formato de registro de trabajo de grado



UNIVERSIDAD
NACIONAL
DE COLOMBIA

La revisión de antecedentes evidencia que, aunque existen estudios que documentan la relación entre la calidad del aire y salud en Manizales, persiste una brecha en la aplicación de técnicas avanzadas de Machine Learning para desarrollar herramientas predictivas que integren estos hallazgos en soluciones tecnológicas prácticas para la toma de decisiones en salud pública.

7. METODOLOGÍA

Se deberá mostrar, en forma organizada y precisa, cómo será alcanzado cada uno de los objetivos específicos propuestos. La metodología debe reflejar la estructura lógica y el rigor científico del proceso de desarrollo del trabajo de grado, empezando por la elección de un enfoque metodológico específico y finalizando con la forma como se van a analizar, interpretar y presentar los resultados. Deben detallarse los procedimientos, técnicas, actividades y demás estrategias metodológicas requeridas para la investigación. Deberá indicarse el proceso a seguir en la recolección de la información, así como en la organización, sistematización y análisis de los datos. Se recomienda considerar los posibles problemas que se pueden presentar al igual de cómo se podrían solucionar.

Recolección de datos:

- Datos históricos de calidad del aire (PM10, PM2.5, NO₂, etc.) provenientes de estaciones locales (Corpocaldas, CDIAC).
- Datos meteorológicos (temperatura, humedad, viento) donde los datos (CDIAC, IDEAM).
- Datos de salud anonimizados sobre enfermedades respiratorias/cardiovasculares (Territorial de la salud).

Preprocesamiento:

- Limpieza de datos, normalización, tratamiento de valores faltantes, codificación de variables categóricas.
- Análisis exploratorio y visualización inicial de relaciones.

Construcción de modelos:

- Selección de algoritmos de Machine Learning (KNN, Árboles de decisión, Redes neuronales, Regresión logística).
- División del dataset en entrenamiento y prueba (por ejemplo, 80/20).
- Validación cruzada y ajuste de hiperparámetros.

Evaluación:

- Métricas: Accuracy, Precision, Recall, F1-score, AUC.
- Comparación de modelos para seleccionar el de mejor desempeño.

Desarrollo de la aplicación:

- Creación de una interfaz interactiva con Streamlit o herramienta similar.
- Integración del modelo predictivo en la aplicación.
- Posibilidad de ingreso de nuevos datos y visualización del nivel de riesgo estimado.

Evaluación de usabilidad:

- Pruebas con usuarios potenciales.
- Encuestas o entrevistas estructuradas.

Análisis y resultados

Redacción final

8. CRONOGRAMA DE ACTIVIDADES

ACTIVIDADES A REALIZAR	Semanas de ejecución de cada actividad															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Recolección de datos	X	X	X													
Preprocesamiento		X	X	X												
Construcción de modelos			X	X	X	X										
Evaluación					X	X	X									
Desarrollo de la aplicación						X	X	X	X	X	X	X	X			
Evaluación de usabilidad												X	X	X		
Análisis y resultados													X	X	X	
Redacción final														X	X	X



9. RESULTADOS ESPERADOS Y APOORTE ESPECÍFICO

Listado de los posibles resultados que se pretende alcanzar con la ejecución de la propuesta. Los resultados se evidencian, típicamente, en forma de productos tangibles (informes, artículos, prototipos de software, etc).

Reflexión sobre el posible uso de los resultados de su trabajo de grado y los potenciales beneficios que puede tener sobre la sociedad, el medio natural, la salud humana u otros aspectos en el corto, mediano y largo plazo.

Resultados esperados:

- Base de datos consolidada de calidad del aire y salud pública en Manizales.
- Modelos de Machine Learning validados para predicción de impacto en salud.
- Aplicación web funcional para visualizar niveles de riesgo en tiempo real.
- Informe técnico con hallazgos y recomendaciones.

Aportes específicos:

- Herramienta práctica para toma de decisiones en salud pública.
- Fomento de la concienciación ciudadana sobre la calidad del aire.
- Contribución metodológica para investigaciones similares en otras ciudades.
- Posibilidad de escalamiento e integración con sistemas locales de monitoreo ambiental.

10. BIBLIOGRAFÍA CITADA

Relacione aquí únicamente las referencias citadas en las secciones precedentes. Se recomienda observar las normas de citación APA o, alternativamente, el estilo IEEE.

Saffon, I., Costa, Y. J., & Betancur, J. F. (2022). Calidad del aire y su relación con enfermedades respiratorias en la ciudad de Manizales. *Naturaleza Y Sociedad. Desafíos Medioambientales*, 2, 129-142 <https://doi.org/10.53010/nys2.06>

Vásquez, J. (2022). Beneficios de los modelos basados en algoritmos de machine learning para la predicción. Medellín. Obtenido de <https://hdl.handle.net/10495/29252>

Li-Ramírez, Jade Alexandra, Pérez-Zapata, Ángela María, Duque-Méndez, Néstor Darío, & Aristizábal-Zuluaga, Beatriz Helena. (2016). Generación y representación de Indicadores de calidad de aire: caso de estudio aplicado a Manizales. *Iteckne*, 13(2), 174-184. Retrieved June 04, 2025, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-17982016000200008&lng=en&tlng=es.

Universidad Nacional de Colombia, & CORPOCALDAS. (2015). CDIAC - Centro de Datos e Indicadores Ambientales de Caldas. Recuperado de cdiac.manizales.unal.edu.co/ el 3 de Junio de 2025.

Organización Mundial de la Salud. (2021). *Ambient (outdoor) air pollution*. Geneva: WHO Press.

"IDEAM. (2020). SISAIRE - Sistema de Información de Calidad del Aire en Colombia." SISAIRE - IDEAM." Recuperado de <http://sisaire.ideam.gov.co/ideam-sisaire-web/consultas.xhtml/> el 3 de Junio de 2025."

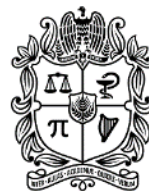
Secretaría de Salud de Caldas. (2025). [Base de datos]. PQRS Recuperado el 4 de junio de 2025, de <https://saluddecaldas.gov.co/ruta-especifica>

11. REFERENCIAS BIBLIOGRÁFICAS ADICIONALES

Relacione documentos que no fueron citados en las secciones precedentes pero que serán de utilidad para desarrollar la propuesta. Se recomienda observar las normas de citación APA o, alternativamente, el estilo IEEE.

"Universidad Nacional de Colombia, & CORPOCALDAS. (2015). CDIAC - Centro de Datos e Indicadores Ambientales de Caldas. Recuperado de cdiac.manizales.unal.edu.co/ el 3 de Junio de 2025."

Hastie, Trevor & Tibshirani, Robert & Friedman, Jerome & Franklin, James. (2004). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Math. Intell.. 27. 83-85. 10.1007/BF02985802.

Macroproceso**Proceso****Título: Formato de registro de trabajo de grado**

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Aval del futuro director			
Nombre:	Diógenes Ramírez RamírezRamírez	Unidad Académica Básica:	Depto. Matemáticas
Correo electrónico:	ddramirezra@unal.edu.co	Teléfono:	+57 3122805596
Firma:	 <p>Nota: con la firma el profesor manifiesta que participó en la construcción de la propuesta y avala su presentación. Además, se compromete a realizar el proceso de acompañamiento durante el período académico inmediatamente siguiente.</p>		

Firma del estudiante