



ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS

REPO-AURA

Proyecto de curso PTIA

Grupo 02

Autor:

Andrés Felipe Chavarro Plazas

Fecha de entrega: 16 de septiembre del 2025

Declaración firmada

“Declaro que he escrito este trabajo de investigación por mí mismo, y que no he utilizado otras fuentes o recursos que los indicados para su preparación. Declaro que he indicado claramente todas las citas directas e indirectas, y que este documento no se ha presentado en otro lugar para fines de examen o publicación”

Nombre del estudiante & Firma:

Andrés Felipe Chavarro Plazas

Fecha: 16 de septiembre del 2025

...

Fecha: 16 de septiembre del 2025

Índice de contenido

1. Introducción	3
2. Trabajos relacionados	4
3. Descripción del problema	5

1. Introducción

En el ámbito de la ingeniería de software actual, la gestión de proyectos se enfrenta a una brecha crítica: la dependencia de métricas de resultados que actúan como indicadores de éxito dentro de este, que a menudo revelan disfunciones cuando su impacto en el cronograma y la calidad es ya ineludible. Este proyecto se postula como una respuesta a dicha brecha, proponiendo un sistema proactivo que aproveche la riqueza de datos generados durante el ciclo de vida del desarrollo para ofrecer una visión predictiva de la salud del mismo. El enfoque se centra en la resolución de tres problemáticas interconectadas y fundamentales:

Primero, la **alta tasa de fracaso en proyectos de software**, impulsada por la acumulación de deuda técnica y la erosión de la colaboración dentro de los equipos, factores que las métricas de rendimiento convencionales no logran capturar de manera efectiva. Segundo, la dificultad inherente a la detección temprana de riesgos centrados en el capital humano, como la sobrecarga de personal y la degradación del ambiente compartido de trabajo, los cuales son precursores directos de la ineficiencia y la rotación. Tercero, la **subjetividad en la evaluación de la salud integral de un proyecto**, que actualmente recae en la intuición, en lugar de en un análisis cuantitativo, continuo y basado en evidencia.

La naturaleza de estos desafíos excede las capacidades de la computación tradicional, posicionando a este proyecto como una aplicación idónea para la materia **Principios de Técnicas de Inteligencia Artificial (PTIA)**. La solución exige una arquitectura de IA correcta para el correcto funcionamiento. Por un lado, se requiere la implementación de un sistema basado en conocimiento para codificar la heurística experta en la identificación de patrones de riesgo conocidos. Por otro, es indispensable el uso del aprendizaje automático, específicamente de redes neuronales recurrentes, para discernir patrones complejos y tendencias en secuencias de datos heterogéneos como código, texto y metadatos. Así, este trabajo no solo aborda una problemática de alta relevancia, sino que también funciona como una síntesis práctica de los paradigmas fundamentales de la inteligencia artificial.

2. Trabajos relacionados

El análisis de repositorios de software para extraer conocimiento accionable es un campo consolidado, conocido como Minería de Repositorios de Software (MSR). Las herramientas y trabajos existentes han abordado la predicción de la salud de un proyecto desde diferentes ángulos, enfocándose en la calidad del código estático o en métricas de productividad. Este proyecto, **RepoAura**, se posiciona en la intersección de estos enfoques, proponiendo un análisis holístico que integra la dimensión técnica con la humana y temporal.

Predicción de Defectos y Deuda Técnica

- **Herramienta - SonarQube:** Es la plataforma líder para el análisis estático de código. Su enfoque es identificar bugs y code smells basándose en reglas predefinidas sobre el estado actual del código. Su principal limitación es que ofrece una "fotografía" estática, no una predicción de tendencias. RepoAura se diferencia al analizar la evolución temporal para predecir problemas futuros, no solo para describir los actuales.
- **Trabajo Académico - "Predicting Defects for Eclipse":** Este trabajo seminal demostró que las métricas históricas como el "churn" de código son fuertes predictores de defectos, usando modelos estadísticos. RepoAura se diferencia al extender esta premisa integrando factores humanos y usando modelos más especializados con RNNs para capturar la dinámica temporal de forma más sofisticada.

Análisis de Factores Humanos y de Colaboración

- **Herramienta - Pluralsight Flow:** Plataforma de "Developer Intelligence" que mide la del equipo. Su enfoque es la eficiencia y la velocidad de desarrollo. RepoAura se diferencia al centrarse en el riesgo y la salud social y técnica. No se pregunta "¿cuán rápido vamos?", sino "¿estamos en riesgo de fallar?".
- **Trabajo Académico - "The Impact of Sentiment on Software Development":** Esta línea de investigación utiliza Procesamiento de lenguaje natural para correlacionar el sentimiento negativo en las comunicaciones con la introducción de bugs. RepoAura se diferencia al no solo mostrar una correlación, sino al integrar el sentimiento como una característica clave dentro de un modelo predictivo holístico que también considera métricas de código y actividad.

3. Descripción del problema

La elección del problema se fundamentó en un conjunto de criterios diseñados para garantizar su viabilidad, relevancia y rigor académico, asegurando que la solución sea técnicamente desafiante y aplicable a la industria del software.

Criterios de Selección

- **Relevancia Industrial:** Se priorizó un problema cuya solución tuviera un impacto tangible en la industria, abordando la necesidad real de mitigar los altos índices de fracaso en proyectos. El objetivo es superar las métricas reactivas tradicionales mediante la detección proactiva de riesgos que preceden a desviaciones en costo y calidad.
- **Alineación Académica y Complejidad Técnica:** El problema posee una complejidad inherente que justifica un enfoque de IA híbrido. Requiere la aplicación integrada de sistemas basados en reglas para codificar heurísticas expertas y de aprendizaje automático para identificar patrones complejos en los datos que los métodos tradicionales no pueden capturar, justificando su desarrollo en el marco del curso.
- **Disponibilidad de Datos:** Un criterio pragmático fue la existencia de conjuntos de datos públicos y ricos, disponibles en repositorios de código abierto. Esta accesibilidad es un pilar indispensable para el entrenamiento y la validación de cualquier modelo de aprendizaje supervisado, eliminando una barrera crítica del proyecto.
- **Potencial de Innovación:** Se buscó un enfoque que, si bien se basa en trabajos existentes, ofrece una perspectiva holística novedosa. La innovación radica en combinar el análisis de las dimensiones técnica, colaborativa y humana en un único modelo, tratando el proyecto como un sistema socio-técnico integrado.

Problema Seleccionado

Considerando los criterios expuestos, **el problema seleccionado es la detección proactiva del riesgo y la predicción de la salud integral en proyectos de desarrollo de software**. Este desafío se abordará mediante la construcción de un modelo de aprendizaje automático multifactorial, entrenado para identificar las señales débiles y los patrones sutiles que preceden a los fallos. El objetivo final es desarrollar un sistema de alerta temprana que permita pasar de una gestión de proyectos reactiva a una gobernanza proactiva y guiada por datos.