

Universidad de Costa Rica  
Escuela de Ingeniería Eléctrica

Bitácora del proyecto

Clasificación de fallos funcionales en procesadores  
para servidores utilizando Aprendizaje Automático

Jesús Zuñiga Méndez

Tutor: Erick Carvajal

II Semestre 2024

## Índice

Objetivo General: .....	5
Objetivos Específicos: .....	5
Semana 1: del 26 de agosto al 1 de septiembre .....	6
• Objetivo .....	6
• Actividades realizadas .....	6
• Dificultades:.....	6
• Resultados:.....	6
• Próxima actividad: .....	7
Semana 2: del 2 de septiembre al 8 de septiembre .....	7
• Objetivo: .....	7
• Actividades realizadas .....	7
• Dificultades:.....	7
• Resultados:.....	7
• Próxima actividad: .....	7
Semana 3: del 9 de septiembre al 15 de septiembre .....	7
• Objetivo: .....	7
• Actividades realizadas: .....	8
• Dificultades:.....	8
• Resultados:.....	8
• Próxima actividad: .....	8
Semana 4: del 15 de septiembre al 22 de septiembre .....	8
• Objetivo: .....	8
• Actividades realizadas .....	8
• Dificultades:.....	9
• Resultados:.....	9
• Próxima actividad: .....	9
Semana 5: del 23 de septiembre al 29 de septiembre .....	9
• Objetivo .....	9

• Actividades realizadas .....	9
• Dificultades:.....	9
• Resultados:.....	9
• Próxima actividad: .....	9
Semana 6: del 30 de septiembre al 6 de octubre .....	10
• Objetivo .....	10
• Actividades realizadas .....	10
• Próxima actividad: .....	10
Semana 7: del 7 de octubre al 13 de octubre .....	10
• Objetivo .....	10
• Actividades realizadas .....	10
• Dificultades:.....	10
• Resultados:.....	10
• Próxima actividad: .....	10
Semana 8: del 14 de octubre al 20 de octubre .....	11
• Objetivo .....	11
• Actividades realizadas .....	11
• Dificultades:.....	11
• Resultados:.....	11
• Próxima actividad: .....	11
Semana 9: del 21 de octubre al 27 de octubre .....	11
• Objetivo .....	11
• Actividades realizadas .....	11
• Dificultades:.....	12
• Resultados:.....	12
• Próxima actividad: .....	12
Semana 10: del 28 de octubre al 3 de noviembre .....	12
• Objetivo .....	12

• Actividades realizadas .....	12
• Dificultades:.....	12
• Resultados:.....	12
• Próxima actividad: .....	12
Semana 11: del 4 de noviembre al 10 de noviembre.....	12
• Objetivo .....	12
• Actividades realizadas .....	13
• Dificultades:.....	13
• Resultados:.....	13
• Próxima actividad: .....	13
Semana 12: del 11 de noviembre al 17 de noviembre .....	13
• Objetivo .....	13
• Actividades realizadas .....	13
• Dificultades:.....	14
• Resultados:.....	14
• Próxima actividad: .....	14
Semana 13: del 18 de noviembre al 24 de noviembre .....	14
• Objetivo .....	14
• Actividades realizadas .....	14
• Dificultades:.....	14
• Resultados:.....	15
• Próxima actividad: .....	15
Semana 14: del 25 de noviembre al 1 de diciembre .....	15
• Objetivo .....	15
• Actividades realizadas .....	15
• Dificultades:.....	15
• Resultados:.....	15
• Próxima actividad: .....	15

## Objetivo General:

Implementar un modelo de aprendizaje automático para la clasificación de fallos funcionales en procesadores de servidores.

## Objetivos Específicos:

- Desarrollar un algoritmo capaz de normalizar los fallos específicos de distintas familias de procesadores con el fin de asegurar la escalabilidad del modelo implementado

- Explorar diversos modelos de aprendizaje automático para la clasificación efectiva de los fallos en procesadores
- Evaluar el rendimiento de al menos dos modelos de aprendizaje automático implementados para la clasificación de fallos en procesadores, utilizando métricas estándar de precisión y exactitud

## Semana 1: del 26 de agosto al 1 de septiembre

- **Objetivo:** Buscar material bibliográfico enfocado en aprendizaje automático.
- **Actividades realizadas:** Búsqueda y lectura de bibliografía sobre aprendizaje automático
- **Dificultades:** En la bibliografía encontrada se mencionan muchos algoritmos, como regresión logística, árboles de decisión, redes neuronales convolucionales, entre otros, que pueden solucionar el problema, por lo que, es necesario discriminar mucha información para que de esta forma se pueda elegir el algoritmo más apropiado.
- **Resultados:** Se encontraron muchos algoritmos apropiados para la resolución del problema, por lo que, uno de los pasos más importantes es definir el set de datos a clasificar,

hasta este momento de acuerdo con lo leído existe una inclinación a las redes neuronales convolucionales, ya que el set de datos existente contiene muchas entradas para una misma unidad (procesador), por lo que, si se diseña un algoritmo que pueda mapear estas entradas a un arreglo bidimensional, se puede tratar cada unidad como una imagen RGB.

- **Próxima actividad:** Depurar el set de datos a clasificar.

## Semana 2: del 2 de septiembre al 8 de septiembre

- **Objetivo:** Depurar el set de datos ya existente para que sea apto para el aprendizaje con redes neuronales convolucionales.
- **Actividades realizadas:** Revisión y eliminación de entradas innecesarias en el set de datos, instalación y prueba de Python 12 con la biblioteca TensorFlow
- **Dificultades:** El set de datos con el que se trabajará contiene entradas que no se relacionan propiamente a fallos, debido a que, estas entradas pueden contener datos sobre la ubicación, proyecto, presupuesto, entre otros; estas entradas son datos irrelevantes para el objetivo del proyecto, por lo que, se debe proceder con una filtración de estos.
- **Resultados:** Parte de la información que contiene el set de datos depurado es alfanumérica, por lo que, es recomendable buscar la forma de tratarla como numérica. Por otro lado, el set de datos depurado parece ser apto para el desarrollo de un algoritmo que permita mapear los datos como un arreglo bidimensional, por lo tanto, la idea de utilizar redes neuronales convolucionales toma mayor fuerza.
- **Próxima actividad:** Crear un algoritmo para el mapeo del set de datos en un arreglo bidimensional.

## Semana 3: del 9 de septiembre al 15 de septiembre

- **Objetivo:** Desarrollar un algoritmo que permita el mapeo del set de datos en un arreglo bidimensional

- **Actividades realizadas:**

**Martes (tiempo dedicado 5horas):** inicio del desarrolló del algoritmo y las necesidades que este debe cumplir, ya que es necesario que para cada familia de procesador se definan los pesos de prioridad de cada registro perteneciente a cada unidad.

**Jueves (tiempo dedicado 5horas):** Gracias al análisis realizado el martes sobre el set de datos, se definió que los registros de tipo de prueba y la configuración de la unidad para la prueba realizada deben tener mayor peso que los códigos de fallo para cada unidad, por lo que se continúa realizando el algoritmo en función de estas prioridades.

- **Dificultades:** La asignación de prioridades para cada registro del set de datos requiere una inversión de tiempo y conocimientos preadquiridos, por lo tanto, para la próxima semana se buscará orientación por parte de algunos compañeros.
- **Resultados:** Se obtuvo el primer prototipo de algoritmo a utilizar, pero se debe mejorar para la próxima semana.
- **Próxima actividad:** Buscar orientación por parte de expertos en los datos que tengan mayor experiencia para que sea aplicada al algoritmo.

## Semana 4: del 15 de septiembre al 22 de septiembre

- **Objetivo:** Fijar reuniones formativas con expertos en el tema y continuar con el desarrollo del algoritmo
- **Actividades realizadas:** Se fijaron dos reuniones formativas martes y viernes en horario de proyecto, para solicitar ayuda técnica para el estudio de los datos.

**Martes (tiempo dedicado 5 horas):** se discutió temas relevantes con respecto a la configuración de la unidad a la hora de ejecutar las pruebas.

**Viernes (tiempo dedicado 4 horas):** se abarcaron los temas de código de error y prioridad de los mismos. En ambas semanas después de la reunión se continuó con la definición del algoritmo.



- **Dificultades:** en las reuniones se abordaron muchos datos técnicos sobre las pruebas de las unidades, por lo que, es importante realizar un buen filtrado de los aspectos más importantes para incluir en el algoritmo
- **Resultados:** se lo logró determinar el peso de importancia para cada registro del set de datos y se continuó con la definición del algoritmo, el algoritmo se va a centrar en codificar cada registro de texto y número a su análogo hexadecimal según el código ASCII para convertir estos datos a números que permitan obtener un color RGB para convertir los datos de cada unidad a una imagen que permita clasificarse mediante redes neuronales convolucionales
- **Próxima actividad:** Se planea iniciar con la codificación en PYTHON del algoritmo.

## Semana 5: del 23 de septiembre al 29 de septiembre

- **Objetivo:** Iniciar la codificación del algoritmo en PYTHON
- **Actividades realizadas:** Se inicio con la codificación del algoritmo en PYTHON.  
**Martes (tiempo dedicado 6horas):** se exploraron algunas librerías como PILLOW y THINKER, se escogió utilizar la versión más reciente de PYTHON (3.12).  
**Viernes (tiempo dedicado 4horas):** se continua con la codificación del algoritmo.
- **Dificultades:** La primera dificultad que se encontró fue recorrer el archivo de Excel, pero la librería PANDAS resultó de utilidad, otra de las dificultades experimentadas es el tiempo que tarda en procesarse un archivo con tantos registros.
- **Resultados:** Se avanzó el algoritmo, ya se recorre el archivo y se codifican los registros alfanuméricos a su análogo hexadecimal en ASCII.
- **Próxima actividad:** Continuar con el desarrollo del algoritmo.

## Semana 6: del 30 de septiembre al 6 de octubre

- **Objetivo:** Continuar con el desarrollo del algoritmo.
- **Actividades realizadas:** Durante esta semana sufrí un accidente que me produjo una lumbalgia que me mantiene en cama por indicación médica, por lo que las actividades del proyecto se postergaran para la próxima semana
- **Próxima actividad:** Continuar con el desarrollo del algoritmo.

## Semana 7: del 7 de octubre al 13 de octubre

- **Objetivo:** Continuar con el desarrollo del algoritmo.
- **Actividades realizadas:** Se realizaron correcciones en el código.  
**Martes (tiempo dedicado 3horas):** Se detecto un problema con el código escrito, ya que algunos caracteres de los registros alfanuméricos no son soportados por ASCII por lo que en esta sesión se decidió migrar a Unicode para tener un soporte más robusto, la sesión fue más corta ya que aun continúo un poco afectado de la espalda. **Viernes (tiempo dedicado 4horas):** Se termino de migrar el código para su soporte en Unicode, además se migro de la librería PANDAS a openpyxl ya que resulto más útil para la lectura del archivo.
- **Dificultades:** La principal dificultad se presentó al reconocer el error de los resultados con los caracteres ASCII, por lo que se invirtió gran cantidad de tiempo en la depuración del código.
- **Resultados:** Se obtuvo una versión más robusta del código que logra soportar cualquier caracter presente en los registros alfanuméricos, además se mejoro un poco el tiempo de lectura de los archivos.
- **Próxima actividad:** Utilizar el archivo resultante de la codificación en Unicode para convertir los datos a vectores validos mapeables a arreglos bidimensionales de RGB

## Semana 8: del 14 de octubre al 20 de octubre

- **Objetivo:** Utilizar el archivo resultante de la codificación en Unicode para convertir los datos a vectores validos mapeables a arreglos bidimensionales de RGB
- **Actividades realizadas:** Se utiliza el archivo resultante codificado a valores Unicode para mapear los datos a arreglos bidimensionales  
**Martes (tiempo dedicado 5horas):** Utilizando los valores Unicode que son en su totalidad numéricos se crea el código para mapear cada registro de cada unidad, buscando obtener firmas características de cada fallo  
**Viernes (tiempo dedicado 6horas):** Se continua con el trabajo iniciado el martes, se continúa escribiendo el código y revisando resultados validando que sean correctos
- **Dificultades:** El mapeo a arreglos bidimensionales genera resultados que pueden ser dificiles de leer, por lo que la depuración de errores puede tardar mucho tiempo
- **Resultados:** Ya se obtuvieron los primeros registros convertidos a arreglos bidimensionales convertibles a valores RGB
- **Próxima actividad:** Continuar con la depuración de errores y mejora del código obtenido

## Semana 9: del 21 de octubre al 27 de octubre

- **Objetivo:** Continuar con la depuración de errores y mejora del código obtenido
- **Actividades realizadas:** Durante la depuración del código se encontraron algunos errores que se corrigieron tales como fallos a la hora de definir el tamaño de los arreglos provocando errores de indexación, además se encontró la necesidad de definir una escala de prioridades para algunas headers por lo que se incorpora esta funcionalidad  
**Martes (tiempo dedicado 3horas):** Con el tiempo restante después de la exposición en clase se realiza la corrección de errores de indexación  
**Viernes (tiempo dedicado 5horas):** Se agrego al código la función de prioridades en los headers, logrando con esto crear archivos con gamas de colores mas marcadas que faciliten la identificación automática

- **Dificultades:** Definir el tamaño de los arreglos de forma dinámica represento cierto grado de dificultad ya que parte de las necesidades es mantener el tamaño de los arreglos reducido
- **Resultados:** Se obtuvieron las primeras imágenes RGB mapeadas desde los registros, con la cuales se puede empezar el entrenamiento de los modelos
- **Próxima actividad:** Definir los modelos para la clasificación automática

## Semana 10: del 28 de octubre al 3 de noviembre

- **Objetivo:** Definir los modelos para la clasificación automática
- **Actividades realizadas:** Se inicio con la definición en Python de un modelo basado en redes neuronales convolucionales  
**Martes (tiempo dedicado 4horas):** Se inicio con la programación del modelo basado en CNN usando la librería tensorflow en el lenguaje de programación Python  
**Viernes (tiempo dedicado 4horas):** Se continuo con la definición del modelo obteniendo las primeras pruebas de entrenamiento
- **Dificultades:** El inicio de la programación se complicó un poco ya que el encontrar las fuentes se extendió un poco, pero se obtuvieron buenos resultados con las pruebas iniciales
- **Resultados:** Se logro entrenar el modelo y se realizaron las primeras pruebas de clasificación, se detecto que es necesario optimizar el proceso.
- **Próxima actividad:** Optimizar el entrenamiento del modelo y verificar si es necesario realizar algún cambio en el algoritmo de mapeo

## Semana 11: del 4 de noviembre al 10 de noviembre

- **Objetivo:** Optimizar el entrenamiento del modelo. Validar posibles cambios en el algoritmo de mapeo

- **Actividades realizadas:** Se realizaron ajustes al proceso de entrenamiento procurando aumentar el número de décadas de aprendizaje para mejorar el porcentaje de error  
**Martes (tiempo dedicado 3horas):** Se realizaron varias pruebas cambiando el número de décadas y el rango de aprendizaje para buscar una mejora en el porcentaje de error  
**Viernes (tiempo dedicado 5horas):** Se realizaron algunos ajustes en el algoritmo de mapeo para mejorar la escala de colores mapeados con el fin de que el algoritmo sea más eficiente
- **Dificultades:** El tiempo de aprendizaje suele ser extenso por lo que se consume mucho tiempo esperando que el entrenamiento se complete
- **Resultados:** Se obtuvieron los primeros resultados con porcentajes de error cercanos al 70% por lo que se debe seguir mejorando y explorando otros algoritmos de clasificación
- **Próxima actividad:** Definir un algoritmo de clasificación basado en redes neuronales profundas DNN

## Semana 12: del 11 de noviembre al 17 de noviembre

- **Objetivo:** Definir un algoritmo de clasificación basado en redes neuronales profundas DNN
- **Actividades realizadas:** Se definió un modelo basado en redes neuronales profundas y se tuvo una reunión con el profesor guía donde se recomienda cambiar el enfoque de los algoritmos de clasificación usados hasta el momento  
**Martes (tiempo dedicado 4horas):** Se escribió el código de una DNN basada en los ejemplos disponibles en el manual de TensorFlow con capas convolucionales y capas densas, además se definió un modelo conocido como completamente conectado compuesto solo de capas densas  
**Viernes (tiempo dedicado 5horas):** Se tuvo la reunión con el profesor guía, por recomendación el modelo de capas completamente conectadas se eliminó y se centro en

buscar modelos de CNN mas profundos con algoritmos conocidos, logrando encontrar un ejemplo de implementación para el algoritmo AlexNet y probandolo

- **Dificultades:** Realizar el cambio de los algoritmos definidos a la fecha represento una inversión en tiempo considerable
- **Resultados:** Se logro escribir el código para un modelo de CNN basado en AlexNet obteniendo resultados superiores al 90% de exactitud
- **Próxima actividad:** Buscar documentación sobre algún otro algoritmo basado en CNN e iniciar la recopilación de resultados

## Semana 13: del 18 de noviembre al 24 de noviembre

- **Objetivo:** Buscar documentación sobre algún otro algoritmo basado en CNN e iniciar la recopilación de resultados
- **Actividades realizadas:** Se realizo la investigación correspondiente para la implementación de un nuevo algoritmo, encontrándose un ejemplo de implementación para el algoritmo Vgg,  
**Martes (tiempo dedicado 5 horas):** Se escribió el código correspondiente a la arquitectura de Vgg, sin embargo, el algoritmo parece ser muy complejo para la capacidad computacional del equipo donde realizo las pruebas generando algunos reinicios, no se logró completar el entrenamiento  
**Viernes (tiempo dedicado 5 horas):** Se encontraron algunas arquitecturas basadas en Vgg que reducen el modelo, utilizando menos capas convolucionales y menos filtros, se implemento modelo siguiendo esta lógica y se logro entrenar evitando los reinicios, se obtuvieron resultados similares a los obtenidos con AlexNet
- **Dificultades:** El algoritmo Vgg-16 supera ampliamente la capacidad computacional disponible por lo que surgió la necesidad de modificarlo, además se descubrió que mediante la librería sklearn es posible obtener los cálculos de precisión y exactitud para el análisis de resultados

- **Resultados:** Se obtuvieron resultados de presión altos para el algoritmo vgg implantado superando el 94%
- **Próxima actividad:** Realizar las pruebas finales, obtener resultados e incorporarlos a los entregables finales

## Semana 14: del 25 de noviembre al 1 de diciembre

- **Objetivo:** Realizar las pruebas finales, obtener resultados e incorporarlos a los entregables finales
- **Actividades realizadas:** Esta semana se dedicó a la confección de los entregables finales, agregando los resultados obtenidos al reporte, validando el código fuente y creando el poster para la presentación  
**Martes (tiempo dedicado 7horas):** Se realizaron las pruebas del código final, se obtuvieron los resultados y se adjuntaron al informe, además se inició con la confección del poster  
**Viernes (tiempo dedicado 8horas):** Se completaron los entregables y se hizo una revisión detallada de los documentos
- **Dificultades:** Se tuvo que invertir mayor cantidad de tiempo en algunas ocasiones para reprobador los modelos ya que se detectaron algunos fallos en el código
- **Resultados:** Se logran completar los entregables y se obtuvieron resultados finales de la clasificación automática, alcanzando valores superiores al 90% de precisión, exactitud, recuperación y F1 Score
- **Próxima actividad:** Definir un algoritmo de clasificación basado en redes neuronales profundas DNN