

## Avance #1: Creación de herramienta de aprendizaje automático basada en imágenes satelitales, como insumo para futuros casos de uso relacionados a penetración de paneles solares en Costa Rica

Autor: Carlos Andrés Rodríguez Trigueros  
Máster universitario en computación.  
Inteligencia Artificial.

Docente guía: Carlos Gaitán Potayos.  
Noviembre de 2022.

## Índice

1. Introducción.....	3
2. Descripción del avance del proyecto.....	3
2.1 Grado de cumplimiento de los objetivos y resultados previstos en el plan de trabajo. ....	3
2.2 Justificación de los cambios en caso necesario .....	5
3. Relación de las actividades realizadas.....	5
3.1 Actividades previstas en el plan de trabajo .....	6
3.2 Actividades no previstas y realizadas o programas.....	6
4. Relación de las desviaciones en la temporización y acciones de mitigación si procede y actualización del cronograma si procede .....	6
5. Listado de los resultados parciales obtenidos hasta el momento (entregables que se adjuntan).....	7
6. Comentarios de vuestro director particular si lo consideran necesario .....	7
7. Anexos .....	7

## 1. Introducción

El siguiente documento tiene como objetivo mostrar los avances obtenidos resultado de los primeros esfuerzos para completar el trabajo final del máster, en el mismo se mostrarán cumplimiento de objetivos en base a cronograma original presentado, variaciones en planeamiento y recursos generados a ser entregados como parte del trabajo final de máster y por ende del avance.

## 2. Descripción del avance del proyecto

### 2.1 Grado de cumplimiento de los objetivos y resultados previstos en el plan de trabajo.

Con respecto al avance del proyecto y con respecto al cronograma presentado durante el primer PEC, se detalla que los objetivos correspondientes a:

- Crear conjunto de datos basado en imágenes satelitales utilizando Google Maps API
- Etiquetar conjunto de datos para identificar paneles solares
- Investigar conjunto de técnicas y estrategias disponibles sobre las que entrenar un modelo de aprendizaje automático.

Que todos fueron completados en su totalidad, reuniendo un total al menos 100 imágenes con paneles solares y generando sus respectivas anotaciones de forma manual. En cuanto a la investigación sobre técnicas y estrategia, se realiza una investigación sobre documentos relacionados al estado del arte para este tipo de problemas y se ha adjuntado los resultados como parte de avance en la minuta.

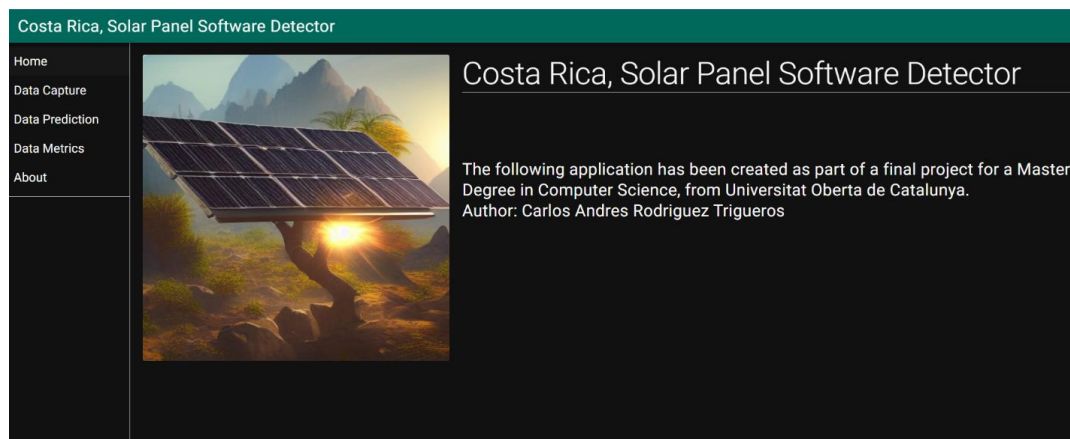
En los que respecta al objetivo:

- Generar de sitio web y servidor que usando un modelo de aprendizaje automático sea capaz de identificar paneles solares modelo de aprendizaje automático

Se ha iniciado de forma prematura este objetivo y se ha adjuntado un proyecto basado en React y Typescript donde la herramienta creada para la recolección de datos se ha incorporado a esta aplicación como una utilidad extra, refiérase a la Figura 1 y Figura 2.

Figura 1.

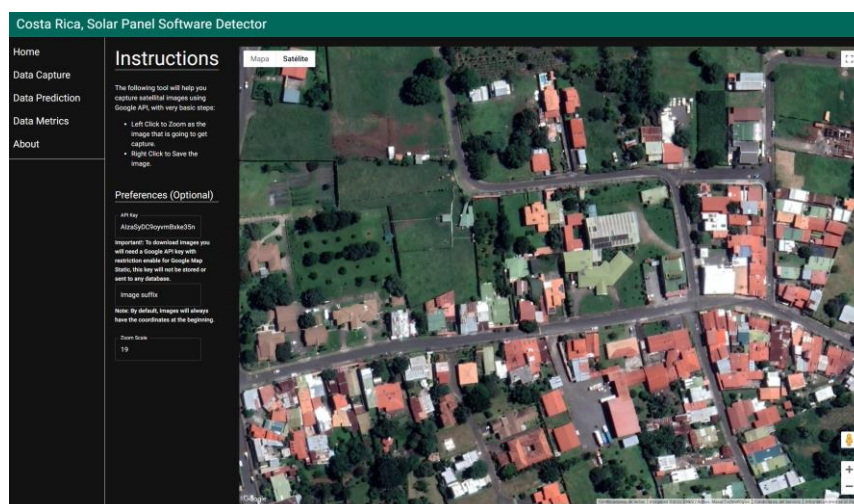
Aplicación web, página principal de aplicación



Nota: Creación propia, 2022.

Figura 2.

Herramienta de captura de imágenes satelitales.



Nota: Creacion propia, 2022

En las figuras, se muestra claramente parte de las interfaces con las que contara la aplicación web, se evaluara eventualmente si cambiar la experiencia de usuario cuando se avance con los demás objetivos.

Para la Figura 2, se integró una librería que adapta la aplicación Google Maps API y se configuro de tal forma que, para rápida captura de imágenes baste con utilizarse el ratón del computador y sus botones para generar un zoom y generar la descarga de la imagen.

En cuanto al objetivo:

- Exploración temprana de modelo de segmentación con aumentación de datos

Se hizo una exploración temprana para conocer si se pueden conseguir resultados con un set de datos tan limitado en comparación con lo investigado en el estado del arte y en efecto se comprueba se pueden obtener buenos resultados, se adjunta en los recursos, productos generados como parte de este, aclarando que aún se trabajara en ellos, pues las métricas parecen mostrar sobreentrenamiento.

En cuanto a los objetivos (Este objetivo se da como adelantamiento, por temas laborales):

- Creación de modelo de aprendizaje automático y refinamiento de modelo de aprendizaje automático.

Se ha empezado ambas fases en paralelo como adelanto en disponibilidad de tiempo, se hace entrega de modelo ya bastante avanzado a nivel de código fuente y con reporte de métricas en anexo 7.3, se espera seguir trabajando en el mismo para refinar aspectos gramaticales y de forma que en segundo PEC, se incluyan resultados en la minuta.

## 2.2 Justificación de los cambios en caso necesario

Se ha considera expandir las funcionalidades de la aplicación e indagar que tan factible es dentro del tiempo, utilizar el objetivo: "Creación de modelo de aprendizaje automático" para incluir un modelo de clasificación, aparte del modelo de detección de objetos u segmentación que se había considerado.

La justificación viene dada a la investigación del estado del arte realizada, que por la naturaleza del área es una investigación que conlleva un tiempo significativo y que sus resultados resultan complejos de comunicar y entender; por ende, se detectó que la utilización de modelos de clasificación suele ser utilizada como primera forma de detección de paneles solares, este punto sin embargo, podría llegar a ser excluido de los resultados, pues es un objetivo extra que se estaría incluyendo en caso de ser exitoso , por lo tanto se espera abordar como un entregable extra y no como algo requerido, pues se tendría que revalorizar el riesgo que implica incluirlo y se buscara una exploración temprana iniciando el segundo avance de proyecto, por lo tanto, se adjunta un nuevo análisis de riesgo, a fin de abordar como riesgo la inclusión de este modelo de clasificación, no se hace sin embargo cambios en resultados esperados de momento.

## 3. Relación de las actividades realizadas

### 3.1 Actividades previstas en el plan de trabajo

- Creación de set de datos utilizando Google Maps API
- Etiquetado manual de set de datos utilizando AWS Sagemaker Ground Truth.
- Investigación de estado del arte, a través de búsqueda bibliográfica en internet.
- Se ha avanzado en la Minuta, y se adjunta el adelanto de varias secciones. Se evita duplicar información, por ende este informe se mantiene ligero.

### 3.2 Actividades no previstas y realizadas o programas

- Se ha cambiado una de las tareas del cronograma, donde el punto de la Figura 3 que corresponde a “Aumentación de datos”, se ha cambiado a: “Exploración temprana de modelo de segmentación con aumentación de datos”, esto ya que se encontró que la aumentación de los datos se puede lograr de forma sencilla con el uso de una librería especializada.
- Se ha agregado la pieza de captura de datos a la aplicación web, para enriquecer la aplicación web.
- Se ha recolectado más de 400 imágenes correspondientes a imágenes sin paneles solares, esto para evaluar la posibilidad de un modelo de clasificación aparte del modelo de detección de paneles (detección de objetos o segmentación).
- Se ha reconsiderado tareas en cronograma; se adjunta anexo #2, que presenta cronograma actualizado, básicamente se ha reestimado el tiempo de algunas de las tareas a fin de habilitar la posible exploración de agregación de un modelo de clasificación en la última tarea.

## 4. Relación de las desviaciones en la temporización y acciones de mitigación si procede y actualización del cronograma si procede

No se detectan acciones de mitigación, en su lugar se este procedimiento a la expansión de los productos finales, al explorar de forma temprana la posibilidad de incluir un modelo extra de clasificación, cambio que se verá reflejado en anexo.

## 5. Listado de los resultados parciales obtenidos hasta el momento (entregables que se adjuntan)

- Se hace entrega de un repositorio, que contiene, el código fuente obtenido hasta el momento, de la aplicación web, creada en React y Typescript.
- Un set de datos comprimido, que contiene dentro, dos carpetas llamadas test y train, compuesto en su totalidad por alrededor de 120 imágenes.
- Un set de datos comprimido, que contiene dentro, una carpeta llamada noPanel, compuesto por alrededor de 400 imágenes,
- Jupyter Notebooks, que contienen código fuente utilizado para entrenamiento de modelos, en etapa avanzada y de refinamiento o tuning.
- Cronograma y análisis de riesgos actualizado en anexos.

## 6. Comentarios de vuestro director particular si lo consideran necesario

Se da la recomendación de incluir análisis de riesgo actualizado y de explorar parametrización de modelo.

## 7. Anexos

Anexo #1: Análisis de riesgos:

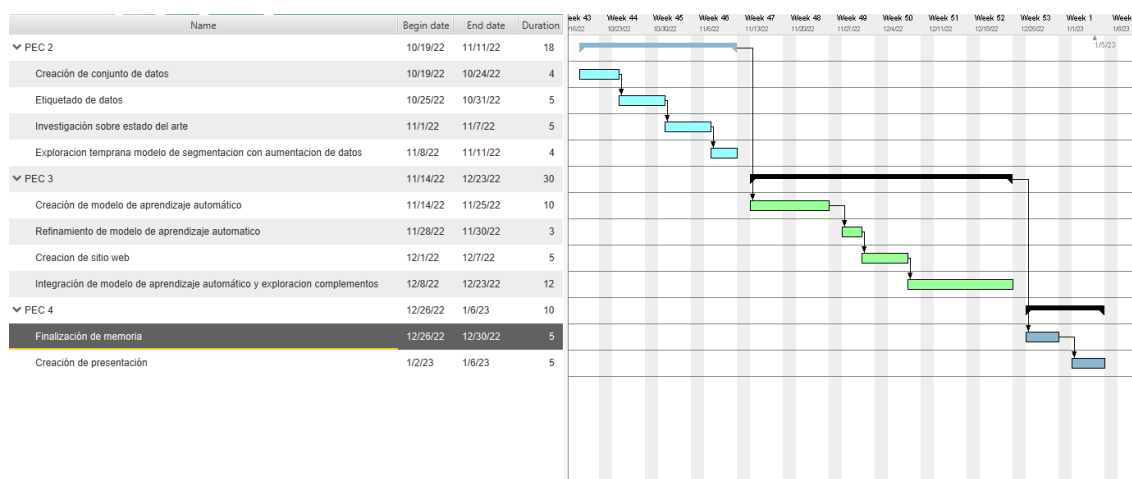
- Métricas de calidad satisfactorias:
  - Factores: Calidad de conjunto de datos recolectado
  - Descripción: Datos recolectados no sean suficientes como para producir buenas métricas en el modelo de aprendizaje automático.
  - Nivel: Medio
  - Mitigación: Modificar variables de modelo de aprendizaje automático, expansión de set de datos de forma artificial.
- Imágenes satelitales, en país con clima tropical y en desarrollo.
  - Factores: Popularidad de paneles solares en áreas residenciales
  - Descripción: Se espera que no se puedan llegar a encontrar muchos paneles solares y que existan muchas imágenes, donde cuestiones relacionadas a economía y clima afecten negativamente al modelo de aprendizaje automático, afectando por lo tanto predicciones.
  - Nivel: Medio
  - Mitigación: Expansión de set de datos de forma artificial, delimitar o reconocer variables que afectan negativamente el modelo, por ende, asumir casos donde predicción no sea exitosa.

- Actualización: Se ha recolectado y se ha encontrado que no muchas edificaciones cuentan con paneles solares, se ha hecho uso de librerías de aumentación de datos y ha resultado exitoso.
- Entendimiento y uso de estado del arte:
  - Factores: Investigación, conocimiento e implementación de las técnicas
  - Descripción: Generar solución que se alinee al estado del arte con respecto al modelo de aprendizaje automático, donde el estado del arte consuma mucho tiempo y sea muy difícil de implementar.
  - Nivel de riesgo: Alto
  - Mitigación: Lectura y exploración de diferentes alternativas y hacer pequeños experimentos en caso de ser necesario.
  - Actualización: Mitigado de forma efectiva, no se espera variaciones altas.
- Uso de recursos computacionales:
  - Factores: Recursos computacionales y propiedades de imágenes.
  - Descripción: Dependiendo de conjunto de datos y calidad de imágenes, tiempo para entrenamiento puede ser significativo y requerir uso de recursos computacionales con GPU, que poseen un costo significativo si se llega a extender por varias horas.
  - Nivel de riesgo: Medio
  - Mitigación: Aceptar posibles repercusiones a nivel de métricas o revisar planes gratuitos o opciones para casos académicos. Explorar posibilidad de generar entrenamiento en computadora propia por múltiples días para así disminuir costo.
- Posible Inclusión de modelo de clasificación.
  - Factores: Recursos computacionales y propiedades de imágenes.
  - Descripción: Estado del arte recomienda la incorporación de modelo de clasificación, aparte de modelo de segmentación.
  - Nivel de riesgo: Alto
  - Mitigación: Captura de imágenes sin paneles solares a fin de dar posibilidad a esta opción, adelanto de actividades de cronograma e inclusión de modelo de clasificación para actividades de segundo PEC en caso de contar con disponibilidad de tiempo.
  - Actualización: Se capturaron imágenes extras que serán parte de entrega parcial de resultados y se planea iniciar modelo de clasificación como parte de la tarea actualizada llamada: "integración de modelo de aprendizaje automático y exploración complementos"



## Anexo #2: Cronograma actualizado

Name	Begin date	End date	Duration
▼ PEC 2	10/19/22	11/11/22	18
Creación de conjunto de datos	10/19/22	10/24/22	4
Etiquetado de datos	10/25/22	10/31/22	5
Investigación sobre estado del arte	11/1/22	11/7/22	5
Exploracion temprana modelo de segmentacion con aumentacion de datos	11/8/22	11/11/22	4
▼ PEC 3	11/14/22	12/23/22	30
Creación de modelo de aprendizaje automático	11/14/22	11/25/22	10
Refinamiento de modelo de aprendizaje automatico	11/28/22	11/30/22	3
Creacion de sitio web	12/1/22	12/7/22	5
Integración de modelo de aprendizaje automático y exploracion complementos	12/8/22	12/23/22	12
▼ PEC 4	12/26/22	1/6/23	10
Finalización de memoria	12/26/22	12/30/22	5
Creación de presentación	1/2/23	1/6/23	5



## Anexo #3. Reporte temprano de resultados modelo de clasificacion

<https://wandb.ai/carlosan1708/Segmentation-TFM/reports/Summary-Solar-Panel-Segmentation--VmIldzoyOTk3NTk0>