**Proyecto APT Hortiscan**

***Capstone***

***Fecha: 18/09/2024***

**Profesor: ROCIO CONTRERAS AGUILA**

**Asignatura: CAPSTONE**

**Sección:** 002D

**Alumno: Lucas Rodríguez**

**Resumen**

El proyecto HortiScan aborda la problemática de la digitalización de formularios agrícolas mediante un sistema automatizado basado en reconocimiento óptico de caracteres (OCR) y algoritmos de Machine Learning y Deep Learning. Este sistema está diseñado para mejorar la precisión y velocidad en el procesamiento de grandes volúmenes de datos en sectores como la agricultura, reduciendo la dependencia de métodos manuales. El propósito de este trabajo es desarrollar una aplicación móvil que capture imágenes de formularios en terreno, los procese automáticamente y los almacene en formato digital, facilitando su acceso y edición a través de una plataforma web. Para lograrlo, se utiliza un marco teórico centrado en tecnologías de procesamiento de imágenes y aprendizaje automático. La metodología empleada incluye el desarrollo modular de la aplicación móvil y web, la implementación de un OCR avanzado y la integración de APIs dedicadas para cada plataforma. Aún en la fase inicial, los resultados esperados del proyecto incluyen la optimización de los procesos de captura y digitalización de datos, proporcionando una herramienta adaptable y escalable que puede ser implementada en otros sectores con necesidades similares.

**Palabras Clave:** OCR, Digitalización, Machine Learning, Deep Learning, Agricultura

***Abstract***

*The HortiScan project addresses the problem of digitizing agricultural forms through an automated system based on Optical Character Recognition (OCR) and Machine Learning and Deep Learning algorithms. This system is designed to improve the accuracy and speed of processing large volumes of data in sectors such as agriculture, reducing reliance on manual methods. The purpose of this work is to develop a mobile application that captures images of forms in the field, processes them automatically, and stores them in digital format, facilitating access and editing through a web platform. To achieve this, the theoretical framework focuses on image processing technologies and machine learning. The methodology involves the modular development of the mobile and web applications, the implementation of advanced OCR, and the integration of dedicated APIs for each platform. Although the project is still in its initial phase, the expected outcomes include the optimization of data capture and digitization processes, providing a scalable and adaptable tool that can be applied in other sectors with similar needs.*

***Keywords:*** *OCR, Digitization, Machine Learning, Deep Learning, Agriculture.*

**Descripción de proyecto APT**

HortiScan tiene como finalidad optimizar la captura y el procesamiento de datos mediante la implementación de un sistema automatizado de digitalización, basado en el reconocimiento óptico de caracteres (OCR) mejorado con algoritmos de Machine Learning y Deep Learning. La propuesta contempla el desarrollo de una aplicación móvil para la captura de imágenes de formularios completados en los campos de vegetación hortícola dentro de la Región Metropolitana, con validación automática de datos y almacenamiento seguro. Estas imágenes serán procesadas por un sistema de OCR que generará documentos en formato Word. Asimismo, se desarrollará una plataforma web que permitirá a los usuarios visualizar y editar los formularios procesados.

HortiScan es relevante en el campo laboral de la Ingeniería en Informática porque aborda un problema real que muchas industrias enfrentan, la necesidad de digitalizar grandes volúmenes de datos de manera eficiente y precisa. En sectores como la agricultura, logística, salud y administración pública, el manejo de documentos físicos es un desafío continuo que retrasa procesos, genera errores y demanda recursos significativos. El proyecto ofrece una solución de transferencia tecnológica al automatizar la captura de datos y su digitalización mediante OCR avanzado, permitiendo que la información se transforme rápidamente de un formato físico a uno digital sin intervención manual extensa. Esto no solo mejora la velocidad de los procesos empresariales, sino que también optimiza el uso de recursos humanos, permitiendo que el personal se concentre en tareas de mayor valor agregado.

Este proyecto, al facilitar la digitalización y gestión remota de datos, responde directamente a esas necesidades empresariales y ayuda a las organizaciones a ser más competitivas en un mundo cada vez más digital.

**Relación del proyecto APT con las competencias del perfil de egreso.**

Este proyecto de digitalización automatizada mediante OCR mejorado requiere aplicar las competencias del perfil de egreso de manera integrada, lo que refleja una relación directa entre el proyecto y las habilidades adquiridas en la carrera de Ingeniería en Informática. El desarrollo del sistema demanda no solo habilidades técnicas en el diseño y desarrollo de software, sino también la capacidad de gestionar proyectos informáticos complejos, planificando y organizando cada etapa desde la captura de datos en la aplicación móvil hasta su procesamiento en la plataforma web, asegurando que los plazos y recursos se ajusten a los requerimientos organizacionales. Además, es necesario diseñar un modelo de datos eficiente y escalable que soporte el almacenamiento de los formularios digitalizados, garantizando la integridad de la información y asegurando la sostenibilidad del sistema. Las pruebas de certificación del sistema son fundamentales para validar el funcionamiento del OCR y la interfaz web, aplicando buenas prácticas definidas por la industria, lo que asegura que cada componente cumpla con los estándares de calidad exigidos. Finalmente, el desarrollo del software debe seguir una metodología sistemática que permita integrar de manera eficiente todos los componentes, asegurando el mantenimiento y evolución del sistema en el tiempo. En conjunto, este proyecto permite aplicar las competencias adquiridas durante la carrera para enfrentar desafíos reales y entregar una solución tecnológica que optimiza procesos críticos en diversos sectores, alineándose claramente con los objetivos del perfil de egreso.

**Relación del proyecto con tus intereses profesionales.**

El proyecto APT HortiScan se alinea de manera integral con los intereses profesionales del equipo, abarcando áreas clave como la gestión de proyectos informáticos, el desarrollo de software de calidad, la ciencia de datos y la creación de soluciones web y móviles. En gestión de proyectos, les permite aplicar sus habilidades en planificación, organización y liderazgo, asegurando que cada fase del proyecto se ejecute de manera eficiente y contribuyendo al desarrollo de soluciones innovadoras. La identificación y análisis de requerimientos, junto con la implementación de estándares de calidad en el software, son fundamentales para asegurar el éxito del sistema OCR avanzado que están desarrollando.

Asimismo, la aplicación de algoritmos de machine learning y deep learning les permite mejorar la precisión del OCR, optimizando la digitalización y gestión de grandes volúmenes de datos, lo que se conecta con su interés en el uso de tecnologías avanzadas para mejorar la competitividad empresarial. En el ámbito del desarrollo web y móvil, el proyecto les brinda la oportunidad de crear soluciones multiplataforma que integren funciones críticas tanto en la plataforma web como en la aplicación móvil.

Finalmente, la gestión y almacenamiento eficiente de los datos digitalizados les permite aplicar sus conocimientos en bases de datos robustas, optimizando el procesamiento y análisis de información para la toma de decisiones estratégicas en la agricultura. Este proyecto no solo refuerza sus competencias actuales, sino que también les ofrece la oportunidad de desarrollar nuevas habilidades y contribuir de manera significativa a la digitalización de documentos en el sector agrícola.

**Argumento del por qué el proyecto es factible a realizarse dentro de la asignatura.**

Este proyecto es factible de realizar dentro de la asignatura porque, aunque presenta un desafío técnico en la implementación de un OCR avanzado, el resto de las tareas son manejables en términos de tiempo, materiales y recursos disponibles. La mayoría de los componentes, como el desarrollo de la aplicación móvil y la plataforma web, pueden implementarse utilizando tecnologías y frameworks accesibles dentro del tiempo de la asignatura. Además, el equipo cuenta con las herramientas necesarias para el desarrollo y la integración de las diferentes partes del sistema.

La mayor complejidad radica en la integración del OCR avanzado, que requiere trabajar con algoritmos de Machine Learning para mejorar la precisión del reconocimiento de caracteres. Sin embargo, el proyecto se plantea de manera modular, lo que permite abordar esta tarea en paralelo con el desarrollo de otros componentes. En caso de encontrar dificultades técnicas con el OCR, como problemas de precisión o limitaciones en los modelos, se podrían emplear técnicas de preprocesamiento de imágenes y ajustar los parámetros del algoritmo para mejorar los resultados.

**Reflection**

As a team, reflecting on the early stages of the HortiScan project, we have recognized the importance of having a well-structured and modular approach to manage complex systems. This project has collectively pushed us to think beyond basic software development and explore areas such as Optical Character Recognition (OCR) and Machine Learning. While we are still in the initial phases, the progress we have made thus far gives us confidence in our ability to overcome the technical challenges that lie ahead. These early stages of planning and development have reinforced the value of adaptability and continuous iteration, both of which are essential when working with technologies that require constant refinement. Moving forward, we are optimistic that the foundation we are building will support the successful implementation of the full system, while also enhancing our team's collective skills in project management, software development, and data processing.