

# Seminario de Proyectos I 2023

## U1 - 1C

### Resumen. Selección y descripción del tema de investigación

**Nombre:** David Aaron Ramirez Olmeda

**Programa:** Maestría en Ciencia de Datos e Información



### Nombre de la problemática: Predicción de Enfermedades a través de Patrones de Alimentación y Sellos Alimenticios

#### Descripción del problema:

El objetivo de este proyecto es desarrollar un modelo de ciencia de datos que pueda predecir el riesgo de enfermedades en individuos en función de sus patrones de alimentación y la información proporcionada por los sellos nutricionales en los alimentos. La idea es identificar las relaciones entre los hábitos alimenticios, los valores nutricionales de los alimentos y la aparición de ciertas enfermedades. Esto permitirá a las personas tomar medidas preventivas y adoptar una dieta más saludable basada en la información de los sellos alimenticios.

#### Abordaje del problema desde la Ciencia de datos:

La ciencia de datos proporciona un enfoque cuantitativo y analítico para analizar grandes conjuntos de datos relacionados con la alimentación y la salud. Se recopilarn y analizarán datos de patrones de alimentación, preferencias alimenticias, hábitos de vida, información nutricional y registros de salud. Estos datos se utilizarán para identificar patrones y tendencias, y luego se integrarán con los valores de los sellos alimenticios para entrenar un modelo predictivo. Este modelo podrá predecir el riesgo de enfermedades en función de los patrones alimenticios y la información de los sellos nutricionales.

#### Herramientas y/o algoritmos necesarios:

- Procesamiento de lenguaje natural (PLN) para analizar descripciones de alimentos y sellos nutricionales.
- Algoritmos de minería de datos para identificar patrones y relaciones en los datos.
- Modelos de aprendizaje automático, como regresión logística o clasificación, para predecir enfermedades en función de los patrones alimenticios y la información de los sellos alimenticios.

#### Descripción de los datos necesarios:

Se requieren datos que incluyan información detallada sobre los hábitos alimenticios de los individuos, preferencias alimenticias, información nutricional de los alimentos consumidos, información de los sellos nutricionales y registros de salud históricos. Estos datos permitirán identificar relaciones entre la dieta, los valores nutricionales y la aparición de enfermedades.

**Fuente(s) de los datos necesarios:** Los datos pueden provenir de encuestas alimenticias, registros médicos, aplicaciones de seguimiento de la salud, fuentes de datos abiertos gubernamentales sobre salud y alimentación, así como plataformas en línea que albergan recetas y detalles nutricionales. El uso de datos personales y de salud debe abordarse con cuidado para proteger la privacidad de los individuos involucrados.

**Limitaciones generales de tu propuesta:**

- Disponibilidad y calidad de los datos: La calidad de las predicciones depende en gran medida de la calidad y cantidad de los datos disponibles.
- Complejidad de los patrones alimenticios: Los patrones alimenticios son influenciados por una variedad de factores, lo que puede dificultar la identificación de relaciones claras.
- Limitaciones éticas y de privacidad: El uso de datos personales y de salud debe manejarse de manera ética y legal para proteger la privacidad de los individuos.

Este proyecto tiene el potencial de proporcionar información valiosa para ayudar a las personas a tomar decisiones informadas sobre su dieta y estilo de vida, y así prevenir la aparición de enfermedades.

---

---

---

**Nombre de la problemática: Predicción de Rendimientos Agrícolas para la Optimización de la Agricultura en Sinaloa**

**Descripción del problema:**

Este proyecto tiene como objetivo desarrollar un modelo de ciencia de datos que pueda predecir los rendimientos agrícolas en la región de Sinaloa, México. Dada la importancia de la agricultura en esta región, el proyecto busca brindar a los agricultores herramientas para anticipar y planificar de manera más eficiente las cosechas. La predicción precisa de los rendimientos agrícolas es crucial para la toma de decisiones en la gestión de cultivos y la distribución de recursos.

**Abordaje del problema desde la Ciencia de datos:**

La ciencia de datos proporciona una metodología poderosa para analizar datos históricos de cultivos, condiciones climáticas y prácticas agrícolas. Se recopilarán y analizarán datos relacionados con factores como el clima, el tipo de cultivo, las prácticas agrícolas y los datos históricos de rendimientos. Estos datos permitirán entrenar un modelo predictivo que pueda prever los rendimientos agrícolas en función de los diferentes factores.

**Herramientas y/o algoritmos necesarios:**

- Modelos de regresión, como regresión lineal o regresión logística, para predecir los rendimientos agrícolas en función de múltiples variables.
- Técnicas de análisis temporal para comprender las tendencias y patrones a lo largo de las temporadas agrícolas.

**Descripción de los datos necesarios:**

Los datos requeridos incluyen información detallada sobre los cultivos plantados, las fechas de siembra y cosecha, las condiciones climáticas (temperatura, humedad, precipitación, etc.), y las prácticas agrícolas utilizadas en la región de Sinaloa.

**Fuente(s) de los datos necesarios:**

Los datos podrían provenir de registros históricos de cultivos en la región, estaciones meteorológicas locales, bases de datos agrícolas y organizaciones gubernamentales relacionadas con la agricultura. Afortunadamente Sinaloa cuenta con una sección de datos abiertos, separada de la nacional, en donde se podrían recolectar estos datos.

**Aplicación práctica en Sinaloa:**

La implementación exitosa de este modelo podría brindar beneficios significativos a la comunidad agrícola de Sinaloa. Los agricultores podrían anticipar con mayor precisión los momentos óptimos para la siembra y la cosecha, gestionar los recursos de manera más eficiente y reducir los riesgos económicos asociados con la variabilidad de los rendimientos agrícolas.

**Limitaciones generales de tu propuesta:**

- Disponibilidad y calidad de los datos: La precisión de las predicciones depende de la calidad y cantidad de los datos históricos disponibles.
- Influencia de factores externos: Aunque el modelo puede predecir rendimientos en función de datos pasados, eventos inesperados como desastres naturales pueden afectar los resultados.
- Adaptabilidad: La implementación en el mundo real debe considerar la variabilidad en las condiciones y prácticas agrícolas.

Este proyecto tiene el potencial de impulsar la eficiencia y sostenibilidad de la agricultura en