

4-11-19

# Sistema de recomendación para la asignación de recursos financieros y selección de cultivos en Jalisco

**Trabajo Terminal No. 2019-B082**

**Alumnos:** Acosta Rosales Jair Sebastián, \*Mercado Núñez Sergio

**Directores:** Dra. Martínez Seis Bella Citlali, M. en C. Buitrón Dámaso Israel

\*sergiomercadonu@gmail.com

**Resumen** - En el Estado de Jalisco, los conjuntos de datos abiertos y disponibles, correspondientes al sector agrícola, otorgan la oportunidad de que, por medio de minería de datos y el aprendizaje máquina, se pueda extraer conocimiento útil y comprensible. La información obtenida, que se procesará, tendrá relación con los productos cultivados, los estímulos gubernamentales otorgados y el clima. La unificación de los conjuntos de datos y su posterior análisis, haciendo uso de algoritmos de clasificación y agrupamiento, se generará un sistema de recomendaciones al gobierno orientado a la asignación de recursos financieros que que podría brindar a través de programas agrícolas, así como sugerencias respecto a dónde, cuándo y qué sembrar dependiendo de los datos recabados [3].

**Palabras clave** - Agricultura, Aprendizaje automático, Ciencia de Datos, Minería de datos

## 1. Introducción

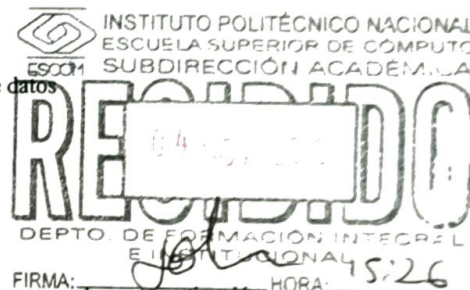
### Aprendizaje automático

El aprendizaje automático ("Machine learning", en inglés) es una herramienta para la automatización, predicción, reconocimiento y recomendación de acciones. Este se ha convertido en un tema de investigación importante en los últimos años. Dicho conjunto de técnicas, aplicado a una gran y diversa cantidad de campos de estudio, permite encontrar formas más sofisticadas para lograr que un sistema aprenda de un tópico deseado, y con base en ese aprendizaje inducido, sea capaz de tomar decisiones que permitan cumplir con una tarea; para este proyecto de investigación, el campo de estudio elegido es la agricultura en México, particularmente el estado de Jalisco debido a que es en este estado donde se encuentra una gran cantidad de datos recabados acerca de los cultivos que se producen allí [3].

### Panorama histórico de la agricultura en México

Se sabe que la agricultura es una de las actividades más importantes para el país, que permitieron tener en el año 2018 un 10% de PIB [1] solo en este sector, gracias a la destacable diversidad de productos que son cosechados, durante todo el año y en cada una de las diferentes temporadas. Desafortunadamente, es la agricultura empresarial la que, en los últimos 30 años, ha recibido los estímulos, en detrimento de los pequeños y medianos productores.

Desde fines de los años 1980, el Estado puso en marcha una política de liberalización aunada a la privatización, a la liquidación de empresas y organismos del sector paraestatal, tales como Fertilizantes Mexicanos (Fertimex), mientras que en el sector comercial, la Compañía Nacional de Subsistencias Populares (Conasupo) y sus filiales industriales y comerciales también desaparecieron. En el sector financiero, el Estado retiró su apoyo de Anagsa, así como de numerosos fideicomisos, con tareas en materia de fomento, financiamiento y organización campesina. La finalidad, la disminución y desaparición de diversas formas de subsidios en el sector agropecuario. [4]



Uno de los principales puntos de inflexión, para este sector, fue la firma del Tratado de Libre Comercio de América del Norte (TLCAN); dicho tratado se negoció teniendo en mente una agricultura moderna, mecanizada y orientada a las exportaciones o a producir insumos industriales. Se considera que, en el caso de la agricultura, el TLCAN constituyó un complemento de la política de modernización del sector. El objetivo era lograr un acceso más seguro al mercado de Estado Unidos, utilizar los acuerdos comerciales para consolidar las reformas políticas internas; atraer las inversiones extranjeras; establecer un mecanismo de solución de controversias; asegurar un crecimiento económico más acelerado y sostenido; y con ello desalentar la emigración a los Estados Unidos. Sin embargo, su principal impulsor, el presidente Carlos Salinas de Gortari, ya no tuvo tiempo y a sus sucesores no les interesó; como consecuencia aumentó la migración, y la miseria en el campo mexicano, junto con los cinturones de pobreza alrededor de las grandes ciudades. [4]

Los productores manifiestan necesidades de capacitación, asesoría técnica y servicios, mismos que sólo pueden brindarse a través de la extensión agropecuaria y que el Estado está obligado por ley a resolver. Sin embargo, también es evidente que, históricamente, dicho servicio de extensión en México no ha encontrado un modelo eficiente, por varias razones: se ha referenciado y supeditado a las organizaciones de productores legalmente constituidas, pero regularmente controladas por grupos políticos, lo que invariablemente ocasiona la simulación y el abuso, que no ayudan a aumentar la productividad y la calidad de vida en las comunidades. [4]

#### **La aportación del proyecto**

Ante esta situación, es por demás urgente apoyar los trabajos de extensión agropecuaria y el primer paso es el diagnóstico; el principal objetivo es identificar los recursos naturales disponibles, y es ahí donde se pretende participar. A través del análisis y uso de múltiples bases de datos [3] con información de carácter agrícola, se persigue transformar dichos datos en conocimiento de manera automática, por medio de la ciencia de datos, para que consecuentemente se extiendan recomendaciones racionales basadas en modelos computacionales, con respecto a la distribución de recursos del sector agrícola. Con toda la información anterior es recomendable que se defina el principal sistema-producto (refiriéndose a los recursos naturales vistos no sólo como un activo económico, sino como parte integral de las cadenas productivas, lo que implica conocimiento previo, manejo, aprovechamiento y conservación de los mismos) o cadena productiva donde pueda ayudar al desempeño de la producción agrícola.

Un sistema similar que ha sido desarrollado es:

El Sistema de recomendación de cultivos para precisión agrícola ("Crop Recommendation System for Precision Agriculture", en inglés) es un trabajo cuyo enfoque es similar con el sistema propuesto, en este se hace uso de data mining y pruebas con distintos algoritmos de clasificación y agrupamiento, como lo son K-NN, CHAID, árboles de decisión y el clasificador de Bayes.

Nombre del sistema	Características
Crop Recommendation System for Precision Agriculture [2].	<ul style="list-style-type: none"> <li>El Sistema fue conformado con múltiples algoritmos de clasificación y agrupamiento, de esta forma, todos los algoritmos predicen la pertenencia a una clase y al final la clase que haya sido elegida por la mayoría de algoritmos es la elegida para el sistema en general.</li> <li>Cuenta con una interfaz de usuario que permite el ingreso de las respectivas entradas necesarias para poder clasificar un espacio que se desea usar para cultivar.</li> </ul>



	<ul style="list-style-type: none"> <li>Las características que usaron para definir su vector característico de entrada respecto al suelo son: profundidad, textura, ph, color, permeabilidad, drenaje, retención de agua y erosión.</li> </ul>
--	--

Como podemos ver este sistema considera diferentes técnicas para obtener una mejor precisión en cuanto a la predicción de resultados, este trabajo fue hecho para el distrito Madurai en la India, por lo que está centrado en estudios del suelo y productos cultivados en ese país. En nuestro caso de estudio nos enfocaremos en los recursos que da el gobierno, a través de los programas que desarrolla la Secretaría de Agricultura y Desarrollo Rural (SADER), como el Programa de Producción para el Bienestar y el Programa de Fomento a la Agricultura. [6], es importante recalcar que en los conjuntos de datos consultados para nuestra propuesta no se cuenta con información tan descriptiva acerca de las características del suelo, en nuestro caso contamos con características como el clima, la cantidad que se produjo, el dinero necesario para realizarlo, el tipo de riego que requiere dicho cultivo, entre otras cosas [3], por lo que las recomendaciones de sembrado irían más enfocadas a las ganancias económicas que estás podrían generar.

## 2. Objetivo

**General:** Desarrollar un sistema que permita generar automáticamente recomendaciones basadas en contenido para la distribución de recursos en la producción agrícola y la selección de cultivo, haciendo uso de la ciencia de datos o en el análisis de grandes volúmenes de información.

**Específicos:**

1. Realizar una selección y análisis de los datos encontrados en las bases de datos del portal **Datos Abiertos de México** [3].
2. Realizar un preprocesamiento de los datos extraídos de las bases de datos para evitar nulos o duplicidad.
3. Realizar un estudio de algoritmos que permitan la clasificación, generación o agrupamiento de clases.
4. Generar recomendaciones a partir de la clasificación.
5. Por medio de un sistema de recomendaciones, accesible a través de una página web, dar recomendaciones al sector del gubernamental, al que le competa, acerca de la distribución del dinero a los productores así como lo productos que se pueden cultivar en ciertos terrenos y en qué momento de acuerdo a la temporada.

## 3. Justificación

El sistema de recomendación para asignación de recursos financieros y selección de cultivos en Jalisco, surge como una necesidad de saber cuáles son los productos agrícolas que pueden ser sembrados en un terreno particular, así como la temporada más apta en que se puede cultivar dicho producto en específico, ya no solo centrándose únicamente en el suelo dónde se realizará el cultivo sino también en cómo es que el clima de la temporada puede afectar o beneficiar a lo que será producido.

Por otro lado, el sector gubernamental podrá tener una idea acerca del dinero que va a ser necesarios para poder llevar a cabo el cultivo de los productos, que han sido recomendados como los indicados para ser cultivados, esto permitirá que el recurso sea aprovechado en el sector que lo necesite.

Es importante recalcar que el enfoque en el estado de Jalisco es debido a que, además de ser un de los estados que más producción agrícola tienen en el país, también cuentan con el registro [3] de lo que se ha cultivado, el costo, el tipo de riego, entre otras cosas hasta el año 2017, por lo que este sistema se puede tomar como base y ejemplo para que en otros estados se comience a hacer el mismo registro de información para un futuro uso similar.

#### 4. Productos o resultados esperados

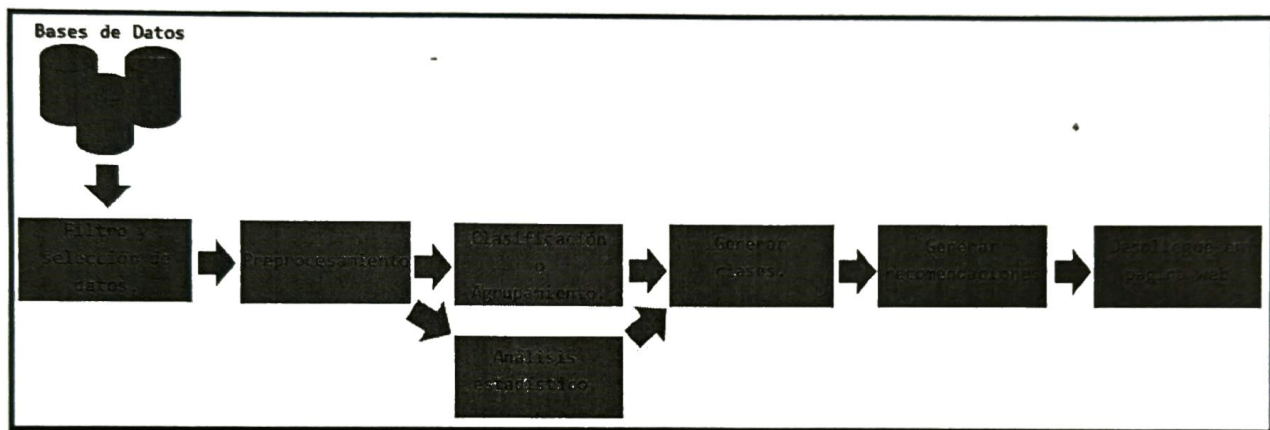


Figura 1. Arquitectura del sistema.

El proyecto será dividido en siete módulos, los cuales son:

1. **Filtro y selección de datos:** en este módulo se elegirán de las Bases de Datos aquellos campos que sean permitan dar con la solución del problema.
2. **Preprocesamiento de datos:** Se realizará la limpieza correspondiente a los datos para evitar futuros inconvenientes que puedan generar, algunos de estos inconvenientes son la duplicidad, nulos ó características ruidosas.
3. **Clasificación o agrupamiento:** Se implementarán algoritmos que permitan la clasificación o agrupamiento en relación a los datos procesados, como primera propuesta se piensa en el uso de redes neuronales convolucionales, una técnica ampliamente aplicada en el deep learning.
4. **Análisis estadístico:** Con el análisis previo de los datos se harán uso de series de tiempo que permita la clasificación del tiempo climatológico.
5. **Generar clases:** Gracias al los módulos de **Clasificación o agrupamiento** y **Análisis estadístico** se podrán encontrar patrones que ayudarán a generar las clases que serán la base para dar las recomendaciones.

6. **Generar recomendaciones:** Una vez generadas las clases, se podrá generar las recomendaciones que permitan al usuario final conocer el tipo de cultivo recomendado y los recursos necesarios para llevarlo a cabo.
7. **Despliegue en página web:** Finalmente las recomendaciones se darán en una página web.

## 5. Metodología

La metodología propuesta para el desarrollo de este proyecto es la llamada "Team Data Science Process" [6] una metodología que empresas como Microsoft emplean en proyectos de Ciencia de Datos, está es una metodología ágil iterativa que permite tener un mejor manejo de cómo es llevar a cabo un proyecto de Ciencia de Datos que puede tener la necesidad de que los datos sean analizados en repetidas ocasiones para llegar a resultados relacionados con los objetivos buscados, la metodología cuenta con cuatro bloques de actividades mostrados en la siguiente imagen:

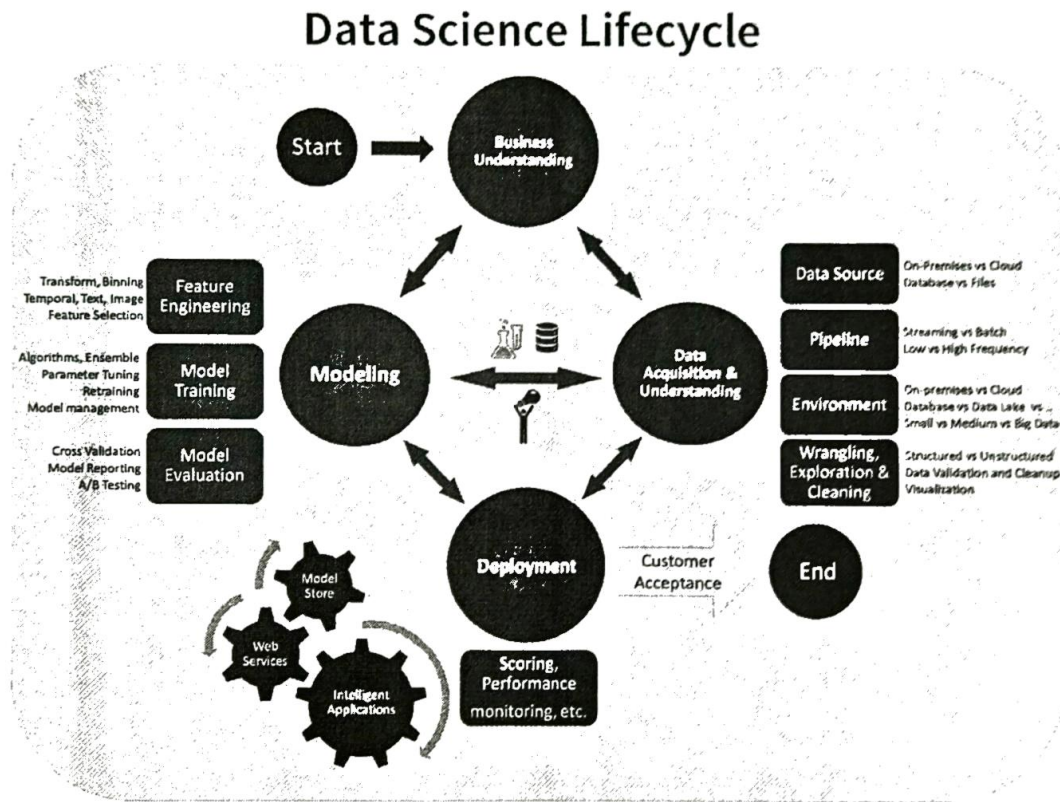


Imagen 1. Data Science Lifecycle [6]

Como podemos observar, gran parte de los bloques se interconectan, esto se justifica dado que problemas derivados en un bloque pueden ser ocasionados por alguno de los bloques anteriores, en donde, por ejemplo, no se han considerado los modelos adecuados o no se ha hecho una correcta limpieza de datos y por ende hay ruido en ello, o en el peor de los casos, el problema no ha sido entendido correctamente y no se sabe qué es lo que se va a resolver, es así como los cuatro bloques pueden ser definidos como:



- **Business Understanding:** En este bloque se analizará el problema, lo que se debe resolver y el cómo se va a resolver, es importante recalcar que si no se sabe cómo resolver el problema, es muy difícil hacer que una computadora lo haga por si misma.
- **Data acquisition & understanding:** Este bloque es implementado para obtener los datos que se desea usar y realizar la respectiva limpieza de estos, los datos pueden ser obtenidos de conjuntos de datos en internet o inclusive el mismo científico de datos puede darse a la tarea de generarlos por si mismo, esta parte es de suma importancia dado que se debe realizar una gran exploración para encontrar y hallar los datos que permitan ayudar a resolver el problema.
- **Modeling:** Con los datos obtenidos y limpios se comienza una fase de modelado de arquitecturas e implementación de algoritmos de clasificación para poder tener resultados que permitan comparar el desempeño de estos y seleccionar aquel o aquellos para la fase de despliegue.
- **Deployment:** En esta fase los algoritmos usados son implementados y monitoreados constantemente para ver su desempeño final en aplicación, para ello se despliegan en apps móviles, servicios web, etc.

Esta metodología nos permitirá mucha flexibilidad en cuanto a la retroalimentación que se pueda obtener de cada bloque, pues si se considera necesario volver a fases anteriores, es posible hacerlo, con el objetivo de que cada fase tenga un desempeño que ayude a conseguir el objetivo deseado.

## 6. Cronograma

Nombre del alumno: Mercado Nuñez Sergio

TT No.:

Título del TT: Sistema de recomendación para la asignación de recursos financieros y selección de cultivos en Jalisco.

[illegible]

TT No.:

[illegible]



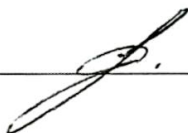
## 7. Referencias

- [1] Seminis, El Gran Potencial De La Industria Agrícola Mexicana, 2018. [En línea]. Disponible en: <https://www.seminis.mx/el-gran-potencial-de-la-industria-agricola-mexicana/>
- [2] S.Pudumalar, E.Ramanujam, R.Harine Rajashree, C.Kavya, T.Kiruthika y J.Nisha, Crop Recommendation System for Precision Agriculture, 2016. [En línea]. Disponible en: <http://agri.ckcest.cn/ass/8185d605-6c4d-4d8a-b280-c867c2304d42.pdf>
- [3] SIAP, Datos Abiertos de México, 2017. [En línea]. Disponible en: [https://datos.gob.mx/busca/dataset/datos-estadisticos-de-la-produccion-agricola-generada-a-nivel-nacional/resou/rce/584a209f-c5cb-4a2d-b5ee-312cc32cada1?inner\\_span=True](https://datos.gob.mx/busca/dataset/datos-estadisticos-de-la-produccion-agricola-generada-a-nivel-nacional/resou/rce/584a209f-c5cb-4a2d-b5ee-312cc32cada1?inner_span=True)
- [4] S. Mena Munguía, M. Ramírez Martínez, Panorama de la agricultura en México, Jalisco, México: Editorial Universitaria | Libros UDG, 2014.
- [5] Programas Agricultura 2019, 2019, [En línea]. Disponible en: <https://www.gob.mx/agricultura/acciones-y-programas/programas-sader-2019-194732>
- [6] Microsoft, What is the Team Data Science Process?, 2017. [En línea]. Disponible en: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>

## 8. Alumnos y Directores

Sergio Mercado Núñez.- Estudiante de la carrera de Ingeniería en Sistemas Computacionales. Boleta: 2015630303, Teléfono: 5527569024, Correo electrónico: [sergiomercadonu@gmail.com](mailto:sergiomercadonu@gmail.com)

Firma: \_\_\_\_\_



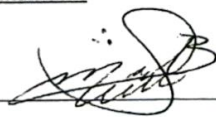
Jair Sebastián Acosta Rosales.- Estudiante de la carrera de Ingeniería en Sistemas Computacionales. Boleta: 2015630004. Teléfono: 5568823983. Correo electrónico: [jSebastian\\_AR@outlook.com](mailto:jSebastian_AR@outlook.com)

Firma: \_\_\_\_\_



Bella Citlali Martínez Seis. Dra. en Ciencias en Computación del CINVESTAV-IPN en 2018, M. en C. de la Computación del CIC-IPN en 2010, Ing. en Sistemas Computacionales de ESCOM-IPN en 2008, Profesor de UPIITA-IPN desde 2010. Áreas de interés: DM. Ext. 56912 Correo electrónico: [bellims@gmail.com](mailto:bellims@gmail.com)

Firma: \_\_\_\_\_



Israel Buitrón Dámaso.- M en C. en Computación por el CINVESTAV-IPN e Ing. en Sistemas Computacionales por la ESCOM-IPN. Realizó sus estudios de posgrado en temas relacionados con la criptografía y la combinatoria aplicada a protocolos de autenticación, Áreas de interés: criptografía, teoría de grafos, combinatoria, dinero electrónico. Datos de contacto: Teléfono: 57296000, Extensión: 54021, Correo electrónico: [ibuitron@ipn.mx](mailto:ibuitron@ipn.mx), Página web: [www.comunidad.escom.ipn.mx/ibuitron](http://www.comunidad.escom.ipn.mx/ibuitron).

Firma: \_\_\_\_\_



CARÁCTER: Confidencial

FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública.

PARTES CONFIDENCIALES: Número de boleta y teléfono.