**Propuesta Proyecto Final de Graduación**

1. **Integrantes del Grupo:**

* Juan Martín Preciado Rodriguez
* Celular +52 66228193338
* Oficina +52 6622819338 ext 259

1. **Título:** Determinación de perfiles productivos de tres estados del sur de México.
2. **Descripción:** Dar una explicación general de su proyecto en no más de una página. Mencionar al menos los siguientes elementos:
   * El estudio exploratorio obedece a un proyecto de intervención en zonas marginadas de México, los datos se relacionan con producción agrícola, desarrollo social y tenencia de la tierra.
   * Los datos de producción agrícola comprenden desde 2006-2016, los cuales al tener distinta cardinalidad respecto a los municipios, se plantean dos alternativas: 1) utilizar el promedio anual; 2) utilizar la tendencia histórica de los indicadores (lo cual permitirá expresar el cambio a través de los once años).
   * El problema a solucionar es explorar de forma multidimensional el desarrollo rural de las región, con el propósito de generar conocimiento que permita elevar la eficiencia y eficacia de un marco muestral para intervención*.*

**Objetivo General:** Identificar visualmente los perfiles productivos de los municipios (844) de cuatro estados del sur de México (Michoacán, Guerrero, Oaxaca y Chiapas) con base a su historial de producción agrícola, marginación, uso de la tierra.

1. **Objetivos Específicos: 1)** Conformación de base de datos a partir de la integración de tres fuentes de datos (SAGARPA: Secretaría de Agricultura Ganadería Desarrollo Rural, Pesca y Alimentación; CONEVAL: Consejo Nacional de Evaluación de la Política de Desarrollo Soncial y RAN: Registro Agrario Nacional, INEGI: Instituto Nacional de Estadística y Geografía); 2) Obtención de los perfiles característicos a partir de la aplicación del algoritmo k-medias y su validación; 3) Formación de capas por perfil y mapeado de muncipios.
2. **Explicación de los Datos:**

* SAGARPA: Historial de producción agrícola, CONEVAL: Índices de marginación y población total por municipio, RAN: Tenencia y uso de la tierra.
* SAGARPA: Estado, Municipio, Cultivo, Año, Superficie sembrada, Superficie Cosechada, Rendimiento, Volumen, Precio y Valor de la producción; CONEVAL: Estado, Municipio, Población total e Índice de Marginación; RAN: Estado, Municipio, Superficie Parcelada, % Agrícola, % Ganadero, % Agropecuario, Forestal y Otros usos; INEGI: Archivo SHP, donde se incluyen los polígonos para cada estado y municipio (Datos que conforman el SpatialPolygonsDataFrame).
* SAGARPA: 399056 Renglones, 22 Columnas; CONEVAL: 14646 Renglones, 23 Columnas; RAN en conformación; INEGI: ESRI Shapefile con los polígonos de todos los municipios de México.

1. **Tipo(s) de Problema:** Seleccione una o varias de las siguientes opciones:

|  |  |
| --- | --- |
| Clustering (aprendizaje no supervisado) | |
| Predictivo (aprendizaje supervisado) | |
| Regresión | Cubos OLAP |
| Series de tiempo |  |
| Otro (especifique) Análisis espacial de los perfiles productivos. | |

1. **Método(s) a Utilizar:** Seleccione una o varias de las siguientes opciones:

|  |  |
| --- | --- |
| OLAP | Árboles de decisión |
| ACP | Máquinas vectoriales de soporte |
| Agrupación jerárquica | Bosques aleatorios |
| K-medias | Métodos de potenciación |
| K vecinos más cercanos | Regresión lineal |
| Regresión logística | Regresión RIDGE |
| Análisis discriminante lineal | Regresión LASSO |
| Análisis discriminante cuadrático | Regresión Elastic Net |
| Métodos bayesianos | Series de tiempo |
| Redes neuronales | Métodos simbólicos |
| Otro (especifique) Análisis espacial y mapeo a partir de la segmentación obtenida con el algoritmo K-Medias (Se descarta el uso de algoritmo jerárquico por dos razones: 1) la cantidad de municipios 884, lo cual visualmente es poco informativo; 2) expresar los grupos a través de un mapa permite identificar fácilmente el agrupamiento de los municipios, además de brindar información del acceso si se utilizan mapas *raster* a través de la librería “*leaflet*”, brindado información adicional para la definición del marco muestral, que es otros usos que se le dará al proyecto. ). | |

1. **Software(s) a Utilizar:**

|  |  |
| --- | --- |
| R directamente | FactoMineR |
| Rattle | Paquetes para “Big Data” |
| Microsoft Analysis Services | |
| Otro (especifique) | |

1. **Comentarios Adicionales:** Este proyecto es parte del análisis exploratorio de la región para la definición de un marco muestral para acopio de información..
2. **Evaluación:**

* El proyecto debe ser entregado en un documento auto-reproducibles o en un archivo PDF.
* El documento lo que debe incluir es un detalle paso a paso sobre la aplicación de la metodología CRISP-DM al problema seleccionado.
* La evaluación será la siguiente:
  + Selección correcta del o los métodos = 15%
  + Selección correcta del software = 15%
  + Aplicación adecuada de CRISP-DM = 15%
  + Calidad del documento = 15%
  + Defensa del proyecto = 40%

1. Números de teléfono celular con código internacional de los integrantes del grupo para llamarlos en caso de problemas de comunicación el día de la defensa.
2. **Citas para la Presentación del Proyecto:** Seleccione mínimo tres fechas (en diferentes días y horas) en las que podría hacer la defensa del proyecto, la defensa tendrá una duración máxima de 1:30 minutos.

***NOTA: Escoger muy bien esas fechas porque por ninguna razón habrán cambios o postergaciones, quienes no se presenten el día de la defensa perderán el curso.***

|  |  |
| --- | --- |
| 1 de noviembre 6pm | 5 de noviembre 6pm |
| 1 de noviembre 7:30pm | 5 de noviembre 7:30pm |
| 6 de noviembre 6pm | 7 de noviembre 6pm |
| 6 de noviembre 7:30pm | 7 de noviembre 7:30pm |
|  | 8 de noviembre 6pm |
|  | 8 de noviembre 7:30pm |
| 12 de noviembre 6pm | 13 de noviembre 6pm |
| 12 de noviembre 7:30pm | 13 de noviembre 7:30pm |
| 14 de noviembre 6pm | 15 de noviembre 6pm |
| 14 de noviembre 7:30pm | 15 de noviembre 7:30pm |