Quien suscribe, estudiante regular de la Maestría en Ciencia de Datos e Información

|  |  |
| --- | --- |
| **Nombre:** | **Pablo Gómez García** |

Solicitamos ante el Coordinador Académico de la Maestría en Ciencia de Datos e Información, la autorización y registro para titulación en la Maestría, en la modalidad:

|  |  |
| --- | --- |
|  | Reporte Analítico de Experiencia Laboral |
|  | Propuesta de Intervención |
| **X** | Implementación de un Proyecto |
|  | Solución Estratégica |

**1.- INFORMACIÓN REFERIDA AL TEMA**

|  |  |
| --- | --- |
| **TITULO TENTATIVO DEL PROYECTO** | Segmentación de Municipios de la República Mexicana |

**2.- RESUMEN**

2.1. **Introducción y antecedentes del tema** (completar por el alumno en digital, máximo 20líneas)

|  |
| --- |
| En la actualidad existen en México diversos estudios a nivel municipio, en su mayoría muestran características poblacionales de cada uno de ellos, también muestran estados de opinión, preferencias electorales, índices de Pobreza y Marginación etc.  Toda la información que se tiene acerca de los municipios es valiosa para distintos fines, sin embargo a la fecha hay muy pocos estudios que describen a los municipios por medio de varias de sus características, si es que existe tales estudios lo hacen a nivel descriptivo simple.  Dentro de este proyecto planteamos la posibilidad de aplicar una técnica no supervisada para encontrar una segmentación muy característica de los municipios del País. |

2.2. **Planteamiento del problema**. (Completar por el alumno en digital, máximo 20 líneas)

Aquí se debe explicar muy claramente el problema que se pretende investigar.

|  |
| --- |
| Tomando como punto de partida la información pública que existe de los municipios en diversos temas, plateamos la siguiente Hipótesis:  “*Existen variables entre los datos públicos de los municipios que permiten diferenciarlos en segmentos cuyos elementos que los conforman poseen características homogéneas entre sí pero heterogéneas con los elementos de otros segmentos*”.  Bajo el supuesto anterior se propone recaudar información pública de fuentes como el INEGI, la PGR, la CNBV entre otras más. Con esta información se procedería a construir un Modelo no supervisado para encontrar los segmentos planteados en la hipótesis. Cabe mencionar que este Modelo no tendrá como objetivo clasificar municipios que sean creados en el futuro pues esto no sucede a menudo, sin embargo busca mostrar de manera descriptiva el valor y aporte de cada segmento para los fines planteados en el objetivo del proyecto. |

2.3. **Objetivo General** (máximo 5 líneas) y **Objetivos Específicos** (máximo 10 líneas), (completar por el alumno en digital)

Aquí se debe indicar muy claramente el objetivo general y los objetivos específicos

|  |
| --- |
| **OBJETIVO GENERAL** |
| Plantear una solución desde un enfoque Analítico basada en un Modelo no supervisado que logre segmentar a los municipios de la República Mexicana. Se pretende que dicha segmentación resulte de utilidad para estrategias de implementación de Políticas Públicas, de Marketing y para estrategias de Seguridad Pública. |

|  |
| --- |
| **OBJETIVOS ESPECÍFICOS** |
| •Recabar información histórica de al menos 2 años de profundidad para las variables que se van a considerar con la finalidad de que el Modelo tenga un periodo de observación lo suficientemente robusto.  •Al finalizar el proyecto se pretende presentar a los segmentos encontrados de manera descriptiva por medio de las variables que los hacen diferentes.  •Plantear conclusiones referentes al aporte que da cada segmento a las posibles estrategias para Políticas Públicas, Marketing y de Seguridad Pública (siendo conscientes de que los segmentos encontrados pueden aportar a una o a varias estrategias o bien a ninguna). |

**3. Resultados Esperados**

3.1. Resultados a los que se pretende llegar con éste trabajo (completar por el alumno, digital, máximo 10 líneas)

|  |
| --- |
| Como lo planteamos en el objetivo del proyecto lo que se espera es encontrar una segmentación de los municipios que resulte de utilidad para estrategias de implementación de Políticas Públicas, de Marketing y para estrategias de Seguridad Pública. Tenemos claro que las técnicas no supervisadas no son predecibles en cuanto al número de segmentos que pueden arrojar, sin embargo pretendemos que sean al menos 7 perfectamente diferentes y claros de explicar referente a sus características. |

**4. Programa de Trabajo**

4.1 Cronograma de actividades estimado para el desarrollo del tema.

Aquí deben especificarse las fechas estimativas para el desarrollo del proyecto

|  |
| --- |
|  |

**5. Metodología**

5.1. Metodología a utilizar para el desarrollo del tema (Completar por el Alumno, en digital, máximo 10 líneas).

|  |
| --- |
| La metodología a utilizar para este proyecto es la propuesta por SAS Institute conocida como ***SEMMA*** (Sample, Explore, Model, Modify and Assess). Las etapas se emplearan de la siguiente manera:   * **Muestreo (Sample):** Se realizará un muestreo representativo previo al análisis de los datos. * **Exploración (Explore):** Se explorarán los datos a través de técnicas estadísticas para identificar comportamientos que puedan impactar en el desempeño del modelo. * **Modificación (Modify)**: Se realizará una selección y transformación de los datos de acuerdo con las variables seleccionadas para el proceso de modelado. * **Modelación (Model):** Se realizará el entrenamiento de varios algoritmos de modelación estadística para entrenar el modelo buscado. * **Evaluación (Assess)**: Se valorarán los resultados obtenidos en el paso anterior en datos distintos a aquellos con los que se construyó para elegir el mejor modelo. |

**6. Fuentes de Información**

6.1 Bibliografía física y electrónica mínima a utilizar (completar por el alumno, en digital, mínimo citar 5 referencias).

|  |
| --- |
| 1. Mitchell, Tom, “Machine Learning”, Ed. McGraw-Hill (1997), cap 6 pp 154-199.  2. Everitt, B.S. (2011). Cluster analysis, 5th Edition. Wiley.  3. Peña Sánchez de Rivera, D. “Estadística. Modelos y Métodos. Volumen 2” Ed. Alianza. Madrid, 1987  4. Introduction to machine learning, Third Edition. Ethem Alpaydin. MIT Press  5. Understanding Machine Learning: From Theory to Algorithms. Shai Shalev-Shwartz and Shai Ben-David. Cambridge University Press. |

**7. Índice Tentativo**  (Máximo 6 líneas)

|  |
| --- |
| Introducción  Capítulo 1: Planteamiento de la Problemática.  Capítulo 2: Técnicas de modelación no supervisadas.  Capítulo 3: Metodología aplicada al problema.  Capítulo 5: Resultados y Conclusiones. |