**TALENTO TECH 2024-MINTIC**

**FORMATO DE PRESENTACIÓN “PLAN DE PROYECTO TI”**

**Contexto específico de aplicación del proyecto** *(Marque con una X)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **AGRO** | **EDUCACIÓN** | **TURISMO** | **GOBIERNO** | **FINANZAS** | **MARKETING** | **SALUD** | **OTRO** |
|  |  |  |  |  |  |  | **X** |

**Cohorte #: \_\_01\_\_ Año: \_\_2024\_\_\_ Tutor: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Nombre del Proyecto (y del producto/servicio):**

|  |
| --- |
| **Análisis del dataset Life expectancy, para encontrar patrones y conseguir una predicción y segmentación de datos** |

**Departamento de residencia del estudiante:**

|  |
| --- |
| Cauca |

**Municipio de residencia del estudiante:**

|  |
| --- |
| **Popayán** |

**Rural:** (Marque con una X)

|  |  |  |  |
| --- | --- | --- | --- |
| **SI** |  | **NO** | **X** |
| **Vereda o Corregimiento:** | | | | |  |

**Autor (es):**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Nombres y Apellidos** | **Tipo de identificación** | **No. identificación** | **Curso: Programación, Inteligencia Artificial, Análisis Datos, Block Chain, Arquitectura Nube** | **Nivel: Explorador, Integrador, Innovador** | **Modalidad: Virtual, Semipresencial o Presencial** |
| 1 | Eduardo Castro | cc | 1058789149 | Análisis de Datos | Integrador | Virtual |

**Palabras clave:**

|  |  |
| --- | --- |
| **Palabra clave 1** | Regresion Lineal |
| **Palabra clave 2** | Clustering |
| **Palabra clave 3** | Life expectancy |
| **Palabra clave 4** | Reduccion de dimensionalidad |

**Planteamiento del problema que solucionará el producto/servicio:**

|  |
| --- |
| Qué sucede?  R/  La expectativa de vida es algo que siempre ha sido una métrica que distingue a un país de otro, en este caso queremos averiguar que variables del dataset son las que mas afectan a la expectativa de vida de la población de una nación  Por qué sucede?  R/  La desigualdad en los países hace que una parte de la población cuente con mejores privilegios en relación con otros habitantes que poseen menos recursos económicos  A quiénes afecta?  R/  A todas las personas que habitamos una nación  De qué manera?  R/  Enfermedades, disminución de la expectativa de vida, mala cobertura en salud. |

**Pertinencia del proyecto TI:**

|  |
| --- |
| **Pertinencia:**  Cómo funciona el producto/servicio a desarrollar?  Un análisis de datos de un conjunto de datos tiene muchas fases antes de su implementación, entre ellas el análisis exploratorio ahí encontramos la distribución de las variables, datos atípicos, Categorización etc , después de esto se realiza un modelo de regresión lineal, el cual mediante los datos reales del dataset va predecir unos nuevos datos, los cuales van a estar muy relacionados con el dato real.  En qué beneficia a los usuarios?  Cuando queremos hacer una predicción sobre la expectativa de vida de una nación, cuando queremos ver los factores donde no se pone mucha atención.  **Mercado:**  Qué tamaño tiene el mercado y la oportunidad?  R/  Empresas de análisis de datos, entidades gubernamentales  Es un mercado en crecimiento  R/  Si, el análisis de datos va en un aumento exponencial y con el nuevo auge de la inteligencia artificial será mucho mas relevante  Cuáles son las tendencias?  R/  Análisis de datos para escuelas, colegios, universidades, para evitar la deserción estudiantil, generar predicciones de datos climatograficos.etc |

**Objetivos:**

|  |
| --- |
| **General**  ***Analizar el dataset encontrado en Kagle sobre la expectativa de vida para generar un modelo de predicción y clusterizacion y encontrar tendencias sobre la expectativa de vida***  **Específicos**   1. Realizar el análisis exploratorio de datos 2. Aplicar algoritmos de imputación para el tratamiento de datos 3. Desarrollar la reducción de dimensionalidad aplicando el método PCA 4. Aplicar regresión lineal múltiple con variables reducidas 5. Generar una predicción en la expectativa de vida 6. Segmentar el dataset mediante el algoritmo KMeans |

**Metodología**:

|  |
| --- |
| **Propia de cada área (Revisar los documentos anexos a este documento)** |

**Plazo:** Duración del proyecto.

|  |  |
| --- | --- |
| **SEMANAS** | **DIAS** |
| 4 |  |

**CRONOGRAMA DE ACTIVIDADES** (Diagrama de Gantt): *Detalle las actividades lo más preciso posible e indique para cada una su duración en semanas, y la secuencialidad.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Actividad** | **S1** | **S2** | **S3** | **S4** | **S5** | **S6** | **Responsable** |
| 1 | Búsqueda de información | X |  |  |  |  |  | Eduardo Castro |
| 2 | Análisis exploratorio de datos |  | X |  |  |  |  | Eduardo Castro |
| 3 | Tratamiento de datos |  | x | x |  |  |  | Eduardo Castro |
| 4 | Construcción de modelo de regresión lineal |  |  |  | x |  |  | Eduardo Castro |
| 5 | Construcción de modelo de clusterizacion |  |  |  | x | x |  | Eduardo Castro |
| 6 | Evaluación de los modelos |  |  |  |  | x |  | Eduardo Castro |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |