

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES
DE MONTERREY



Especificación de Datos

Version 0.1

Inteligencia artificial avanzada para la ciencia de datos II
(Gpo 501)

Camila P. Cusicanqui Padilla A00571258

Suemy Aquino Zumaya A00828585

Javier de Golferichs García A01139500

Franco Quintanilla Fuentes A00826953

Jesús David Núñez Rodríguez A01634928

Facundo Vecchi A01283666

Ricardo Andres Arriaga Quezada A01570553

Emilia Victoria Jácome Iñiguez A00828347

1 HERRAMIENTAS Y TECNOLOGÍAS

Expliquen qué herramientas y tecnologías van a usar para trabajar con los datos, y por qué creen que estas sean la mejor combinación.

Las herramientas que se estarán utilizando en este proyecto corresponden a Apache Spark dado que la INEGI almacena sus datos en este formato, para la parte del proceso del lenguaje natural se utilizará tanto Neuraan como Amazon Lex para aprovechar al máximo lo mejor de ambas APIs, finalmente para la realización del front end, se realizará en React.

Se siguen los tutoriales otorgados por el INEGI para su correcta descarga y uso:

Parte 1:

<https://abxda.medium.com/geo-big-data-desde-cero-parte-1-e48d7ac2c7eb>

Parte 2:

<https://abxda.medium.com/geo-big-data-desde-cero-parte-2-85121f5d654d>

Parte 3:

<https://abxda.medium.com/geo-big-data-desde-cero-parte-3-a5180a86133d>

Ejemplo:

<https://abxda.medium.com/auto-machine-learning-geo-f540406c525a>

Neuraan API:

<https://neuraan.com>

2 BIG DATA

De acuerdo con la cantidad de datos que se tiene determine si es necesario utilizar un enfoque orientado a Big Data o no.

Juzgando por la cantidad de datos que se tiene, se podría considerar que se deberán aplicar técnicas de Big Data, así como de visualización y análisis de los datos del INEGI. Sin embargo, para el desarrollo de este proyecto, no se aplicarán dichas técnicas ya que únicamente se manejarán los datos sin analizarlos para un contexto específico.

El uso de NLP (Natural Language Processing) no se puede considerar como Big Data ya que es el área encargada de la interacción entre humano y computadora, la cual funcionara para obtener la solicitud del usuario y filtrar los datos proporcionados por el INEGI.