**PREGUNTA PROBLEMA.**

**¿Como mediante un algoritmo de programación se puede determinar la dificultad de compresión y legibilidad textual de un artículo científico?**

**METODOLOGIA:**

**Fase de Planeación.**

Se define que la densidad léxica equivale a que mayor número de palabras diferentes por texto mayor dificultad. Medir el número de palabras por oración, obteniendo así el índice de longitud oracional, y el número de frases complejas que hay por oración para obtener la complejidad de oración. El numero promedio de signos de puntuación se utiliza como uno de los indicadores de complejidad. El índice SRR, Se centra en medir el vocabulario y la estructura de oraciones para predecir la dificultad relativa de legibilidad.

**Fase de Análisis.**

Las herramientas para utilizar para el análisis de texto y minería de texto ejercen una carga de tiempo y recursos a la hora de ejecutarse, mediante el análisis presente se automatizaron las funciones, ciclos y procedimientos del algoritmo para que resultara de manera óptima y sin ninguna Interversión manual por parte del usuario.

**Fase de Desarrollo.**

Técnicas de PLN para la extracción y minería de datos en Python. La herramienta NLTK, utiliza el texto del articulo y con ello podemos definir mediante distintos métodos; enumeración de palabras, longitud, signos, modificadores, letras, oraciones, frases complejas. Se utilizo el paquete de recursos para idioma inglés de Spacy, Esta es una canalización en inglés entrenada en texto web escrito (blogs, noticias, comentarios), que incluye vocabulario, sintaxis y entidades.