**Reflection Journal: Text Representation**

**Key Insights**

**Reflective Journal: A Deep Dive into Text Representation**

**Key Insights**

This exercise has provided more insights into the learning about the foundational concepts of converting words into numbers with different techniques that support modern artificial intelligence systems. Here are high level takeaways that I have learned from the exercise

* **Numerical Representation:** Of interest, converting text to numbers is fascinating to me and the application of vector in classification using the nearest neighbor concept in defining relationships between words. This has provided machine learning the opportunity to use mathematical constructs in determining the semantics around text representation for further analysis.
* **Evolution of a-contextual to Contextual Representations:** The exercise provides the concepts of Bag of Words to more sophisticated method such as term frequency inverse - inverse document frequency in measuring how rare a word is across all documents and assigning score to it, and the semantic richness of word embeddings. Text progression shows trade-off between simplicity and the ability to capture the nuances of human language.
* **Distributional Hypothesis:** This is a concept where nearest neighbor is applied in defining theoretical underpinning of word embeddings which allow the capturing of semantic relationship between words in a structure that is surprisingly intuitive.
* **The "Magic" of Word Arithmetic:** The "king - man + woman ≈ queen" analogy was a real "aha!" moment for me. It's a powerful demonstration of how word embeddings can capture the underlying structure and meaning of language in a way that is both tangible and powerful.

**Challenges**

* **Vector Spaces:** Vector space analysis is numerical and it was challenging understanding the concept of high-dimensional vector space in relation to text. The analysis with the results obtained after running the code really made something abstract to concrete evidence.
* **Trade-offs:** The selection of the right application to use for any project require understanding the intents of the users and what are the trade-offs between the different representation methods. For example, while word embeddings are powerful, they are also more complex and less interpretable than simpler methods like BOW and TF-IDF.

**Connections to Real-World Applications**

* **Search Engines:** This is one of the real-world applications where the importance of TF-IDF is heavily used to identify rare words and score the relevance of documents perfectly.
* **Recommendation Systems:** The notion of using word embeddings to represent items in a recommendation system is a powerful one as being used by Netfliz or Amazon making it easier for these companies to project what customers will like based on what has happened before. This pre-trained model is highly beneficial as marketing strategy.
* **Sentiment Analysis:** The text classification exercise was a great introduction to the world of sentiment analysis. This application is commonly used on customer reviews from social media.

**Questions That Arose**

* **How do we handle the nuances of language?** Embedding words are required in text representation but how can sarcasm, cultural sentiment be understood in a text without creating ethical issues?
* **Ethical implications of biased data?** The exercise showed impacts of bias in training data, but how can this be addressed in the future to eliminate different biases that can raise questionable issues in the future?

**Future Applications**

* **Health:** Text representation could be used to analyze a person's health issues which can be a game changer in the health industry.
* **Smart cities:** Applications of this technology will definitely accelerate the drive for smart home and cities globally.

In conclusion, this exercise is very informative and influence my curiosity to learn more on the applications of artificial intelligence in every sphere of life.