

# Clasificación de Texto sobre notas periodísticas sobre el uso del Aguacate en el Super Bowl (2024 y 2025)

Cynthia Martinez Espinoza

Matricula 1011238

Correo electrónico: [cynthia.martineze@uanl.edu.mx](mailto:cynthia.martineze@uanl.edu.mx)

**Index Terms**—Palabras clave: Descarte Palabras, tokenizar , palabras frecuentes.

- Distribución de las palabras más comunes
- Gráficos de relaciones (bigramas, trigramas)

## I. INTRODUCCIÓN

Este trabajo consiste en categorizar y organizar un texto para ayudar en la búsqueda, filtrado y análisis de información.

Analizar y comparar noticias relacionadas con el consumo y la relevancia del aguacate durante los Super Bowls de 2024 y 2025, utilizando técnicas de procesamiento de texto.

El aguacate es un producto destacado en eventos deportivos como el Super Bowl debido al consumo de guacamole. Este informe analiza dos noticias de diferentes años para identificar patrones y cambios en el tratamiento del tema.

## II. PLANTEAMIENTO DEL PROBLEMA

Analizar el texto de las diferentes notas periodísticas de periodico milenio sobre la importación de aguacate de Estados Unidos desde México en el periodo del super bowl.



Fig. 1. Objetivo Clasificacion Texto

### III. SOLUCION PROPUESTA

- Preparación de los datos: se seleccionan los notas periodísticas de milenio monterrey y se cargan los datos correspondientes.
- Preparación de los datos: Se extraen los datos de la página directa de milenio noticias con la librería de newspaper3k.
- Se procesan los datos simplificando el texto (tokenizar)
- Se aplica el stemming
- Identificación de palabras clave
- Se realiza análisis descriptivo de cada fuente.

## IV. EXPERIMENTACIÓN

Se identificaron patrones, palabras clave y relaciones significativas en las noticias seleccionadas.

Total de palabras Nota del 2024: 515

Palabras únicas Nota del 2024: 304

Longitud promedio de palabras Nota del 2024: 7.04

Total de palabras Nota del 2025: 189

Palabras únicas Nota del 2025: 155

Longitud promedio de palabras Nota del 2025: 7.34



Fig. 2. Grafo Nota 2024

## V. ANALISIS DE RESULTADOS

En ambas noticias, las palabras más comunes giran entorno a términos como "aguacate", "México", "super", "bowl", y "exportacion".

Los bigramas revelaron las relaciones más relevantes en el texto: Super Bowl 2024: Relaciones como "super bowl", "bowl aguacate" y "guacamole consumo".

Super Bowl 2025: Bigramas como "aguacate produccion" y "incremento demanda" reflejan un enfoque más en la producción y consumo.

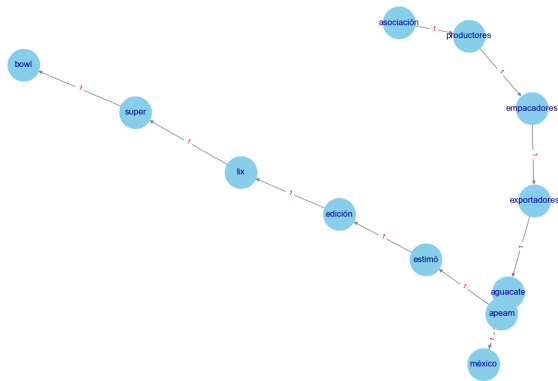


Fig. 3. Grafo Nota 2025

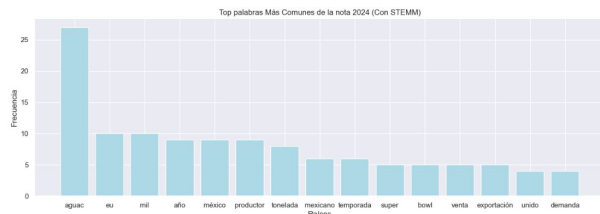


Fig. 4. Palabras Frecuentes Nota 2024

Mientras que la noticia de 2024 se enfoca más en el consumo general durante el evento, la de 2025 resalta el impacto económico y la creciente demanda de aguacates.

## VI. CONCLUSIÓN

El análisis mostró que el tema del aguacate en el Super Bowl evoluciona con el tiempo, pasando de un enfoque en el consumo hacia un interés por los aspectos económicos y de producción.

Los bigramas y las palabras clave proporcionan una visión clara de cómo las noticias estructuran la narrativa en torno al aguacate.

## REFERENCES

- [1] Bird, S., Klein, E., Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
- Manning, C. D., Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.
- Jurafsky, D., Martin, J. H. (2019). Speech and Language Processing. Pearson Education.

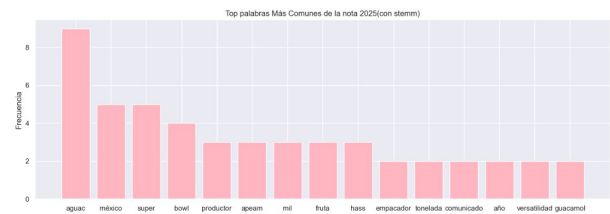


Fig. 5. Palabras Frecuentes Nota 2025