

Clasificación de Texto sobre Notas Periodísticas

Uso del Aguacate en el Super Bowl (2024 y 2025)

Cynthia Martínez Espinoza
Matrícula: 1011238
cynthia.martineze@uanl.edu.mx

16 de enero de 2025

Resumen

Este trabajo consiste en categorizar y organizar un texto para ayudar en la búsqueda, filtrado y análisis de información. Se analizan y comparan noticias relacionadas con el consumo y la relevancia del aguacate durante los Super Bowls de 2024 y 2025, utilizando técnicas de procesamiento de texto.



Figura 1: Objetivo Clasificación de Texto

1. Introducción

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.

2. Planteamiento del Problema

Analizar el texto de las diferentes notas periodísticas del periódico *Milenio* sobre la

importación de aguacate de Estados Unidos desde México durante el período del Super Bowl.

3. Solución Propuesta

- Preparación de los datos: selección de notas periodísticas de *Milenio Monterrey* y carga de datos.
- Extracción de datos con la librería `newspaper3k`.
- Procesamiento de texto: simplificación, tokenización y stemming.
- Identificación de palabras clave y análisis descriptivo de cada fuente.
- Visualización de distribución de palabras más comunes.
- Generación de gráficos de relaciones (bigramas y trigramas).

4. Experimentación

- 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.

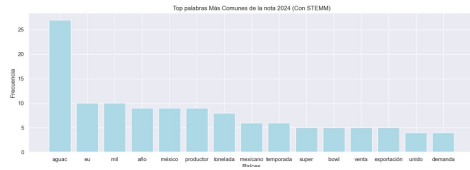


Figura 2: Palabras frecuentes Nota del 2024



Figura 3: Palabras frecuentes Nota del 2025

5. Análisis de Resultados

En ambas noticias, las palabras más comunes giran en torno a términos como “aguacate”, “México”, “super”, “bowl” y “exportación”.

5.1. Super Bowl 2024

Relaciones identificadas en los textos:

- Bigramas: “super bowl”, “bowl aguacate”, “guacamole consumo”.

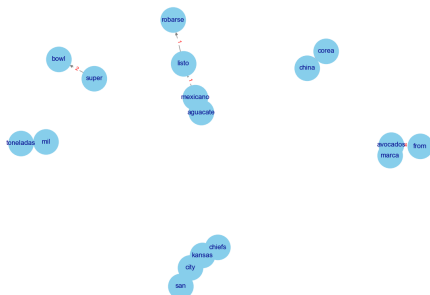


Figura 4: Grafo Nota 2024

5.2. Super Bowl 2025

Relaciones destacadas:

- Bigramas: “aguacate producción”, “incremento demanda”.

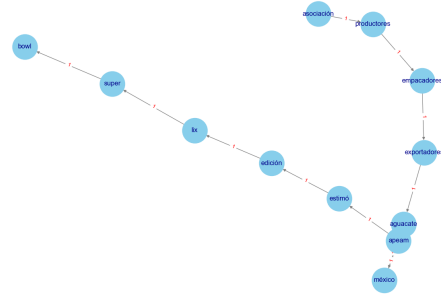


Figura 5: Grafo Nota 2025

6. 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.

Referencias

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