



UNIVERSIDAD ESAN
FACULTAD DE INGENIERÍA
INGENIERÍA DE TECNOLOGÍAS DE INFORMACIÓN Y SISTEMAS

**Desarrollo de una herramienta inteligente para la generación de clips informativos sobre
tendencias en videos de noticias mediante clustering de frames representativos**

Oscar Eduardo Salazar García
Asesor: Junior Fabian Arteaga

Lima, 24 de noviembre de 2024

Índice general

Índice de Figuras	6
Índice de Tablas	9
1. PLANTEAMIENTO DEL PROBLEMA	14
1.1. Descripción de la Realidad Problemática	14
1.1.1. Problema General	16
1.1.2. Problemas Específicos	17
1.2. Objetivos de la Investigación	17
1.2.1. Objetivo General	17
1.2.2. Objetivos Específicos	17
1.3. Justificación de la Investigación	18
1.3.1. Teórica	18
1.3.2. Práctica	18
1.3.3. Metodológica	18
1.4. Delimitación del Estudio	19
1.4.1. Espacial	19
1.4.2. Temporal	19
1.4.3. Conceptual	19
1.5. Hipótesis	20

1.5.1. Hipótesis General	20
1.5.2. Hipótesis Específicas	20
1.5.3. Matriz de Consistencia	20
2. MARCO TEÓRICO	21
2.1. Antecedentes de la investigación	21
2.1.1. Efficient Video Classification Using Fewer Frames	21
2.1.2. An improved deep convolutional neural network-based YouTube video classification using textual features	23
2.1.3. TV News Database Indexing System with Video Structure Analysis, Representative Images Extractions and OCR for News Titles	26
2.1.4. Shear Detection and Key Frame Extraction of Sports Video Based on Machine Learning	28
2.1.5. Research on Key Frame Extraction of Digital Image Based on Unsupervised Clustering Algorithm	31
2.1.6. NewsVideo Summarization Combining SURF and Color Histogram Features	33
2.2. Bases Teóricas	36
2.2.1. Imagen Digital	36
2.2.2. Video	38
2.2.3. Deep Learning	40
2.2.4. Análisis de Componentes Principales (PCA)	45
2.2.5. Clustering	48
2.2.6. Inercia en clustering	49
2.3. Marco Conceptual	51
2.3.1. Scraping	51
2.3.2. Automatización	51
2.3.3. Curaduría de contenido	52

2.3.4. Preprocesamiento de datos	52
2.3.5. Extracción de características	53
2.3.6. Clustering	53
2.3.7. Reducción de dimensionalidad	53
2.3.8. Evaluación de clusters	53
2.3.9. Selección de frames clave	54
2.3.10. Metadatos	54
2.3.11. Etiquetado de datos	54
2.3.12. Concatenación de características	54
2.3.13. Extracción de Frames	55
2.3.14. Aplicación de Línea de Comando	55
2.3.15. DataFrame	55
2.3.16. Concatenación de Videos	55
3. METODOLOGÍA DE LA INVESTIGACIÓN	56
3.1. Diseño de la investigación	56
3.1.1. Alcance de la investigación	56
3.1.2. Tipo de investigación	56
3.1.3. Enfoque de investigación	57
3.2. Población y muestra	57
3.3. Operacionalización de variables	57
3.4. Técnicas de recolección	58
3.5. Técnicas para el procesamiento y análisis de la información	59
3.5.1. Metodología de la implementación de la solución	59
3.5.2. Metodología para la medición de resultados	62
3.6. Cronograma de Actividades	62

3.7. Presupuesto	63
4. DESARROLLO DE LA SOLUCIÓN	65
4.1. Equipo utilizado	65
4.2. Especificaciones de Python	66
4.3. Librerías	66
4.4. Extracción de datos	68
4.4.1. Automatización de la búsqueda de videos en YouTube	68
4.4.2. Descarga de videos	69
4.4.3. Organización de videos	70
4.5. Preprocesamiento de datos	71
4.5.1. Extracción de frames	71
4.5.2. Etiquetado de frames	72
4.5.3. Operaciones sobre los frames	72
4.5.4. Guardado de frames	74
4.5.5. HOG (Histogram of Oriented Gradients)	75
4.5.6. Extracción de características con CNN	78
4.5.7. Concatenación de características	79
4.6. Clasificación de datos	80
4.6.1. Método del codo	80
4.6.2. Aplicación de K-means	81
4.6.3. Exploración de clústeres	82
4.6.4. Selección de frames clave	83
4.7. Generación de clips	84
4.7.1. Descripción del proceso	85
4.7.2. Visualización de Resultados	85

5. CONCLUSIONES Y RECOMENDACIONES	88
5.1. Conclusiones	88
5.2. Recomendaciones	89
BIBLIOGRAFÍA	91
Anexos	93
A. Anexo I: Matriz de Consistencia	94
B. Anexo II: Árbol de problemas	96
C. Anexo III: Links de videos extraídos	97

Índice de Figuras

1.1. Porcentaje de usuarios en base a su consumo de noticias en redes sociales en 2023	14
1.2. Top webs más visitadas en Perú en 2023	16
2.1. Metodología del antecedente. Fuente:Bhardwaj et al., 2019.	23
2.2. Metodología de investigación. Fuente:Raza et al., 2024.	25
2.3. Segmentación del video. Fuente:Rozsa y Mocofan, 2022.	28
2.4. Metodología del estudio. Fuente:X. Wang y Jiang, 2023.	30
2.5. Metodología del estudio. Fuente:Z. Wang y Chen, 2023.	33
2.6. Representación de la transformación digital de una imagen. Fuente:Gonzalez y Woods, 2021.	36
2.7. Imagen RGB y sus componenetes. Fuente:Gonzalez y Woods, 2021.	37
2.8. Composición de imagen RGB. Fuente:Gonzalez y Woods, 2021.	37
2.9. Representación del video en frames. Fuente:Tekalp, 2015.	38
2.10. Proceso de Histogram of Oriented Gradients (HOG). Fuente:Bhattarai et al., 2023.	40
2.11. Estructura de una red neuronal. Fuente: Elaboración propia.	41
2.12. Proceso de convolución. Fuente: Goodfellow et al., 2016. <i>Recuperado de Deep Learning.</i>	42
2.13. Tipos de operaciones pooling. Fuente: Sze et al., 2017.	43
2.14. Arquitectura de ResNet50. Fuente:Mukherjee, 2022.	44

2.15. Ejemplo de Eigenfaces generados mediante PCA para el reconocimiento facial. Fuente: Turk y Pentland (1991) Turk y Pentland, 1991.	48
2.16. Ilustración de K-means. Fuente: Bishop, 2006. <i>Recuperado de Pattern Recognition and Machine Learning.</i>	49
2.17. Ilustración del método del codo para seleccionar el número de clusters óptimo en K-means. Fuente: Humaira y Rasyidah, 2018.	51
3.1. Metodología propuesta. Fuente: Elaboración propia	59
3.2. Cronograma de actividades del proyecto.	63
4.1. Página de partida de scrapping.	68
4.2. Filtros aplicados a los resultados de video.	69
4.3. Instrucción de descarga de video.	70
4.4. Descripción final de carpeta de videos a procesar.	71
4.5. Frames etiquetados en la carpeta de frames.	72
4.6. Proceso de redimensión de la imagen.	73
4.7. Proceso de recorte de la imagen.	74
4.8. Características principales de la carpeta de frames generada.	75
4.9. Proceso de transformación de imagen utilizando HOG.	76
4.10. Cantidad de lotes y tiempo utilizado para HOG.	77
4.11. Estructura del DataFrame con los vectores de características HOG generados.	78
4.12. Dataframe generado por la red convolucional entrenada.	79
4.13. Concatenación de los vectores de características.	80
4.14. Representación de la Inercia utilizando el método del codo.	81
4.15. Asignación de clusters y distancias al centroide correspondiente.	82
4.16. Frames más cercanos al centroide en un clúster específico.	83
4.17. Frames más cercanos al centroide en un clúster específico.	83
4.18. Frames representativos en los clusters con k = 6.	84

4.19. Frames representativos en los clusters con k = 10.	84
4.20. Muestra de frames clave y sus frames adyacentes.	85
4.21. Muestra de frames del clip generado.	87
A.1. Matriz de consistencia Completo. Fuente: Elaboración propia	95
B.1. Árbol de problemas. Fuente: Elaboración propia	96
C.1. Links de videos extraídos. Fuente: Elaboración propia	97

Índice de Tablas

2.1.	Resultados de la Clasificación de Videos Utilizando Menos Fotogramas	23
2.2.	Resultados experimentales de los modelos	26
2.3.	Resultados experimentales de la segmentación de video en noticias televisivas .	28
2.4.	Resultados de la extracción de fotogramas clave y detección de cortes en videos	33
2.5.	Resultados de la detección de límites de escenas por tipo de video	35
2.6.	Comparación de métodos de detección de límites de escenas	35
2.7.	Comparación de métodos en el dataset TVSum50	36
3.1.	Descripción del estudio	57
3.2.	<i>Operacionalización de Variables</i>	58
3.3.	<i>Presupuesto del Proyecto</i>	64
4.1.	Especificaciones del equipo utilizado	65
4.2.	Especificaciones de Python utilizado	66
4.3.	Descripción de librerías usadas en el proyecto	66
4.4.	Especificaciones del video generado a partir de frames Clave	86

Resumen

En respuesta a la creciente necesidad de sintetizar grandes volúmenes de contenido audiovisual, esta tesis propone una herramienta inteligente para la generación automatizada de clips informativos a partir de videos de noticias. Utilizando técnicas avanzadas de visión por computadora, como HOG y ResNet50, se desarrolló un modelo capaz de extraer características visuales relevantes. A partir de 70,000 frames procesados de 40 videos de noticias, los vectores de características fueron integrados y optimizados mediante reducción dimensional con PCA, preservando la información clave y reduciendo significativamente el costo computacional. El algoritmo de K-means permitió agrupar los frames en clústeres con valores óptimos de k definidos mediante su inercia. Los frames más representativos de cada clúster se seleccionaron en función de su proximidad al centroide, generando clips que reflejan patrones visuales predominantes. La metodología demuestra ser efectiva para identificar tendencias visuales y organizar grandes volúmenes de datos. Estos resultados validan la viabilidad de esta herramienta en el análisis automatizado de noticias, destacando su capacidad para sintetizar contenido relevante de manera eficiente.

Palabras claves: Clustering, Video, Unsupervised Learning, Key Frames

Abstract

In response to the growing need to synthesize large volumes of audiovisual content, this thesis proposes an intelligent tool for the automated generation of informative clips from news videos. Using advanced computer vision techniques such as HOG and ResNet50, a model was developed to extract relevant visual features. Based on 70,000 frames processed from 40 news videos, the feature vectors were integrated and optimized through dimensionality reduction using PCA, preserving key information while significantly reducing computational costs. The K-means algorithm was used to cluster the frames, with optimal k values determined by inertia. The most representative frames from each cluster were selected based on their proximity to the centroid, generating clips that reflect predominant visual patterns. The methodology proves effective in identifying visual trends and organizing large datasets. These results validate the feasibility of this tool in automated news analysis, highlighting its ability to efficiently synthesize relevant content.

Introducción

En el contexto actual de crecimiento exponencial de contenido audiovisual en redes sociales y plataformas digitales, los medios de comunicación enfrentan el desafío de seleccionar y presentar información relevante de manera ágil y concisa. Este estudio responde a la necesidad de una herramienta que permita procesar grandes volúmenes de datos de video y extraer clips representativos de noticias mediante técnicas de clustering de frames, ofreciendo una solución automatizada para la identificación de tendencias y la generación de resúmenes informativos en un formato atractivo para las audiencias.

La proliferación de plataformas digitales como YouTube, Facebook y TikTok ha transformado no solo la forma en que se produce y distribuye contenido, sino también las expectativas de los consumidores, quienes demandan información breve, precisa y accesible. En este contexto, los avances en aprendizaje automático y procesamiento de imágenes se han convertido en herramientas esenciales para abordar los desafíos asociados con el análisis y la curaduría de grandes volúmenes de datos audiovisuales. Tecnologías como el clustering, el análisis de características visuales y la automatización de procesos han demostrado su eficacia para extraer patrones significativos y sintetizar información relevante en formatos consumibles.

Este proyecto se desarrolla en el cruce de estas tendencias, aplicando técnicas avanzadas de inteligencia artificial para abordar una necesidad específica de los medios de comunicación: identificar automáticamente momentos destacados en contenido noticioso. Al aprovechar el poder del aprendizaje automático, se busca construir una solución escalable y adaptable que no solo facilite el trabajo de los profesionales en este campo, sino que también optimice la experiencia del consumidor, permitiéndole acceder a información clave de manera más rápida y efectiva.

El primer capítulo abordará el planteamiento del problema, estableciendo la relevancia de esta investigación y definiendo los objetivos específicos que guían el desarrollo de la solución propuesta. En el segundo capítulo, se presentará el marco teórico, que brindará un contexto conceptual y técnico necesario para comprender los fundamentos de la herramienta desarrollada, situando el trabajo dentro del campo de estudio y revisando los enfoques y tecnologías clave utilizados.

En el tercer capítulo, se expondrá la metodología empleada en el proyecto, describiendo

detalladamente el proceso de diseño de la solución, desde la selección de datos hasta las etapas de análisis y validación. Se hará énfasis en los principios metodológicos que guiaron cada fase, como la recopilación automatizada de videos, la extracción de frames clave y la aplicación de algoritmos avanzados de clustering. Además, se explicará cómo se integraron herramientas tecnológicas específicas para garantizar la reproducibilidad, la escalabilidad y la eficiencia del sistema propuesto, abordando los desafíos técnicos y las decisiones que serán tomadas en cada etapa.

En el cuarto capítulo, se presentará un análisis exhaustivo del desarrollo técnico de la herramienta. Esto incluye la implementación de técnicas de procesamiento de imágenes, la configuración de parámetros en los algoritmos de agrupamiento y las pruebas realizadas para garantizar un desempeño óptimo. También se documentará cómo las configuraciones de hardware y software contribuyeron a optimizar los tiempos de ejecución y mejorar la precisión en la selección de clips representativos. Este capítulo busca proporcionar una visión completa del flujo de trabajo técnico y destacar las innovaciones introducidas para enfrentar las limitaciones inherentes al manejo de datos audiovisuales complejos. Finalmente, en el quinto capítulo se presentarán las conclusiones alcanzadas y se ofrecerán recomendaciones basadas en los hallazgos del estudio, incluyendo posibles líneas de investigación futura y mejoras que podrían integrarse en la herramienta para extender su aplicabilidad en otros contextos.

Capítulo 1

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la Realidad Problemática

En la era digital actual, el consumo de contenido audiovisual en redes sociales ha crecido exponencialmente, convirtiéndose en una fuente invaluable de información y entretenimiento. Una de los principales viene a ser la búsqueda de la actualidad del mundo, de acuerdo a una encuesta realizada por Pew Research Center al sector adulto de EEUU, 3 de cada 10 adultos utilizan regularmente Youtube como plataforma de noticias, siguiendo Instagram y Tiktok. La cantidad de información que se genera día a día es abrumadora y el tiempo escaso.

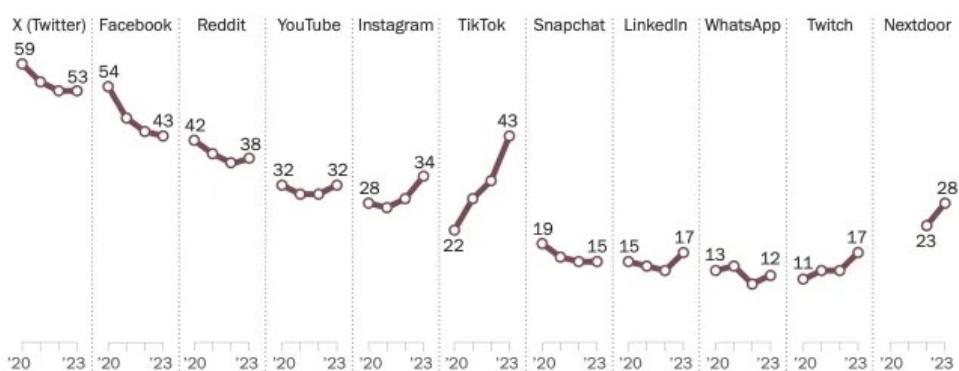


Figura 1.1: Porcentaje de usuarios en base a su consumo de noticias en redes sociales en 2023.

Fuente: [PewResearchCenter, 2023. Encuesta a adultos de EEUU realizada en 2023.](#)

De acuerdo a la [2.10](#), se puede observar que plataformas con tendencia a ‘Más información en menos tiempo’ empiezan a ser las preferidas entre el público, observando el sorprendente aumento de plataformas como Instagram y Tiktok. Los medios informativos están tomando en cuenta estos cambios y han iniciado a informar en pocos segundos lo más relev-

vante del día mediante videos cortos en sus redes sociales, para llegar a este tipo de público. Pero el principal desafío diario es cómo seleccionar el contenido que represente los hechos más importantes ocurrido en la actualidad.

Dentro de esta problemática se encuentra la gran cantidad de información audiovisual generada cada minuto y la limitada capacidad para analizar y extraer información significativa de manera manual. Este desafío requiere soluciones tecnológicas que permitan procesar eficientemente grandes volúmenes de datos audiovisuales y sintetizar contenido relevante.

En este contexto, las técnicas avanzadas de visión por computadora y aprendizaje automático surgen como soluciones efectivas. Estas metodologías permiten no solo procesar y organizar datos audiovisuales, sino también identificar patrones visuales significativos mediante la extracción de características de los frames. Herramientas como HOG y ResNet50 son fundamentales en este proceso, ya que proporcionan vectores de características que encapsulan tanto información local como abstracta de cada frame. Para optimizar el manejo de estos datos de alta dimensionalidad, se utiliza PCA, una técnica de reducción dimensional que facilita el análisis y reduce el costo computacional.

El uso de estas herramientas automatiza la selección y análisis de contenido, superando las limitaciones de los métodos tradicionales que no son adecuados para manejar la escala y complejidad del contenido audiovisual generado en plataformas como YouTube. Esto asegura que la información relevante pueda ser identificada y sintetizada de manera precisa y eficiente, respondiendo a las demandas actuales de procesamiento y curación de contenido informativo.

En particular, YouTube juega un papel crucial en este análisis. Según Data Reportal en la [1.2](#), en el contexto peruano, YouTube es la segunda página más visitada, con un promedio de 227 millones de visitas mensuales y 11.4 millones de visitantes únicos al mes en 2023. YouTube se convierte así en una fuente inagotable de contenido relevante para el análisis de noticias y eventos actuales. La plataforma no solo alberga una vasta biblioteca de videos, sino que también proporciona metadatos valiosos, como descripciones, etiquetas y estadísticas de visualización, que enriquecen el análisis del contenido. Aplicar herramientas de big data en YouTube permite automatizar la extracción y el procesamiento de contenido, facilitando la selección de los clips más relevantes que capturan los eventos importantes del día.



Figura 1.2: Top webs más visitadas en Perú en 2023.

Fuente: [DataReportal, 2024](#). Basado en el tráfico web de Diciembre de 2022 a Noviembre del 2023.

Esta automatización optimiza tanto el tiempo como los recursos necesarios para curar contenido, mejorando la precisión en la selección de información significativa.

La generación de clips informativos a partir de videos largos presenta múltiples beneficios para los medios de comunicación. Primero, facilita la creación de resúmenes concisos y precisos que destacan los puntos clave de las noticias, permitiendo a los usuarios acceder a la información relevante de manera rápida. Este enfoque es crucial en un entorno donde la atención del usuario es limitada y la demanda de contenido digerible es alta. Esto no solo aumenta el alcance y el impacto de las noticias, sino que también mejora la participación del público y fomenta una mayor interacción con el contenido.

La automatización en la generación de clips también contribuye a la eficiencia operativa de los medios. Al reducir la necesidad de intervención manual en la selección y edición de contenido, los recursos se liberan para enfocarse en otras áreas críticas, como la investigación y el análisis en profundidad. Esto permite a los medios mantenerse competitivos en un entorno en constante cambio y con una demanda creciente de contenido actualizado y relevante.

1.1.1. Problema General

¿De qué manera el desarrollo de una herramienta inteligente basada en técnicas de clustering de frames puede facilitar la identificación automática de tendencias a partir de videos de noticias para crear clips de videos representativos?

1.1.2. Problemas Específicos

- ¿Cómo obtener de manera eficiente los videos de noticias referentes a un día en específico, asegurando que sean representativos de los eventos más importantes del día?
- ¿Qué técnicas de preprocesamiento de imágenes y normalización son más efectivas para mejorar la calidad y relevancia de los frames extraídos de los videos?
- ¿Qué técnicas de extracción de características permiten representar adecuadamente los frames para identificar patrones visuales relevantes?
- ¿Cuál es el algoritmo de clustering más adecuado para agrupar los frames en base a las similitudes de las características extraídas que posean entre ellas?
- ¿Cómo generar clips informativos concisos, asegurando que reflejen los eventos clave discutidos en las noticias?

1.2. Objetivos de la Investigación

1.2.1. Objetivo General

Desarrollar una herramienta inteligente que identifique de manera automática tendencias en videos de noticias mediante el clustering de frames representativos y genere clips de videos que resuman los eventos clave.

1.2.2. Objetivos Específicos

- Desarrollar un método eficiente para la recolección de videos de noticias, asegurando que estos reflejen los eventos más importantes del día.
- Implementar técnicas de preprocesamiento de imágenes y normalización que mejoren la calidad de los frames extraídos, optimizando su relevancia para el análisis.
- Desarrollar técnicas de extracción de características que permitan representar adecuadamente los frames, facilitando la identificación de patrones visuales relevantes.
- Seleccionar e implementar un algoritmo de clustering que agrupe los frames basándose en la similitud de características, garantizando que los grupos formados representen tendencias claras.

- Desarrollar un sistema automatizado para generar clips informativos concisos que resuman los eventos clave identificados en los videos de noticias.

1.3. Justificación de la Investigación

1.3.1. Teórica

Esta investigación busca explorar cómo el uso de técnicas avanzadas de visión por computadora y aprendizaje automático puede optimizar la selección y generación automatizada de clips informativos a partir de contenido audiovisual en YouTube. La base teórica se enfoca en la capacidad de estas herramientas para procesar grandes volúmenes de datos de manera eficiente, identificando patrones visuales relevantes y eliminando información redundante. Además, se fundamenta en buenas prácticas en la extracción de características, reducción dimensional y agrupamiento de datos, resaltando la importancia de la integración de estas técnicas para mejorar la precisión y representatividad en la curación de contenido. Este estudio no solo aporta una solución práctica al desafío de manejar la sobrecarga de información en el análisis de videos, sino que también establece un marco teórico aplicable a la industria de los medios de comunicación, promoviendo el desarrollo de tecnologías inteligentes en la síntesis y organización de contenido audiovisual.

1.3.2. Práctica

Al culminar la investigación, se espera que los medios de comunicación puedan utilizar el sistema automatizado basado en Big Data y aprendizaje automático para seleccionar y generar clips informativos relevantes de manera eficiente. Esto facilitará la creación de resúmenes concisos que destaque los puntos clave de las noticias, permitiendo a los usuarios acceder a información relevante rápidamente. La implementación práctica de esta solución tecnológica ayudará a los medios a adaptarse a las nuevas tendencias de consumo de contenido, donde la rapidez y concisión son cruciales. Además, mejorará la participación del público y fomentará una mayor interacción con el contenido.

1.3.3. Metodológica

. El desarrollo del sistema automatizado de análisis de contenido audiovisual, basado en técnicas avanzadas de visión por computadora y aprendizaje automático, permitirá procesar

grandes volúmenes de datos de video de manera eficiente. La metodología incluye el uso de herramientas específicas para la extracción de características visuales, como HOG y ResNet50, que serán optimizadas mediante reducción dimensional utilizando PCA. Además, se implementarán algoritmos de clustering, como K-means, para agrupar los datos y detectar patrones relevantes en los videos. Esta metodología automatiza procesos clave, optimizando el tiempo y los recursos necesarios para analizar contenido audiovisual, y garantiza una selección precisa y representativa de la información significativa. De esta forma, se ofrece una solución tecnológica eficiente y práctica para abordar los desafíos actuales en el análisis y síntesis de contenido informativo.

1.4. Delimitación del Estudio

1.4.1. Espacial

El estudio se llevará a cabo en el contexto de plataformas digitales, enfocándose específicamente en YouTube como la principal fuente de contenido audiovisual para el análisis. Se seleccionarán videos de noticieros, priorizando aquellos que cubran eventos actuales de relevancia informativa. La elección de YouTube se debe a su amplia popularidad y uso en Perú, donde es una de las plataformas más visitadas, lo que garantiza un acceso continuo a una gran cantidad de contenido audiovisual relevante. La investigación se centrará en la capacidad de YouTube para proporcionar tanto contenido audiovisual como metadatos valiosos para el análisis.

1.4.2. Temporal

Los datos que se extraerán corresponderán a la última semana de noviembre del 2024. Durante este período se llevará a cabo la recopilación, procesamiento y análisis de los videos seleccionados. Dado que la herramienta está diseñada para ser de uso recurrente, los resultados se generarán de manera automatizada pocos minutos después de la finalización del proceso de recopilación, garantizando así una síntesis eficiente y oportuna del contenido audiovisual.

1.4.3. Conceptual

La presente investigación se centrará en el desarrollo de un sistema automatizado para la selección y generación de clips informativos a partir de contenido audiovisual, utilizando

técnicas avanzadas de visión por computadora y aprendizaje automático. El enfoque estará en el análisis de videos de noticias en YouTube, aplicando métodos como HOG y ResNet50 para la extracción de características visuales, complementados con reducción dimensional mediante PCA. Además, se implementarán algoritmos de clustering, como K-means, para identificar eventos relevantes y sintetizar contenido visual representativo.

1.5. Hipótesis

1.5.1. Hipótesis General

Mediante el uso de técnicas de modelos de aprendizaje automático, se logrará mejorar la identificación de tendencias y la generación de clips informativos representativos de videos de noticias.

1.5.2. Hipótesis Específicas

- La recolección automatizada de videos de noticias referentes a un día específico permitirá capturar de manera eficiente los contenidos más representativos para el análisis de tendencias.
- El uso de técnicas avanzadas de preprocessamiento y normalización de imágenes mejorará significativamente la calidad de los frames, facilitando su clasificación y análisis posterior.
- La extracción de características adecuadas de los frames permitirá identificar correctamente patrones visuales que reflejen temas importantes y eventos clave.
- El uso de un algoritmo de clustering optimizado permitirá agrupar los frames con precisión en función de sus similitudes, identificando adecuadamente las tendencias en los noticias.
- La generación automatizada de clips informativos basados en las tendencias identificadas ofrecerá resúmenes concisos y efectivos de los eventos más relevantes del día.

1.5.3. Matriz de Consistencia

A continuación se presenta la matriz de consistencia elaborada para la presente investigación (véase Anexo A.1).

Capítulo 2

MARCO TEÓRICO

2.1. Antecedentes de la investigación

En esta sección se presentarán diversos artículos de investigación o tesis las cuales abordarán diversas técnicas y enfoques que se emplearon para afrontar problemas similares al de esta tesis.

2.1.1. Efficient Video Classification Using Fewer Frames

[Bhardwaj et al., 2019](#)

2.1.1.1. Planteamiento del problema

La necesidad de reducir las operaciones de punto flotante (FLOPs) y la huella de memoria. Aunque los modelos compactos actuales son más ligeros en términos de memoria, todavía requieren un número significativo de FLOPs ya que procesan todos los cuadros de un video. La propuesta de esta investigación es desarrollar modelos de clasificación de videos que procesen menos cuadros, disminuyendo así el número de FLOPs.

2.1.1.2. Objetivos

- Desarrollar modelos de clasificación de video que tengan una huella de memoria pequeña (menor a 1 GB) y que sean eficientes

- Crear modelos que procesen menos cuadros de video para reducir el número de FLOPs, manteniendo una alta eficiencia computacional.
- Utilizar un modelo maestro computacionalmente intensivo que procesa todos los cuadros para entrenar a un modelo estudiante más eficiente que procesa solo una fracción de los cuadros.
- Realizar una evaluación exhaustiva con tres tipos de modelos de clasificación de video: modelos recurrentes, modelos de agrupación y agregación, y modelos de agrupación y agregación eficientes en memoria.

2.1.1.3. Metodología

La metodología del documento incluye varias etapas, a partir de la recolección de datos hasta la implementación del modelo. A continuación, se presenta un resumen de la metodología:

1. **Recopilación de datos:** Se utilizó el conjunto de datos YouTube-8M (versión 2017), que contiene 8 millones de videos con múltiples clases asociadas a cada video. Los videos tienen una longitud promedio de 200 segundos.
2. **Entrenamiento del profesor y estudiante:** Se entrena una red neuronal compleja (profesor) que procesa todos los fotogramas de un video para generar una representación detallada del mismo. Se seleccionan solo algunos fotogramas del video (cada j-ésimo fotograma) para ser procesados por la red neuronal del estudiante. Se entrena una red neuronal más simple (estudiante) para que genere representaciones similares a las del profesor, utilizando solo los fotogramas seleccionados.
3. **Minimización de pérdidas:** Se minimizan las diferencias entre las representaciones del profesor y del estudiante mediante funciones de pérdida específicas, como la pérdida de error cuadrático.
4. **Evaluación y Comparación:** Se evalúa el rendimiento del estudiante en términos de precisión y eficiencia computacional, comparándolo con el modelo del profesor y otros métodos base.

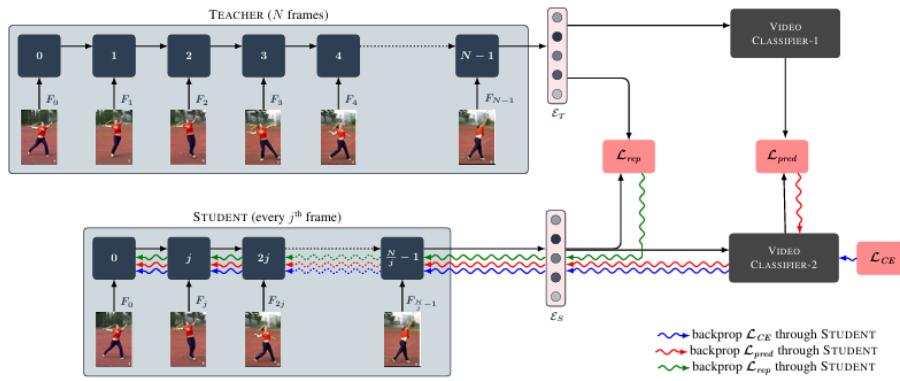


Figure 1: Architecture of TEACHER-STUDENT network for video classification

Figura 2.1: Metodología del antecedente. Fuente:[Bhardwaj et al., 2019](#).

2.1.1.4. Resultados

Los resultados obtenidos se presentan en un cuadro comparativo entre los diferentes modelos utilizados en el estudio. Los modelos comparativos 1 y 2 representan otras configuraciones de modelos de aprendizaje profundo probadas por los autores para comparar la efectividad de DeepASL.

Tabla 2.1: Resultados de la Clasificación de Videos Utilizando Menos Fotogramas

Modelo	k	GAP	mAP	FLOPs (Billones)	Tiempo de Evaluación (hrs.)
Teacher-Skyline	N/A	0.811	0.414	5.058	13.00
Uniform-k	10	0.759	0.324	0.167	7.61
Uniform-k	20	0.785	0.363	0.268	8.20
Uniform-k	30	0.795	0.378	0.520	9.11

Nota. GAP = Precision Global Ajustada; mAP = Media de Precisión. FLOPs se refiere a las operaciones de punto flotante.

2.1.2. An improved deep convolutional neural network-based YouTube video classification using textual features

[Raza et al., 2024](#)

2.1.2.1. Planteamiento del problema

El artículo aborda el desafío del crecimiento exponencial del contenido de video en plataformas como YouTube, donde se suben más de 30,000 videos por hora, lo que crea la necesidad de categorizar automáticamente este contenido para hacerlo accesible a los usuarios. Las técnicas actuales de clasificación de videos, que se basan en el análisis de texto (como títulos, descripciones y etiquetas), son limitadas en precisión y no han sido suficientemente estudiadas para manejar la vasta cantidad de categorías disponibles. Además, el procesamiento de video e imagen es computacionalmente costoso, lo que impulsa el interés en enfoques basados en texto, que son más eficientes pero subdesarrollados. El artículo sugiere que se requiere un enfoque avanzado de inteligencia artificial que pueda mejorar la categorización de videos utilizando información textual de manera más eficaz y precisa.

2.1.2.2. Objetivos

- Realizar un análisis exploratorio de datos de YouTube para obtener información sobre los videos y sus categorías, identificando patrones y tendencias clave.
- Desarrollar un modelo mejorado de red neuronal convolucional profunda (DCNN) que permita categorizar videos de YouTube con alta precisión y eficiencia.
- Comparar el rendimiento del modelo DCNN con otros enfoques de aprendizaje profundo, como redes neuronales recurrentes (RNN) y unidades de memoria a largo plazo (GRU), así como con modelos de aprendizaje automático tradicionales, como la regresión logística, máquinas de soporte vectorial (SVM), árboles de decisión y bosques aleatorios.

2.1.2.3. Fundamento Teórico

Dado que el procesamiento de imágenes y videos es computacionalmente costoso, el estudio explora el uso de características textuales (títulos, descripciones y etiquetas) para clasificar los videos de manera más eficiente. En este contexto, se propone un modelo de red neuronal convolucional profunda (DCNN) como enfoque principal para la categorización de videos, destacando su capacidad para manejar grandes volúmenes de datos con alta precisión. El estudio compara el desempeño del modelo DCNN con otros modelos de aprendizaje profundo, como las redes neuronales recurrentes (RNN) y las unidades de memoria a largo plazo (GRU), demostrando que el DCNN supera a estos enfoques en términos de precisión y eficiencia en la clasificación de videos.

2.1.2.4. Metodología:

1. Adquisición de datos:

- Este conjunto de datos contiene 20,000 videos categorizados en 9 temas como aventura, arte y música, ciencia, deportes, entre otros.

2. Preprocesamiento textual:

- Limpieza de ruido en los textos de los títulos y descripciones de los videos. Se transforman los textos a minúsculas, se eliminan números, puntuaciones y espacios en blanco adicionales. El proceso incluye tokenización, eliminación de tokens no alfabéticos y palabras vacías, finalizando con la lematización para convertir las palabras a sus formas raíz.
- Análisis estadístico tanto a nivel observacional como de corpus. Se evalúan medidas como el tamaño promedio del vocabulario y la riqueza léxica de las descripciones.
- El conjunto de datos se divide en un 80 % para entrenamiento y un 20 % para pruebas, con el objetivo de entrenar

3. Entrenamiento del modelo:

- Se implementan redes neuronales recurrentes (RNN) y unidades de memoria controlada (GRU) para procesar secuencias textuales. Se explica cómo estas redes manejan datos secuenciales mediante conexiones entre capas y se optimizan a través de capas de dropout y activación.

4. Categorización de los videos:

- Se desarrolla una red neuronal convolucional profunda (DCNN) para la clasificación de videos. El modelo utiliza capas de convolución, agrupamiento y capas densas, optimizando sus parámetros y estructura para obtener un alto rendimiento en la categorización.

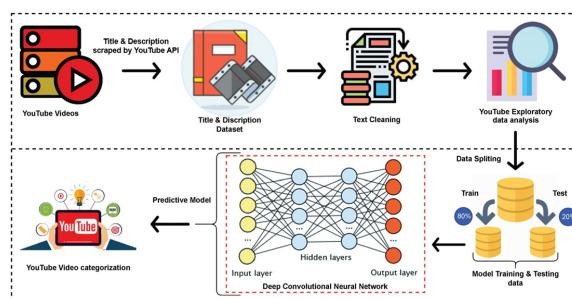


Figura 2.2: Metodología de investigación. Fuente:[Raza et al., 2024](#).

2.1.2.5. Resultados:

Los resultados del artículo se presentan en la siguiente tabla:

Tabla 2.2: Resultados experimentales de los modelos

Modelo	Precisión (%)	AUC (%)	Puntuación F1
DCNN	96	99	0.95
RNN	92	97	0.91
GRU	94	98	0.93

2.1.3. TV News Database Indexing System with Video Structure Analysis, Representative Images Extractions and OCR for News Titles

[Rozsa y Mocofan, 2022](#)

2.1.3.1. Planteamiento del Problema

El problema abordado en este trabajo es la falta de sistemas eficientes para la indexación y recuperación de contenido en archivos de video, particularmente en los programas de noticias televisivas. Muchas cadenas de televisión acumulan grandes cantidades de material de video, tanto en formatos analógicos como digitales, que no están organizados en sistemas de gestión de contenido que permitan búsquedas rápidas y reutilización del material. Esto dificulta la recuperación eficiente de información para la creación de nuevos contenidos o la supervisión del cumplimiento de las transmisiones.

2.1.3.2. Objetivos

- Desarrollar un sistema semiautomático para analizar la estructura temporal de los programas de noticias televisivas.
- Implementar métodos de segmentación de video y extracción de imágenes representativas para mejorar la indexación de contenido.
- Aplicar reconocimiento óptico de caracteres (OCR) para extraer títulos de noticias y facilitar su búsqueda y recuperación.
- Probar el sistema propuesto en diferentes programas de noticias y evaluar su eficiencia en comparación con otros métodos.

2.1.3.3. Fundamento Teórico

El trabajo se basa en la necesidad de sistemas de gestión de contenido multimedia que puedan indexar y categorizar de manera eficiente los programas de televisión, especialmente las noticias. Se hace uso de tecnologías como la segmentación de video, la extracción de fotogramas clave y el OCR, que permite reconocer texto en imágenes. La segmentación del video se realiza analizando secuencias de fotogramas y detectando cambios significativos mediante histogramas de color y nivel de gris. Posteriormente, se extraen imágenes representativas para facilitar la navegación y búsqueda dentro de los archivos. La integración del OCR permite reconocer los títulos de las noticias, lo que facilita su búsqueda en la base de datos.

2.1.3.4. Metodología

1. Segmentación de video:

- Se identifican puntos de inicio y fin de las tomas y se extraen imágenes representativas para cada toma. Para ello, se calcula la diferencia entre los histogramas de color de fotogramas sucesivos, utilizando métricas como la distancia de Bhattacharyya.

2. Extracción de imágenes representativas:

- se seleccionan imágenes representativas para cada toma con el fin de facilitar la búsqueda. Este proceso se realiza adaptando los umbrales de discriminación dependiendo de la complejidad del material.

3. Reconocimiento óptico de caracteres (OCR):

- Se extrae títulos de noticias a partir de áreas predefinidas en la imagen. Se aplica preprocessamiento a las imágenes (conversión a escala de grises, umbralización) para mejorar la precisión de la extracción de texto.

4. Evaluación del sistema:

- El sistema fue probado en programas de noticias de varias cadenas de televisión de Rumania, midiendo el tiempo de análisis y el número de imágenes útiles extraídas para la indexación en la base de datos.

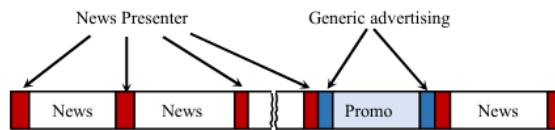
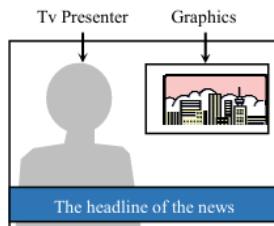


Fig. 1. The standard TV news program time format

**Figura 2.3:** Segmentación del video. Fuente:[Rozsa y Mocofan, 2022](#).

2.1.3.5. Resultados:

Tabla 2.3: Resultados experimentales de la segmentación de video en noticias televisivas

Canal de TV	Duración del programa (min)	Imágenes extraídas
ProTV, TVR1, Kanal D, Prima TV, Intermedia	40–45	800–900

Nota. Los resultados muestran la duración promedio de los programas de noticias y el número de imágenes extraídas por programa en varios canales de televisión.

Descripción de los resultados: Los experimentos se realizaron con programas de noticias de 40 a 45 minutos de duración de cinco canales de televisión distintos. El sistema propuesto para la segmentación de video extrajo entre 800 y 900 imágenes representativas por programa, que son utilizadas para la indexación en una base de datos de noticias televisivas. La cantidad de imágenes extraídas varía según el contenido y el estilo de producción de cada canal, pero en todos los casos se logró una segmentación precisa del contenido, facilitando la búsqueda y recuperación de material archivado.

2.1.4. Shear Detection and Key Frame Extraction of Sports Video Based on Machine Learning

[X. Wang y Jiang, 2023](#)

2.1.4.1. Planteamiento del Problema

El trabajo aborda el problema del crecimiento exponencial de los datos de video, lo que ha generado la necesidad de herramientas eficientes para procesar, analizar y recuperar el contenido. En particular, los videos deportivos presentan desafíos debido a la diversidad de tomas y cambios rápidos en los ángulos de cámara, lo que dificulta la detección de cortes y la extracción de fotogramas clave que representen de manera precisa el contenido de las escenas.

2.1.4.2. Objetivos

- Proponer un método basado en el análisis del movimiento para detectar los cortes de escena en videos deportivos.
- Desarrollar un algoritmo que permita la extracción automática de fotogramas clave representativos de las escenas deportivas.
- Reducir el volumen de datos en los índices de video mediante la selección de fotogramas clave sin comprometer la representación visual.
- Crear una plataforma de navegación de videos basada en fotogramas clave que facilite la búsqueda y recuperación del contenido deportivo.

2.1.4.3. Fundamento Teórico

El trabajo se fundamenta en la teoría del aprendizaje automático (ML) y el procesamiento de video, con énfasis en la eficiencia computacional. El análisis de video basado en el contenido (Content-Based Video Analysis) permite una mejor comprensión de los datos visuales mediante técnicas de visión por computadora y procesamiento de imágenes. La detección de cortes y la extracción de fotogramas clave son aspectos esenciales para la organización y recuperación de videos. Los modelos de ML, como el SVM, se utilizan para clasificar patrones en los datos visuales, mientras que los algoritmos de estimación de movimiento, basados en el bloque de coincidencia, permiten identificar transiciones entre fotogramas y escenas de manera eficiente, reduciendo el tiempo y los recursos computacionales necesarios para analizar videos extensos.

2.1.4.4. Metodología

1. Estimación del movimiento basado en bloques:

- El algoritmo divide el video en bloques de píxeles y calcula la similitud de los bloques entre fotogramas consecutivos para identificar cambios significativos de movimiento. Esto permite detectar los cortes de escena basados en el movimiento.

2. Detección de cortes mediante SVM:

- El modelo de máquina de soporte vectorial (SVM) clasifica las transiciones entre fotogramas como cortes graduales o no graduales, utilizando características extraídas de los vectores de movimiento.

3. Extracción de fotogramas clave:

- Se utiliza un modelo de atención del usuario para determinar la relevancia de cada fotograma dentro de una escena. Los fotogramas con mayor valor de atención se seleccionan como representativos, permitiendo así una navegación más eficiente del video.

4. Evaluación experimental:

- El sistema fue probado en videos deportivos, midiendo su eficiencia en la detección de cortes y la capacidad de reducir el volumen de datos en los índices de video. La eficiencia se evaluó en términos del tiempo de procesamiento requerido y la precisión de la detección de transiciones graduales y no graduales. La plataforma generada permite navegar a través de los fotogramas clave para visualizar resúmenes del contenido, optimizando tanto la recuperación de la información como la experiencia de usuario.

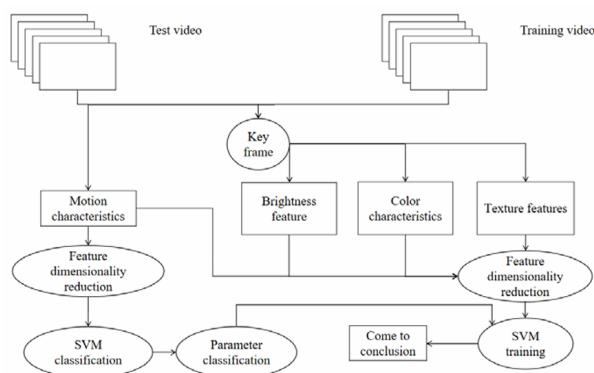


Figura 2.4: Metodología del estudio. Fuente:[X. Wang y Jiang, 2023](#).

2.1.4.5. Resultados

El sistema de detección de cortes y extracción de fotogramas clave propuesto demuestra ser eficiente y preciso. El uso del modelo basado en aprendizaje automático ha permitido una mejor detección de transiciones graduales y no graduales en videos deportivos, además de la extracción precisa de fotogramas clave. Este sistema puede aplicarse a múltiples tipos de videos sin necesidad de entrenamientos específicos para cada tipo, lo que mejora la eficiencia en el análisis y la recuperación de videos en entornos de alto volumen de datos.

2.1.5. Research on Key Frame Extraction of Digital Image Based on Unsupervised Clustering Algorithm

Z. Wang y Chen, 2023

2.1.5.1. Planteamiento del Problema

Con el crecimiento exponencial de la información multimedia, el procesamiento eficiente de grandes volúmenes de datos de video se ha convertido en un desafío clave. Los métodos tradicionales para la extracción de fotogramas clave son inefficientes para manejar videos a gran escala y presentan limitaciones en cuanto a la precisión y velocidad de procesamiento. Esto genera una necesidad urgente de desarrollar técnicas más avanzadas que permitan extraer automáticamente los fotogramas más representativos de un video, facilitando su análisis y almacenamiento.

2.1.5.2. Objetivos

- Proponer un método de extracción de fotogramas clave basado en un algoritmo de clustering no supervisado.
- Mejorar la precisión en la detección de fotogramas clave en videos digitales mediante el uso de algoritmos de aprendizaje automático.
- Reducir el tiempo de procesamiento en la extracción de fotogramas clave, optimizando la eficiencia del sistema.
- Validar la efectividad del algoritmo propuesto mediante simulaciones y comparaciones con modelos tradicionales de extracción de fotogramas clave.

2.1.5.3. Fundamento Teórico

El artículo se basa en el uso de algoritmos de clustering no supervisados para clasificar fotogramas de video según su similitud, permitiendo identificar aquellos más representativos de una escena. La extracción de fotogramas clave es crucial para la recuperación y análisis de contenido en videos, ya que permite generar resúmenes concisos. A diferencia de los métodos tradicionales, que dependen de intervalos de tiempo o reglas de calidad de imagen, el clustering no supervisado se enfoca en descubrir automáticamente patrones de similitud en los datos, lo que mejora la precisión de la extracción de fotogramas. La combinación con redes neuronales RBF permite modelar y ajustar mejor las trayectorias y patrones de movimiento en los videos.

2.1.5.4. Metodología

1. Extracción de características basadas en segmentación:

- El video se segmenta en bloques de imagen, de los cuales se extraen características como el histograma de color, la textura y la forma. Estas características se agrupan en un vector para cada bloque de imagen, lo que permite realizar una comparación entre fotogramas.

2. Cálculo de distancias de similitud:

- Se utiliza la distancia Euclíadiana para medir la similitud entre fotogramas, normalizando previamente los vectores de características. Esto permite identificar fotogramas que contienen información novedosa y los clasifica como candidatos a fotogramas clave.

3. Algoritmo de clustering no supervisado:

- El algoritmo realiza un análisis de movimiento y características de contenido entre los fotogramas. Agrupa fotogramas similares en clusters y selecciona aquellos que mejor representan el video. El proceso se optimiza mediante el uso de redes neuronales RBF, que ajustan las trayectorias de movimiento y optimizan la selección de fotogramas clave.

4. Evaluación experimental:

- Se realizaron simulaciones con diferentes tipos de videos para comparar el rendimiento del algoritmo propuesto con modelos tradicionales. Los resultados mostraron una mayor precisión en la extracción de fotogramas clave y una reducción en el tiempo de procesamiento.

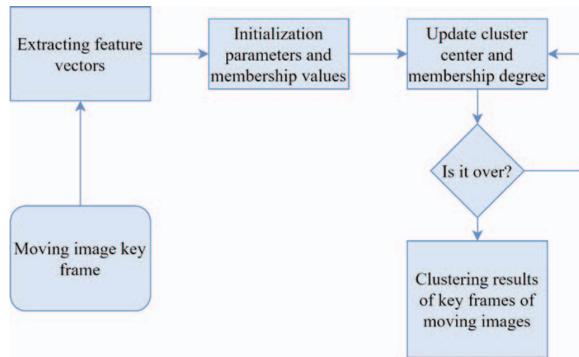


Figura 2.5: Metodología del estudio. Fuente:[Z. Wang y Chen, 2023](#).

2.1.5.5. Resultados

Tabla 2.4: Resultados de la extracción de fotogramas clave y detección de cortes en videos

Método	Precisión (%)	Tiempo de procesamiento (segundos)
Clustering no supervisado	95.45	120
Modelo de mezcla gaussiana	89.32	150
Algoritmo de movimiento	92.11	135

Resumen de los resultados: El algoritmo de clustering no supervisado propuesto demostró ser más preciso y eficiente que los métodos comparados, lo que lo convierte en una solución efectiva para la extracción de fotogramas clave en grandes volúmenes de datos de video. Esto sugiere que este enfoque tiene un gran potencial para aplicaciones en análisis de video en tiempo real y en escenarios donde se requiere procesamiento rápido de datos.

2.1.6. NewsVideo Summarization Combining SURF and Color Histogram Features

[Liang et al., 2021](#)

2.1.6.1. Planteamiento del Problema

El volumen de videos noticiosos ha crecido exponencialmente, lo que hace necesario contar con métodos eficientes para resumirlos y facilitar su consumo. Los métodos actuales de resumen de video no siempre capturan la complejidad de los videos debido a la variabilidad en las transiciones entre escenas. Además, muchos de los enfoques existentes fallan en la detección

precisa de los límites entre escenas debido a los cambios en la complejidad visual, lo que resulta en resúmenes de video poco representativos.

2.1.6.2. Objetivos

- Proponer un método novedoso para la detección de límites de escenas basado en características SURF.
- Desarrollar un algoritmo mejorado de clustering que no requiera predefinir el número de clusters y pueda adaptarse a la complejidad visual de las escenas.
- Reducir la redundancia en los marcos clave seleccionados para representar mejor el contenido del video noticioso.
- Validar la eficacia del método propuesto mediante experimentos en datasets públicos y autoconstruidos.

2.1.6.3. Fundamento Teórico

El artículo se basa en el uso de características locales, como las características SURF (Speeded Up Robust Features), que son invariantes ante rotaciones, cambios de escala y variaciones de iluminación. Estas características permiten una mayor precisión en la detección de límites entre escenas, en comparación con métodos que dependen de características globales como los histogramas de color.

El algoritmo SURF se utiliza para identificar puntos clave en los fotogramas, lo que permite comparar la similitud entre ellos. Además, el uso de un algoritmo de clustering mejorado basado en histogramas HSV permite extraer los marcos clave más representativos de cada escena.

2.1.6.4. Metodología

1. Preprocesamiento de datos:

- Se eliminan elementos estáticos, como logotipos o subtítulos rodantes, que pueden interferir con la detección de límites de escena.

2. Extraer características:

- Se extraen puntos de interés locales y se comparan entre fotogramas adyacentes utilizando el algoritmo FLANN para acelerar el proceso.

3. Similitud entre fotogramas:

- A partir de las coincidencias entre características SURF, se genera una curva de similitud que permite identificar transiciones abruptas y graduales entre escenas.

4. Clustering basado en histogramas:

- Se agrupan los fotogramas de cada escena según sus histogramas de color. El algoritmo ajusta dinámicamente el número de clusters para reflejar la complejidad visual de la escena.

5. Selección de marcos clave:

- Se elige el fotograma más cercano al centro de cada cluster como representante clave de esa escena, garantizando que los marcos seleccionados representen fielmente el contenido del video.

2.1.6.5. Resultados:

Tabla 2.5: Resultados de la detección de límites de escenas por tipo de video

Tipo de Video	Recall (%)	Precisión (%)
News in 30 Minutes	96.57	97.24
CCTV News	100.00	94.29
Sports Express	96.05	85.88
Total	97.22	93.33

Tabla 2.6: Comparación de métodos de detección de límites de escenas

Método	Recall (%)	Precisión (%)
Ours	97.22	93.33
Feng	86.48	93.20
Wang	90.41	94.28
Rachida	96.49	95.87
TransNetV2	92.68	98.52

Nota. Comparación de los resultados de distintos métodos para la detección de límites de escenas.

Tabla 2.7: Comparación de métodos en el dataset TVSum50

Método	Recall (%)	Precisión (%)
Ours	95.93	87.91
TransNetV2	92.56	92.02

2.2. Bases Teóricas

En esta sección se presentan los conceptos teóricos fundamentales que sustentan el desarrollo de este proyecto.

2.2.1. Imagen Digital

Una imagen digital se define como una representación visual constituida por una matriz de píxeles, en la cual cada píxel contiene datos sobre la intensidad de luz y, en ocasiones, sobre el color en un punto específico. Según [Gonzalez y Woods, 2021](#), esta matriz permite almacenar y manipular información visual de manera estructurada, facilitando el análisis computacional de contenido visual. Las imágenes digitales constituyen el formato base en el análisis de datos visuales y son aplicadas en diversas áreas de la visión por computadora, la medicina, la robótica y la inteligencia artificial.

Las imágenes en escala de grises contienen un solo canal que representa la intensidad de luz en cada píxel, con valores que varían entre el negro (valor mínimo) y el blanco (valor máximo). Este tipo de representación se centra en los contrastes de brillo y es útil en aplicaciones de análisis donde el color no es esencial, como la detección de bordes o la segmentación de formas.

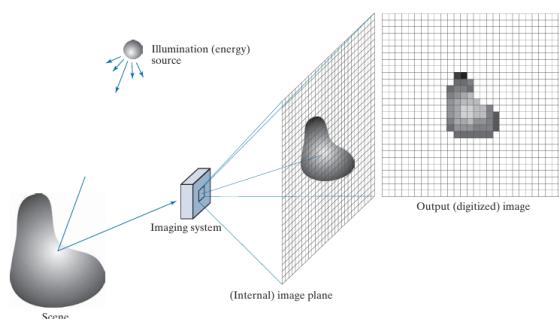


Figura 2.6: Representación de la transformación digital de una imagen. Fuente:[Gonzalez y Woods, 2021](#).



Figura 2.7: Imagen RGB y sus componentes. Fuente:[Gonzalez y Woods, 2021](#).

En cambio, las imágenes en color suelen estar compuestas por tres canales: rojo, verde y azul (RGB). Cada canal almacena la intensidad de uno de estos colores, y su combinación permite la reproducción de una amplia gama de tonos. [Pratt, 2007](#) destaca que esta separación en canales RGB proporciona un control detallado de cada componente cromático, lo cual es fundamental en aplicaciones que requieren precisión en la representación de colores, como el análisis de patrones complejos o la clasificación de objetos.

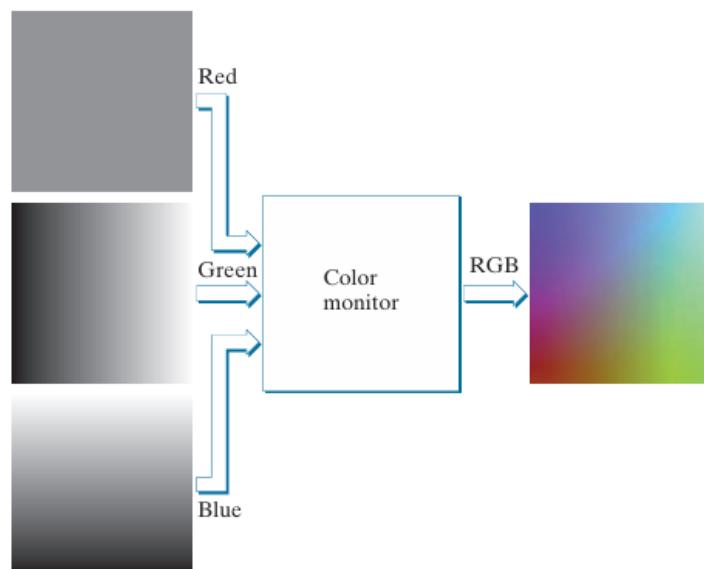


Figura 2.8: Composición de imagen RGB. Fuente:[Gonzalez y Woods, 2021](#).

2.2.2. Video

Un video es una secuencia de imágenes digitales o fotogramas que se muestran en rápida sucesión, generando la percepción de movimiento continuo. Cada fotograma en un video es una imagen digital individual que contiene información sobre la intensidad y el color organizada en una matriz de píxeles. Este flujo de fotogramas, al ser reproducido en orden y a una velocidad específica, medida en fotogramas por segundo (fps), permite al cerebro humano interpretar los cambios entre las imágenes como movimiento, debido al fenómeno de la persistencia visual.[Smith, 2012](#)

La estructura de un video, al combinar múltiples fotogramas, facilita el análisis de patrones de cambio y permite detectar eventos en tiempo real, lo cual es fundamental en aplicaciones de visión por computadora y procesamiento de video. Según [Tekalp, 2015](#), los videos ofrecen una representación temporal y espacial de datos visuales que permite a los algoritmos de procesamiento analizar variaciones de intensidad y movimiento en una secuencia continua, aplicándose en áreas como la vigilancia, la detección de objetos y la reconstrucción tridimensional de escenas.

Este modelo de datos secuenciales se ha convertido en un estándar en sistemas de visión artificial y análisis de imágenes en movimiento, siendo clave para el desarrollo de técnicas avanzadas de procesamiento de video y aprendizaje profundo.

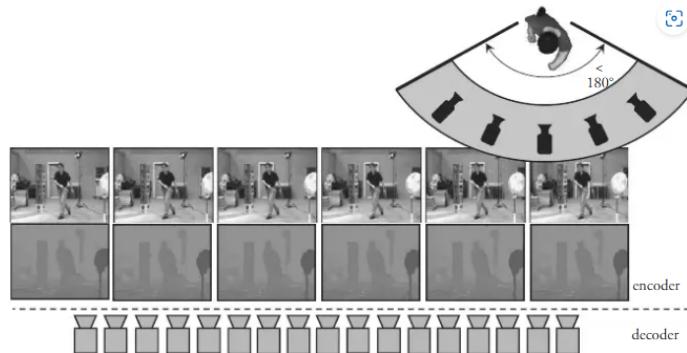


Figura 2.9: Representación del video en frames. Fuente:[Tekalp, 2015](#).

Este modelo de datos secuenciales se ha convertido en un estándar en sistemas de visión artificial y análisis de imágenes en movimiento, siendo clave para el desarrollo de técnicas avanzadas de procesamiento de video y aprendizaje profundo.

2.2.2.1. Preprocesamiento de Imágenes

El preprocesamiento de imágenes es una etapa inicial en el procesamiento digital que busca preparar los datos visuales para un análisis posterior. Este proceso implica una serie de transformaciones que ajustan las imágenes para optimizar su calidad y formato, mejorando así la consistencia y precisión en las tareas de análisis. Al normalizar los datos y reducir el ruido visual, el preprocesamiento permite obtener una representación más uniforme de las imágenes, facilitando la detección de patrones visuales relevantes.

2.2.2.2. Extracción de Características

La extracción de características es un procedimiento mediante el cual las imágenes se transforman en vectores de datos que describen sus aspectos más representativos. Este proceso permite simplificar la información visual y destacar los elementos esenciales de cada imagen, facilitando su análisis y clasificación en aplicaciones de visión por computadora. Entre los métodos más comunes de extracción de características se encuentra el *Histogram of Oriented Gradients* (HOG).

- *Histogram of Oriented Gradients* (HOG): Introducido por [Dalal y Triggs, 2005](#), HOG es un descriptor ampliamente utilizado en el procesamiento de imágenes y la visión por computadora, principalmente en la detección de objetos. Este método permite capturar las características locales de una imagen mediante la distribución de los gradientes de intensidad, lo que resulta especialmente útil para delinear bordes y contornos.

Para calcular el HOG, la imagen se divide en pequeñas celdas, dentro de las cuales se calcula un histograma de orientaciones de gradientes. La magnitud y la dirección del gradiente en cada píxel se obtienen aplicando filtros de derivadas en las direcciones horizontal y vertical, calculados como:

$$G_x = I(x+1, y) - I(x-1, y)$$

$$G_y = I(x, y+1) - I(x, y-1)$$

donde G_x y G_y representan los gradientes en las direcciones horizontal y vertical, respectivamente.

La magnitud del gradiente $M(x, y)$ y su orientación $\theta(x, y)$ se calculan con las siguientes expresiones:

$$M(x,y) = \sqrt{G_x^2 + G_y^2}$$

$$\theta(x,y) = \arctan2(G_y, G_x)$$

Posteriormente, se construye un histograma de orientaciones para cada celda, el cual se normaliza para incrementar la robustez ante cambios de iluminación y contraste. Los histogramas de todas las celdas se concatenan para formar el descriptor final de HOG.

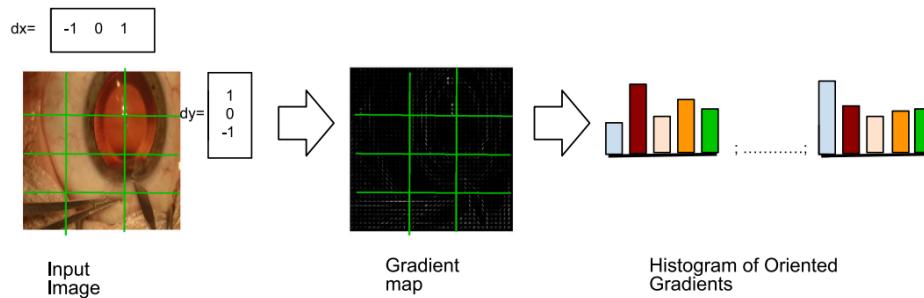


Figura 2.10: Proceso de Histogram of Oriented Gradients (HOG). Fuente:[Bhattarai et al., 2023](#).

2.2.3. Deep Learning

El Deep Learning es un subcampo del aprendizaje automático que se basa en redes neuronales profundas, estructuras compuestas por múltiples capas de procesamiento no lineales que aprenden representaciones jerárquicas de los datos. En base a [Goodfellow et al., 2016](#), cada capa sucesiva transforma la información recibida de la capa anterior, lo que permite que la red construya representaciones progresivamente más abstractas y complejas. Esta capacidad para aprender características de alto nivel de manera automática y sin necesidad de intervención humana ha posicionado al Deep Learning como una herramienta central en inteligencia artificial y procesamiento de datos complejos.

El aprendizaje en redes profundas se realiza mediante técnicas de optimización, como el algoritmo de retropropagación, que ajusta los pesos de las conexiones entre neuronas en función del error obtenido en las predicciones, permitiendo a la red mejorar su precisión iterativamente. De acuerdo con [LeCun et al., 2015](#), este enfoque permite que las redes neuronales profundas descubran patrones y relaciones complejas en los datos, incluso en presencia de grandes volúmenes de información. La combinación de capas profundas y grandes conjuntos de datos permite que los modelos de Deep Learning superen a los métodos tradicionales en tareas complejas como la visión por computadora, el procesamiento de lenguaje natural y el reconocimiento de voz.

A diferencia de los métodos de aprendizaje automático convencionales, el Deep Learning es capaz de manejar datos de alta dimensionalidad y construir modelos robustos y generalizables, lo cual lo convierte en una solución eficaz para problemas donde las relaciones entre variables no son fácilmente definibles.

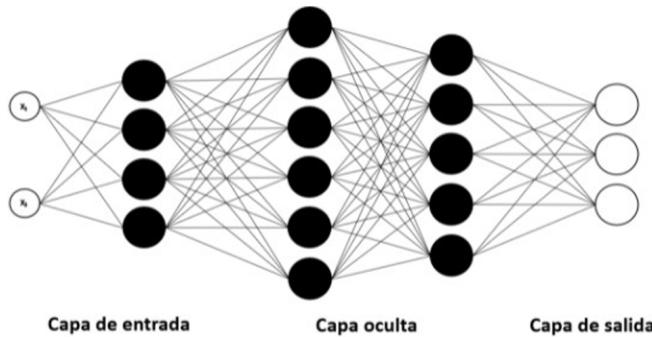


Figura 2.11: Estructura de una red neuronal. Fuente: Elaboración propia.

2.2.3.1. Redes Neuronales Convolucionales (CNNs)

Las Redes Neuronales Convolucionales (CNNs) son una arquitectura central en Deep Learning, especialmente reconocida por su rendimiento superior en el análisis y procesamiento de datos visuales. Su eficacia proviene de su capacidad para aprender representaciones jerárquicas y espaciales a través de la aplicación de capas convolucionales, que extraen automáticamente características clave de las imágenes al aprovechar la estructura espacial inherente de los datos ([Goodfellow et al., 2016](#)).

Las CNNs están compuestas por múltiples capas especializadas que procesan la información de manera progresiva, permitiendo que el modelo aprenda patrones visuales de diferente complejidad. En las capas iniciales, las CNNs detectan características de bajo nivel, como bordes y texturas, mientras que las capas más profundas capturan patrones abstractos y complejos, integrando la información aprendida en las primeras capas. Según [Goodfellow et al., 2016](#), esta organización en capas jerárquicas permite que las CNNs optimicen su capacidad de reconocimiento y clasificación, adaptándose a la variabilidad de los datos visuales sin necesidad de preprocesamiento extenso.

- **Capa Convolutacional:** La capa convolucional es un componente fundamental en las redes neuronales convolucionales (CNNs) y se destaca por su capacidad para extraer características locales de los datos visuales mediante el uso de filtros (o kernels). Estos filtros se aplican a la imagen de entrada a través de una operación de convolución, permitiendo detectar patrones específicos como bordes, texturas y estructuras complejas. Esta

arquitectura permite a las CNNs capturar dependencias espaciales en los datos, ya que cada filtro actúa sobre una pequeña región de la imagen, conocida como campo receptor, lo que permite al modelo aprender patrones locales esenciales para el análisis visual ([Goodfellow et al., 2016](#)).

En una capa convolucional, cada filtro genera un mapa de características al pasar sobre diferentes regiones de la imagen y capturar su respuesta a ciertos patrones. Al aplicar un filtro K sobre una imagen de entrada X , el resultado es un mapa de características F , calculado mediante la siguiente operación de convolución:

$$F(i, j) = \sum_m \sum_n X(i + m, j + n) \cdot K(m, n),$$

donde i y j representan las posiciones en el mapa de salida, mientras que m y n son los índices de los elementos del filtro K . Esta estructura de capas permite que las CNNs identifiquen características de bajo nivel en las primeras capas, como bordes y texturas, y evolucionen hacia patrones más abstractos en capas profundas, integrando la información aprendida en cada nivel jerárquico ([LeCun et al., 2015](#)).

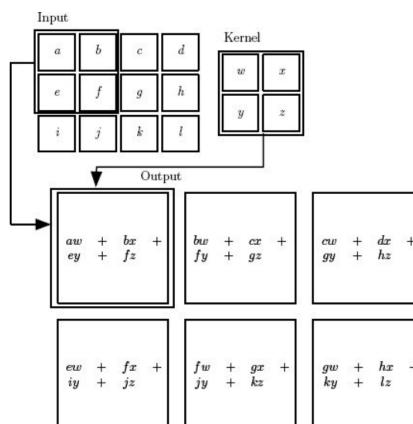


Figura 2.12: Proceso de convolución. Fuente: [Goodfellow et al., 2016](#). Recuperado de *Deep Learning*.

La compartición de parámetros en la capa convolucional, donde un mismo filtro se aplica a diferentes regiones de la imagen, permite reducir significativamente el número de parámetros del modelo, lo cual mejora la eficiencia y facilita su generalización. Esta capacidad de capturar relaciones espaciales clave con menos parámetros convierte a la capa convolucional en un elemento esencial en el análisis de imágenes, permitiendo una extracción de características visuales relevantes de manera escalable y eficiente.

- **Capa de Pooling:** La capa de pooling es un componente crucial en las redes neuronales convolucionales (CNNs) que se utiliza para reducir la dimensionalidad espacial de los

mapas de características, manteniendo las características más relevantes y reduciendo la sensibilidad de la red a pequeñas variaciones en la posición y escala de los patrones dentro de los datos visuales. Esta capa mejora la eficiencia del modelo y contribuye a evitar el sobreajuste al reducir el número de parámetros y complejidad del modelo ([LeCun et al., 2015](#)).

Existen varios tipos de operaciones de pooling, siendo las más comunes el max pooling y el average pooling. En el max pooling, la capa selecciona el valor máximo dentro de una región definida (o ventana) del mapa de características, lo cual preserva las características más destacadas de cada región, permitiendo que la red se enfocue en las señales más fuertes. Por otro lado, el average pooling calcula el valor promedio en cada región, proporcionando una representación suavizada de las características visuales. Estas técnicas permiten resumir la información de manera compacta, manteniendo las características esenciales y reduciendo el tamaño de los mapas de características ([Scherer et al., 2010](#); [Sze et al., 2017](#)).

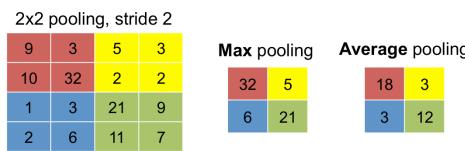


Figura 2.13: Tipos de operaciones pooling. Fuente: [Sze et al., 2017](#).

Matemáticamente, el max pooling de un mapa de características F con una ventana de pooling de tamaño $p \times p$ se expresa como:

$$P(i, j) = \max_{m,n} \{F(i + m, j + n)\},$$

donde $P(i, j)$ es el valor en el mapa de salida después de aplicar max pooling, y m, n representan los desplazamientos dentro de la ventana de pooling.

La capa de pooling también ayuda a que las CNNs sean más robustas ante variaciones en la posición y orientación de los patrones dentro de las imágenes, ya que se enfoca en las características más dominantes y evita que la red dependa de detalles específicos de la ubicación ([Scherer et al., 2010](#)). Este proceso de reducción de dimensionalidad no solo mejora la eficiencia computacional, sino que también facilita la generalización en los datos visuales, haciéndola fundamental en aplicaciones de visión por computadora.

2.2.3.2. ResNet50

La ResNet50 es una arquitectura de red neuronal profunda que pertenece a la familia de las Redes Residuales (Residual Networks o ResNet), introducida por He et al. en 2015 [He et al., 2016](#). Esta red es conocida por su capacidad para entrenar modelos de gran profundidad sin sufrir problemas de degradación del rendimiento, que afectan a las redes convolucionales tradicionales cuando se incrementa el número de capas. Para lograr esto, ResNet50 implementa bloques residuales, los cuales permiten una mejor propagación del gradiente a lo largo de las capas y facilitan el entrenamiento de redes más profundas.

ResNet50 está compuesta por 50 capas, que incluyen capas convolucionales, de *pooling* y totalmente conectadas. La clave de su arquitectura reside en los bloques residuales, en los que la salida de una capa se suma directamente a la salida de una capa posterior, lo que permite que los datos "salten" una o más capas. Esta operación de "salto" conserva la información a lo largo de la red y mitiga la pérdida del gradiente, un problema común en redes profundas ([He et al., 2016](#)). La estructura de ResNet50 puede dividirse en:

- **Bloques residuales:** Cada bloque consiste en una serie de capas convolucionales, seguidas de una conexión residual que permite sumar la entrada del bloque a la salida de una capa más profunda, facilitando el flujo de información. Este mecanismo de conexión residual es fundamental para evitar el problema de desaparición del gradiente en redes profundas y permite a la red aprender representaciones más complejas sin perder precisión en el entrenamiento.
- **Capas convolucionales y pooling:** La red comienza con una capa convolucional inicial y una operación de pooling, seguidas de varios bloques residuales. Al final de la red, se realiza un global average pooling para reducir la dimensionalidad antes de pasar a la capa totalmente conectada de salida. Este diseño permite a ResNet50 procesar eficientemente características de bajo y alto nivel, lo que la hace adecuada para tareas avanzadas de clasificación y detección en conjuntos de datos complejos.

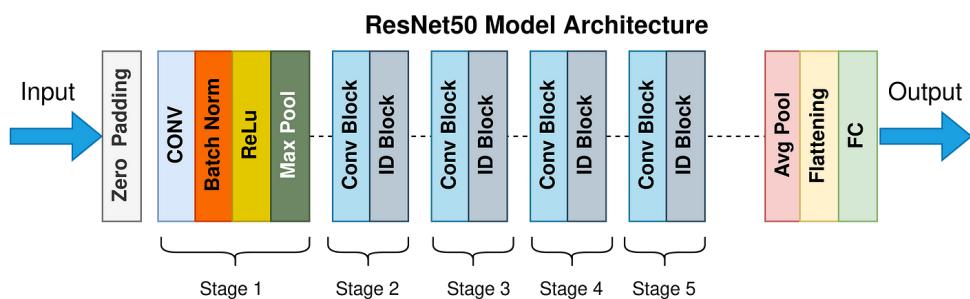


Figura 2.14: Arquitectura de ResNet50. Fuente:[Mukherjee, 2022](#).

Funcionamiento de los bloques residuales Los bloques residuales son el componente esencial que diferencia a ResNet50 de otras arquitecturas. En cada bloque, una entrada x se combina con la salida de una convolución $F(x)$ mediante una suma:

$$y = F(x) + x,$$

donde $F(x)$ es la salida de las capas convolucionales del bloque. Esta suma permite que el bloque aprenda la función de residuo en lugar de una transformación completa, lo que facilita el entrenamiento y previene el desvanecimiento del gradiente.

2.2.4. Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales (PCA) es una técnica estadística ampliamente utilizada para la reducción de dimensionalidad en conjuntos de datos de alta dimensión. Mediante PCA, los datos se transforman en un espacio de menor dimensión que conserva la mayor cantidad de variabilidad posible, lo cual resulta útil en el preprocesamiento de datos para reducir la redundancia y simplificar el análisis sin perder información crítica. Según [Jolliffe y Cadima, 2016](#), PCA es especialmente relevante en aplicaciones donde el número de variables es elevado, y es deseable representar los datos en un espacio de menor dimensión que mantenga la información esencial.

2.2.4.1. Funcionamiento de PCA

A continuación se describen los pasos principales de PCA, con las fórmulas correspondientes para su implementación.

1. **Centramiento de los datos:** Dado un conjunto de datos con n muestras y d características, se calcula la media de cada característica y se centra cada variable restando su media. Este centrado es fundamental porque asegura que los componentes principales representen direcciones de variación en torno al origen, lo cual simplifica los cálculos y asegura una correcta interpretación de la varianza en los datos en base a [Jolliffe y Cadima, 2016](#).

Sean \mathbf{X} la matriz de datos original de tamaño $n \times d$, y $\bar{\mathbf{X}}$ el vector de medias, donde cada elemento \bar{X}_j es la media de la característica j . La matriz centrada \mathbf{X}_c se obtiene restando la media de cada columna:

$$\mathbf{X}_c = \mathbf{X} - \mathbf{1}_n \bar{\mathbf{X}}^T,$$

donde $\mathbf{1}_n$ es un vector columna de unos de longitud n . Este paso de centrado, mencionado también por [Bishop, 2007](#), asegura que la varianza se calcule en torno al origen.

2. Cálculo de la matriz de covarianza: La matriz de covarianza, denotada como \mathbf{C} , representa las relaciones lineales entre las características y permite identificar patrones de variabilidad conjunta. La matriz de covarianza, de tamaño $d \times d$, refleja las variaciones compartidas entre las variables, facilitando la identificación de relaciones lineales entre ellas ([Jolliffe y Cadima, 2016](#)).

La matriz de covarianza \mathbf{C} se calcula como:

$$\mathbf{C} = \frac{1}{n-1} \mathbf{X}_c^T \mathbf{X}_c.$$

Según [Bishop, 2007](#), esta matriz permite capturar la estructura interna de las relaciones entre las variables, proporcionando una base sólida para la reducción de dimensionalidad al identificar las direcciones de máxima varianza.

3. Descomposición en valores y vectores propios: Para encontrar los componentes principales, se realiza una descomposición en valores propios de la matriz de covarianza \mathbf{C} . Este proceso permite obtener los *vectores propios* y *valores propios* de \mathbf{C} , los cuales representan las direcciones de máxima varianza y la cantidad de varianza explicada en esas direcciones, respectivamente. Este enfoque garantiza que los componentes principales maximicen la varianza en los datos ([Jolliffe y Cadima, 2016](#)).

La descomposición de la matriz de covarianza se representa como:

$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i,$$

donde λ_i es el valor propio asociado al vector propio \mathbf{v}_i , y representa la varianza explicada por ese componente. Los vectores propios se ordenan en función de los valores propios de mayor a menor, asegurando que los primeros componentes capturen la mayor parte de la variabilidad en los datos ([Bishop, 2007](#)).

4. Selección de componentes principales: Para reducir la dimensionalidad, se seleccionan los primeros k componentes principales que explican la mayor cantidad de variabilidad. La proporción de la varianza explicada por cada componente se calcula como:

$$\text{Varianza Explicada} = \frac{\lambda_i}{\sum_{j=1}^d \lambda_j}.$$

Este cálculo permite evaluar la importancia relativa de cada componente, utilizando métodos como el criterio del "codo" para identificar el número óptimo de componentes k ([Jolliffe y Cadima, 2016](#)).

5. Proyección en el espacio reducido: Finalmente, los datos originales se proyectan en el espacio de menor dimensión definido por los primeros k componentes principales. Dado

\mathbf{V}_k , una matriz de tamaño $d \times k$ que contiene los k vectores propios seleccionados como columnas, los datos proyectados $\mathbf{X_p}$ se calculan como:

$$\mathbf{X_p} = \mathbf{X_c} \mathbf{V}_k.$$

La matriz $\mathbf{X_p}$ representa los datos originales en un espacio de menor dimensión, conservando la información más relevante y simplificando su estructura para aplicaciones posteriores .

2.2.4.2. PCA en Imágenes

Se aplica en el ámbito de imágenes digitales para reducir la dimensionalidad de los datos, lo cual es especialmente útil debido a la gran cantidad de información contenida en cada imagen. En el contexto de imágenes, cada imagen se convierte en un vector de características, representando cada píxel como una dimensión en el espacio. Esto genera un espacio de alta dimensionalidad, especialmente en imágenes de alta resolución, donde el número de dimensiones es igual al número de píxeles.

Uno de los usos más destacados de PCA en visión por computadora es el método de Eigenfaces, introducido por [Turk y Pentland, 1991](#). En este método, cada imagen facial es transformada en un vector de características y PCA se utiliza para identificar las direcciones principales de variación en el conjunto de imágenes. Esto permite representar cada imagen como una combinación lineal de componentes principales o eigenfaces, que capturan las variaciones esenciales en los rostros humanos.

Para aplicar PCA en imágenes, el conjunto de imágenes se representa en una matriz de datos \mathbf{X} , donde cada columna corresponde a una imagen vectorizada (es decir, un vector de longitud igual al número de píxeles). La matriz \mathbf{X} tiene dimensiones $d \times n$, donde d es el número de píxeles y n es el número de imágenes.

Esta reconstrucción aproxima la imagen original utilizando solo los primeros k componentes, lo que reduce significativamente la cantidad de datos necesarios para representar cada imagen, manteniendo las características principales.

Al aplicar PCA en un conjunto de imágenes faciales, las primeras eigenfaces representan las características más comunes y generales de los rostros, como la forma general de la cara o la posición de los ojos y la boca. Las eigenfaces subsiguientes capturan detalles específicos o variaciones menores. Esta reducción de dimensionalidad facilita la clasificación y el reconocimiento facial al proyectar cada imagen en el espacio de eigenfaces, lo cual permite comparar y distinguir entre rostros en función de sus representaciones en este espacio reducido.

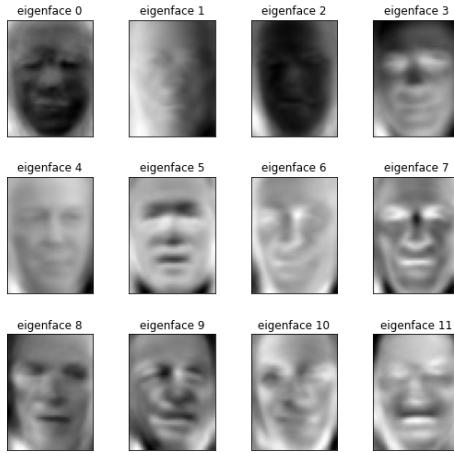


Figura 2.15: Ejemplo de Eigenfaces generados mediante PCA para el reconocimiento facial. Fuente: Turk y Pentland (1991) [Turk y Pentland, 1991](#).

2.2.5. Clustering

Es una técnica de aprendizaje no supervisado que agrupa un conjunto de datos en subconjuntos llamados *clusters*, de manera que los objetos dentro de un mismo cluster sean más similares entre sí que con los de otros clusters. Es una herramienta clave en la minería de datos, el análisis de datos y el aprendizaje automático.

A diferencia del aprendizaje supervisado, donde se dispone de etiquetas de clase, el clustering busca encontrar estructuras ocultas en los datos sin conocimiento previo de las etiquetas.

2.2.5.1. K-means

Su objetivo es dividir un conjunto de datos en k grupos o *clusters*, minimizando la varianza dentro de cada grupo. La esencia de K-means es encontrar k centroides, uno para cada cluster, de manera que los puntos de datos dentro de un cluster estén más cerca de su propio centroide que de cualquier otro.

Dado un conjunto de datos $X = \{x_1, x_2, \dots, x_n\}$ donde cada $x_i \in \mathbb{R}^d$, el algoritmo K-means intenta particionar los datos en k clusters $C = \{C_1, C_2, \dots, C_k\}$, tal que se minimice el criterio de suma de los cuadrados de las distancias (SSE, *Sum of Squared Errors*) entre los puntos y sus respectivos centroides.

La función objetivo que el algoritmo K-means intenta minimizar es la siguiente:

$$J(C) = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2,$$

donde:

- C_i es el i -ésimo cluster,
- $\mu_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j$ es el centroide del cluster C_i , calculado como el promedio de los puntos de datos asignados a ese cluster,
- $\|x_j - \mu_i\|$ es la distancia Euclídea entre un punto x_j y el centroide μ_i .

Esta función mide la compacidad de los clusters. El objetivo del algoritmo es minimizar $J(C)$, es decir, minimizar la suma de las distancias al cuadrado entre los puntos de datos y sus centroides correspondientes.

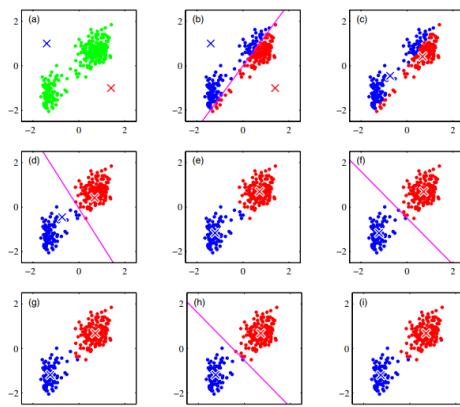


Figura 2.16: Ilustración de K-means. Fuente: [Bishop, 2006](#). Recuperado de *Pattern Recognition and Machine Learning*.

2.2.6. Inercia en clustering

En el análisis de clustering, la **inercia** es una medida de la dispersión interna de los datos dentro de un conjunto de clusters. En términos prácticos, la inercia cuantifica cuán estrechamente agrupados están los puntos de datos alrededor de los centroides de sus respectivos clusters. Esta medida es particularmente relevante en el algoritmo de *K-means*, donde se busca minimizar la inercia para asegurar que cada cluster esté lo más compacto posible.

2.2.6.1. Definición de inercia

La inercia, también conocida como *within-cluster sum of squares* (WCSS), se calcula como la suma de las distancias cuadradas entre cada punto de datos y el centroide de su cluster asignado. Matemáticamente, si tenemos un conjunto de datos $X = \{x_1, x_2, \dots, x_n\}$ agrupado en k clusters $C = \{C_1, C_2, \dots, C_k\}$, la inercia se define como:

$$\text{Inercia} = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2$$

donde:

- C_i es el i -ésimo cluster,
- x_j representa cada punto de datos perteneciente al cluster C_i ,
- μ_i es el centroide del cluster C_i , calculado como el promedio de los puntos de datos dentro de C_i ,
- $\|x_j - \mu_i\|$ es la distancia euclíadiana entre el punto x_j y el centroide μ_i .

La inercia mide, por tanto, la compactación de los clusters: a menor inercia, mayor cohesión interna en cada cluster.

2.2.6.2. Optimización de clusters mediante Inercia(Método del codo)

La minimización de la **inercia** es uno de los objetivos fundamentales en el algoritmo de *K-means*, ya que una inercia baja indica que los puntos de datos están más cerca de sus respectivos centroides, sugiriendo que los clusters son compactos y bien definidos. Sin embargo, la inercia tiende a disminuir conforme se incrementa el número de clusters, lo cual puede llevar a un sobreajuste si no se establece un límite adecuado en el número de clusters.

Para determinar el número óptimo de clusters, se utiliza el **método del codo**, una técnica gráfica ampliamente reconocida en el análisis de clusters. Según Humaira y Rasyidah [Humaira y Rasyidah, 2018](#), este método consiste en representar la inercia en función de diferentes valores de k (número de clusters). A medida que k aumenta, la inercia disminuye debido a que los clusters se vuelven más pequeños y específicos. Sin embargo, después de cierto punto, el beneficio de añadir clusters adicionales se reduce y la curva de inercia comienza a estabilizarse, formando un “codo”.

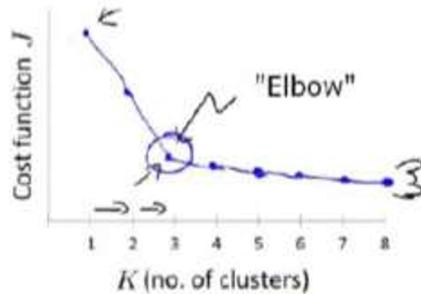


Figura 2.17: Ilustración del método del codo para seleccionar el número de clusters óptimo en K-means. Fuente: [Humaira y Rasyidah, 2018](#).

El *codo* en la curva representa el punto en el cual el número de clusters comienza a añadir solo mejoras marginales en la inercia. Elegir el valor de k en este codo permite identificar una estructura de clusters adecuada, logrando un equilibrio entre una baja inercia y un número razonable de clusters. Este enfoque evita tanto la fragmentación excesiva del conjunto de datos como el sobreajuste al seleccionar un número óptimo de clusters que mejor represente la estructura de los datos.

2.3. Marco Conceptual

2.3.1. Scraping

Scraping es el proceso de extracción automática de información desde páginas web. Este método permite recopilar grandes volúmenes de datos de manera eficiente y organizada, especialmente útil en contextos donde se necesita obtener y estructurar datos no disponibles de forma directa en una API o base de datos. Para su implementación, se utilizan herramientas y librerías, como BeautifulSoup y Selenium, que facilitan la navegación y extracción de elementos específicos dentro de una página web, como textos, imágenes o enlaces. El scraping es particularmente relevante en el análisis de datos y la curaduría de contenido, ya que permite alimentar sistemas de análisis y machine learning con datos en tiempo real o en grandes cantidades.

2.3.2. Automatización

La automatización es el proceso de emplear tecnología para realizar tareas de forma autónoma, con mínima intervención humana. En el ámbito de procesamiento de datos y análi-

sis de contenido, la automatización facilita la ejecución de procesos repetitivos y complejos, como la extracción y organización de grandes volúmenes de datos, mejorando tanto la precisión como la eficiencia. En el contexto de contenido audiovisual, la automatización permite aplicar algoritmos que identifican patrones y seleccionan datos relevantes sin necesidad de intervención manual en cada paso, asegurando que el sistema pueda manejar el flujo constante de información. Además, la automatización es fundamental para lograr consistencia en el tratamiento de los datos, reduciendo errores y tiempo en cada fase del proceso, desde la recolección de datos hasta la generación de resultados finales, optimizando el análisis en contextos de alta carga informativa.

2.3.3. Curaduría de contenido

La curaduría de contenido es un proceso que involucra la selección, organización y presentación de la información más relevante dentro de un amplio conjunto de datos. En el contexto de la información digital, la curaduría se convierte en una herramienta clave para filtrar y destacar elementos específicos que son de interés para el usuario. Esto es especialmente útil en medios audiovisuales, donde se busca identificar los momentos clave en grandes volúmenes de contenido para crear resúmenes informativos. La curaduría automatizada utiliza algoritmos y técnicas de machine learning para analizar y seleccionar segmentos de interés, optimizando así el consumo de información. Esta automatización de la curaduría también ayuda a personalizar la información en función de patrones específicos de uso, haciendo que la presentación del contenido sea más precisa y atractiva para audiencias específicas.

2.3.4. Preprocesamiento de datos

El preprocesamiento es una fase crítica en el análisis de datos, especialmente en el tratamiento de imágenes y videos. Involucra una serie de técnicas que mejoran la calidad de los datos y preparan el contenido visual para su posterior análisis. Entre los procesos de preprocesamiento comunes se incluyen la normalización de colores, que ajusta los valores de color para que sean coherentes en todas las imágenes; el redimensionado, que ajusta el tamaño de los frames; y la eliminación de ruido, que remueve elementos no deseados en los datos visuales. Estas técnicas permiten que los datos sean interpretados de manera uniforme por los algoritmos de aprendizaje automático, asegurando una mayor precisión y consistencia en el análisis de grandes volúmenes de información. El preprocesamiento también contribuye a reducir el tiempo de procesamiento, optimizando el flujo de trabajo en proyectos de análisis masivo.

2.3.5. Extracción de características

La extracción de características es el proceso mediante el cual se identifican y capturan los atributos visuales más representativos de una imagen o un video. Este proceso convierte el contenido visual en vectores de datos numéricos que pueden ser fácilmente manipulados y analizados por algoritmos de machine learning.

2.3.6. Clustering

El clustering es una técnica de aprendizaje no supervisado que organiza datos en grupos o clusters basados en sus similitudes. Esta técnica es particularmente útil en el análisis de contenido visual, donde el clustering permite agrupar frames de video que presentan características visuales similares, facilitando la identificación de segmentos relevantes en el contenido. El clustering ayuda a reducir la complejidad en el análisis de datos masivos, permitiendo una estructuración que optimiza la selección de frames clave y mejora la eficiencia de los resúmenes automáticos.

2.3.7. Reducción de dimensionalidad

La reducción de dimensionalidad es una técnica que simplifica la estructura de los datos, eliminando variables redundantes o irrelevantes sin perder la esencia de la información. En el contexto de análisis visual, la reducción de dimensionalidad permite optimizar el uso de los datos al extraer solo aquellas características que aportan mayor valor informativo. Esto no solo reduce el tiempo y recursos necesarios para el procesamiento, sino que también permite mejorar la precisión de los modelos de machine learning al reducir el ruido en los datos.

2.3.8. Evaluación de clusters

La evaluación de clusters es el proceso de analizar la calidad y coherencia de los grupos formados en un conjunto de datos, asegurando que los datos agrupados sean internamente homogéneos y estén bien diferenciados de otros clusters. En proyectos de análisis visual, la evaluación de clusters es crucial para garantizar que los grupos formados representen de manera precisa distintos segmentos del contenido. Esto permite optimizar la selección de frames clave y asegurar que el contenido sea adecuado para la creación de resúmenes automáticos.

2.3.9. Selección de frames clave

La selección de frames es el proceso de identificar y elegir momentos específicos o representativos de un video para su análisis posterior. Esta técnica permite reducir significativamente el número de frames que se deben procesar, optimizando el uso de recursos computacionales y mejorando la eficiencia en el análisis de contenido. La selección de frames clave se realiza en función de métricas de representatividad, permitiendo que los segmentos seleccionados reflejen de manera precisa los momentos importantes del video. En la creación de resúmenes automáticos, la selección de frames clave facilita la generación de una representación visual concisa sin perder detalles importantes.

2.3.10. Metadatos

Los metadatos son datos descriptivos que acompañan al contenido multimedia y proporcionan información adicional sobre el archivo, como su fecha de creación, duración, etiquetas y descripción. En el análisis de videos, los metadatos permiten organizar, buscar y filtrar contenido de manera estructurada, mejorando la accesibilidad y facilitando la clasificación y curaduría de contenido audiovisual. Los metadatos son fundamentales para gestionar grandes volúmenes de datos, ya que permiten una organización eficiente de la información y ofrecen contexto sobre cada archivo de video, facilitando su identificación y análisis.

2.3.11. Etiquetado de datos

El etiquetado de datos es el proceso de asignar identificadores o etiquetas a elementos de un conjunto de datos para facilitar su clasificación y análisis. En el caso de los frames de video, el etiquetado permite identificar y organizar los diferentes momentos en función de su contenido, ayudando a estructurar los datos para un análisis más eficiente.

2.3.12. Concatenación de características

El stacking de características es un proceso que combina múltiples vectores de características en una estructura unificada, permitiendo representar diversas propiedades de los datos de manera conjunta. En el análisis de contenido visual, el stacking permite integrar características obtenidas a través de diferentes métodos de extracción en una sola estructura. Esto mejora la riqueza descriptiva de los datos y optimiza la precisión de los algoritmos en tareas de clasificación y clustering, proporcionando al modelo una visión más completa de los patrones y

atributos presentes en el contenido visual.

2.3.13. Extracción de Frames

La extracción de frames consiste en seleccionar imágenes de un video a intervalos específicos, lo que permite obtener una muestra representativa del contenido visual. Este proceso es crucial para análisis posteriores, ya que reduce el volumen de datos y facilita la manipulación de imágenes.

2.3.14. Aplicación de Línea de Comando

La aplicación de línea de comando permite ejecutar scripts y comandos específicos directamente en el sistema operativo. Esta técnica es útil en el procesamiento de datos y facilita la automatización de tareas como la descarga y organización de archivos en proyectos de gran escala.

2.3.15. DataFrame

El *DataFrame* es una estructura de datos en formato tabular, que permite organizar y manejar grandes volúmenes de datos de manera eficiente. Cada columna puede contener diferentes tipos de datos, y esta organización es ideal para el análisis estructurado en proyectos de procesamiento de datos y machine learning.

2.3.16. Concatenación de Videos

La concatenación de videos es el proceso de unir varios segmentos de video en uno solo. Este procedimiento es particularmente útil para crear resúmenes visuales de clips relevantes, permitiendo una presentación continua de los eventos clave sin interrupciones.

Capítulo 3

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. Diseño de la investigación

En esta sección se describirá el diseño de la investigación, el cual incluirá el alcance, enfoque y tipo de estudio que se realizará. Asimismo, se detallarán la población y muestra que serán consideradas en este trabajo.

3.1.1. Alcance de la investigación

El presente estudio tendrá un alcance explicativo, ya que se analizará cómo la aplicación de técnicas avanzadas de procesamiento de imágenes y algoritmos de clustering contribuirá a la automatización en la generación de clips informativos a partir de videos de noticias. Se buscará identificar patrones visuales y explicar cómo estos representan eventos clave en las noticias. El uso de técnicas como HOG para la extracción de características y K-means para el agrupamiento de frames permitirá optimizar la selección de contenido relevante en un entorno con sobrecarga informativa.

3.1.2. Tipo de investigación

Según [Hernández et al., 2021](#), una investigación de tipo experimental se caracteriza por la manipulación y control de variables con el objetivo de observar sus efectos en una o más variables dependientes, permitiendo establecer relaciones de causalidad. Este estudio será

de tipo experimental, ya que se intervendrá en las condiciones de análisis al aplicar técnicas específicas de procesamiento de imágenes y algoritmos de clustering sobre los frames de los videos de noticias. Aunque los datos (videos) son obtenidos de plataformas externas como YouTube, la investigación involucra una manipulación deliberada en el tratamiento de dichos datos a través de métodos de extracción de características y agrupamiento, buscando validar la efectividad de estos procedimientos en la identificación de patrones representativos.

3.1.3. Enfoque de investigación

El enfoque de esta investigación será cuantitativo, dado que se basará en la recolección y análisis de datos numéricos derivados de los videos de noticias procesados. A través de la implementación de técnicas de procesamiento de imágenes y algoritmos de clustering, se buscará medir, agrupar y analizar objetivamente los frames más representativos para la generación automática de clips informativos.

3.2. Población y muestra

Tabla 3.1: Descripción del estudio

Categoría	Descripción
Población	Videos de noticias disponibles en plataformas como YouTube que cubrirán eventos de actualidad a nivel nacional e internacional.
Muestra	<ul style="list-style-type: none"> ■ 40 videos de noticias en formato MP4 que serán extraídos automáticamente de YouTube utilizando la herramienta Youtube-Dlp.
Unidad de análisis	Un video de noticias completo que contendrá información sobre eventos relevantes de un día específico.

3.3. Operacionalización de variables

La operacionalización de variables permite establecer una relación clara entre los conceptos teóricos y los indicadores medibles utilizados en el presente estudio. En la Tabla 3.2 se

detallan las variables independientes y dependientes, sus respectivas dimensiones, indicadores y los cálculos necesarios para su evaluación.

Tabla 3.2: Operacionalización de Variables

Variable	Dimensión	Indicador	Cálculo
Independiente: Algoritmo de Clustering	Datos de entrada	Volumen de datos	Número de videos procesados y total de frames extraídos
	Proceso de agrupamiento	Inertia	Suma de distancias cuadradas de los puntos a su centroide
Dependiente: Creación automática de clips informativos	Indicadores de desempeño	Representatividad de los clips	Análisis de frames clave seleccionados respecto a su proximidad al centroide

3.4. Técnicas de recolección

En esta investigación se emplearán técnicas de recopilación de datos automatizadas que permitirán la obtención eficiente y estructurada de videos de noticias. La herramienta principal que se utilizará será Youtube-DLP, un software de código abierto que facilitará la descarga automatizada de videos en formato MP4 desde plataformas como YouTube. Esta técnica asegurará la recopilación de grandes volúmenes de datos de manera precisa, permitiendo definir criterios específicos como la relevancia del contenido, la fecha de publicación y la duración de los videos, con el objetivo de garantizar que el material recopilado sea pertinente para el análisis de eventos clave. Asimismo, los videos recopilados se almacenarán y organizarán de manera estructurada, incluyendo metadatos como el título y la fuente de los videos, lo que facilitará tanto el acceso como el procesamiento posterior para el análisis automatizado.

3.5. Técnicas para el procesamiento y análisis de la información

3.5.1. Metodología de la implementación de la solución

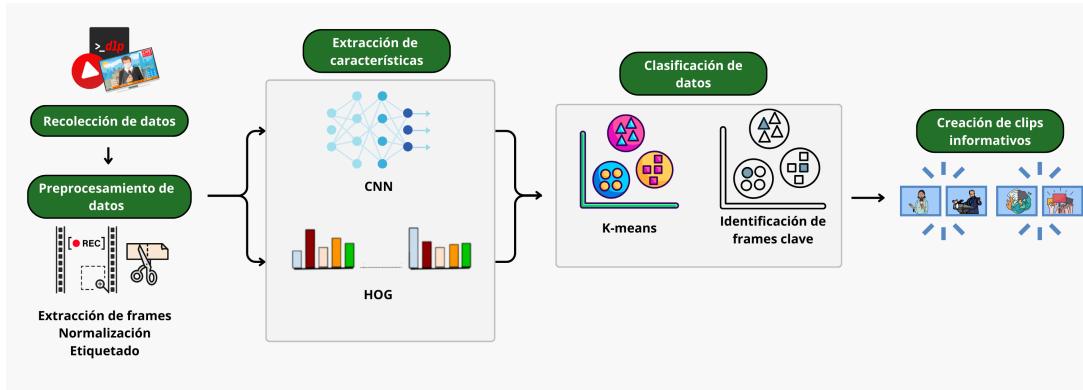


Figura 3.1: Metodología propuesta. Fuente: Elaboración propia

3.5.1.1. Recolección de datos

La recolección de datos constituye un componente esencial para el desarrollo de esta investigación, ya que proporciona el insumo principal para las etapas de análisis y procesamiento. En este proyecto, se recopilarán un total de 40 videos en formato MP4, descargados de manera automatizada desde la plataforma YouTube. Estos videos serán seleccionados específicamente de noticieros, con el objetivo de garantizar que el contenido refleje temas y tendencias de interés informativo.

Para asegurar que los videos recopilados sean pertinentes y cumplan con los objetivos del proyecto, se aplicarán filtros específicos durante el proceso de búsqueda en YouTube. Estos filtros seleccionarán exclusivamente videos publicados el día de la consulta, garantizando que el contenido refleje información actual y relevante. Asimismo, solo se incluirán videos cuya duración sea mayor a 20 minutos, asegurando que los datos recopilados contengan suficiente material audiovisual para el análisis posterior y pluralidad en los temas tocados dentro de las noticias.

Para la descarga de los videos se empleará la herramienta Youtube-DLP, que permite la extracción de contenido y su gestión eficiente a través de comandos personalizados. Esta herramienta no solo asegura la uniformidad en los formatos descargados, sino que también permite automatizar la organización y almacenamiento de los archivos en una estructura predefinida.

3.5.1.2. Preprocesamiento de datos

El preprocesamiento de datos será crucial para preparar los videos y sus frames recopilados antes de aplicar los modelos de análisis y clustering. Las tareas de preprocesamiento incluirán:

- **Extracción de frames:** Se extraerán frames de los videos con el fin de capturar representaciones visuales relevantes de los eventos y elementos clave presentes en las noticias. Los intervalos de captura se establecerán en función de los frames por segundo.
- **Reducción de frames:** Los frames extraídos serán evaluados para eliminar aquellos que sean irrelevantes o redundantes, como transiciones o escenas sin valor informativo, garantizando que el análisis se enfoque en el contenido más significativo.
- **Normalización:** Los frames serán normalizados en términos de resolución y formato, asegurando que los datos visuales sigan un estándar homogéneo, lo que permitirá que los algoritmos de procesamiento funcionen de manera más eficiente.
- **Etiquetado de frames procesados:** Los frames procesados serán almacenados y etiquetados para facilitar su organización y posterior análisis durante la fase de generación de clips informativos.

3.5.1.3. Extracción de Características

La extracción de características será esencial para capturar la información visual relevante de los frames. Se utilizarán dos técnicas complementarias: Histogram of Oriented Gradients (HOG) y Redes Neuronales Convolucionales (CNN), cuyos vectores de características se integrarán mediante un proceso de stacking para obtener una representación única y representativa de cada frame. Dado que estos vectores pueden tener una alta dimensionalidad, se aplicará una técnica de reducción dimensional utilizando PCA, optimizando así el análisis posterior sin comprometer la representatividad de los datos.

- **Extracción con HOG:** El descriptor HOG capturará características locales de los frames, como bordes y contornos, proporcionando una representación detallada de los gradientes.
- **Extracción con CNN:** Simultáneamente, se aplicará una CNN para extraer características más profundas y abstractas de los frames. La arquitectura CNN, basada en ResNet50, generará vectores de características que encapsulan información jerárquica de las imágenes procesadas.

- **Stacking de características:** Los vectores de características obtenidos con HOG y CNN se combinarán mediante un proceso de stacking para formar un único vector por frame. Este vector integrado proporciona una representación robusta que combina información tanto local como abstracta.
- **Reducción dimensional con PCA:** Para abordar la alta dimensionalidad de los vectores combinados, se utilizará Principal Component Analysis (PCA). Este método permitirá reducir las dimensiones a un espacio más manejable, preservando al menos el 95 % de la variabilidad de los datos originales. La aplicación de PCA no solo optimiza el costo computacional de los algoritmos posteriores, como K-means, sino que también minimiza el riesgo de redundancia en los vectores característicos.

3.5.1.4. Clasificación de datos

Una vez obtenidos los vectores de características, se agruparán los frames en función de su similitud para identificar aquellos que representen los eventos más significativos en los videos de noticias. Esta clasificación de datos se llevará a cabo en dos fases principales: la determinación del número óptimo de clústeres mediante el método del codo y la posterior aplicación del algoritmo K-Means para agrupar los frames.

- **Determinación del número óptimo de clusters:** Para seleccionar el valor óptimo de k (número de clusters), se utilizará el método del codo. Este enfoque evalúa la inercia del modelo, que representa la suma de las distancias cuadradas entre los puntos y sus centroides dentro de cada cluster. Al graficar la inercia en función de k , se identificará el punto de inflexión donde añadir más clusters no produce una mejora significativa, determinando así el valor adecuado de k .
- **Agrupamiento con K-Means:** Una vez definido el valor de k , el algoritmo K-Means se aplicará para agrupar los frames en clusters, de acuerdo con las características extraídas mediante HOG y CNN. Cada cluster contendrá frames con patrones visuales similares, lo que permitirá identificar escenas relacionadas dentro de un video.
- **Identificación de frames clave:** Despues del agrupamiento, se seleccionarán los frames más representativos de cada cluster. Estos serán aquellos más cercanos al centroide del cluster, asegurando que las características principales del grupo estén reflejadas en los frames seleccionados. Los frames clave servirán como base para la generación de clips informativos automáticos.

3.5.1.5. Creación de clips informativos

La etapa final de la metodología consistirá en la generación automática de clips informativos a partir de los frames clave previamente seleccionados. El objetivo principal será crear resúmenes visuales concisos y representativos que capturen los eventos más relevantes de los videos de noticias, asegurando una narrativa visual coherente.

- **Selección de secuencias de frames clave:** A partir de los frames clave identificados en cada clúster, se extraerán secuencias de frames adyacentes que mantengan una continuidad temporal. Esta selección garantizará que los momentos más significativos de los videos sean representados de manera eficiente, preservando el contexto de las escenas.
- **Generación de clips informativos:** Las secuencias de frames seleccionadas serán ensambladas para crear clips informativos breves. Este proceso asegurará la fluidez visual y narrativa de cada clip, permitiendo que los eventos destacados sean presentados de forma clara y concisa, maximizando su relevancia informativa.

3.5.2. Metodología para la medición de resultados

El rendimiento de las técnicas de clustering y clasificación será evaluado mediante métricas adecuadas para modelos no supervisados, enfocándose en la identificación de eventos clave en los videos.

Inercia (Within-cluster Sum of Squares)

La inercia medirá la compacidad de los clusters formados, calculando la suma de las distancias cuadradas entre los puntos y sus centroides.

$$\text{Inercia} = \sum_{i=1}^n \sum_{j=1}^k \|x_i - c_j\|^2 \quad (\text{Ecuación 3.1})$$

3.6. Cronograma de Actividades

La representación gráfica de este cronograma en la Figura 3.2 permite visualizar la secuencia y la duración de cada actividad, destacando la interdependencia entre fases clave, y garantizando una gestión eficaz del tiempo a lo largo del proyecto.

Actividades/Tareas	Meses/Semanas																						
	Agosto				Septiembre				Octubre				Noviembre				Diciembre						
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4			
1 Planificación de investigación		1	2	3	4																		
2 Asesoría					1	2	3	4															
3 Revisión y ajustes del marco teórico						1	2	3	4														
4 Desarrollo de metodología							1	2	3	4													
5 Recolección de datos								1	2	3	4												
6 Entrega 1									1	2	3	4											
7 Preprocesamiento de datos										1	2	3	4										
8 Prototipo de modelo de clasificación											1	2	3	4									
9 Desarrollo de métricas de evaluación												1	2	3	4								
10 Asesoría													1	2	3	4							
11 Mejorado del prototipo														1	2	3	4						
12 Documentación de resultados															1	2	3	4					
13 Entrega 2																1	2	3	4				
14 Desarrollo del Informe																	1	2	3	4			
15 Ajuste de parámetros de la herramienta																		1	2	3	4		
16 Asesoría																			1	2	3	4	
17 Desarrollo de presentación																			1	2	3	4	
18 Entrega Final																			1	2	3	4	
19 Sustentación																				1	2	3	4

Figura 3.2: Cronograma de actividades del proyecto.

3.7. Presupuesto

El presupuesto del proyecto, como se detalla en la Tabla 3.3, incluye los recursos materiales, técnicos y de servicios necesarios para su ejecución exitosa. Los costos están distribuidos en diferentes categorías: Hardware, que abarca equipos de cómputo esenciales; Software, centrado en herramientas de código abierto como Python; y Servicios, que incluye los gastos de conectividad a internet y suministro eléctrico. Este desglose permite visualizar los recursos disponibles y asegurar que el proyecto disponga de los medios necesarios para cumplir con sus objetivos de manera eficiente.

Tabla 3.3: Presupuesto del Proyecto

Tipo	Categoría	Recurso	Descripción	Fuente Financiadora	Monto (Soles)
Hardware	Computadoras	Laptop	Computadora portátil con capacidad de procesamiento avanzado	Propio	6,000
Software	Procesamiento	Python	Lenguaje de programación de código abierto	-	-
Servicio	Conectividad	Internet	Conexión a internet de alta velocidad para uso de herramientas y recursos online	Propio	900
Servicio	Conectividad	Electricidad	Suministro de energía eléctrica para los equipos durante el procesamiento de datos	Propio	1,200
Total					8,100

Capítulo 4

DESARROLLO DE LA SOLUCIÓN

Este capítulo describe el desarrollo técnico y metodológico aplicado para llevar a cabo el análisis de los datos en este proyecto. Cada sección detalla los pasos realizados, las herramientas empleadas y los enfoques metodológicos utilizados para lograr una representación visual y analítica de los datos de video obtenidos.

4.1. Equipo utilizado

A continuación se presenta una descripción detallada de las características del equipo de cómputo utilizado en el desarrollo del proyecto. Esto es relevante para replicar el análisis en condiciones similares.

Tabla 4.1: Especificaciones del equipo utilizado

Característica	Especificación
Nombre del dispositivo	LAPTOP-OSCAR
Procesador	AMD Ryzen 7 6800H con Radeon Graphics, 3.20 GHz
RAM instalada	16.0 GB (15.3 GB utilizable)
Tipo de sistema	Sistema operativo de 64 bits, procesador basado en x64
Sistema operativo	Windows 11

4.2. Especificaciones de Python

Para este proyecto se utilizó la versión 3.12.4 de Python, cuyo entorno fue configurado para cumplir con los requisitos de las bibliotecas y módulos necesarios.

Tabla 4.2: Especificaciones de Python utilizado

Característica	Especificación
Versión de Python	3.12.4
IDE utilizado	Visual Studio Code

4.3. Librerías

Las librerías descritas a continuación fueron esenciales para las distintas etapas de procesamiento, análisis y visualización en este proyecto.

Tabla 4.3: Descripción de librerías usadas en el proyecto

Librería	Descripción
Web Scraping y automatización	
time	Proporciona funciones para gestionar el tiempo, como retrasos y mediciones de intervalos, útiles en sincronización de procesos y espera entre solicitudes.
os	Facilita la interacción con el sistema operativo, como gestionar archivos y directorios, útil en automatización.
subprocess	Ejecuta comandos del sistema desde Python, permitiendo interactuar con otros programas o scripts.
psutil	Ofrece una interfaz para monitorear recursos del sistema (CPU, memoria) y gestionar procesos en segundo plano.
selenium	Herramienta para automatización de navegadores, ideal para pruebas web y scraping de sitios dinámicos.
tqdm	Crea barras de progreso en loops, útil para monitorear procesos largos.
Procesamiento de datos	
numpy	Biblioteca fundamental para operaciones matemáticas y estructuras de datos de alto rendimiento, como arreglos multidimensionales.

Librería	Descripción
pandas	Ofrece estructuras de datos como DataFrames, ideales para manipulación y análisis de datos tabulares.
gc	Biblioteca para gestionar y liberar memoria mediante la recolección de basura.
Procesamiento de imágenes	
cv2	OpenCV, biblioteca de visión artificial para procesamiento y análisis de imágenes y video, común en proyectos de visión.
hog	Función en skimage.feature para extracción de características en imágenes usando Histogramas de Gradientes Orientados (HOG).
Redes neuronales y Deep learning	
ResNet50	Arquitectura de red preentrenada, común para extracción de características y clasificación en imágenes. Se importa desde tensorflow.keras.applications.
preprocessing	Módulo en Keras para preprocessamiento de datos e imágenes, útil en preparación de datos para redes neuronales.
models	API de Keras para crear y gestionar modelos, permitiendo el diseño de redes neuronales profundas.
Clustering y Machine learning	
KMeans	Algoritmo de agrupamiento en sklearn.cluster, útil para clasificar datos en grupos similares.
metrics.pairwise	Módulo en sklearn para calcular distancias entre vectores, como pairwise_distances_argmin_min.
PCA	Método de reducción de dimensionalidad en sklearn.decomposition, que simplifica conjuntos de datos complejos.
Matemáticas y utilidades	
math	Biblioteca estándar para operaciones matemáticas básicas, como redondeo y cálculos geométricos.
Visualización	
matplotlib	Biblioteca principal para visualización de datos, permite crear gráficos de varios tipos en Python.

4.4. Extracción de datos

La extracción de datos fue un paso fundamental en este proyecto, ya que permitió obtener información relevante de videos publicados en YouTube para su posterior análisis. El proceso se llevó a cabo en diferentes etapas que abarcaron desde la recopilación de los enlaces de los videos, la descarga de los mismos, hasta la extracción de frames que fueron utilizados para generar descriptores de características mediante técnicas de visión computacional.

4.4.1. Automatización de la búsqueda de videos en YouTube

Para obtener los videos de interés, se realizó una búsqueda automatizada en YouTube utilizando la librería Selenium. El objetivo era filtrar noticias en vivo y limitar la cantidad de videos a descargar. La siguiente secuencia de pasos describe el proceso:

- Se inició una sesión de navegador automatizada con Selenium, controlando el navegador Chrome.
- A través de interacciones programadas, se realizaron búsquedas con palabras clave específicas y se aplicaron filtros para obtener resultados de videos de un único día y de la duración suficiente.
- Los enlaces de los videos obtenidos fueron almacenados en una lista para ser descargados posteriormente.

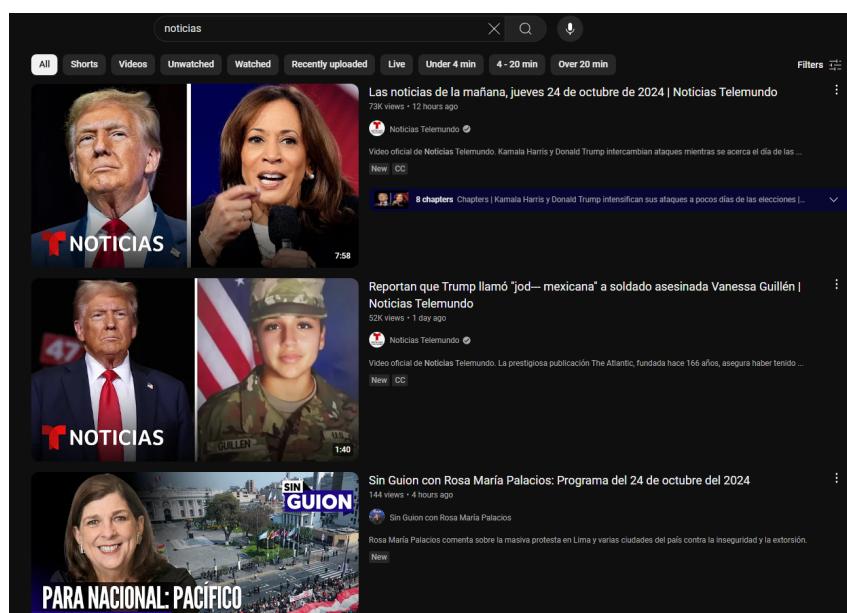


Figura 4.1: Página de partida de scrapping.

Search filters				
UPLOAD DATE	TYPE	DURATION	FEATURES	SORT BY
Last hour	Video X	Under 4 minutes	Live	Relevance
Today X	Channel	4 - 20 minutes	4K	Upload date
This week	Playlist	Over 20 minutes X	HD	View count
This month	Movie		Subtitles/CC	Rating
This year			Creative Commons	
			360°	
			VR180	
			3D	
			HDR	
			Location	
			Purchased	

Figura 4.2: Filtros aplicados a los resultados de video.

En la [Figura C.1](#), se presenta una recopilación de los enlaces de los videos obtenidos mediante el proceso de scraping. Esta tabla muestra información clave como el título, canal, duración y el enlace directo a cada video, los cuales fueron utilizados como parte del análisis en este proyecto.

4.4.2. Descarga de videos

Para la recopilación de datos, se automatizó la descarga de videos de YouTube que cumplieran con los criterios. Para ello, se utilizó la herramienta `yt-dlp` estableciéndole los parámetros de descarga necesarios como el tamaño máximo de archivo y la calidad del video establecida en 720p.

El código de descarga inicializa una lista de enlaces de YouTube y ejecuta la descarga de cada video de forma secuencial, con un intervalo entre descargas para evitar bloqueos por parte de YouTube. A continuación, se detallan las opciones clave usadas con `yt-dlp`:

- `--format "bestvideo[height<=720]"`: Especifica que se descargue la mejor calidad de video disponible con una altura máxima de 720 píxeles.

En la [Figura 4.3](#), se muestra un ejemplo de una instrucción empleada durante el proceso de extracción de videos utilizando la herramienta `youtube-dlp`, configurando parámetros específicos como la calidad del video y el formato.

```
PS C:\Users\edu08> yt-dlp https://www.youtube.com/watch?v=2zQKUqbR5NY --format "bestvideo[height<=720]" --max-filesize 400M
[youtube] Extracting URL: https://www.youtube.com/watch?v=2zQKUqbR5NY
[youtube] 2zQKUqbR5NY: Downloading webpage
[youtube] 2zQKUqbR5NY: Downloading ios player API JSON
[youtube] 2zQKUqbR5NY: Downloading player fb725ac8
[WARNING: [youtube] 2zQKUqbR5NY: nsig extraction failed: You may experience throttling for some formats
n = uAbtHHTmN12bFVAU ; player = https://www.youtube.com/s/player/fb725ac8/player_ias.vflset/en_US/base.js
[WARNING: [youtube] 2zQKUqbR5NY: nsig extraction failed: You may experience throttling for some formats
n = m13z10pf68TZhki ; player = https://www.youtube.com/s/player/fb725ac8/player_ias.vflset/en_US/base.js
[youtube] 2zQKUqbR5NY: Downloading m3u8 information
[info] 2zQKUqbR5NY: Downloading 1 format(s): 609
[hlsnative] Downloading m3u8 manifest
[hlsnative] Total fragments: 4233
[download] Destination: ● #ENVIVO | Todo sobre el PARO DE TRANSPORTISTAS del 23 de octubre | 23/10/2024 [2zQKUqbR5NY].mp4
[download] 0.1% of ~ 2.81GiB at 1.25MiB/s ETA 25:08 (frag 6/4233)
```

Figura 4.3: Instrucción de descarga de video.

4.4.3. Organización de videos

Para organizar los videos descargados, se empleó un script que clasifica los archivos en función de su tipo y ubicación en directorios específicos. Este proceso permite gestionar eficientemente los archivos de video y eliminar los archivos no necesarios generados durante la descarga. El procedimiento se detalla a continuación.

- 1. Creación del directorio de videos:** Inicialmente, se define un directorio específico para almacenar los archivos de video. Si la carpeta de destino no existe, el código se encarga de crearla utilizando el comando `os.makedirs`.
- 2. Clasificación de archivos:** Se definen las extensiones de archivo de video permitidas (`.mp4` y `.part`), las cuales son revisadas en cada archivo del directorio actual. Cualquier archivo con una de estas extensiones se mueve al directorio de videos.
- 3. Etiquetado de archivos:** Cada video movido al directorio es renombrado automáticamente utilizando un esquema de etiquetado consecutivo, como `"video_1"`, `"video_2"`, etc. Este etiquetado facilita la identificación y referencia de los videos durante las etapas posteriores de análisis.
- 4. Eliminación de archivos no necesarios:** Se eliminan del directorio todos aquellos archivos que no cumplen con los criterios establecidos para el procesamiento, asegurando que únicamente los videos relevantes y el archivo principal `Tesis.ipynb` se mantengan.

En la investigación se obtuvieron 9.06 GB de contenido dividido en 39 videos de noticias. La estructura final del directorio de videos se muestra en la [Figura 4.4](#).

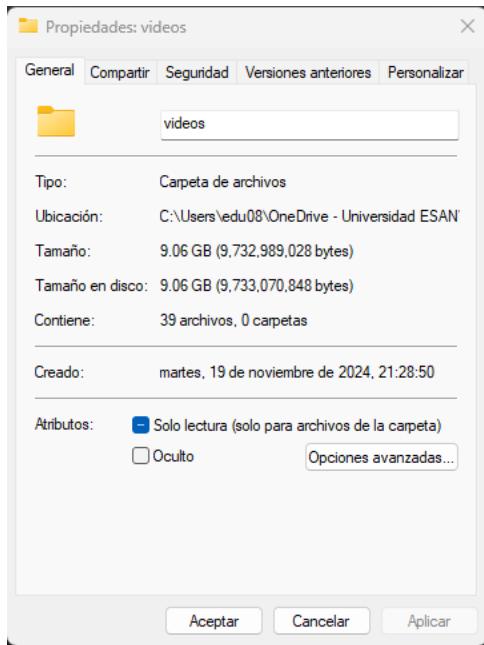


Figura 4.4: Descripción final de carpeta de videos a procesar.

Durante el proceso de descarga, se presentó la corrupción de un elemento. Este archivo, al no poder ser reproducido ni procesado correctamente, fue identificado y descartado durante la etapa de clasificación.

4.5. Preprocesamiento de datos

El preprocesamiento de datos es un paso fundamental en este proyecto y consiste en tres etapas principales: la extracción de frames, el etiquetado de frames y las operaciones de transformación de los frames. Estas etapas permiten obtener representaciones visuales homogéneas y estructuradas, facilitando el análisis posterior mediante técnicas de machine learning.

4.5.1. Extracción de frames

La primera etapa del preprocesamiento es la extracción de frames de los videos. Los videos almacenados en la carpeta de entrada se procesan para obtener un conjunto de imágenes a intervalos de tiempo específicos, lo que permite captar información visual a lo largo de la duración del video. La extracción de frames se realiza a una tasa de 3 frame por segundo, lo cual proporciona una representación balanceada entre el contexto y la carga de procesamiento. Este proceso se logra utilizando la biblioteca OpenCV.

4.5.2. Etiquetado de frames

Cada frame extraído se etiqueta de manera única para mantener una referencia clara de su video de origen y posición temporal. La nomenclatura de los frames es de la forma Video{número}_Frame{número}.png, donde:

- Video{número} hace referencia al identificador del video de origen.
- Frame{número} indica el número de frame dentro del video.

Esta estructura de nombres facilita el seguimiento de cada frame y permite una asociación directa entre la imagen y el video de origen. En etapas posteriores se hará uso de la etiqueta para la identificación de frames clave y la generación de clips.

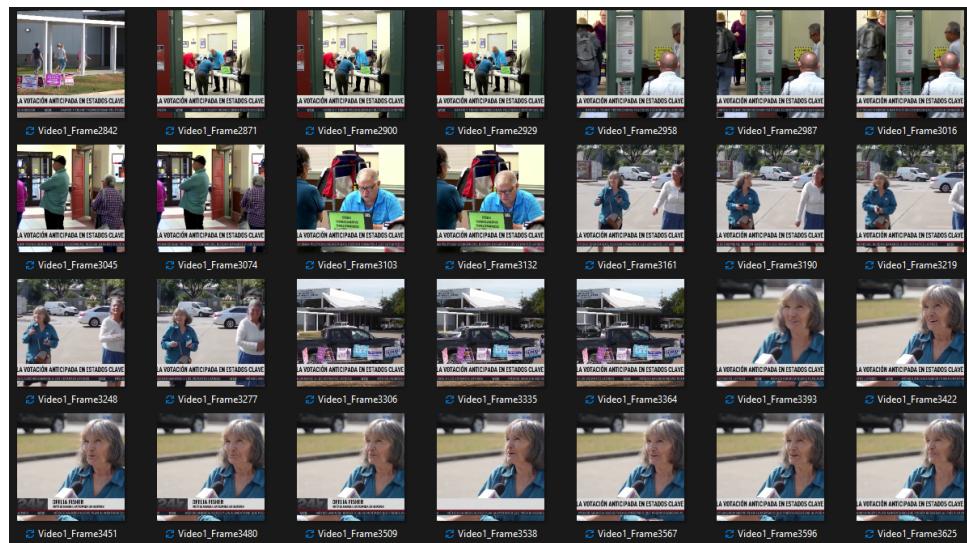


Figura 4.5: Frames etiquetados en la carpeta de frames.

4.5.3. Operaciones sobre los frames

Para garantizar la uniformidad en los datos y preparar los frames para su posterior análisis, cada imagen extraída de los videos pasa por un proceso de redimensionamiento y recorte. Estos ajustes son esenciales para normalizar el tamaño de los frames y preservar la proporción original de la imagen, evitando así posibles distorsiones que puedan afectar los resultados en etapas posteriores del procesamiento.

El tamaño final de cada frame se define en 224x224 píxeles, un formato ampliamente compatible con redes neuronales convolucionales (CNNs) y modelos de aprendizaje profundo en general. Este proceso de preprocesamiento se realiza en dos pasos clave:

- **Redimensión:** Inicialmente, la imagen es escalada para que su dimensión menor (ancho o alto) se ajuste al tamaño objetivo de 224 píxeles. Esto asegura que la imagen completa se mantenga dentro del marco sin deformaciones y preservando la proporción de los elementos visuales originales.
- **Recorte:** Una vez redimensionada, la imagen es recortada para ajustar la dimensión restante a 224 píxeles, logrando así un encuadre preciso sin distorsionar los contenidos importantes.



Figura 4.6: Proceso de redimensión de la imagen.

La Figura 4.6 ilustra la etapa de redimensión, mientras que la Figura 4.9 muestra el proceso de recorte, donde se enmarca la sección central de la imagen para asegurar consistencia en el análisis de los frames.

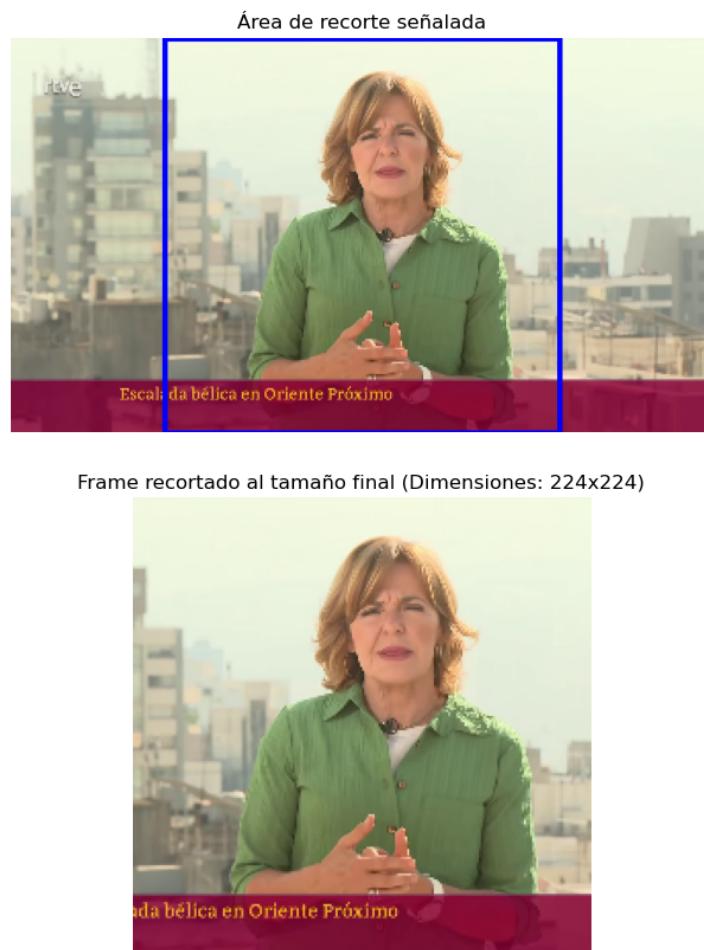


Figura 4.7: Proceso de recorte de la imagen.

4.5.4. Guardado de frames

Finalmente, los frames procesados se almacenan en un directorio de salida específico, asegurando que cada imagen esté etiquetada de forma clara y sea fácilmente localizable para su análisis posterior. En total, se generaron **73,005 frames**, ocupando un espacio de **5.40 GB** en disco. Los frames se guardaron en formato .png, manteniendo una alta calidad de imagen sin compresión excesiva, lo que asegura la precisión necesaria para las etapas de análisis posteriores.

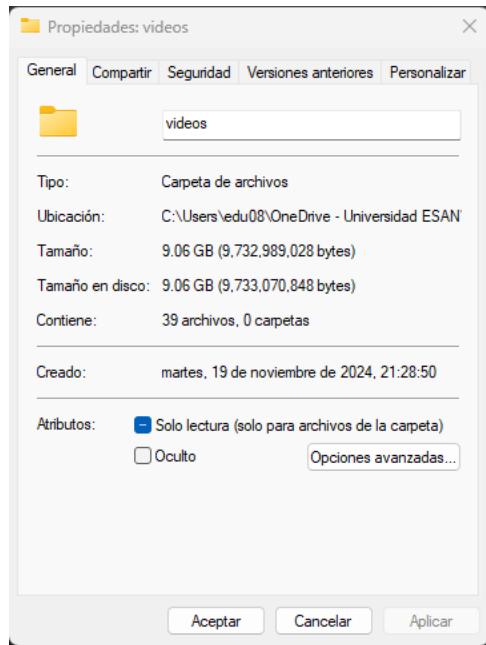


Figura 4.8: Características principales de la carpeta de frames generada.

La estructura final del directorio de frames generados, presentada en la [Figura 4.8](#), es el resultado de este proceso. Cada imagen fue uniformemente etiquetada y organizada, quedando lista para el procedimiento de Extracción de características.

4.5.5. HOG (Histogram of Oriented Gradients)

El método HOG se implementa en este proyecto como una técnica avanzada de extracción de características visuales, transformando cada frame en histogramas de gradientes de orientación. Este enfoque permite identificar contornos y texturas esenciales en los datos de imagen, evitando la necesidad de redundar en descripciones del proceso detallado, el cual ya ha sido expuesto en la sección teórica sobre *extracción de características* (ver sección [2.2.2.1](#)). En esta fase de análisis, HOG contribuye de manera significativa al modelado de patrones visuales, resultando especialmente efectivo en la representación de información estructural crítica para el análisis de patrones y el agrupamiento de datos visuales.

Para la implementación práctica, cada frame es procesado mediante un algoritmo que convierte la imagen a escala de grises, calcula gradientes locales en celdas de **8x8 píxeles** y normaliza los histogramas utilizando el método L2-Hys. Los bloques están configurados en **2x2 celdas**, y el descriptor final de cada imagen contiene **9 orientaciones** por celda, lo que resulta en vectores de alta dimensionalidad. Este proceso asegura que cada frame sea transformado en una estructura numérica que encapsula las características visuales fundamentales de manera eficiente.

La Figura 4.9 ilustra el procedimiento aplicado a un frame del dataset, destacando cómo los gradientes son calculados y organizados para generar un descriptor HOG representativo.



Figura 4.9: Proceso de transformación de imagen utilizando HOG.

4.5.5.1. Reducción de dimensionalidad con PCA

El volumen de datos generado por los descriptores HOG, dada su alta dimensionalidad, presentó un desafío computacional significativo, especialmente al manejar **73,005 frames** en el proyecto. Para abordar este problema y optimizar el almacenamiento y procesamiento, se implementó el método *Principal Component Analysis* (PCA), una técnica estadística de reducción de dimensionalidad. Este enfoque permitió conservar la mayor cantidad de información relevante posible mientras se reducía el tamaño de los vectores de características.

La implementación de PCA en este proyecto se dividió en las siguientes etapas:

- Selección de un conjunto inicial:** Para inicializar el modelo PCA, se seleccionaron los primeros **100 frames** del conjunto de datos. Este lote inicial sirvió como muestra representativa para capturar la estructura general de los datos.
- Ajuste del modelo PCA:** Los vectores HOG de estos frames iniciales fueron utilizados para entrenar el modelo PCA. En este caso, el número de componentes principales fue configurado en **50**, asegurando una reducción significativa de dimensionalidad sin comprometer la calidad de la información retenida.
- Transformación de datos:** Una vez ajustado, el modelo PCA se aplicó a los vectores HOG de todos los frames del conjunto. Esto permitió transformar los descriptores originales de miles de dimensiones a vectores compactos de solo **50 componentes**, lo que facilitó tanto el almacenamiento como el procesamiento en etapas posteriores.

4.5.5.2. Procesamiento por lotes

El manejo de **73,005 frames** en el proyecto representó un desafío significativo debido al alto volumen de datos y los recursos computacionales requeridos. Para garantizar la eficiencia y estabilidad durante la extracción y reducción de características, se implementó un sistema de procesamiento por lotes. Este enfoque permitió gestionar el conjunto de datos de manera incremental, minimizando el consumo de memoria y evitando sobrecargas en el sistema.

1. **Definición del tamaño del lote:** Cada lote fue configurado para incluir un máximo de **1,000 frames**, asegurando un equilibrio entre el uso de memoria y el tiempo de ejecución. Esto resultó en un total de **74 lotes** para procesar todos los frames.
2. **Extracción y reducción de características por lote:** Cada lote fue cargado en memoria, y los vectores HOG correspondientes a los frames fueron generados utilizando la metodología descrita previamente. Estos vectores de alta dimensionalidad fueron inmediatamente transformados mediante el modelo PCA, previamente ajustado, reduciendo su tamaño a **50 componentes principales**.
3. **Almacenamiento incremental:** Los resultados de cada lote fueron exportados directamente a un archivo CSV llamado `hog_feat.csv`, utilizando el modo de escritura incremental (`append`). Este enfoque evitó la acumulación de datos en memoria y permitió mantener el progreso del procesamiento incluso en caso de interrupciones.
4. **Liberación de recursos:** Tras procesar cada lote, se liberaron los recursos asociados, incluyendo los datos temporales y las imágenes cargadas, mediante el uso del recolector de basura con la función `gc.collect()`. Este paso fue esencial para garantizar la eficiencia en el manejo de memoria durante la ejecución prolongada del programa.

La Figura 4.10 presenta una visualización del número total de lotes procesados y el tiempo requerido para completar el análisis mediante HOG y la reducción de dimensionalidad con PCA.

```
Inicializando PCA...
Procesando y guardando lotes con PCA: 100%|██████████| 74/74 [26:57<00:00, 21.86s/it]
Procesamiento completo. Resultados guardados en 'hog_feat.csv'
```

Figura 4.10: Cantidad de lotes y tiempo utilizado para HOG.

Los descriptores HOG obtenidos se integran en un DataFrame, en el cual cada frame está asociado con su correspondiente vector de características reducidas mediante PCA. Esta

estructura tabular, mostrada en la Figura 4.11, no solo facilita la organización y accesibilidad de los datos, sino que también optimiza significativamente el almacenamiento y el procesamiento.

hog_features	
0	[0.026657101656209183, 0.0, 0.0077347234922594...]
1	[0.019755457171115317, 0.0, 0.0091349473256150...]
2	[0.3045512257884793, 0.0, 0.0, 0.1361431886203...]
3	[0.1060905630263379, 0.0, 0.0, 0.0, 0.44411738...]
4	[0.05177980937663794, 0.23095137295797685, 0.0...]
...	...
10768	[0.0030156632191483943, 0.0, 0.004268331090326...]
10769	[0.0015685775208479217, 0.0, 0.0, 0.0130828519...]
10770	[0.02925901987584034, 0.0, 0.00487613239748424...]
10771	[0.02453043706578999, 0.0, 0.00231275203540140...]
10772	[0.0017905083147485153, 0.0, 0.0, 0.0, 0....]

Figura 4.11: Estructura del DataFrame con los vectores de características HOG generados.

4.5.6. Extracción de características con CNN

En esta etapa del proyecto, se implementó una red neuronal convolucional (CNN) pre-entrenada para extraer características visuales de alto nivel a partir de los frames procesados. Se utilizó el modelo ResNet50, entrenado previamente en el conjunto de datos *ImageNet*, configurado como un extractor de características mediante la eliminación de su capa final de clasificación. Esto permitió generar vectores que encapsulan información discriminativa sobre formas, texturas y estructuras complejas presentes en las imágenes.

La arquitectura de ResNet50 está compuesta por 50 capas que incluyen bloques residuales, diseñados para mejorar el aprendizaje profundo y evitar problemas de desvanecimiento del gradiente. En este proyecto, la capa avg_pool, ubicada al final de la red antes de la clasificación, fue seleccionada para generar los vectores de características. Esta capa produce un vector de **2,048 dimensiones** que resume la información más relevante de cada imagen procesada.

4.5.6.1. Reducción de dimensionalidad con PCA y procesamiento por lotes

El proceso de reducción de dimensionalidad y manejo por lotes en la extracción de características con CNN sigue una secuencia similar al descrito en el apartado de HOG (ver sección 4.5.5.1). Ambos métodos emplean el modelo PCA para reducir vectores de alta dimensionalidad y un enfoque de procesamiento por lotes para optimizar el manejo de grandes volúmenes de datos. Sin embargo, dado que las características extraídas con CNN tienen una naturaleza distinta, algunas etapas del proceso presentan diferencias específicas.

Una vez generado el vector de características de cada imagen, este fue transformado mediante PCA, reduciendo su tamaño de **2,048 dimensiones a 50 componentes principales**. La implementación de PCA siguió el mismo esquema previamente establecido, pero con vectores de mayor dimensionalidad, lo que aumentó el costo computacional durante su inicialización y transformación.

El flujo de trabajo en los lotes mantuvo su estructura general, pero adaptado a las particularidades del modelo CNN. La salida de cada lote fue almacenada incrementalmente en un archivo `cnn_feat.csv`, asegurando que los datos estuvieran disponibles para su análisis posterior sin requerir un almacenamiento excesivo en memoria.

La Figura 4.12 muestra una vista de los datos generados por ResNet50 a cada imagen dentro del dataset.

		cnn_features
0		[1.678071, 0.0, 0.14191155, 0.0, 0.0, 0.012353...
1		[1.696687, 0.0, 0.1509949, 0.0, 0.0, 0.0068190...
2		[0.1450273, 0.25749272, 0.44002354, 0.8941569,...
3		[0.29323852, 0.07147341, 0.3167117, 0.48115057...
4		[0.27542925, 0.1919094, 0.17766139, 0.8014795,...
...		...
10768		[0.16416429, 1.7469567, 0.21899797, 2.601019, ...
10769		[0.27553532, 0.8345362, 0.0, 2.0761902, 0.2544...
10770		[0.2632856, 1.8207201, 0.014081388, 1.8049905,...
10771		[0.43951535, 0.38442245, 0.0, 0.072890356, 0.0...
10772		[0.301641, 0.4499437, 0.0, 0.20329915, 0.12526...

Figura 4.12: Dataframe generado por la red convolucional entrenada.

4.5.7. Concatenación de características

Luego de generados los vectores de características de ambas técnicas, se integran en un único vector mediante un proceso de concatenación. Esta integración se realiza alineando ambos conjuntos de características de acuerdo con el identificador del frame y concatenando los vectores en una nueva columna denominada `combined_features` dentro del DataFrame final, generando así una representación robusta de cada frame.

La Figura 4.13 ilustra gráficamente este procedimiento.

	frame_name	hog_features	cnn_features	combined_features
0	Video10_Frame0.png	[0.026657101656209183, 0.0, 0.0077347234922594...]	[1.678071, 0.0, 0.14191155, 0.0, 0.0, 0.012353...	[0.026657101656209183, 0.0, 0.0077347234922594...]
1	Video10_Frame100.png	[0.019755457171115317, 0.0, 0.0091349473256150...]	[1.696687, 0.0, 0.1509949, 0.0, 0.0, 0.0068190...	[0.019755457171115317, 0.0, 0.0091349473256150...]
2	Video10_Frame1000.png	[0.3045512257884793, 0.0, 0.0, 0.1361431886203...	[0.1450273, 0.25749272, 0.44002354, 0.8941569...	[0.3045512257884793, 0.0, 0.0, 0.1361431886203...
3	Video10_Frame1025.png	[0.1060905630263379, 0.0, 0.0, 0.0, 0.44411738...	[0.29323852, 0.07147341, 0.3167117, 0.48115057...	[0.1060905630263379, 0.0, 0.0, 0.0, 0.44411738...
4	Video10_Frame1050.png	[0.05177980937663794, 0.23095137295797685, 0.0...	[0.27542925, 0.1919094, 0.17766139, 0.8014795...	[0.05177980937663794, 0.23095137295797685, 0.0...

Figura 4.13: Concatenación de los vectores de características.

La estructura del DataFrame resultante, `merged_df`, incluye las siguientes columnas, junto con las dimensiones reducidas de los vectores generados mediante PCA:

- **frame_name:** Identificador único del frame correspondiente.
- **hog_features:** Vector de características extraído mediante HOG, reducido a **50 dimensiones** mediante PCA.
- **cnn_features:** Vector de características extraído mediante CNN (ResNet50), reducido a **50 dimensiones** mediante PCA.
- **combined_features:** Vector unificado que integra las características de HOG y CNN, con una dimensión total de **100 dimensiones**, resultante de la concatenación de ambos vectores.

4.6. Clasificación de datos

Para organizar y analizar los datos obtenidos mediante extracción de características, se aplica un proceso de reducción de dimensionalidad seguido de un algoritmo de clustering. A continuación, se detallan los pasos para la clasificación y organización de datos.

4.6.1. Método del codo

El método del codo se empleó para determinar el número óptimo de clusters (k) al aplicar el algoritmo de clustering sobre los vectores concatenados de características `combined_features`. Este método evalúa la inercia, definida como la suma de las distancias cuadradas desde cada punto hasta su centroide, variando k dentro de un rango predefinido. En este caso, los valores de k analizados oscilaron entre 1 y 15.

La inercia mostró un comportamiento decreciente a medida que se incrementaba k , lo cual es característico de este método. Sin embargo, la identificación del codo en el gráfico no

fue sencilla, dado que no hubo un punto de estabilización abrupto. A partir del análisis visual de la Figura 4.14, se identificaron los valores $k = 6$ y $k = 10$ como candidatos viables, ya que representan puntos donde la tasa de disminución de la inercia comienza a estabilizarse.

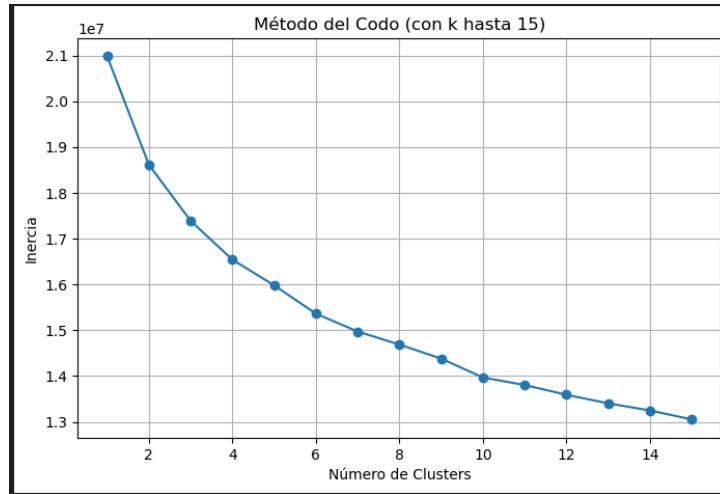


Figura 4.14: Representación de la Inercia utilizando el método del codo.

4.6.2. Aplicación de K-means

Con los valores $k = 6$ y $k = 10$ identificados como candidatos viables a partir del método del codo, se aplicó el algoritmo de K-means al conjunto de datos que contiene los vectores concatenados de características combinadas (`combined_features`). Este proceso permitió agrupar los frames en clústeres específicos basados en patrones de similitud en sus características visuales, proporcionando una organización estructurada de los datos.

El algoritmo asignó una etiqueta de clúster a cada frame en el conjunto de datos, indicando el grupo al cual pertenece según su cercanía al centroide correspondiente. Estas etiquetas fueron integradas como una nueva columna en el `DataFrame` original, facilitando el análisis y la categorización de los frames dentro de cada clúster. Además, se calculó la distancia de cada frame al centroide de su clúster, lo cual permitió determinar el grado de representatividad de cada frame dentro de su grupo.

	frame_name	hog_features	cnn_features	combined_features	cluster_label	distance_to_centroid
0	Video10_Frame0.png	[0.026657101656209183, 0.0, 0.077347234922594...]	[1.678071, 0.0, 0.14191155, 0.0, 0.0, 0.012353...]	[0.026657101656209183, 0.0, 0.077347234922594...]	4	35.950113
1	Video10_Frame100.png	[0.019755457171115317, 0.0, 0.0091349473256150...]	[1.696687, 0.0, 0.1509949, 0.0, 0.0, 0.0068190...]	[0.019755457171115317, 0.0, 0.0091349473256150...]	4	35.724369
2	Video10_Frame1000.png	[0.3045512257884793, 0.0, 0.0, 0.1361431886203...]	[0.1450273, 0.25749272, 0.44002354, 0.8941569,...]	[0.3045512257884793, 0.0, 0.0, 0.1361431886203...]	3	29.857784
3	Video10_Frame1025.png	[0.1060905630263379, 0.0, 0.0, 0.44411738...]	[0.29323852, 0.07147341, 0.3167117, 0.48115057...]	[0.1060905630263379, 0.0, 0.0, 0.44411738...]	3	31.120892
4	Video10_Frame1050.png	[0.05177980937663794, 0.23095137295797685, 0.0...]	[0.27542925, 0.1919094, 0.17766139, 0.8014795...]	[0.05177980937663794, 0.23095137295797685, 0.0...]	3	30.131808
...
10768	Video9_Frame9860.png	[0.0030156632191483943, 0.0, 0.004268331090326...]	[0.16416429, 1.7469567, 0.21899797, 2.601019, ...]	[0.0030156632191483943, 0.0, 0.004268331090326...]	3	27.635337
10769	Video9_Frame9889.png	[0.0015685775208479217, 0.0, 0.0, 0.0130828519...]	[0.27553532, 0.8345362, 0.20761902, 0.2544...]	[0.0015685775208479217, 0.0, 0.0, 0.0130828519...]	3	28.308464
10770	Video9_Frame9918.png	[0.02925901987584034, 0.0, 0.00487613239748424...]	[0.2632856, 1.8207201, 0.014081388, 1.8049905...]	[0.02925901987584034, 0.0, 0.00487613239748424...]	3	29.617391
10771	Video9_Frame9947.png	[0.02453043706578999, 0.0, 0.00231275203540140...]	[0.43951535, 0.38442245, 0.0, 0.072890356, 0.0...]	[0.02453043706578999, 0.0, 0.00231275203540140...]	2	26.385246
10772	Video9_Frame9976.png	[0.0017905083147485153, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...]	[0.301641, 0.4499437, 0.0, 0.20329915, 0.12526...]	[0.0017905083147485153, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...]	2	26.610295

Figura 4.15: Asignación de clusters y distancias al centroide correspondiente.

4.6.3. Exploración de clústeres

Para complementar el análisis de los clústeres generados por el algoritmo K-means, se realizó una exploración visual de los frames más representativos de cada grupo. Este proceso permite examinar de manera detallada las características predominantes dentro de un clúster específico, proporcionando una interpretación visual que facilita la validación y comprensión de los resultados.

El procedimiento se llevó a cabo seleccionando un clúster de interés y visualizando los frames más cercanos a su centroide. Para cada frame, se calculó su distancia al centroide correspondiente, y se ordenaron en función de esta métrica. Los frames más cercanos se consideran representativos de las características centrales del clúster.

La Figura 4.16 muestra un ejemplo de la visualización de **10 frames más representativos** para un clúster específico, permitiendo identificar sus características predominantes. Asimismo, la Figura 4.17 presenta una representación adicional para otro clúster seleccionado.



Figura 4.16: Frames más cercanos al centroide en un clúster específico.



Figura 4.17: Frames más cercanos al centroide en un clúster específico.

La Figura 4.16 muestra una representación clara y diversa de **10 frames más cercanos al centroide**, evidenciando patrones visuales variados que reflejan una mayor heterogeneidad del clúster seleccionado en comparación con la Figura 4.17. Esta última presenta una representación limitada, donde los frames provienen de una única fuente, lo que podría indicar un agrupamiento sesgado o poco representativo de las características generales del clúster.

4.6.4. Selección de frames clave

Para cada cluster, identificamos los frames más representativos, es decir, aquellos más cercanos a los centroides, y los almacenamos en la lista `frames_clave`.

Los frames seleccionados como representativos de cada cluster se muestran en una única imagen combinada. Esto proporciona una vista consolidada de los elementos clave de cada cluster.

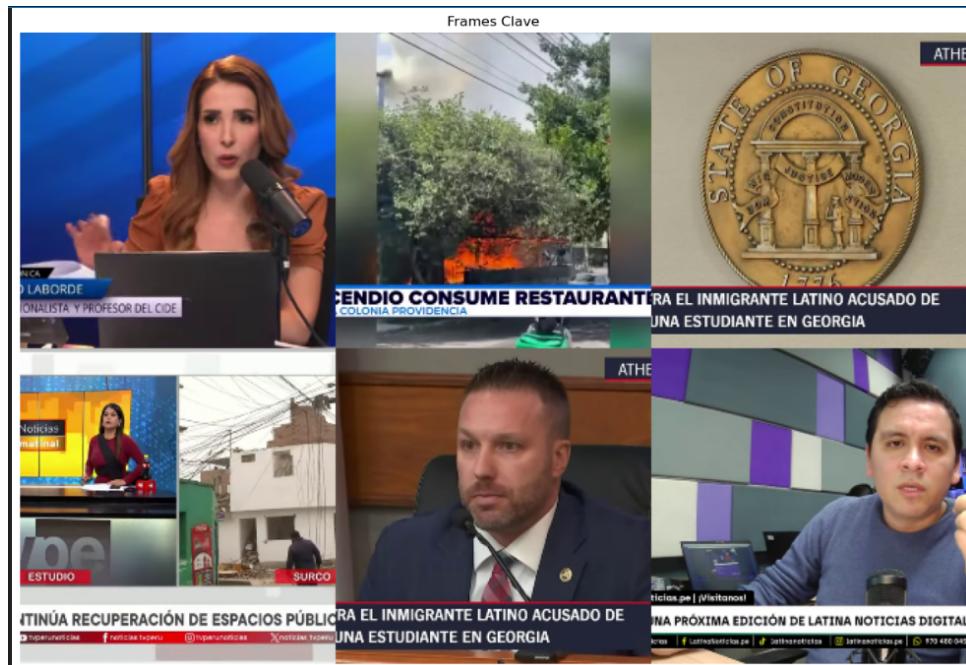


Figura 4.18: Frames representativos en los clusters con $k = 6$.



Figura 4.19: Frames representativos en los clusters con $k = 10$.

La Figura 4.18 ofrece una representación de los elementos clave para $k = 6$, mostrando una agrupación más general. Por otro lado, la Figura 4.19 detalla los frames clave para $k = 10$.

4.7. Generación de clips

En esta sección se detalla el proceso de generación de clips de video basados en frames clave seleccionados en el análisis de clusters. Este proceso permite crear un video final que resalta los eventos representativos, proporcionando un contexto visual adicional para cada frame clave.

4.7.1. Descripción del proceso

El proceso de generación de clips se lleva a cabo en tres etapas principales:

- **Identificación de frame clave:** Cada frame clave se asocia con un video específico y se establece un rango de frames adyacentes a ese frame clave, seleccionado en base a la distancia temporal del evento destacado.
- **Selección de frames adyacentes:** Se seleccionan un conjunto de frames anteriores y posteriores al frame clave. Esto permite capturar el contexto visual necesario para proporcionar una vista completa del evento.
- **Combinación de clips:** Los clips individuales, correspondientes a cada frame clave y sus frames adyacentes, se concatenan en un solo video continuo que representa una síntesis de los eventos clave.

4.7.2. Visualización de Resultados

La Figura 4.21 muestra el clip generado a partir de frames clave y sus frames adyacentes, destacando la coherencia visual que resulta de esta selección.



Figura 4.20: Muestra de frames clave y sus frames adyacentes.

La Tabla 4.4 presenta un resumen de las especificaciones del video generado.

Tabla 4.4: Especificaciones del video generado a partir de frames Clave

Atributo	Valor
Nombre del Video	combined_subclip.mp4
Resolución	854x480
Duración (s)	33.22
FPS	29.97

La Figura 4.21 muestra una selección de frames extraídos del clip generado durante el proceso de análisis automatizado. Esta muestra ilustra la coherencia visual y narrativa lograda mediante la selección de frames clave, los cuales representan los eventos más significativos del contenido original.

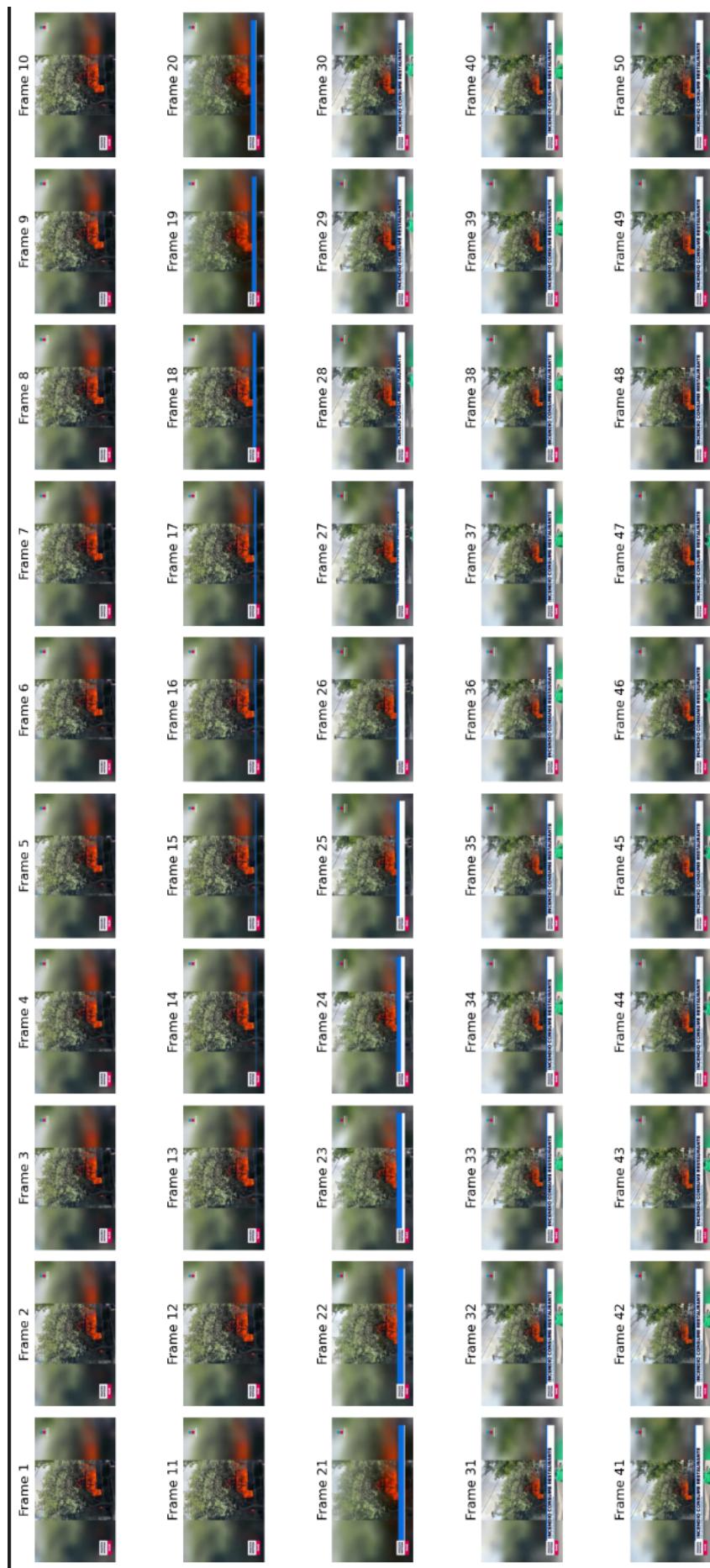


Figura 4.21: Muestra de frames del clip generado.

Capítulo 5

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

A partir de la solución propuesta y los resultados obtenidos, se concluye que es posible desarrollar una herramienta eficiente para la generación de clips informativos sobre tendencias en videos de noticias, cumpliendo con el objetivo general de la investigación. El trabajo realizado demuestra que técnicas avanzadas de extracción de características y algoritmos de *clustering* son capaces de procesar grandes volúmenes de datos audiovisuales y sintetizar información relevante de manera efectiva.

El proceso de recolección de videos, diseñado para garantizar que el contenido refleje eventos actuales y significativos, permitió asegurar la calidad y representatividad del material analizado. Este aspecto resulta esencial, ya que constituye la base para todas las etapas posteriores, desde la extracción de frames hasta la selección de clips clave.

La implementación de descriptores locales y globales, como HOG y CNN, proporcionó una representación robusta del contenido visual. Su combinación a través de concatenación y la reducción dimensional mediante PCA no solo optimizó el costo computacional, sino que también preservó las características esenciales de los datos. Este enfoque resultó clave para abordar el desafío de manejar información de alta dimensionalidad sin perder precisión en el análisis.

La aplicación del algoritmo K-means, con la determinación del número óptimo de clústeres a través del método del codo, permitió agrupar los frames de acuerdo con patrones visuales predominantes. Los resultados obtenidos evidenciaron niveles de granularidad en la

organización de los datos, lo que facilitó la identificación de tendencias relevantes dentro de los videos analizados. Además, la selección de frames clave con base en su proximidad al centroide de cada clúster probó ser una estrategia eficaz para sintetizar información visual significativa, resaltando los eventos más representativos de cada agrupación.

Sin embargo, se identificaron algunos casos de agrupaciones homogéneas, en los que los frames seleccionados provenían principalmente de una única fuente. Esto pone de manifiesto la necesidad de explorar técnicas adicionales que permitan diversificar aún más las agrupaciones y mejorar la representatividad de los clips generados.

Finalmente, la herramienta desarrollada logró cumplir con los objetivos principales del proyecto, demostrando su utilidad como una solución efectiva para la generación automatizada de clips informativos. Este trabajo contribuye al campo de la inteligencia artificial aplicada a la organización de contenido audiovisual, mostrando el potencial de estas técnicas para facilitar el análisis y consumo de grandes volúmenes de información multimedia.

5.2. Recomendaciones

A partir de las observaciones realizadas durante el desarrollo de este trabajo, se plantean recomendaciones que buscan mejorar los resultados obtenidos y explorar nuevas oportunidades de investigación. Estas propuestas están orientadas a optimizar el desempeño de la herramienta y ampliar su aplicabilidad en diferentes contextos.

A futuro, se podría explorar la integración de herramientas de transcripción automática del audio en los videos de noticias, transformando lo que se informa verbalmente en texto procesable. Este enfoque permitiría complementar los datos visuales con una capa adicional de información temática, capturando el contenido informativo de las noticias directamente desde el discurso de los presentadores o narradores. Al incorporar esta transcripción en el análisis, sería posible clasificar los frames no solo por sus características visuales, sino también por su contexto narrativo, mejorando significativamente la precisión del agrupamiento. Además, la transcripción podría ser procesada mediante técnicas de análisis de texto, como la extracción de palabras clave o el modelado de temas, para enriquecer aún más la representatividad de los clips generados.

Por otro lado, se recomienda implementar técnicas avanzadas de segmentación que permitan omitir elementos redundantes en los frames seleccionados, como los presentadores de noticias. Si bien estos elementos son comunes en los videos analizados, no siempre contribuyen significativamente a la representatividad del contenido del clúster, ya que suelen ocupar

áreas centrales del frame sin aportar información temática clave. La segmentación podría enfocarse en resaltar los elementos visuales de mayor relevancia, como gráficos, imágenes de eventos o locaciones.

Adicionalmente, la lectura automatizada de banners informativos presentes en los videos de noticias sería una herramienta poderosa para capturar texto clave que complementa la información visual. Estos banners suelen contener resúmenes o titulares del contenido noticioso, lo que permitiría clasificar los frames con mayor precisión y enriquecer los datos disponibles para el análisis. Este enfoque no solo optimizaría la selección de frames clave, sino que también incrementaría la calidad y relevancia de los clips generados, asegurando que representen de manera fiel el contenido informativo de los videos.

Otra recomendación importante es la implementación de un sistema más robusto para la validación de los resultados. Esto incluye no solo la evaluación cuantitativa de la calidad de los clústeres mediante métricas como el *silhouette score*, sino también la incorporación de validaciones cualitativas con expertos del dominio, quienes podrían aportar perspectivas sobre la coherencia y utilidad de los clips generados en aplicaciones prácticas.

BIBLIOGRAFÍA

- Bhardwaj, S., Srinivasan, M., & Khapra, M. M. (2019). Efficient Video Classification Using Fewer Frames. *arXiv preprint arXiv:1902.10640*.
- Bhattarai, B., Subedi, R., Gaire, R. R., Vazquez, E., & Stoyanov, D. (2023). Histogram of Oriented Gradients meet deep learning: A novel multi-task deep network for 2D surgical image semantic segmentation. *Medical Image Analysis*, 85, 102747. <https://doi.org/10.1016/j.media.2023.102747>
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Science+Business Media.
- Bishop, C. M. (2007). *Pattern Recognition and Machine Learning*. Springer.
- Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 886-893.
- DataReportal. (2024). Top Websites: Similarweb ranking Peru. <https://datareportal.com/reports/digital-2024-peru?rq=Peru>
- Gonzalez, R. C., & Woods, R. E. (2021). *Digital Image Processing* (4th). Pearson.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* [<http://www.deeplearningbook.org>]. MIT Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
- Hernández, R., Fernández, C., & Baptista, P. (2021). *Metodología de la investigación* (7th Edition). McGraw-Hill Education.
- Humaira, H., & Rasyidah, R. (2018). Determining The Appropriate Cluster Number Using Elbow Method for K-Means Algorithm. *WMA-2 2018*.
- Jolliffe, I. T., & Cadima, J. (2016). *Principal Component Analysis* (Third Edition). Springer.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521, 436-444.

- Liang, B., Li, N., He, Z., Wang, Z., Fu, Y., & Lu, T. (2021). News Video Summarization Combining SURF and Color Histogram Features. *Entropy*, 23(8), 982. <https://doi.org/10.3390/e23080982>
- Mukherjee, S. (2022). The Annotated ResNet-50 [Accessed: 2024-10-27]. *Towards Data Science*. <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758>
- PewResearchCenter. (2023). Social media sites by portion of users who regularly get news there. <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>
- Pratt, W. K. (2007). *Digital Image Processing: PIKS Scientific Inside* (4th). Wiley.
- Raza, A., Younas, F., Siddiqui, H. U. R., Rustam, F., Villar, M. G., Alvarado, E. S., & Ashraf, I. (2024). An improved deep convolutional neural network-based YouTube video classification using textual features. *Heliyon*, 10(16), e35812.
- Rozsa, B., & Mocofan, M. (2022). TV News Database Indexing System with Video Structure Analysis, Representative Images Extractions and OCR for News Titles. *2022 International Symposium on Electronics and Telecommunications (ISETC)*, 1-6.
- Scherer, D., Müller, A., & Behnke, S. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. *International Conference on Artificial Neural Networks*, 92-101.
- Smith, A. (2012). *Introduction to Video and Image Processing*. Cambridge University Press.
- Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE*, 105(12), 2295-2329.
- Tekalp, A. M. (2015). *Digital Video Processing*. Prentice Hall.
- Turk, M., & Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1), 71-86.
- Wang, X., & Jiang, J. (2023). Shear Detection and Key Frame Extraction of Sports Video Based on Machine Learning. *2023 Asia-Europe Conference on Electronics, Data Processing and Informatics (ACEDPI)*, 1-6. <https://doi.org/10.1109/ACEDPI58926.2023.00014>
- Wang, Z., & Chen, Y. (2023). Research on Key Frame Extraction of Digital Image Based on Unsupervised Clustering Algorithm. *2023 International Conference on Telecommunications, Electronics and Informatics (ICTEI)*, 1-6. <https://doi.org/10.1109/ICTEI60496.2023.00125>

Anexos

Anexos A

Anexo I: Matriz de Consistencia

	Problemas	Objetivos	Hipótesis	Variables	Indicadores	Índice
General	¿De qué manera el desarrollo de una herramienta inteligente basada en técnicas de clustering de frames puede facilitar la identificación automática de tendencias a partir de videos de noticias para crear clips de videos representativos?	Desarrollar una herramienta inteligente que identifique de manera automática tendencias en videos de noticias mediante el clustering de frames representativos y genere clips de videos que resuman los eventos clave.	Mediante el uso de técnicas de modelos de aprendizaje automático, se logrará mejorar la identificación de tendencias y la generación de clips informativos representativos de videos de noticias.	X · Procesamiento de automatizado de videos	Tiempo promedio de procesamiento de videos	Tiempo de procesamiento en minutos por hora de video.
	¿Cómo obtener de manera eficiente los videos de noticias de un día en específico, asegurando que sean representativos de los eventos más importantes del día?	Desarrollar un método eficiente para la recolección de videos de noticias, asegurando que estos reflejen los eventos más importantes del día.	La recolección automatizada de videos de noticias de un día específico permitirá capturar de manera eficiente los contenidos más representativos para el análisis de tendencias.	Y · Calidad de los clips informativos generados	Porcentaje de clips que reflejan adecuadamente los eventos importantes	Porcentaje de clips aceptados en una evaluación de calidad.
Especificos	¿Qué técnicas de preprocesamiento de imágenes y normalización son más efectivas para mejorar la calidad y relevancia de los frames extraídos de los videos?	Implementar técnicas de preprocesamiento de imágenes y normalización que mejoren la calidad de los frames extraídos, optimizando su relevancia para el análisis.	El uso de técnicas avanzadas de preprocesamiento y normalización de imágenes mejorará significativamente la calidad de los frames, facilitando su clasificación y análisis posterior.	X1 · Captura eficiente de videos de noticieros	Número de videos capturados relevantes capturados por día	Número de videos relevantes capturados
	¿Qué técnicas de extracción de características permiten representar adecuadamente los frames para identificar patrones visuales relevantes?	Desarrollar técnicas de extracción de características que permitan representar adecuadamente los frames, facilitando la identificación de patrones visuales relevantes.	La extracción de características adecuadas de los frames permitirá identificar correctamente patrones visuales que reflejen temas importantes y eventos clave.	Y1 · Representatividad de los videos clave	Porcentaje de videos que cubren eventos clave.	Grado de representatividad en una escala.
	¿Cuál es el algoritmo de clustering más adecuado para agrupar los frames en base a las similitudes de las características extraídas que posean entre ellas?	Seleccionar e implementar un algoritmo de clustering que agrupe los frames basándose en la similitud de características, garantizando que los grupos formados representen tendencias claras.	El uso de un algoritmo de clustering optimizado permitirá agrupar los frames con precisión en función de sus similitudes, identificando adecuadamente las tendencias en los noticieros.	X2 · Algoritmos de clustering	Porcentaje de frames preprocesados correctamente.	Porcentaje de reducción de ruido y mejora de calidad
	¿Cómo generar clips informativos concisos, asegurando que reflejen los eventos clave discutidos en los noticieros?	Desarrollar un sistema automatizado para generar clips informativos concisos que resuman los eventos clave identificados en los videos de noticieros.	La generación automatizada de clips informativos basados en las tendencias identificadas ofrecerá resúmenes concisos y efectivos de los eventos más relevantes del día.	X5 · Técnicas automáticas para generar clips	Número de clips generados por día	Tiempo promedio de generación de clips
				Y5 · Calidad y consistencia	Porcentaje de clips que resumen correctamente los eventos clave	Grado de consistencia de los clips en una escala.

Anexos B

Anexo II: Árbol de problemas

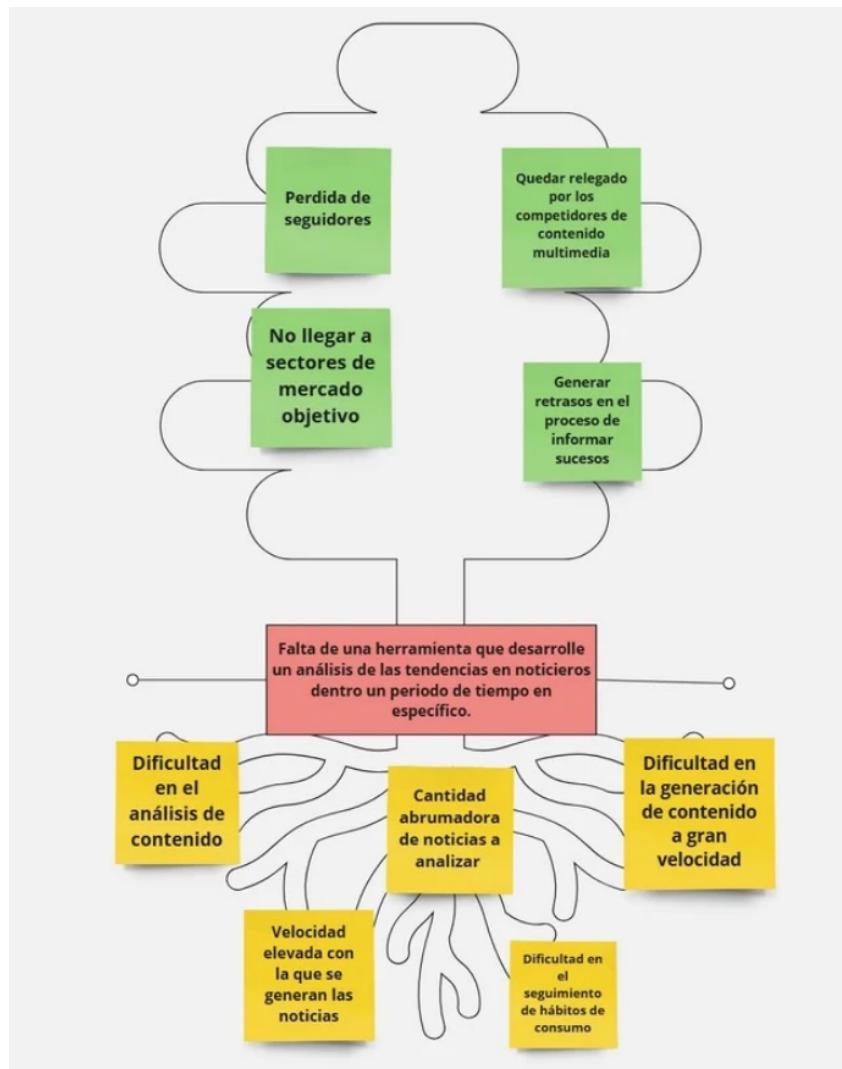


Figura B.1: Árbol de problemas. Fuente: Elaboración propia

Anexos C

Anexo III: Links de videos extraídos

Video ID	Título del video	Canal	Duración
1nI2SGtJbuQ	CASO SHEYLA CÓNDOR: HALLAN MUERTO A POLICÍA ACUSADO DE ASESINATO	Latina Noticias	00:41:40
TbPAIzwlqiU	LATINA NOTICIAS: EDICIÓN MATINAL - MARTES 19 DE NOVIEMBRE 2024	Latina Noticias	04:45:12
FCzsoCU7upE	DW Noticias 18 de noviembre: La guerra en Ucrania rompe el consenso del G2 DW Español	DW Español	00:27:16
TKYjSy5FYOO	EN VIVO 19/11/2024 #PrimeraEmisiónSIN	Noticias SIN	00:47:06
wsAmaa-essM	TVPerú Noticias EN VIVO: Edición Matinal, hoy martes 19 de noviembre del 2024	TVPerú Noticias	05:53:20
j51dClu-EN4	FEMINICIDIO DE SHEYLA CÓNDOR: SALEN A LA LUZ NUEVAS IMÁGENES SOBRE EL MAC	Latina Noticias	00:48:51
eo9EAn87fEg	EN VIVO Teleantioquia Noticias de la 1:00 p.m. 19 de noviembre de 2024 Teleantioquia Noticias	Teleantioquia Noticias	00:57:43
ckPrvDDbmqA	Noticias del 18 de noviembre: TRUMP CONFIRMA PLAN DE DEPORTACIONES	M Diario Gestión	00:29:41
O-GtK97ORvo	Noticias con Fabiola Anaya Programa completo 18 de noviembre de 2024	ImagenTV	00:26:30
F_DutGcltpc	CONSEJO MINISTROS: Aprueban la REFORMA del REGLAMENTO de EXTRANJERÍA R'RTVE Noticias	R'RTVE Noticias	00:46:35
2K8wwW4_8pc	PERÚ VS. ARGENTINA: ¿LA ESPERANZA ES LO ÚLTIMO QUE SE PIERDE CON LA BLANQUA	Latina Noticias	01:11:01
htk4Z5Sn78w	Alex Otaola en vivo, últimas noticias de Cuba - Hola! Ota-Ola (lunes 18 de noviembre)	Cubanos por el Mundo	03:06:10
_r3zRXm8Eqg	EN VIVO: Noticias RCN - Pongámonos al día	Noticias RCN	01:39:14
LTnsjCuSXqa	TVPerú Noticias EN VIVO: Noticias noche, lunes 18 de noviembre del 2024	TVPerú Noticias	01:02:05
M-xnWnW_LtY	Enfrentamientos en instalaciones de la G.N. en Pátzcuaro Las Noticias - 19 de nc NMÁS	NMÁS	01:06:18
akPuTGJXCGU	Noticiero de Guayaquil (Primera Emisión 19/11/24)	Teleamazonas Ecuador	00:53:28
3JtzKCGDqbg	Los costos del plan de deportaciones masivas de Trump Noticias Univision 24/7 Univision Noticias	Univision Noticias	00:58:13
c6oIu3VDlOM	NOTICIAS DEL DÍA: alerta en Rusia por el órdago de Biden, alarma nuclear en Euro	Negocios TV	00:27:21
t2KwOd8qKB4	EN VIVO CHV Noticias - Martes 19 de noviembre de 2024	CHV Noticias	06:37:17
2tz5_En5aoE	Noticias 23 6:00pm: Spirit se declara en bancarrota ¿Qué significa esto?	Univision Miami	00:26:13
E_QuIMqfDLY	LAS NOTICIAS: Rusia cambia su doctrina nuclear, Ucrania ataca con ATACMS y Eur Negocios TV	Negocios TV	01:36:54
ipYWNOhshIo	Noticiero de Ecuador (Emisión Estelar 18/11/24)	Teleamazonas Ecuador	00:24:18
WxoBbjv7T8A	En Vivo Fórmula Noticias con Azucena Uresti Martes 19/11/24	Grupo Fórmula	02:57:39
hWRJ0yhQSIE	Noticias en la Mañana en Vivo	VPItv	02:04:10
-2-5YQQU3zM	Noticiero de Quito (Primera Emisión 19/11/24)	Teleamazonas Ecuador	00:58:33
AHe-nEe6qD4	José Cárdenas, Noticias en Claro En Vivo Lunes 18 de noviembre	UnoTV	01:04:30
x5BMSJVleil	Noticias En Positivo, lunes 18 de noviembre de 2024	Noticias Telemundo	00:28:14
nPAkRF9_M3g	Noticias del 2024/11/18 12h30 • FRANCE 24 Español	FRANCE 24 Español	00:30:00
zP29A82cqPQ	Noticias 23 1230pm: En corte hombre acusado del secuestro y muerte de su hija	Univision Miami	00:30:53
TMzZ8HYSmkw	NOTICIAS DE ISRAEL SECRETOS IRANÍES DE HAMÁS DESCUBIERTOS #israel #hama LabranzaTV	LabranzaTV	01:29:05
d6puUiOur9U	COLAPINTO PALPITA LAS VEGAS: UNA CARRERA QUE SE TRASMITIRÁ DE MADRUGADA	eltrece	01:44:25
iThwO79quZU	Noticias de Venezuela hoy en Vivo	VPItv	00:37:40
PZrw5HmN1YA	Rogelio Ramírez, secretario de hacienda, en exclusiva para Despierta	NMÁS	02:08:27
Nq9Ak3RWdwY	Regresa a corte el migrante indocumentado acusado del asesinato de una estudiante	Noticias Telemundo	07:49:02
-OJ1vKXDniE	Noticiero de Ecuador (Emisión Central 18/11/2024)	Teleamazonas Ecuador	00:23:00
UVuPg5XS2ws	EN VIVO: Noticias Tarde, lunes 18 de noviembre del 2024, en TVPerú Noticias	TVPerú Noticias	02:55:55
kwzeaie2FpU	#EnVivo Sólo les falta la Antártida: Cártel de Sinaloa extiende su dominio (18/1 Azteca Noticias	Azteca Noticias	01:11:28
rdGIO_TsC2o	EN VIVO 18/11/2024 #EmisiónEstelarSIN	Noticias SIN	00:45:56
_KlvqBLvW4E	#EnVivo Exitosa participación de Claudia Sheinbaum en el G20	Noticiero en Redes	01:33:01
_s4acJ0B6Kc	#ENVIVO Hechos de la mañana (19/11/24)	Azteca Noticias	03:10:01

Figura C.1: Links de videos extraídos. Fuente: Elaboración propia