



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

GRADO EN INGENIERÍA MATEMÁTICA E INTELIGENCIA ARTIFICIAL

TRABAJO FIN DE GRADO

Identificación de regímenes de mercado
mediante análisis de noticias financieras y
técnicas de clustering

Autor: Álvaro Fernández Gutiérrez

Director: Guillermo Mestre Marcos

Madrid, junio 2026

Declaración de Originalidad

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título **Identificación de regímenes de mercado mediante análisis de noticias financieras y técnicas de clustering** en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el curso académico 2025/2026 es de mi autoría, original e inédito y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente, y la información que ha sido tomada de otros documentos está debidamente referenciada.

Uso de Inteligencia Artificial¹

Declaro bajo mi responsabilidad que (seleccione la opción correcta):

- No he utilizado Inteligencia Artificial en la elaboración de este documento.
- He utilizado Inteligencia Artificial en la elaboración de este documento y/o Anexo B bajo las condiciones permitidas por la Universidad Pontificia Comillas, i.e. aplicando el Nivel 2 de la [Escala de Evaluación de Perkins et al. \(2024\)](#): *“La IA puede utilizarse para actividades previas a la tarea, como la lluvia de ideas, la descripción y la investigación inicial. Este nivel se centra en el uso de la IA para planificar, sintetizar y generar ideas, pero las evaluaciones deben enfatizar la capacidad de desarrollar y refinar estas ideas de manera independiente”*. Concretamente, la Inteligencia Artificial se ha utilizado para:

Se han utilizado herramientas de Inteligencia Artificial generativa bajo la supervisión del tutor del TFG para apoyar la estructuración y edición del manuscrito, facilitar discusiones metodológicas exploratorias y mejorar la eficiencia en la escritura de código.

¹Esta declaración hace referencia al uso de Inteligencia Artificial generativa para la realización de los documentos del Proyecto. No aplica a Proyectos en los que, por su naturaleza, se deba emplear inteligencia artificial como parte de los mismos (aplicación de técnicas de aprendizaje automático, redes neuronales, análisis de datos. . .).

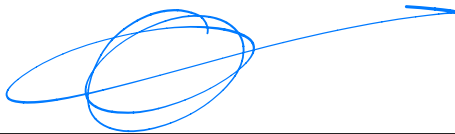
Firmado por FERNANDEZ GUTIERREZ ALVARO -
***3570** el día 15/06/2026 con un
certificado emitido por AC FNMT Usuarios

Firma: Álvaro Fernández Gutiérrez

Fecha: 15/06/26

Autorizada la entrega del proyecto

Director del Proyecto



Firma: Guillermo Mestre Marcos

Fecha: 15/06/26

Agradecimientos

En primer lugar, quiero expresar mi más sincero agradecimiento a mi director, Guillermo, por su orientación, disponibilidad y sus consejos y sugerencias, que han guiado este trabajo desde su planteamiento inicial hasta su forma final. Su dedicación, el tiempo invertido y la confianza depositada en mí han sido fundamentales para mantener el rigor metodológico del proyecto en cada una de sus etapas y para que este trabajo haya salido adelante.

A la Universidad Pontificia Comillas y a la Escuela Técnica Superior de Ingeniería (ICAI), por la formación recibida durante estos cuatro años de carrera, que me han proporcionado las herramientas necesarias para abordar un trabajo de esta naturaleza.

Por último, a mi familia y amigos. Por vuestra paciencia y apoyo incondicional a lo largo de todo el proceso. Gracias por estar siempre presentes cuando más lo necesitaba y por hacer más llevadero cada paso del camino. Este trabajo lleva también un poco de cada uno de vosotros.

Identificación de regímenes de mercado mediante análisis de noticias financieras y técnicas de clustering

Autor: Álvaro Fernández Gutiérrez

Director: Guillermo Mestre Marcos

RESUMEN DEL PROYECTO

La evolución de los mercados financieros se caracteriza por la alternancia de distintos regímenes de mercado. Estos incluyen fases de expansión, asociadas a rentabilidades positivas y menor percepción del riesgo, y fases de contracción o crisis, en las que la incertidumbre aumenta significativamente, la volatilidad se incrementa y las interrelaciones entre los activos pueden experimentar cambios sustanciales. Este Trabajo de Fin de Grado diseña e implementa un pipeline completo que combina indicadores técnicos del S&P 500 con sentimiento sectorial extraído de noticias financieras mediante FinBERT, con el objetivo de identificar regímenes de mercado mediante técnicas de clustering no supervisado.

Palabras clave: regímenes de mercado, clustering, K-Means, análisis de sentimiento, FinBERT, noticias financieras, FNSPID, S&P 500, validación económica.

1. Introducción

Los mercados financieros suelen analizarse mediante un número reducido de regímenes de mercado, habitualmente asociados a fases alcistas (bull markets) y bajistas (bear markets). La identificación temprana de estos regímenes es fundamental para la gestión del riesgo y la asignación de activos. Tradicionalmente, la literatura ha recurrido a variables derivadas del comportamiento histórico de los mercados, como las rentabilidades, la volatilidad o distintos indicadores técnicos. Sin embargo, estos enfoques presentan una limitación inherente: suelen reaccionar a cambios que ya se han producido en el mercado, dificultando la detección anticipada de transiciones entre regímenes. El desarrollo reciente de los modelos de lenguaje basados en arquitecturas transformer, y en particular de modelos especializados en el ámbito financiero como FinBERT, ha permitido extraer de forma sistemática el sentimiento contenido en grandes volúmenes de noticias financieras. Esto proporciona una fuente de información

más inmediata y potencialmente complementaria a los indicadores técnicos tradicionales, con capacidad para captar cambios en las expectativas del mercado antes de que se reflejen plenamente en los precios.

2. Objetivos

El objetivo general es desarrollar y validar una metodología de detección de regímenes de mercado basada en clustering, comparando un modelo que utiliza únicamente indicadores técnicos del S&P 500 frente a modelos que incorporan información de sentimiento sectorial extraída de noticias financieras mediante FinBERT. Los objetivos específicos incluyen: construir un dataset unificado que combine indicadores técnicos diarios con un indicador de sentimiento agregado por sector GICS; diseñar tres modelos de clustering K-Means alternativos y validar económicamente los regímenes obtenidos mediante métricas de rentabilidad, volatilidad, ratio de Sharpe y drawdown por sector GICS, entre otras.

3. Descripción del modelo/sistema

El pipeline desarrollado consta de las siguientes etapas: (1) adquisición y filtrado de datos del corpus FNSPID, restringiendo a constituyentes del S&P 500 en cada fecha; (2) cálculo de sentimiento con FinBERT; (3) agregación sectorial GICS; (4) cálculo de nueve indicadores técnicos del S&P 500; (5) construcción del dataset de clustering con 1.132 observaciones (2009 – 2024); y (6) entrenamiento de tres modelos K-Means ($K = 3$).

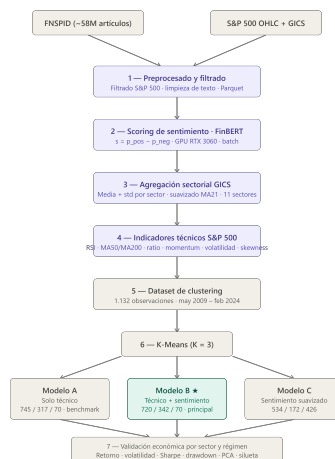


Figura 1: Esquema general del pipeline desarrollado en este proyecto.

4. Resultados

El análisis del coeficiente silhouette (en adelante, silhouette) en un espacio común de componentes principales (PCA) muestra que el Modelo A (técnico) obtiene la mejor cohesión interna ($\approx 0,193$), seguido del Modelo B (técnico + sentimiento) ($\approx 0,178$) y del Modelo C (solo sentimiento suavizado) ($\approx 0,162$). Sin embargo, la validación económica revela que el Modelo B produce regímenes con una diferenciación económica más marcada: el régimen Bear del Modelo B presenta ratios de Sharpe negativos en 10 de los 11 sectores GICS, mientras que el régimen Crisis concentra las rentabilidades anualizadas más elevadas. El sector Energy en régimen Bear presenta el drawdown máximo más severo ($-74,9\%$), así como el mejor ($+16,9\%$) y el peor ($-24,7\%$) día individual de toda la muestra.

5. Conclusiones

Los resultados confirman que los indicadores técnicos constituyen la base más sólida para la detección de regímenes, pero que la incorporación de sentimiento sectorial extraído mediante FinBERT aporta valor incremental en términos de diferenciación económica de los regímenes resultantes. El sentimiento agrega valor como complemento a la información de precio, no como sustituto de ella. Estas conclusiones abren la puerta a líneas de trabajo futuras orientadas a la combinación dinámica de ambas fuentes de información y a la exploración de arquitecturas de deep learning para la detección de regímenes.

6. Referencias

- [1] Hamilton, J.D. “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.” *Econometrica*, 57(2):357–384, 1989.
- [2] Yang, Z. et al. “FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining.” *IJCAI*, 2020.
- [3] Zhang, Y. et al. “FNSPID: A Comprehensive Financial News and Stock Price Integration Dataset.” *arXiv*, 2023.
- [4] Abdollahi, H., Fjesme, S.L., and Sirnes, E. “Measuring Market Volatility Connectedness to Media Sentiment.” *UiT – The Arctic University of Norway*, Working Paper.
- [5] Mudarisov, T., State, R.V., Kraussl, Z., Yakubov, A., and Petrova, T. “Cross-Sector Market Regime Forecasting with LLM-Augmented News Analysis.” In *Proceedings of the 5th ACM International Conference on AI in Finance (ICAIF '24)*, pp. 1–8, Brooklyn, NY, USA, November 2024.

Identificación de regímenes de mercado mediante análisis de noticias financieras y técnicas de clustering

Author: Álvaro Fernández Gutiérrez

Supervisor: Guillermo Mestre Marcos

ABSTRACT

The evolution of financial markets is characterized by the alternation of distinct market regimes. These include expansion phases, associated with positive returns and lower risk perception, and contraction or crisis phases, in which uncertainty increases significantly, volatility rises, and the interrelationships between assets may undergo substantial changes. This Bachelor's Thesis designs and implements a complete end-to-end pipeline combining S&P 500 technical indicators with sector-level sentiment extracted from financial news via FinBERT, with the aim of identifying market regimes through unsupervised clustering techniques.

Keywords: market regimes, clustering, K-Means, sentiment analysis, FinBERT, financial news, FNSPID, S&P 500, economic validation.

1. Introduction

Financial markets are typically analyzed through a small number of market regimes, usually associated with bull and bear market phases. The early identification of these regimes is fundamental for risk management and asset allocation. Traditionally, the literature has relied on variables derived from the historical behaviour of markets, such as returns, volatility, or various technical indicators. However, these approaches present an inherent limitation: they tend to react to changes that have already occurred in the market, making it difficult to anticipate transitions between regimes. The recent development of language models based on transformer architectures, and in particular of finance-specific models such as FinBERT, has made it possible to systematically extract the sentiment contained in large volumes of financial news. This provides a more immediate source of information, potentially complementary to traditional technical indicators, with the capacity to capture shifts in market expectations before they are fully reflected in prices.

2. Objectives

The overall objective is to develop and validate a clustering-based market regime detection methodology, comparing a model relying exclusively on S&P 500 technical indicators against models that additionally incorporate sector-level sentiment extracted from financial news using FinBERT. The specific objectives include: building a unified dataset combining daily technical indicators with a sentiment indicator aggregated by GICS sector; designing three alternative K-Means clustering models; and economically validating the resulting regimes using metrics such as return, volatility, Sharpe ratio, and drawdown by GICS sector, among others.

3. Description of the proposed model/system

The pipeline consists of the following stages: (1) acquisition and filtering of data from the FNSPID corpus, restricting to S&P 500 constituents at each date; (2) sentiment scoring with FinBERT; (3) GICS sector-level aggregation; (4) computation of nine S&P 500 technical indicators; (5) construction of the clustering dataset with 1,132 observations (2009 – 2024); and (6) training of three K-Means models ($K = 3$).

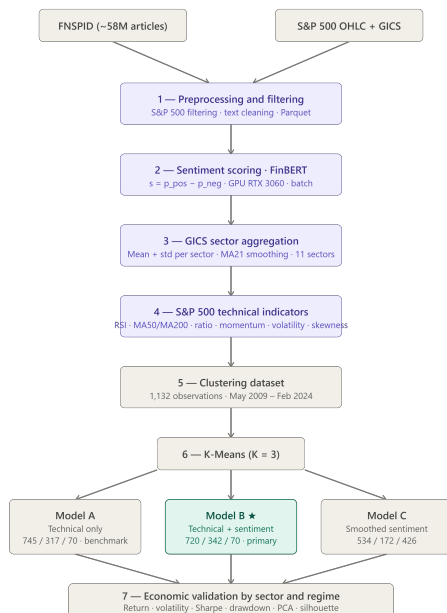


Figura 2: General overview of the pipeline developed in this project.

4. Results

Silhouette analysis in a common principal component space (PCA) shows that Model A (technical) achieves the best internal cohesion (≈ 0.193), followed by Model B (technical + sentiment) (≈ 0.178) and Model C (smoothed sentiment only) (≈ 0.162). However, economic validation reveals that Model B produces regimes with more marked economic differentiation: the Bear regime of Model B presents negative Sharpe ratios in 10 of the 11 GICS sectors, while the Crisis regime concentrates the highest annualized returns. The Energy sector under the Bear regime records the most severe maximum drawdown (-74.9%), as well as the best ($+16.9\%$) and worst (-24.7%) individual day in the entire sample.

5. Conclusions

The results confirm that technical indicators constitute the strongest foundation for regime detection, but that the incorporation of sector-level sentiment extracted via FinBERT adds incremental value in terms of the economic differentiation of the resulting regimes. Sentiment adds value as a complement to price information, not as a substitute for it. These conclusions open the door to future lines of work oriented towards the dynamic combination of both sources of information and the exploration of deep learning architectures for regime detection.

6. References

- [1] Hamilton, J.D. “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.” *Econometrica*, 57(2):357–384, 1989.
- [2] Yang, Z. et al. “FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining.” *IJCAI*, 2020.
- [3] Zhang, Y. et al. “FNSPID: A Comprehensive Financial News and Stock Price Integration Dataset.” *arXiv*, 2023.
- [4] Abdollahi, H., Fjesme, S.L., and Sirnes, E. “Measuring Market Volatility Connectedness to Media Sentiment.” *UiT – The Arctic University of Norway*, Working Paper.
- [5] Mudarisov, T., State, R.V., Kraussl, Z., Yakubov, A., and Petrova, T. “Cross-Sector Market Regime Forecasting with LLM-Augmented News Analysis.” In *Proceedings of the 5th ACM International Conference on AI in Finance (ICAIF '24)*, pp. 1–8, Brooklyn, NY, USA, November 2024.

Contenidos

1	Introducción	1
1.1	Motivación y Contexto	1
1.2	Planteamiento del Problema	2
1.3	Objetivos	2
1.4	Alineación con los Objetivos de Desarrollo Sostenible	3
1.5	Estructura del Documento	4
2	Estado del Arte	5
2.1	Detección de Regímenes de Mercado	5
2.1.1	Antecedentes: Modelos Paramétricos de Cambio de Régimen	5
2.1.2	Clustering No Supervisado para la Detección de Regímenes	5
2.1.3	Reducción de Dimensionalidad y Comparación de Modelos	6
2.2	Análisis de Sentimiento en Finanzas	6
2.2.1	Métodos Basados en Léxico e Índices Agregados	6
2.2.2	Modelos de Lenguaje Transformer y FinBERT	7
2.2.3	Datasets de Noticias Financieras a Gran Escala	7
2.3	Modelos Predictivos para Regímenes de Mercado	8
2.4	Laguna de Investigación	8
3	Sistema/Modelo Desarrollado	9
3.1	Formulación del Problema	9
3.2	Fuentes de Datos	9
3.3	Exploración y Limpieza del Corpus de Noticias	10
3.3.1	Formato de almacenamiento	10
3.3.2	Cobertura temporal y distribución de tickers	10
3.3.3	Filtrado por constituyentes históricos del S&P 500	10
3.3.4	Mapeo sectorial GICS y normalización	11
3.4	Filtrado de Noticias y Puntuación de Sentimiento	11
3.5	Agregación Sectorial y Suavizado	12
3.6	Indicadores Técnicos	13
3.7	Clustering K-Means: Modelos A, B y C	15
3.8	Marco de Validación Económica	17
4	Resultados	19
4.1	Validación Interna: Silhouette	19
4.2	Visualizaciones PCA de los Regímenes	21
4.3	Análisis de la Estructura de Correlación	23
4.4	Importancia de Variables	25
4.5	Distribución Temporal de los Regímenes	27

4.6	Validación Económica a Nivel de Mercado	28
4.7	Validación Económica a Nivel Sectorial	29
4.8	El Modelo C: Un Resultado Negativo Informativo	36
5	Conclusiones y Trabajo Futuro	38
5.1	Conclusiones	38
5.2	Limitaciones	39
5.3	Trabajo Futuro	39
6	Bibliografía	41

Índice de Figuras

1	Esquema general del pipeline desarrollado en este proyecto.	6
2	General overview of the pipeline developed in this project.	9
3	Series de sentimiento sectorial en bruto (claro) frente a suavizado MA21 (oscuro) para los 11 sectores GICS.	13
4	Curvas de codo (inercia) y silueta para $K \in \{2, \dots, 7\}$, Modelo A (solo indicadores técnicos).	15
5	Curvas de codo (inercia) y silueta para $K \in \{2, \dots, 7\}$, Modelo B (indicadores técnicos + sentimiento sectorial bruto).	16
6	Curvas de codo (inercia) y silueta para $K \in \{2, \dots, 7\}$, Modelo C (solo sentimiento sectorial suavizado).	16
7	Silhouette para los Modelos A, B y C, calculado en el espacio PCA común de 9 componentes (82,0% de varianza explicada). La diferencia entre modelos se reduce sustancialmente respecto al espacio propio de cada uno.	20
8	Gráfico de sedimentación del espacio PCA común. Se requieren 9 componentes para alcanzar el 82,0% de varianza explicada sobre las 20 características combinadas, frente a los 4 componentes que bastarían para las 9 características técnicas del Modelo A, evidenciando que el sentimiento aporta dimensiones genuinamente nuevas al espacio de información.	21
9	Modelo A: diagrama de dispersión PCA (izquierda), cargas de las 9 características técnicas sobre CP1 y CP2 (centro) y varianza explicada acumulada (derecha). CP1 y CP2 explican el 43,7% y el 15,8% de la varianza, respectivamente, con un total del 59,5%.	21
10	Modelo B: diagrama de dispersión PCA (izquierda), cargas de las 31 características sobre CP1 y CP2 (centro) y varianza explicada acumulada (derecha). CP1 y CP2 explican únicamente el 23,2% y el 15,7% de la varianza, reflejando el mayor número de dimensiones informativas aportadas por el sentimiento sectorial.	22
11	Matriz de correlación de las 9 características técnicas del Modelo A. Los indicadores de tendencia (MA50_ratio, MA200_ratio, momentum) presentan correlaciones mutuas elevadas, mientras que la volatilidad a corto y largo plazo forman un grupo diferenciado.	23
12	Matriz de correlación de las 31 características del Modelo B. Las variables de sentimiento sectorial presentan correlaciones bajas con los indicadores técnicos (en general $ r < 0,15$), mientras que entre sí forman dos o tres grupos según la naturaleza del sector.	24

13	Estadístico F de ANOVA por característica para el Modelo B. Las seis primeras posiciones están ocupadas por indicadores técnicos, pero cuatro variables de sentimiento sectorial (desviación típica de los sectores IT, Consumer Staples, Consumer Discretionary y Energy) aparecen entre las diez más discriminantes.	26
14	Precio de cierre del S&P 500 con sombreado de fondo por régimen del Modelo A: verde = Alcista, rojo = Bajista, naranja = Crisis.	27
15	Precio de cierre del S&P 500 con sombreado de fondo por régimen del Modelo B: verde = Alcista, rojo = Bajista, naranja = Crisis.	27
16	Ratio de Sharpe anualizado por sector GICS y régimen, Modelo B. El régimen Alcista produce Sharpes positivos en los 11 sectores; el Bajista produce Sharpes negativos en 9 de los 11; el de Crisis presenta los Sharpes más altos en todos los sectores.	30
17	Drawdown máximo por sector GICS y régimen, Modelo B. Energy en régimen bajista presenta el drawdown más severo ($-74,9\%$), mientras que Consumer Staples y Utilities en crisis presentan los drawdowns más contenidos ($-4,7\%$ y $-6,8\%$ respectivamente).	34
18	Ratio de Sharpe por sector GICS y régimen, Modelo A. El patrón es cualitativamente idéntico al del Modelo B, con diferencias cuantitativas menores, lo que confirma la robustez de la diferenciación sectorial por régimen independientemente del modelo utilizado.	35
19	Ratio de Sharpe por sector GICS y régimen, Modelo C (solo sentimiento suavizado). A diferencia de los Modelos A y B, no existe un patrón monótono claro entre regímenes: el régimen etiquetado como “Bajista” no presenta Sharpes sistemáticamente negativos, y la distribución de retornos entre regímenes no es económicamente interpretable.	36
20	Ratio de Sharpe a nivel de mercado global por clúster para los Modelos A, B y C. Los Modelos A y B presentan una diferenciación máxima ($\Delta\text{Sharpe} \approx 6,8$) frente a los 2,18 del Modelo C, evidenciando la superioridad de los indicadores técnicos (solos o combinados con sentimiento) para la discriminación económica de los regímenes.	37

Índice de Tablas

1	Distribución de los constituyentes históricos del S&P 500 por sector GICS.	9
2	Distribución de noticias por sector GICS tras el filtrado completo.	11
3	Reducción de ruido promedio obtenida por diferentes ventanas de suavizado de media móvil, promediada sobre los 11 sectores GICS.	12
4	Resumen de las tres configuraciones de modelos K-Means.	15
5	Modelo A (solo técnico)	17
6	Modelo B (técnico + sentimiento bruto)	17
7	Silhouette para los Modelos A, B y C, en el espacio propio de cada modelo y en el espacio PCA común de 9 componentes (82,0% de varianza explicada).	19
8	Las 10 características más importantes según el estadístico F de ANOVA para el Modelo B. El estadístico F mide la ratio entre la varianza interclúster y la varianza intraclúster para cada variable.	25
9	Perfil económico a nivel de mercado global por régimen, K-Means Modelos A, B y C. El retorno medio diario se expresa en porcentaje; la volatilidad anualizada es la desviación típica diaria de los retornos logarítmicos multiplicada por $\sqrt{252}$; el ratio de Sharpe se anualiza con una tasa libre de riesgo del 2%.	28
10	Métricas económicas por sector y régimen bajo el Modelo B. El retorno anualizado se expresa en tanto por uno; la volatilidad anualizada es $\sigma_{\text{diaria}} \times \sqrt{252}$; el ratio de Sharpe utiliza una tasa libre de riesgo anual del 2%.	32

1 Introducción

1.1 Motivación y Contexto

Los mercados financieros presentan un comportamiento dinámico cuyas características cambian a lo largo del tiempo. Periodos de crecimiento económico y estabilidad suelen estar asociados a rentabilidades positivas y niveles moderados de volatilidad, mientras que las fases de incertidumbre o estrés financiero suelen caracterizarse por un aumento de la volatilidad, una reducción de las rentabilidades esperadas y un incremento de las correlaciones entre activos, lo que dificulta los beneficios asociados a la diversificación. La crisis financiera global de 2008 y el crash de mercado provocado por el COVID-19 en marzo de 2020 son dos de los ejemplos más destacados de este fenómeno en la historia reciente. En ambos episodios, índices bursátiles como el S&P 500 experimentaron caídas superiores al 30% en pocas semanas, los índices de volatilidad alcanzaron máximos históricos y la estructura de correlación entre sectores normalmente poco relacionados (por ejemplo, Energía y Tecnología de la Información) se incrementó bruscamente, rasgo distintivo de lo que la literatura denomina ampliamente régimen de crisis [1].

La idea de que los mercados financieros alternan entre un número reducido de regímenes distintos tiene una larga tradición en econometría y finanzas cuantitativas. El trabajo de Hamilton [2] introdujo los modelos de cambio de régimen de Markov para caracterizar formalmente los cambios en la media y la varianza de series temporales macroeconómicas y financieras. Desde una perspectiva más aplicada, gestores de activos, responsables de riesgo y sistemas de trading algorítmico adaptan habitualmente sus estrategias en función del régimen vigente: el dimensionamiento de posiciones, los ratios de cobertura y la elección de qué factores sobreponderar en una cartera son con frecuencia dependientes del régimen.

Una segunda línea de investigación reciente ha explorado el papel del sentimiento (el tono agregado de noticias, redes sociales o informes de analistas) como factor determinante de, o al menos correlacionado con, la dinámica de mercado. Los primeros trabajos en este ámbito se apoyaban principalmente en métodos basados en diccionarios, en particular las listas de palabras específicas para finanzas desarrolladas por Loughran y McDonald [3]. La llegada de los modelos de lenguaje basados en Transformers [4] y de codificadores bidireccionales preentrenados como BERT [5] ha transformado radicalmente el conjunto de herramientas disponibles para la extracción de sentimiento. Variantes especializadas y adaptadas al dominio financiero como FinBERT [6] han demostrado superar tanto los métodos basados en léxico como los clasificadores de sentimiento de propósito general en texto financiero. Al mismo tiempo, la publicación de corpus y datasets de noticias financieras a gran escala como FNSPID [7] ha hecho factible calcular puntuaciones de sentimiento para millones de noticias que

abarcan más de una década.

1.2 Planteamiento del Problema

La mayoría de los enfoques existentes para la detección de regímenes de mercado se apoyan exclusivamente en variables técnicas derivadas de precios o exclusivamente en índices de sentimiento usados como predictores independientes. Pocos trabajos han combinado de forma sistemática ambas fuentes de información en un único marco de clustering no supervisado y evaluado posteriormente el contenido económico de los grupos resultantes. Esta tesis aborda esta laguna: construyendo un pipeline reproducible que genera un dataset diario que combina indicadores técnicos del S&P 500 con puntuaciones de sentimiento a nivel de sector GICS a partir de FNSPID; entrenando y comparando tres configuraciones de K-Means; y evaluando todas las particiones tanto mediante métricas internas de clustering como mediante una capa de validación económica.

1.3 Objetivos

El objetivo general de este proyecto consiste en desarrollar y validar un marco metodológico que permita identificar de forma automática regímenes de mercado a partir de la integración de información procedente de noticias financieras y variables cuantitativas de mercado, empleando técnicas de procesamiento del lenguaje natural (NLP) y métodos de clustering no supervisado aplicados a series temporales financieras. En particular, se evalúa si la incorporación de indicadores de sentimiento sectorial extraído mediante FinBERT mejora la diferenciación económica de los regímenes detectados respecto a una línea base puramente técnica. Esto se puede dividir en varios objetivos específicos:

1. **Ingeniería de datos.** Construir un dataset unificado, reproducible y libre de fuga de información que integre indicadores técnicos diarios calculados a partir de los precios del S&P 500 con puntuaciones de sentimiento sectorial derivadas del corpus de noticias financieras FNSPID, empleando una tabla de constituyentes históricos para evitar el sesgo de supervivencia.
2. **Extracción y agregación de sentimiento.** Aplicar el modelo FinBERT a cada titular de noticias filtrado para obtener una puntuación de sentimiento continua $s = p_{\text{pos}} - p_{\text{neg}} \in [-1, 1]$, y agregar dicha puntuación a nivel de (fecha, sector GICS) mediante estadísticos brutos y suavizado por media móvil de 21 sesiones (MA21).
3. **Diseño y entrenamiento de modelos de clustering.** Entrenar tres configuraciones de K-Means con $K = 3$ sobre conjuntos de características diferenciados: Modelo A (solo

indicadores técnicos), Modelo B (técnicos más sentimiento sectorial bruto) y Modelo C (solo sentimiento suavizado), seleccionando el número de clústeres mediante el método del codo y el silhouette.

4. **Validación interna.** Comparar los tres modelos de clustering mediante el silhouette en un espacio PCA común, visualizaciones de dispersión en las dos primeras componentes principales y análisis de importancia de variables basado en el estadístico F de ANOVA.
5. **Validación económica.** Analizar el comportamiento de los 11 sectores GICS bajo cada régimen identificado, calculando métricas de retorno anualizado, volatilidad, ratio de Sharpe, drawdown máximo, porcentaje de días positivos y retornos extremos individuales, con el objeto de determinar si los regímenes detectados presentan patrones económicamente consistentes y relevantes para el análisis financiero y la gestión de carteras.

1.4 Alineación con los Objetivos de Desarrollo Sostenible

Este proyecto se enmarca en el contexto de los Objetivos de Desarrollo Sostenible (ODS) definidos por las Naciones Unidas, varios de los cuales guardan relación directa con los retos abordados en este trabajo [8]:

- **ODS 8 – Trabajo Decente y Crecimiento Económico.** La mejora en la comprensión del funcionamiento de los mercados financieros y de los episodios de inestabilidad contribuye a un sistema financiero más eficiente y resiliente. Las herramientas desarrolladas en este trabajo permiten identificar automáticamente periodos de estrés y crisis de mercado, facilitando una gestión del riesgo más informada y contribuyendo así a un crecimiento económico más sostenible.
- **ODS 9 – Industria, Innovación e Infraestructura.** El proyecto aplica técnicas de inteligencia artificial y procesamiento del lenguaje natural de última generación. En particular, modelos Transformer especializados en texto financiero como FinBERT a un pipeline de datos real, completamente reproducible y de código abierto. Este enfoque promueve la innovación tecnológica en el ámbito del análisis financiero cuantitativo y el desarrollo de infraestructuras analíticas avanzadas accesibles.
- **ODS 12 – Producción y Consumo Responsables.** Una identificación más precisa de los regímenes de mercado y de los periodos de riesgo elevado puede contribuir a una gestión más responsable del riesgo financiero, apoyando decisiones de inversión más informadas, prudentes y sostenibles tanto para inversores institucionales como particulares.
- **ODS 17 – Alianzas para Lograr los Objetivos.** Todo el pipeline desarrollado se apoya exclusivamente en datos públicamente disponibles (el corpus FNSPID y los precios

históricos del S&P 500) y en software de código abierto, fomentando la reproducibilidad, la transparencia y la colaboración científica abierta como base para el avance del conocimiento.

1.5 Estructura del Documento

La Sección 2 revisa la literatura relevante sobre detección de regímenes de mercado y análisis de sentimiento en finanzas. La Sección 3 describe el pipeline completo implementado para este proyecto. La Sección 4 presenta y discute los resultados. La Sección 5 sintetiza los hallazgos principales y esboza líneas de trabajo futuro. El documento concluye con una Bibliografía.

2 Estado del Arte

Este capítulo revisa la literatura relevante para los tres pilares principales de este trabajo: la detección de regímenes de mercado mediante técnicas de clustering no supervisado, la extracción y agregación de sentimiento a partir de noticias de ámbito financiero, y la predicción de regímenes mediante modelos de aprendizaje automático supervisado.

2.1 Detección de Regímenes de Mercado

2.1.1 Antecedentes: Modelos Paramétricos de Cambio de Régimen

La idea de que los parámetros que gobiernan la evolución de una serie temporal financiera alternan entre un número finito de estados fue formalizada por Hamilton [2], quien propuso un modelo de cambio de régimen de Markov para caracterizar los ciclos macroeconómicos de expansión y recesión. Este marco fue extendido por Guidolin [9], quien revisó exhaustivamente su aplicación a finanzas empíricas, demostrando que los modelos de cambio de régimen mejoran los retornos ajustados por riesgo en la gestión dinámica de carteras.

Sin embargo, los modelos de cambio de régimen de Markov imponen supuestos distribucionales estrictos y requieren que el número de regímenes sea fijado a priori, lo que limita su flexibilidad cuando la estructura del mercado no es estacionaria o cuando se desea incorporar información de alta dimensionalidad procedente de múltiples fuentes heterogéneas como los indicadores de sentimiento sectorial.

2.1.2 Clustering No Supervisado para la Detección de Regímenes

Una alternativa más flexible a los modelos paramétricos consiste en tratar la detección de regímenes como un problema de clustering no supervisado. El algoritmo K-Means busca asignaciones de clúster $c_t \in \{1, \dots, K\}$ y centroides $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K \in \mathbb{R}^d$ que minimizan la inercia total:

$$J = \sum_{t=1}^N \left\| \mathbf{x}_t - \boldsymbol{\mu}_{c_t} \right\|_2^2. \quad (2.1)$$

La elección del número de clústeres K se guía habitualmente por el método del codo y el

silhouette:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (2.2)$$

donde $a(i)$ es la distancia media intraclúster y $b(i)$ es la distancia media mínima al clúster más próximo. En aplicaciones financieras, $K = 3$ se corresponde naturalmente con los tres regímenes comunes: alcista / bajista / crisis, ampliamente utilizados en la literatura [1], [9].

La aplicación de K-Means a la detección de regímenes de mercado presenta ventajas frente a los modelos paramétricos: no impone supuestos distribucionales sobre las observaciones, escala bien con la dimensionalidad del espacio de características y produce asignaciones deterministas interpretables. Mudarisov et al. [10] adoptan precisamente esta estrategia de dos etapas: clustering para etiquetar los datos seguido de un clasificador supervisado para predecir el régimen, demostrando que la incorporación de análisis de sentimiento sobre noticias financieras mejora la precisión de predicción en un 73% respecto a una línea base puramente temporal.

2.1.3 Reducción de Dimensionalidad y Comparación de Modelos

Cuando se comparan modelos de clustering entrenados sobre espacios de características de distinta dimensionalidad, como un modelo técnico con 9 variables frente a un modelo enriquecido con sentimiento sectorial de 31 variables, el silhouette calculado en el espacio original no es directamente comparable, ya que la geometría del espacio cambia con la dimensionalidad. Una práctica habitual en la literatura consiste en proyectar todos los modelos sobre un espacio común de menor dimensión mediante Análisis de Componentes Principales (PCA) y evaluar las métricas internas en dicho espacio común, reteniendo típicamente las componentes que explican el 80% de la varianza total [10].

2.2 Análisis de Sentimiento en Finanzas

2.2.1 Métodos Basados en Léxico e Índices Agregados

Los primeros trabajos sobre el impacto del sentimiento mediático en los mercados financieros empleaban diccionarios de palabras positivas y negativas de propósito general. Tetlock [11] demostró que un tono mediático inusualmente pesimista predecía presión bajista sobre los

precios de mercado. Loughran y McDonald [3] mostraron que los diccionarios de propósito general clasifican sistemáticamente de forma incorrecta el texto financiero (por ejemplo, la palabra *liability* se considera negativa en un diccionario general pero es neutral en contexto financiero) y propusieron una lista de palabras específica para finanzas que se ha convertido en referencia estándar.

2.2.2 Modelos de Lenguaje Transformer y FinBERT

La llegada de los modelos de lenguaje basados en la arquitectura Transformer [4] y, en particular, de codificadores bidireccionales preentrenados como BERT [5] ha transformado radicalmente el análisis de sentimiento financiero. A diferencia de los métodos basados en léxico, estos modelos capturan el contexto semántico de cada término en función de las palabras que lo rodean, permitiendo distinguir, por ejemplo, entre “el banco subió sus provisiones” (negativo) y “el banco subió sus beneficios” (positivo).

FinBERT [6] es un modelo BERT ajustado sobre texto financiero que produce una distribución de probabilidad sobre tres clases (positivo, negativo y neutro). En la práctica, se emplea la diferencia entre probabilidades como puntuación continua: $s = p_{\text{pos}} - p_{\text{neg}} \in [-1, 1]$, formulación que tanto Abdollahi et al. [12] como el presente trabajo adoptan.

Abdollahi et al. [12] cuantifican la conectividad direccional desde el sentimiento mediático hacia la volatilidad de seis mercados internacionales, y concluyen que (i) el sentimiento induce volatilidad de mercado de forma estadísticamente significativa aunque con una magnitud relativamente reducida en condiciones normales, y (ii) la conectividad sigue un patrón *spiky*: permanece baja en periodos tranquilos y se dispara durante eventos extremos como el COVID-19 o la crisis financiera global. Esta evidencia de comportamiento heterogéneo del sentimiento a lo largo del tiempo motiva el uso de series suavizadas mediante media móvil para capturar la tendencia del sentimiento, reduciendo el ruido de alta frecuencia antes de incorporarlo como variable de clustering.

2.2.3 Datasets de Noticias Financieras a Gran Escala

FNSPID [7] es un corpus público a gran escala que combina varios millones de titulares de noticias financieras con marcas temporales para empresas cotizadas en EE. UU. y los correspondientes datos históricos de precios bursátiles, abarcando el periodo 2009–2024. Su cobertura por ticker permite construir series de sentimiento desagregadas a nivel de sector GICS, lo que constituye una diferencia relevante respecto a los índices de sentimiento agregados al nivel de índice bursátil empleados en trabajos previos.

2.3 Modelos Predictivos para Regímenes de Mercado

Una vez identificados los regímenes mediante clustering, un paso natural consiste en entrenar un modelo supervisado que aprenda a predecir el régimen del día siguiente a partir de las características del día actual. Este planteamiento convierte un problema no supervisado en uno de clasificación multiclase, y tiene valor económico directo: si el régimen futuro puede anticiparse con suficiente precisión, es posible ajustar dinámicamente la composición sectorial de una cartera hacia los sectores que históricamente han obtenido mejores resultados en dicho régimen.

Los algoritmos de ensemble basados en árboles de decisión, como Random Forest y XGBoost son los más empleados en este contexto por su robustez frente al sobreajuste, su capacidad de manejar espacios de características de alta dimensionalidad y la posibilidad de extraer métricas de importancia de variables que permiten interpretar qué factores contribuyen más a la predicción del régimen. Mudarisov et al. [10] demuestran que un clasificador de este tipo, enriquecido con características de sentimiento extraídas mediante FinBERT, mejora en un 73% la precisión de predicción frente a una línea base de series temporales puras, resultado que motiva el diseño del componente predictivo de este trabajo.

2.4 Laguna de Investigación

La revisión anterior permite identificar tres lagunas que motivan esta tesis. En primer lugar, la mayoría de los trabajos sobre detección de regímenes emplean o bien variables técnicas o bien sentimiento, pero rara vez los integran sistemáticamente en un único marco de clustering con comparación explícita de modelos. En segundo lugar, la validación económica de los regímenes detectados suele limitarse al índice de mercado agregado, sin desagregar el análisis por sector GICS. En tercer lugar, la conectividad entre sentimiento sectorial y regímenes de mercado ha sido estudiada principalmente a nivel de índice nacional o internacional, pero no a nivel de sector GICS dentro de un mercado. Esta tesis aborda las tres lagunas de forma conjunta.

3 Sistema/Modelo Desarrollado

3.1 Formulación del Problema

Sea $t = 1, \dots, N$ el índice de días de negociación. Para cada día t , se construye un vector de características $\mathbf{x}_t \in \mathbb{R}^d$ compuesto por características técnicas derivadas de los precios OHLC del S&P 500, y características de sentimiento derivadas de titulares de noticias financieras, agregadas a nivel sectorial. El objetivo es particionar $\{1, \dots, N\}$ en $K = 3$ grupos $\{C_0, C_1, C_2\}$ que representen regímenes de mercado estadística y económicamente significativos.

3.2 Fuentes de Datos

La señal de sentimiento se deriva del corpus FNSPID [7]. Los datos de precios OHLC diarios del S&P 500 y sus constituyentes se obtienen mediante la librería `yfinance`, utilizando una tabla de constituyentes históricos para evitar el sesgo de supervivencia. Cada ticker se asigna a uno de los 11 sectores GICS mediante una tabla de correspondencias construida a partir de los metadatos sectoriales de Yahoo Finance, que cubre 594 tickers tal como se muestra en el Tabla 1.

Tabla 1: Distribución de los constituyentes históricos del S&P 500 por sector GICS.

Sector GICS	Número de tickers
Industriales	94
Servicios Financieros	88
Tecnología de la Información	76
Consumo Discrecional	75
Salud	68
Consumo Básico	39
Inmobiliario	36
Energía	32
Servicios Públicos	31
Materiales	30
Servicios de Comunicación	25
Total	594

3.3 Exploración y Limpieza del Corpus de Noticias

El corpus FNSPID original contiene aproximadamente 58 millones de registros distribuidos en múltiples ficheros. Dado el volumen, la ingesta se realiza mediante procesamiento en streaming por chunks, evitando cargar el dataset completo en memoria. Durante esta fase exploratoria se identificaron varios problemas de calidad que requirieron tratamiento específico.

3.3.1 Formato de almacenamiento

Los datos se almacenan en formato Parquet en lugar de CSV. Esta decisión es relevante porque los titulares de noticias financieras contienen con frecuencia comas y saltos de línea, lo que provocaba corrupción silenciosa de millones de registros al leer ficheros CSV. El formato Parquet resuelve este problema al serializar los datos de forma binaria tipada.

3.3.2 Cobertura temporal y distribución de tickers

El corpus cubre el período 2009–2024. De los 58 millones de registros totales, aproximadamente 9 millones contienen un ticker asignado. La distribución de noticias por ticker es altamente sesgada: un pequeño número de empresas de gran capitalización concentra la mayoría de las noticias, mientras que muchos tickers presentan cobertura muy escasa.

3.3.3 Filtrado por constituyentes históricos del S&P 500

Para garantizar la relevancia de las noticias y evitar el sesgo de supervivencia, cada noticia se conserva únicamente si el ticker correspondiente era constituyente del S&P 500 en la fecha de publicación, según la tabla de constituyentes históricos. Este filtro reduce el corpus de 9 millones a 2.327.890 noticias asociadas a 672 tickers únicos.

Las noticias publicadas en días no hábiles (fines de semana o festivos) se reasignan al siguiente día de mercado abierto. No se aplica ajuste por hora de cierre de mercado dado que el 96% de las noticias presentan timestamp 00:00:00 UTC, indicando que la hora de publicación no es fiable en este dataset. Esta limitación se documenta como trabajo futuro.

3.3.4 Mapeo sectorial GICS y normalización

De los 672 tickers filtrados, 504 se mapean directamente a través de los metadatos sectoriales de los constituyentes actuales del S&P 500. Para los 168 tickers históricos restantes sin sector asignado, se consulta la API de Yahoo Finance, recuperando el sector para 91 tickers adicionales. El total de tickers con sector asignado asciende a 594.

Durante este proceso se detectó una inconsistencia en la nomenclatura sectorial: Yahoo Finance emplea nombres distintos a los del estándar GICS para varios sectores (por ejemplo, Healthcare frente a Health Care, o Financials frente a Financial Services). Se aplica una tabla de normalización explícita para garantizar la coherencia entre ambas fuentes.

Tras aplicar el filtro sectorial, el corpus queda reducido a 1.989.007 noticias (85,4% de las 2.327.890 filtradas por S&P 500), distribuidas entre los 11 sectores GICS tal como muestra la Tabla 2.

Tabla 2: Distribución de noticias por sector GICS tras el filtrado completo.

Sector GICS	Noticias
Servicios Financieros	283.794
Consumo Discrecional	264.829
Tecnología de la Información	261.223
Industriales	252.270
Salud	248.356
Consumo Básico	184.572
Energía	135.309
Servicios Públicos	100.623
Materiales	92.292
Servicios de Comunicación	91.190
Inmobiliario	74.549
Total	1.989.007

3.4 Filtrado de Noticias y Puntuación de Sentimiento

Una noticia publicada en la fecha d para el ticker τ se conserva únicamente si τ era constituyente del S&P 500 en la fecha d . Cada titular filtrado se puntúa con `ProsusAI/finbert` mediante una puntuación de sentimiento continua:

$$s^{(i)} = p_{\text{pos}}^{(i)} - p_{\text{neg}}^{(i)} \in [-1, 1]. \quad (3.1)$$

3.5 Agregación Sectorial y Suavizado

Para cada combinación de día de negociación τ y sector GICS g , se calculan la media y la desviación típica del sentimiento sobre todos los titulares:

$$\bar{s}_{\tau,g} = \frac{1}{|\mathcal{I}_{\tau,g}|} \sum_{i \in \mathcal{I}_{\tau,g}} s^{(i)}. \quad (3.2)$$

Se aplica una media móvil de 21 sesiones (MA21) para obtener la serie suavizada $\tilde{s}_{\tau,g}^{(21)}$, reduciendo el ruido de alta frecuencia en un promedio del 66,5% entre sectores mientras se preservan las tendencias a medio plazo (Tabla 3).

Tabla 3: Reducción de ruido promedio obtenida por diferentes ventanas de suavizado de media móvil, promediada sobre los 11 sectores GICS.

Ventana	Reducción media de ruido	Mín.	Máx.
MA5	47,4%	41,9%	51,7%
MA10	58,4%	51,5%	64,1%
MA21	66,4%	58,9%	72,2%

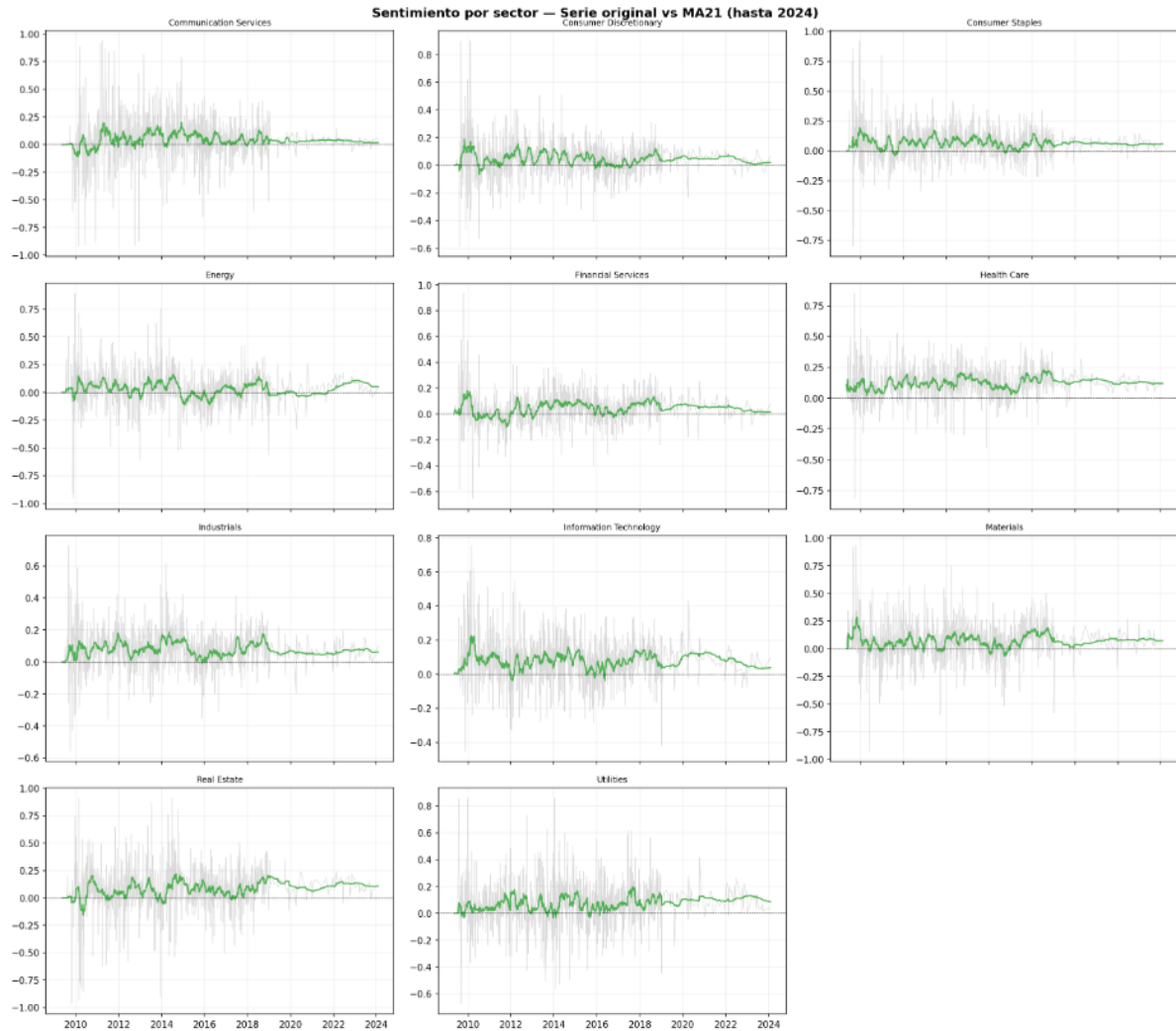


Figura 3: Series de sentimiento sectorial en bruto (claro) frente a suavizado MA21 (oscuro) para los 11 sectores GICS.

3.6 Indicadores Técnicos

Se calculan nueve indicadores técnicos a partir del precio de cierre diario del S&P 500. Sea $r_\tau = \ln(P_\tau/P_{\tau-1})$ el log-retorno diario.

- **RSI (RSI14).** Índice de fuerza relativa calculado sobre una ventana de 14 sesiones:

$$RSI_t = 100 - \frac{100}{1 + RS_t}$$

donde RS_t representa la razón entre las ganancias medias y las pérdidas medias de los últimos 14 días. Este indicador mide la intensidad del movimiento reciente del precio y permite identificar situaciones de sobrecompra o sobreventa. Valores elevados indican una fuerte presión compradora, mientras que valores bajos reflejan predominio de ventas.

- **Ratios de medias móviles (MA50, MA200)**. Se calculan como el cociente entre el precio de cierre y las medias móviles simples de 50 y 200 sesiones:

$$MA50_t = \frac{P_t}{SMA50_t}, \quad MA200_t = \frac{P_t}{SMA200_t}$$

Estas variables permiten medir la posición relativa del precio respecto a su tendencia de medio y largo plazo. Valores superiores a 1 indican que el precio se encuentra por encima de la media móvil correspondiente, mientras que valores inferiores a 1 sugieren una tendencia bajista.

- **Momento (mom21, mom126)**. Variación porcentual acumulada del precio durante las últimas 21 y 126 sesiones:

$$mom_n = \frac{P_t - P_{t-n}}{P_{t-n}}$$

donde n representa el horizonte temporal considerado. Estas variables capturan la persistencia de tendencias alcistas o bajistas a corto y medio plazo, proporcionando información sobre la fuerza y dirección de los movimientos del mercado.

- **Volatilidad realizada (vol30d, vol252d)**. Desviación típica anualizada de los log-retornos calculada sobre ventanas móviles de 30 y 252 sesiones:

$$\sigma_t = \sqrt{252} \cdot std(r_t)$$

La volatilidad realizada constituye una medida del riesgo histórico del mercado. La ventana de 30 sesiones refleja cambios recientes en la incertidumbre, mientras que la de 252 sesiones proporciona una visión más estable del comportamiento anual.

- **Asimetría de retornos (skew30d, skew252d)**. Coeficiente de asimetría de Fisher-Pearson calculado sobre ventanas móviles de 30 y 252 sesiones:

$$Skew = \frac{E[(r - \mu)^3]}{\sigma^3}$$

Esta métrica describe la forma de la distribución de retornos. Valores negativos indican una mayor probabilidad de movimientos extremos a la baja, mientras que valores positivos reflejan una mayor presencia de retornos extremos positivos. La inclusión de

ventanas de corto y largo plazo permite capturar cambios temporales en la distribución de los retornos.

3.7 Clustering K-Means: Modelos A, B y C

El dataset final de clustering abarca desde el 08-05-2009 hasta el 01-02-2024 y contiene 1.132 observaciones diarias y 43 columnas. Se entrenan tres configuraciones de K-Means con $K = 3$, seleccionado mediante el método del codo y el silhouette. El método del codo evalúa la inercia total, definida en la Ecuación (2.1), para valores de $K \in \{2, \dots, 7\}$, identificando el punto a partir del cual las reducciones marginales de inercia son decrecientes. El silhouette, definido en la Ecuación (2.2), complementa este criterio evaluando la cohesión interna y la separación entre clústeres: valores más próximos a 1 indican asignaciones bien definidas. En los tres modelos, ambos criterios convergen en $K = 3$, valor que además tiene una interpretación económica natural al corresponderse con los tres regímenes canónicos de mercado: alcista, bajista y crisis. Las curvas de codo y silhouette para cada modelo se muestran a continuación.

Tabla 4: Resumen de las tres configuraciones de modelos K-Means.

Modelo	Conjunto de características	Columna	d
A	Solo indicadores técnicos	kmeans_tech	9
B	Técnico + sentimiento sectorial bruto (media + desv.)	kmeans_tech_sent	31
C	Solo sentimiento sectorial suavizado (MA21)	kmeans_sent_smooth	11

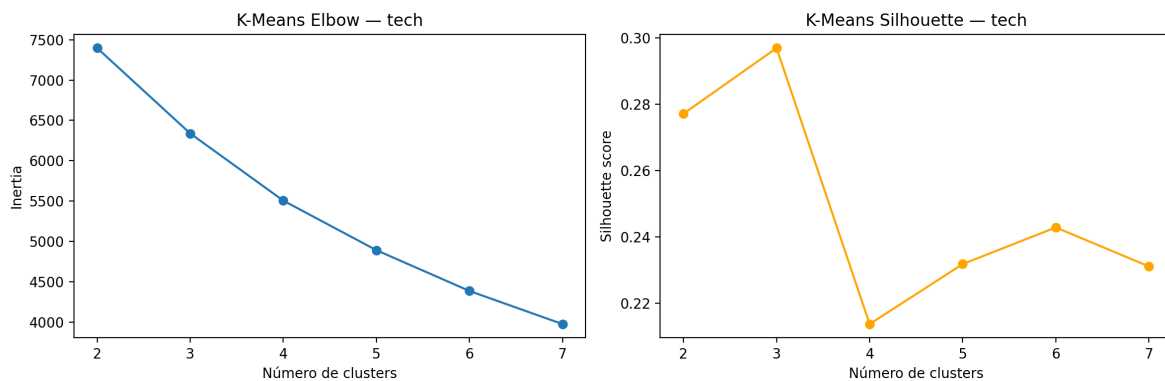


Figura 4: Curvas de codo (inercia) y silueta para $K \in \{2, \dots, 7\}$, Modelo A (solo indicadores técnicos).

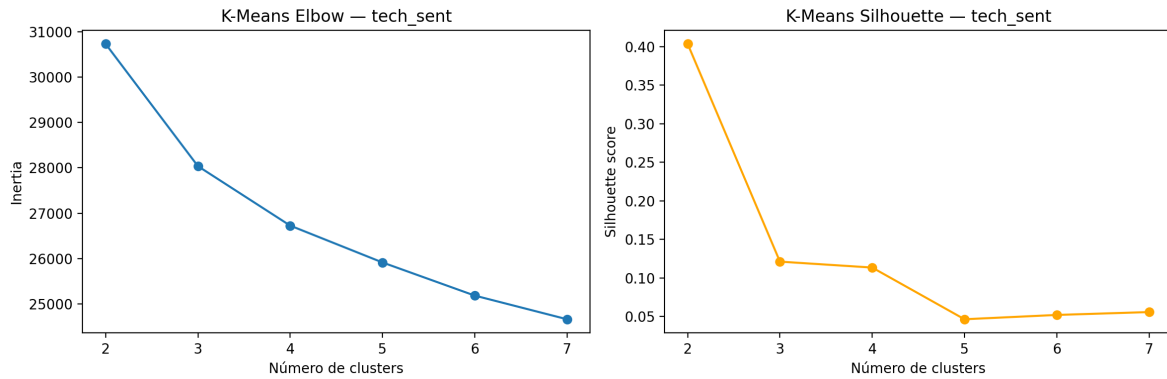


Figura 5: Curvas de codo (inercia) y silueta para $K \in \{2, \dots, 7\}$, Modelo B (indicadores técnicos + sentimiento sectorial bruto).

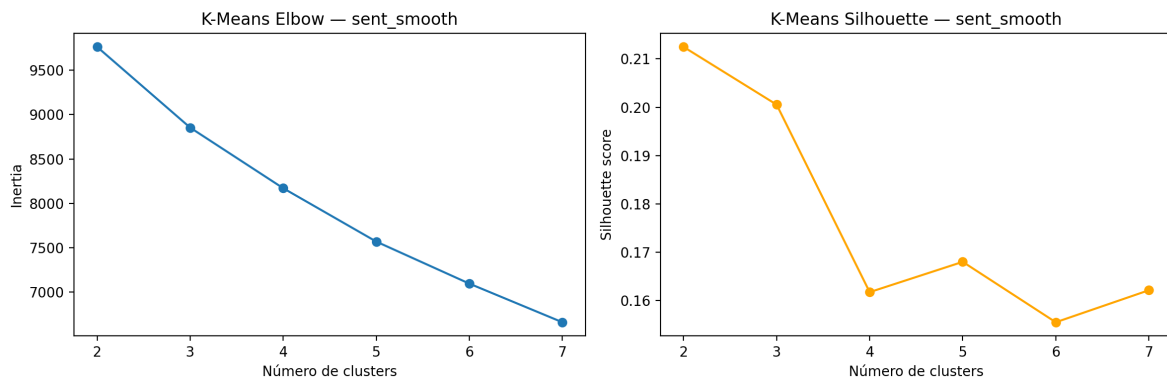


Figura 6: Curvas de codo (inercia) y silueta para $K \in \{2, \dots, 7\}$, Modelo C (solo sentimiento sectorial suavizado).

Los clusters para los Modelos A y B se resumen en la Tabla 5 y la Tabla 6.

Tabla 5: Modelo A (solo técnico)

Característica	Clúster 0 (Alcista)	Clúster 1 (Bajista)	Clúster 2 (Crisis)
RSL_sp500	60,34	42,57	59,25
MA50_ratio_sp500	1,0229	0,9739	1,0392
MA200_ratio_sp500	1,0620	0,9877	1,1337
mom21_sp500	0,0227	-0,0297	0,0281
mom126_sp500	0,0793	-0,0075	0,2373
vol30d_sp500	0,1120	0,2005	0,1765
vol252d_sp500	0,1405	0,1670	0,3718
skew30d_sp500	-0,1417	-0,2704	-0,4073
skew252d_sp500	-0,3879	-0,5259	-0,2129
<i>n</i> (días)	745	317	70
Proporción	65,8%	28,0%	6,2%

Tabla 6: Modelo B (técnico + sentimiento bruto)

Característica	Clúster 0 (Alcista)	Clúster 1 (Bajista)	Clúster 2 (Crisis)
RSL_sp500	60,50	43,34	60,11
MA50_ratio_sp500	1,0235	0,9766	1,0374
MA200_ratio_sp500	1,0647	0,9905	1,1190
mom21_sp500	0,0226	-0,0265	0,0315
mom126_sp500	0,0841	-0,0036	0,1997
vol30d_sp500	0,1110	0,1967	0,1734
vol252d_sp500	0,1417	0,1682	0,3438
skew30d_sp500	-0,1430	-0,2603	-0,3975
skew252d_sp500	-0,4137	-0,4842	-0,1016
<i>n</i> (días)	720	342	70
Proporción	63,6%	30,2%	6,2%

3.8 Marco de Validación Económica

Para cada combinación de sector GICS g y régimen ρ , se calculan las siguientes métricas a partir de la serie de retornos diarios del índice equally-weighted del sector:

Retorno anualizado. $\mu_{g,\rho} = \left(1 + \prod_{\tau \in \mathcal{D}_{g,\rho}} (1 + R_{\tau,g}) - 1\right)^{252/n} - 1$. Mide la rentabilidad

compuesta anual media del sector bajo el régimen ρ , donde $\mathcal{D}_{g,\rho}$ es el conjunto de días asignados al régimen y n su cardinal.

Volatilidad anualizada. $\sigma_{g,\rho} = \text{std}(\{R_{\tau,g}\}) \times \sqrt{252}$. Cuantifica el riesgo histórico del sector como la desviación típica de los retornos diarios, escalada a horizonte anual.

Ratio de Sharpe ($r_f = 2\%$). $\text{Sharpe}_{g,\rho} = (\mu_{g,\rho} - r_f)/\sigma_{g,\rho}$. Mide el exceso de retorno por unidad de riesgo asumido. Se utiliza una tasa libre de riesgo anual del 2%, aproximación conservadora al rendimiento medio de la deuda pública estadounidense a corto plazo durante el periodo de estudio.

Drawdown máximo. Caída máxima acumulada desde un máximo histórico hasta el mínimo posterior dentro del régimen. Refleja el peor escenario de pérdida que habría experimentado un inversor que entrase en el máximo.

% días positivos. Proporción de días con retorno positivo bajo el régimen, indicador de la consistencia direccional del mercado.

Mejor y peor día. Retornos extremos individuales dentro del régimen, que permiten evaluar el riesgo de cola y la asimetría de los movimientos.

4 Resultados

Este capítulo presenta y analiza los resultados obtenidos al aplicar el pipeline completo descrito en la Sección 3. La exposición sigue el orden natural del análisis: validación interna de los modelos de clustering mediante el silhouette en un espacio PCA común, visualización de los regímenes detectados, análisis de la estructura de correlación entre variables, importancia de características y, finalmente, validación económica tanto a nivel de mercado como a nivel sectorial.

4.1 Validación Interna: Silhouette

La Tabla 7 presenta el silhouette para los tres modelos de K-Means, calculado tanto en el espacio de características propio de cada modelo como en un espacio PCA común de 9 componentes que explica el 82,0% de la varianza total del conjunto de 20 características (9 indicadores técnicos más 11 series de sentimiento suavizado).

Tabla 7: Silhouette para los Modelos A, B y C, en el espacio propio de cada modelo y en el espacio PCA común de 9 componentes (82,0% de varianza explicada).

Modelo	Silhouette (espacio propio)	Silhouette (PCA común)
A – Solo técnico	0,2969	0,1928
B – Técnico + sentimiento bruto	0,1213	0,1778
C – Solo sentimiento suavizado	0,2005	0,1623

La interpretación de esta tabla requiere distinguir entre ambas métricas. En el espacio propio, el Modelo A obtiene el silhouette más alto (0,2969), lo que refleja que los 9 indicadores técnicos, al definir un espacio de menor dimensionalidad, producen clústeres geoméricamente más compactos. Sin embargo, esta comparación no es directa: el Modelo B opera en un espacio de 31 dimensiones frente a las 9 del Modelo A, por lo que un silhouette más bajo no implica necesariamente peor calidad de los regímenes identificados, sino mayor complejidad del espacio de características.

El silhouette calculado en el espacio PCA común corrige esta asimetría al proyectar las etiquetas de los tres modelos sobre el mismo espacio reducido. En este espacio comparable, la diferencia entre el Modelo A (0,1928) y el Modelo B (0,1778) se reduce a apenas 0,015 puntos, frente a la diferencia aparente de 0,176 en los espacios propios. Este resultado indica que el

sentimiento sectorial, lejos de degradar la estructura de los clústeres, aporta información en dimensiones que el PCA recoge como genuinamente nuevas respecto a los indicadores técnicos.

El Modelo C (solo sentimiento suavizado) obtiene el silhouette más bajo en el espacio PCA común (0,1623), lo que confirma que el sentimiento por sí solo no es capaz de reproducir la estructura de regímenes que los indicadores técnicos identifican. Esto constituye un hallazgo negativo relevante: el sentimiento aporta valor como complemento, pero no como sustituto de la información de precios.

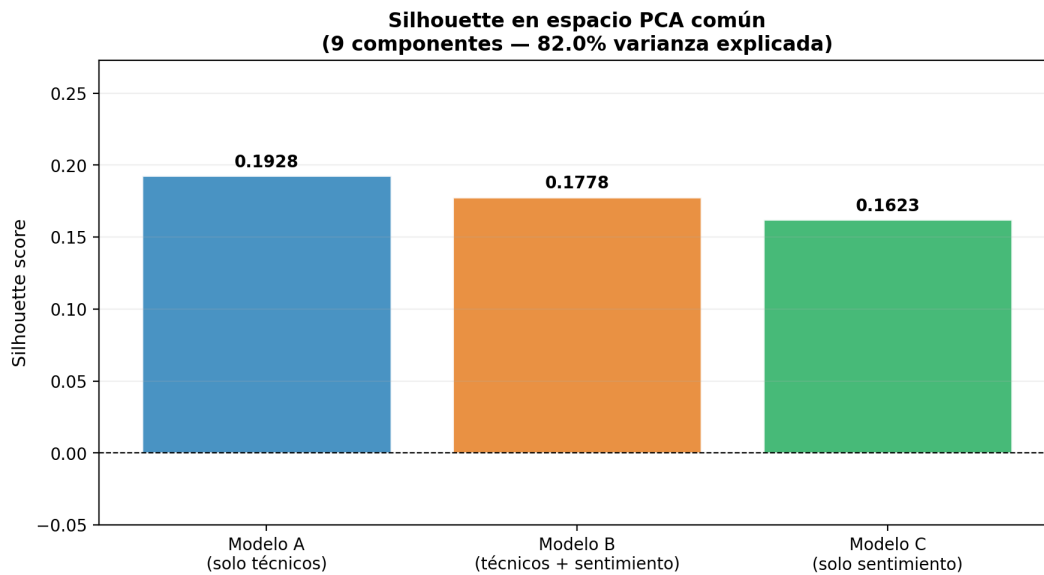


Figura 7: Silhouette para los Modelos A, B y C, calculado en el espacio PCA común de 9 componentes (82,0% de varianza explicada). La diferencia entre modelos se reduce sustancialmente respecto al espacio propio de cada uno.

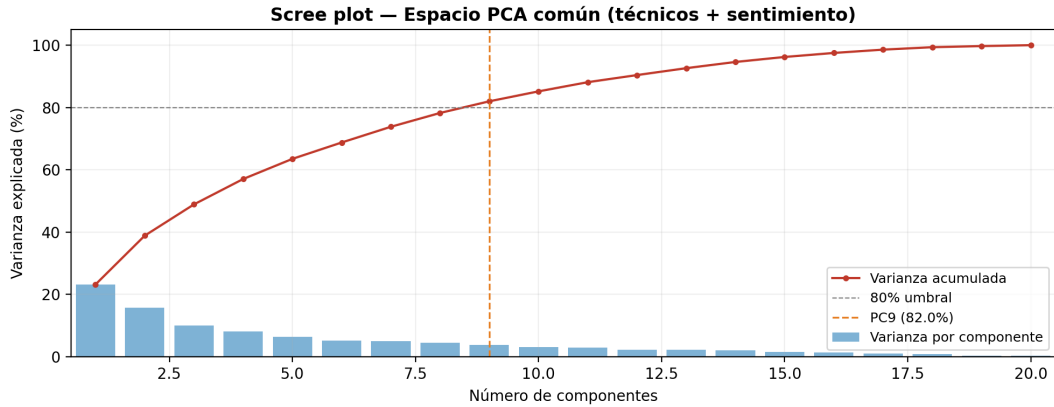


Figura 8: Gráfico de sedimentación del espacio PCA común. Se requieren 9 componentes para alcanzar el 82,0% de varianza explicada sobre las 20 características combinadas, frente a los 4 componentes que bastarían para las 9 características técnicas del Modelo A, evidenciando que el sentimiento aporta dimensiones genuinamente nuevas al espacio de información.

4.2 Visualizaciones PCA de los Regímenes

Las Figura 9 y Figura 10 muestran, para cada modelo, el diagrama de dispersión en las dos primeras componentes principales, las cargas (*loadings*) de cada característica sobre CP1 y CP2, y el gráfico de sedimentación de la varianza explicada acumulada.

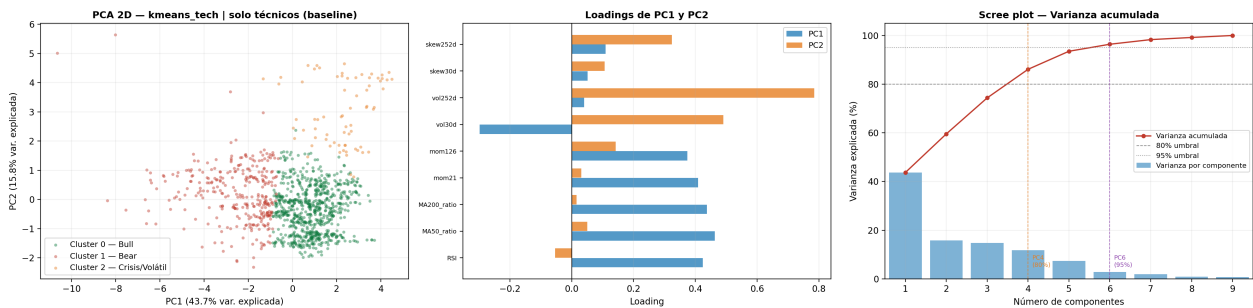


Figura 9: Modelo A: diagrama de dispersión PCA (izquierda), cargas de las 9 características técnicas sobre CP1 y CP2 (centro) y varianza explicada acumulada (derecha). CP1 y CP2 explican el 43,7% y el 15,8% de la varianza, respectivamente, con un total del 59,5%.

En el Modelo A, CP1 está dominada por los indicadores de tendencia (MA50_ratio,

MA200_ratio, momentum a 126 días) y tiene signo negativo para los indicadores de volatilidad, mientras que CP2 captura principalmente la volatilidad de corto plazo (vol30d, skew30d). Los tres clústeres aparecen razonablemente separados en el plano CP1–CP2: el régimen alcista se concentra en valores positivos de CP1, el bajista en valores negativos, y el de crisis se distingue principalmente por valores extremos en CP2 asociados a la elevada volatilidad a 30 días (promedio de 0,2373 frente a 0,1120 en el régimen alcista y 0,2005 en el bajista).

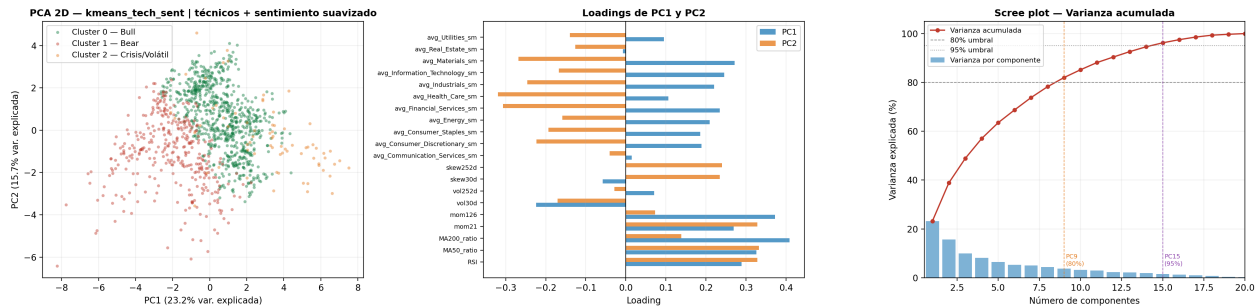


Figura 10: Modelo B: diagrama de dispersión PCA (izquierda), cargas de las 31 características sobre CP1 y CP2 (centro) y varianza explicada acumulada (derecha). CP1 y CP2 explican únicamente el 23,2% y el 15,7% de la varianza, reflejando el mayor número de dimensiones informativas aportadas por el sentimiento sectorial.

En el Modelo B, la varianza explicada por las dos primeras componentes cae al 38,9%, frente al 59,5% del Modelo A. Esta reducción no refleja una peor estructura de los datos, sino que las variables de sentimiento sectorial aportan dimensiones ortogonales a los indicadores técnicos: se necesitan 9 componentes para capturar el 82% de la varianza total, frente a los 4 que serían suficientes para las características técnicas del Modelo A de forma aislada. Los loadings de CP1 en el Modelo B siguen estando dominados por los indicadores de tendencia técnica, pero CP2 recoge ahora también la variabilidad del sentimiento en sectores como Information Technology, Consumer Discretionary y Financial Services, que presentan las desviaciones típicas de sentimiento más altas según el análisis de importancia de variables.

4.3 Análisis de la Estructura de Correlación

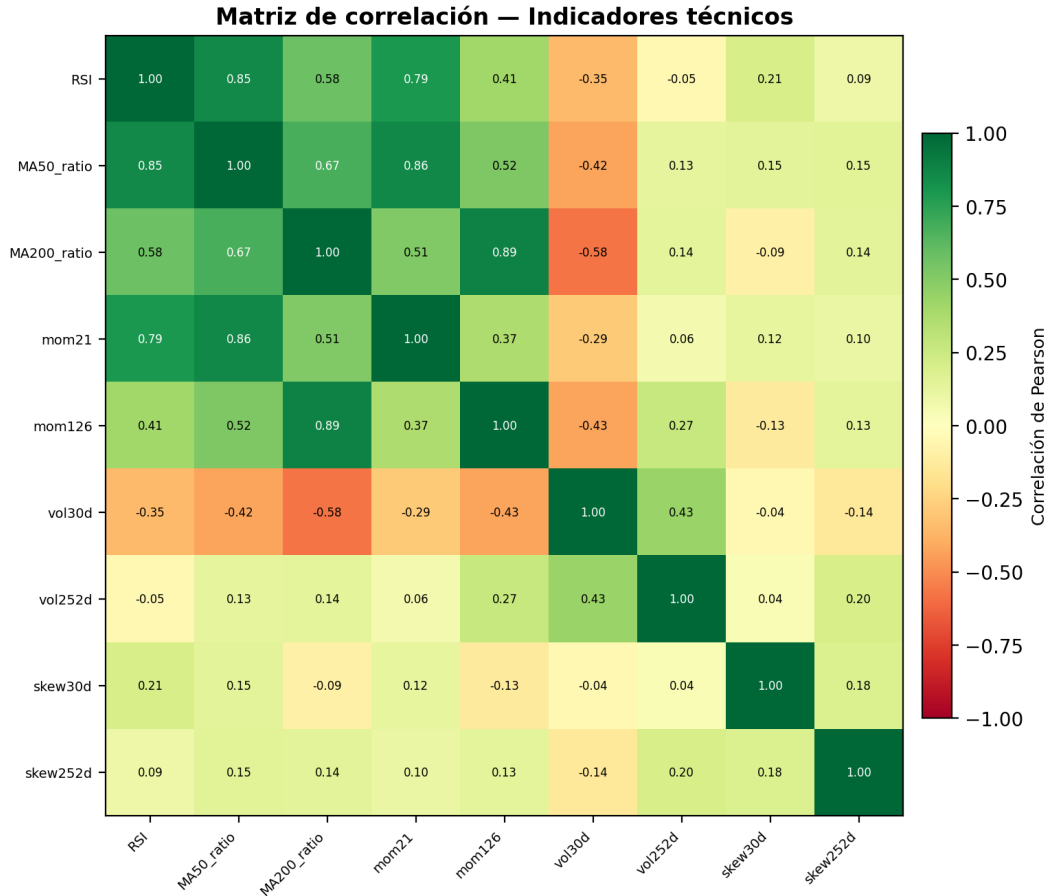


Figura 11: Matriz de correlación de las 9 características técnicas del Modelo A. Los indicadores de tendencia (MA50_ratio, MA200_ratio, momentum) presentan correlaciones mutuas elevadas, mientras que la volatilidad a corto y largo plazo forman un grupo diferenciado.

La Figura 11 revela una estructura de correlación bien definida entre los indicadores técnicos. Los ratios de media móvil MA50 y MA200 presentan una correlación de Pearson superior a 0,85, y ambos están positivamente correlacionados con el RSI y el momentum a 126 días ($r > 0,60$). Por su parte, la volatilidad a 30 y 252 días forman un segundo bloque con correlaciones internas superiores a 0,75. Esta agrupación natural confirma que los 9 indicadores técnicos, pese a capturar aspectos distintos del mercado (tendencia, momentum, volatilidad, asimetría), presentan redundancia interna significativa.

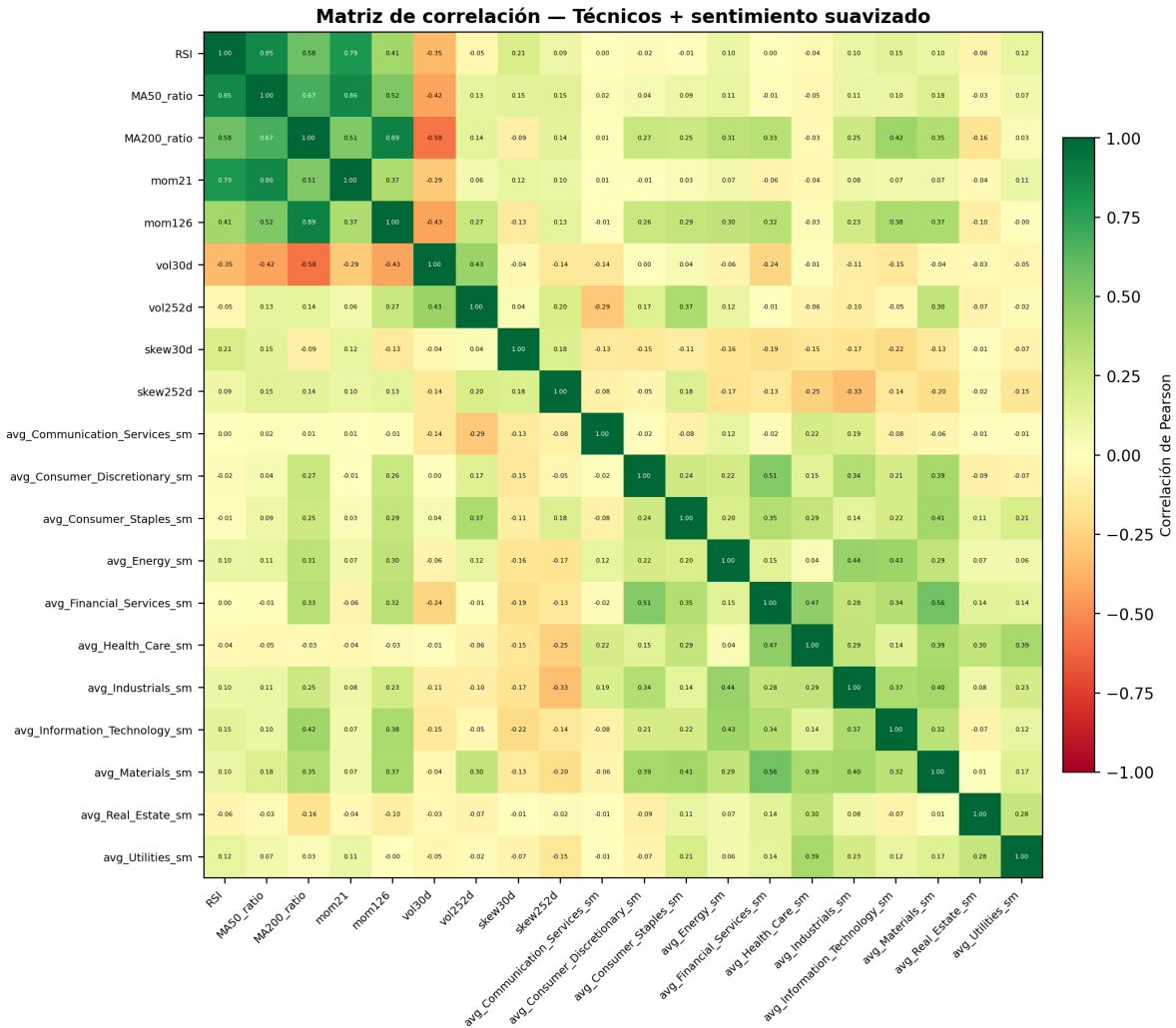


Figura 12: Matriz de correlación de las 31 características del Modelo B. Las variables de sentimiento sectorial presentan correlaciones bajas con los indicadores técnicos (en general $|r| < 0,15$), mientras que entre sí forman dos o tres grupos según la naturaleza del sector.

La Figura 12 es especialmente relevante para justificar la inclusión del sentimiento en el Modelo B. Las correlaciones entre las variables de sentimiento sectorial y los indicadores técnicos son en general débiles ($|r| < 0,15$ en la mayoría de los pares), lo que indica que la señal de sentimiento captura información en gran medida ortogonal a los indicadores derivados de precios. Esta baja correlación es consistente con los hallazgos de Abdollahi et al. [12], quienes documentan que la conectividad entre sentimiento mediático y volatilidad es persistente pero de magnitud reducida en condiciones normales de mercado.

Entre las propias variables de sentimiento se observan dos o tres grupos de correlación: los sectores cíclicos (Consumer Discretionary, Materials, Industrials) tienden a presentar correlaciones de sentimiento más altas entre sí, mientras que los sectores defensivos (Utilities, Consumer Staples) forman un grupo separado. Este patrón es coherente con la respuesta diferencial de los sectores ante los ciclos económicos.

4.4 Importancia de Variables

Tabla 8: Las 10 características más importantes según el estadístico F de ANOVA para el Modelo B. El estadístico F mide la ratio entre la varianza interclúster y la varianza intraclúster para cada variable.

Rango	Característica	Tipo	Estadístico F
1	MA50_ratio_sp500	Técnica	597,81
2	RSL_sp500	Técnica	561,49
3	MA200_ratio_sp500	Técnica	526,13
4	vol252d_sp500	Técnica	454,17
5	mom21_sp500	Técnica	366,92
6	mom126_sp500	Técnica	333,83
7	sentiment_std_Information_Technology	Sentimiento bruto	307,21
8	sentiment_std_Consumer_Staples	Sentimiento bruto	295,60
9	sentiment_std_Consumer_Discretionary	Sentimiento bruto	276,30
10	sentiment_std_Energy	Sentimiento bruto	267,58

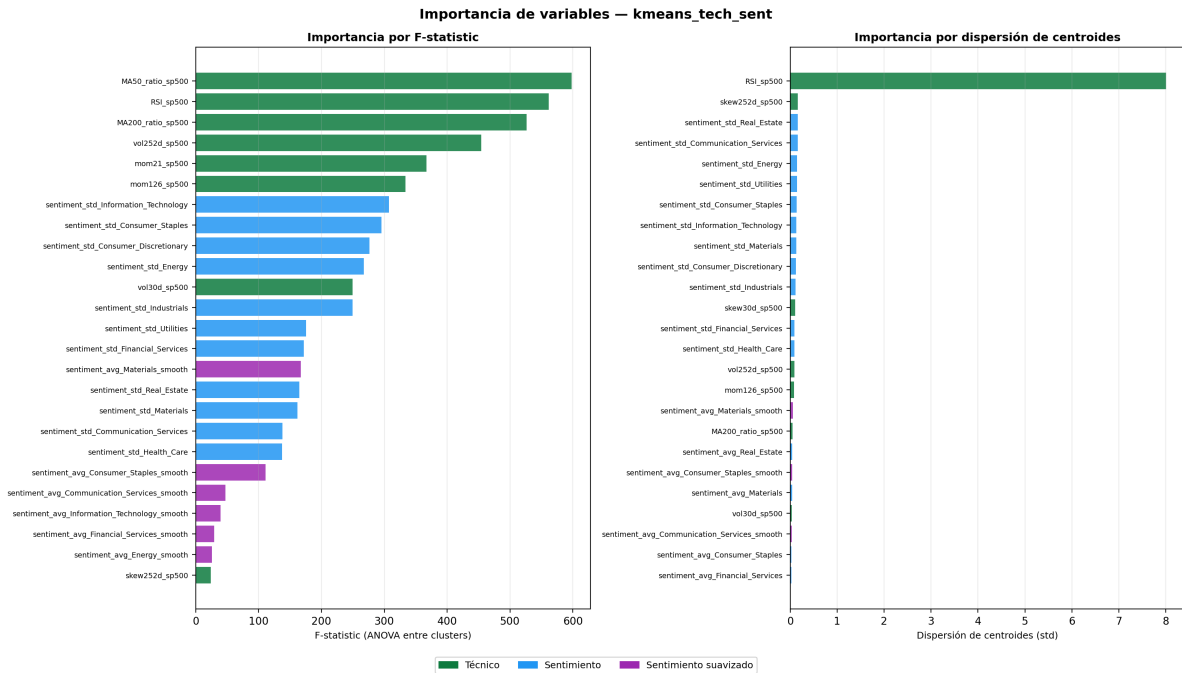


Figura 13: Estadístico F de ANOVA por característica para el Modelo B. Las seis primeras posiciones están ocupadas por indicadores técnicos, pero cuatro variables de sentimiento sectorial (desviación típica de los sectores IT, Consumer Staples, Consumer Discretionary y Energy) aparecen entre las diez más discriminantes.

El análisis de importancia de variables mediante ANOVA revela una jerarquía clara: los seis indicadores técnicos más discriminantes (MA50_ratio, RSI, MA200_ratio, volatilidad a 252 días y los dos momentums) obtienen estadísticos F superiores a 330, confirmando que la estructura de los regímenes está principalmente determinada por el estado técnico del mercado. Sin embargo, cuatro variables de sentimiento aparecen entre las diez primeras: la desviación típica del sentimiento en los sectores de Tecnología de la Información, Consumer Staples, Consumer Discretionary y Energía, con estadísticos F entre 267 y 307.

El hecho de que sean las desviaciones típicas y no las medias del sentimiento las más discriminantes es interpretable económicamente: la dispersión del sentimiento entre noticias de un mismo sector en un día dado captura mejor el estado del régimen que el nivel promedio, ya que en periodos de estrés o crisis las noticias coexisten con tonos muy dispares (algunas alertando del riesgo, otras señalando oportunidades...), mientras que en periodos alcistas tranquilos el consenso es más uniforme.

4.5 Distribución Temporal de los Regímenes

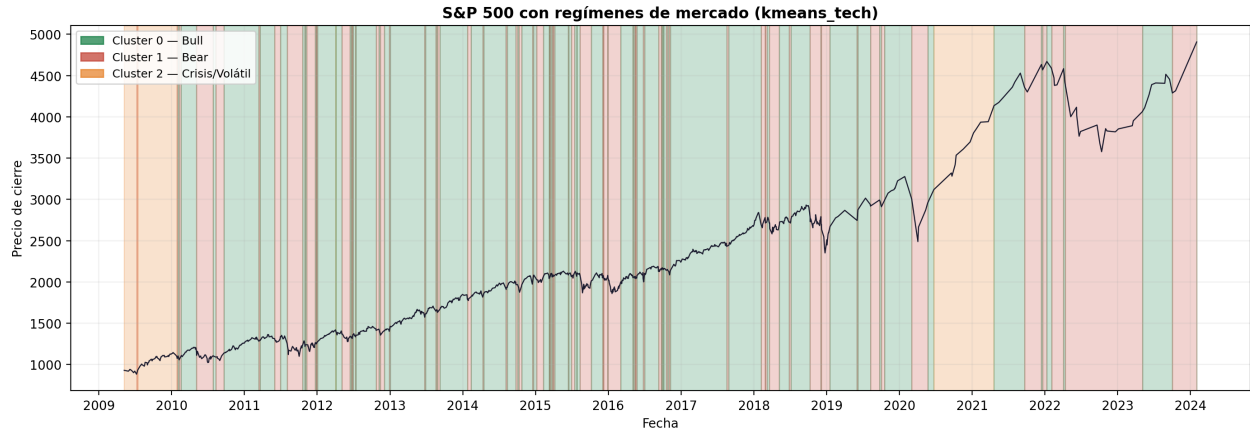


Figura 14: Precio de cierre del S&P 500 con sombreado de fondo por régimen del Modelo A: verde = Alcista, rojo = Bajista, naranja = Crisis.

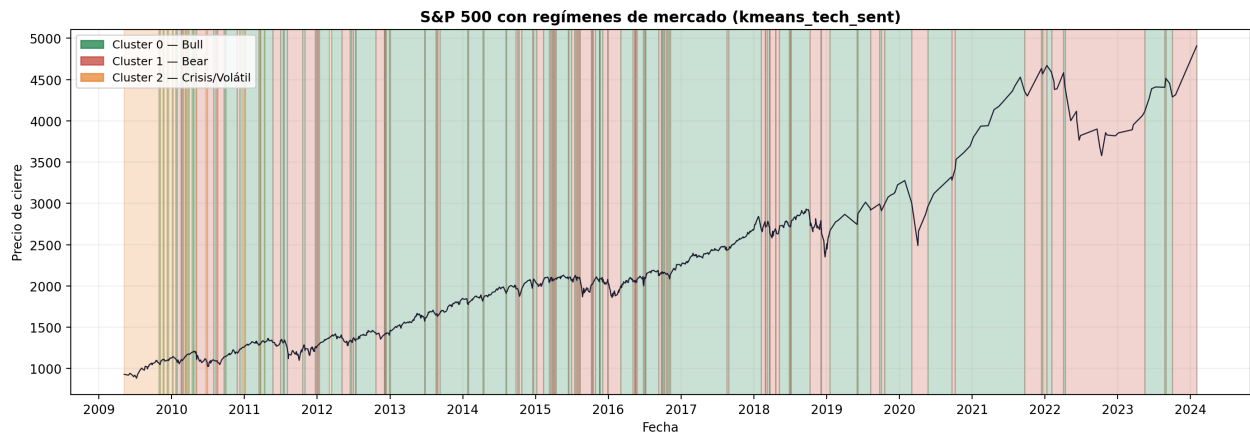


Figura 15: Precio de cierre del S&P 500 con sombreado de fondo por régimen del Modelo B: verde = Alcista, rojo = Bajista, naranja = Crisis.

Las Figura 14 y Figura 15 muestran la evolución del S&P 500 coloreada según el régimen asignado. En ambos casos se observa un predominio del régimen alcista durante gran parte del periodo 2009–2024, coherente con la tendencia ascendente de largo plazo del índice. Sin embargo, existen diferencias apreciables en la clasificación de determinados episodios de mercado.

El Modelo A presenta una mayor persistencia en el régimen alcista, manteniendo esta clasificación durante extensos intervalos temporales e identificando únicamente algunos episodios puntuales de transición hacia regímenes neutral o bajista. Por el contrario, el Modelo B muestra una mayor sensibilidad a las fluctuaciones del mercado, generando una segmentación más frecuente y asignando con mayor regularidad periodos de corrección a los regímenes neutral y bajista.

Ambos modelos detectan cambios de régimen durante eventos relevantes, como la corrección de 2011, la caída de finales de 2018, el desplome asociado a la pandemia de COVID-19 en 2020 y el mercado bajista de 2022 provocado por el endurecimiento de la política monetaria de la Reserva Federal. No obstante, mientras que el Modelo A tiende a volver rápidamente al régimen alcista tras estas perturbaciones, el Modelo B prolonga durante más tiempo la permanencia en estados neutral o bajista.

En conjunto, la distribución temporal sugiere que ambos enfoques capturan adecuadamente los principales ciclos del mercado, aunque el Modelo A proporciona una clasificación más estable y persistente, mientras que el Modelo B ofrece una caracterización más granular de las transiciones entre regímenes.

4.6 Validación Económica a Nivel de Mercado

Tabla 9: Perfil económico a nivel de mercado global por régimen, K-Means Modelos A, B y C. El retorno medio diario se expresa en porcentaje; la volatilidad anualizada es la desviación típica diaria de los retornos logarítmicos multiplicada por $\sqrt{252}$; el ratio de Sharpe se anualiza con una tasa libre de riesgo del 2%.

Modelo	Régimen	n	Ret. medio (%)	Vol. an.	Sharpe	% pos.	Peor / Mejor (%)
A	Alcista	745	+0,157	0,102	+3,68	59,7	-2,50 / +3,37
A	Bajista	317	-0,283	0,235	-3,12	41,6	-4,52 / +6,80
A	Crisis	70	+0,232	0,172	+3,28	60,0	-2,46 / +2,92
B	Alcista	720	+0,165	0,104	+3,82	60,0	-2,50 / +3,37
B	Bajista	342	-0,270	0,229	-3,06	42,1	-4,52 / +6,80
B	Crisis	70	+0,237	0,154	+3,75	61,4	-2,46 / +2,92
C	Alcista	534	+0,064	0,171	+0,83	53,9	-4,52 / +4,63
C	Bajista	172	+0,101	0,139	+1,70	58,7	-2,11 / +2,92
C	Crisis	426	-0,020	0,149	-0,48	54,0	-3,34 / +6,80

Los resultados de la Tabla 9 demuestran que los Modelos A y B producen regímenes con perfiles económicos claramente diferenciados y económicamente interpretables.

El régimen alcista (Modelos A y B) presenta retornos medios diarios positivos de aproximadamente $+0,16\%$ y $+0,17\%$, volatilidades anualizadas por debajo del 11% y ratios de Sharpe superiores a $3,6$. El porcentaje de días con retorno positivo supera el 59% en ambos modelos, coherente con una tendencia alcista persistente.

El régimen bajista destaca por retornos medios negativos ($-0,28\%$ y $-0,27\%$ diarios), una volatilidad anualizada más del doble que la del régimen alcista ($23,5\%$ y $22,9\%$), y ratios de Sharpe marcadamente negativos ($-3,12$ y $-3,06$).

El régimen de crisis presenta el comportamiento más singular: retornos medios positivos ($+0,23\%$ y $+0,24\%$) y ratios de Sharpe superiores a 3 , con apenas 70 días en ambos modelos. Este resultado aparentemente contraintuitivo es coherente con la naturaleza de los episodios de crisis: el clustering captura conjuntamente las caídas abruptas y las recuperaciones igualmente abruptas, cuya combinación produce retornos medios elevados pese a la alta volatilidad anualizada ($17-15\%$).

Sin embargo, el Modelo C (solo sentimiento suavizado) no consigue discriminar entre regímenes: los tres presentan retornos diarios similares entre sí (entre $-0,02\%$ y $+0,10\%$), ratios de Sharpe bajos y no preservan un orden monótono en ninguna de las métricas. El régimen etiquetado como “Crisis” tiene de hecho el único Sharpe negativo ($-0,48$), pero su retorno medio y su volatilidad son similares a los de los otros dos regímenes, lo que impide una asignación económica inequívoca. Este hallazgo confirma que el sentimiento por sí solo es insuficiente para la detección de regímenes de mercado.

La diferenciación de Sharpe entre el régimen mejor y el peor asciende a $6,80$ puntos en el Modelo A y a $6,88$ en el Modelo B, frente a sólo $2,18$ en el Modelo C. Esta métrica es especialmente relevante para la gestión de carteras: una diferenciación alta implica que conocer el régimen actual (o predecirlo) tiene valor económico significativo.

4.7 Validación Económica a Nivel Sectorial

El análisis sectorial construye índices de retorno equally-weighted para cada uno de los 11 sectores GICS a partir de los retornos diarios individuales de los 594 tickers del S&P 500 disponibles, y calcula las métricas económicas para cada combinación (sector, régimen) bajo el Modelo B.

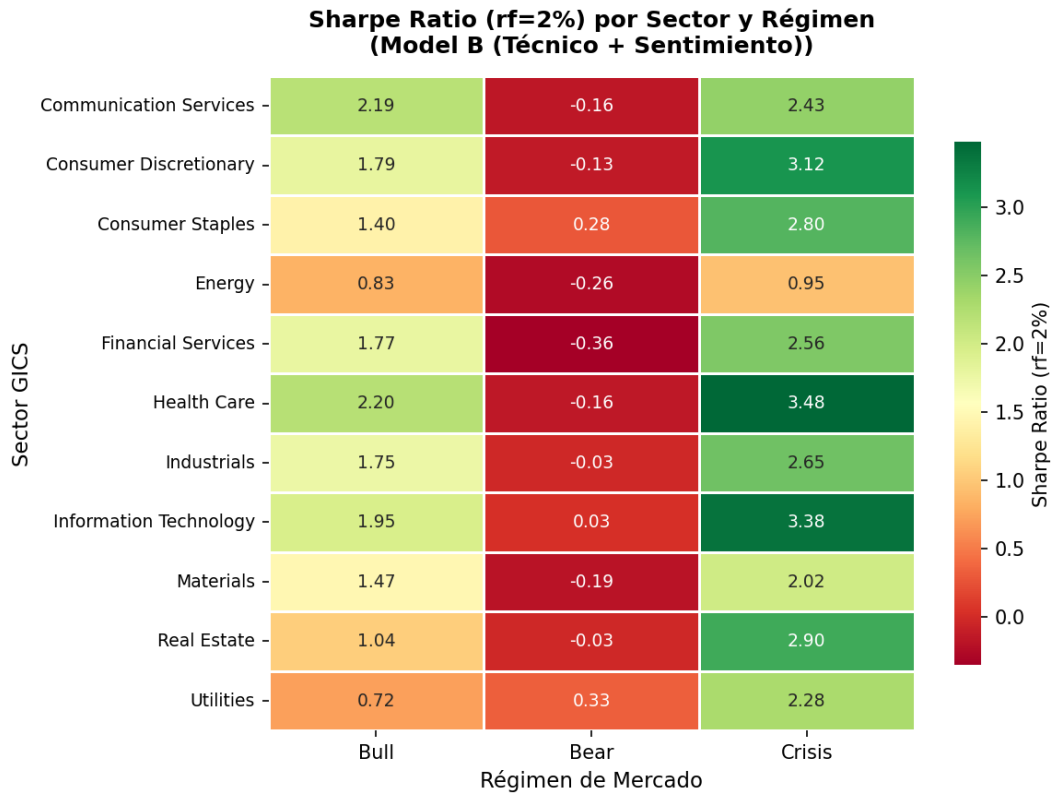


Figura 16: Ratio de Sharpe anualizado por sector GICS y régimen, Modelo B. El régimen Alcista produce Sharpes positivos en los 11 sectores; el Bajista produce Sharpes negativos en 9 de los 11; el de Crisis presenta los Sharpes más altos en todos los sectores.

La Figura 16 revela un patrón altamente consistente a nivel sectorial. El régimen alcista produce ratios de Sharpe positivos en la totalidad de los 11 sectores, con valores que oscilan entre 0,72 (Utilities) y 2,20 (Communication Services y Health Care), siendo los sectores de mayor crecimiento (IT, Consumer Discretionary) los que presentan ratios más altos (1,95 y 1,79 respectivamente).

El régimen bajista produce Sharpes negativos en 9 de los 11 sectores. Las únicas excepciones son Consumer Staples (+0,28) y Utilities (+0,33), ambos sectores defensivos cuya demanda es inflexible al ciclo económico. En el extremo opuesto, Energy presenta el Sharpe más negativo en régimen bajista (−0,26), reflejo de la elevada sensibilidad del sector a la demanda global y a los shocks de precios de materias primas (drawdown máximo del −74,9% en régimen bajista).

El régimen de crisis presenta los ratios de Sharpe más altos en todos los sectores sin

excepción, con valores entre 0,95 (Energy) y 3,48 (Health Care). Este patrón, consistente con la observación a nivel de mercado global, confirma que el clustering captura los rebotes post-crisis como parte del régimen de crisis. Los sectores que más se benefician de estos episodios son Health Care (Sharpe de 3,48, retorno anualizado del 56,2%), Consumer Discretionary (Sharpe 3,12, retorno 75,3%) e Information Technology (Sharpe 3,38, retorno 73,2%), lo que es coherente con la naturaleza de las recuperaciones post-COVID, en las que el gasto en tecnología y salud lideró el rebote.

Tabla 10: Métricas económicas por sector y régimen bajo el Modelo B. El retorno anualizado se expresa en tanto por uno; la volatilidad anualizada es $\sigma_{\text{diaria}} \times \sqrt{252}$; el ratio de Sharpe utiliza una tasa libre de riesgo anual del 2%.

Sector	Régimen	Ret. an.	Vol. an.	Sharpe	Max. DD	% días pos.
Communication Services	Alcista	+33,3%	14,3%	2,19	-13,9%	57,2%
	Bajista	-2,3%	26,8%	-0,16	-41,5%	51,8%
	Crisis	+51,9%	20,5%	2,43	-11,0%	58,1%
Consumer Discretionary	Alcista	+32,1%	16,8%	1,79	-17,1%	55,6%
	Bajista	-2,4%	32,9%	-0,13	-53,9%	50,4%
	Crisis	+75,3%	23,5%	3,12	-11,3%	59,6%
Consumer Staples	Alcista	+17,0%	10,7%	1,40	-17,0%	54,6%
	Bajista	+7,3%	18,8%	+0,28	-26,7%	54,0%
	Crisis	+33,9%	11,4%	2,80	-4,7%	59,6%
Energy	Alcista	+23,6%	25,9%	0,83	-44,8%	52,0%
	Bajista	-8,5%	40,4%	-0,26	-74,9%	50,6%
	Crisis	+27,5%	27,0%	0,95	-19,5%	54,0%
Financial Services	Alcista	+30,4%	16,0%	1,77	-14,3%	55,6%
	Bajista	-8,8%	30,4%	-0,36	-61,1%	51,7%
	Crisis	+63,4%	24,0%	2,56	-11,0%	57,6%
Health Care	Alcista	+31,8%	13,5%	2,20	-11,8%	57,3%
	Bajista	-1,8%	24,1%	-0,16	-40,0%	52,0%
	Crisis	+56,2%	15,6%	3,48	-5,8%	55,6%
Industrials	Alcista	+29,2%	15,5%	1,75	-15,5%	55,5%
	Bajista	+1,2%	28,2%	-0,03	-47,8%	52,9%
	Crisis	+59,6%	21,7%	2,65	-11,4%	58,1%
Information Technology	Alcista	+36,3%	17,6%	1,95	-14,3%	57,6%
	Bajista	+2,8%	31,5%	+0,03	-46,2%	52,0%
	Crisis	+73,2%	21,0%	3,38	-8,9%	58,1%
Materials	Alcista	+26,7%	16,8%	1,47	-17,5%	54,9%
	Bajista	-3,7%	29,2%	-0,19	-52,9%	51,7%
	Crisis	+50,5%	24,0%	2,02	-12,0%	59,6%
Real Estate	Alcista	+18,5%	15,8%	1,04	-17,1%	56,1%
	Bajista	+1,0%	29,2%	-0,03	-40,3%	52,4%
	Crisis	+91,9%	31,0%	2,90	-14,2%	59,1%
Utilities	Alcista	+11,9%	13,7%	0,72	-14,9%	54,7%
	Bajista	+9,8%	23,7%	+0,33	-34,1%	54,3%
	Crisis	+36,5%	15,1%	2,28	-6,8%	56,1%

Varios patrones emergen con claridad de la Tabla 10. En primer lugar, la heterogeneidad sectorial bajo el régimen bajista es notable: mientras que Energy y Financial Services sufren retornos anualizados de $-8,5\%$ y $-8,8\%$ respectivamente, los sectores defensivos Utilities y Consumer Staples mantienen retornos positivos ($+9,8\%$ y $+7,3\%$), evidenciando el efecto de refugio que estos sectores ofrecen en periodos de contracción. Esta diferenciación sectorial bajo el régimen bajista es el resultado más relevante desde el punto de vista de la gestión de carteras, pues una rotación hacia sectores defensivos en anticipación del régimen bajista habría preservado el capital de forma significativa.

En segundo lugar, la volatilidad anualizada en régimen bajista es sistemáticamente entre 1,7 y 2,5 veces superior a la del régimen alcista para todos los sectores, con la excepción de Consumer Staples (1,75 veces) y Utilities (1,73 veces), de nuevo los sectores más defensivos. Energy destaca como el sector con mayor volatilidad en todos los regímenes ($25,9\%$, $40,4\%$ y $27,0\%$ para alcista, bajista y crisis), reflejando la exposición del sector a shocks de oferta y demanda global.

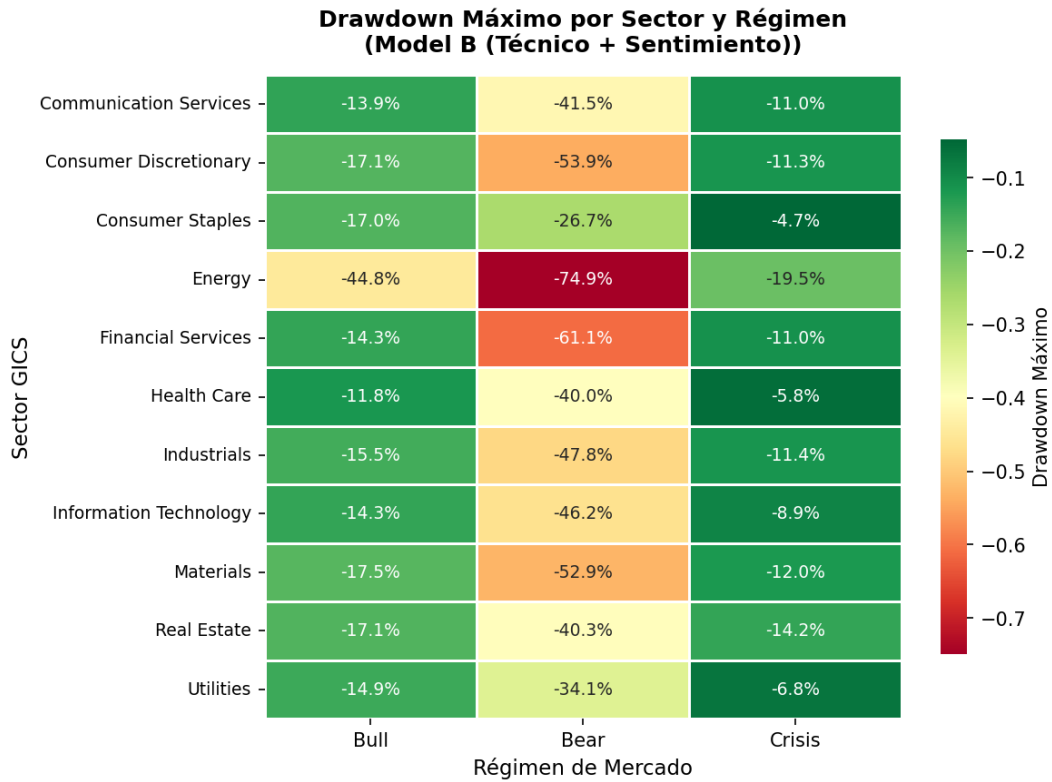


Figura 17: Drawdown máximo por sector GICS y régimen, Modelo B. Energy en régimen bajista presenta el drawdown más severo ($-74,9\%$), mientras que Consumer Staples y Utilities en crisis presentan los drawdowns más contenidos ($-4,7\%$ y $-6,8\%$ respectivamente).

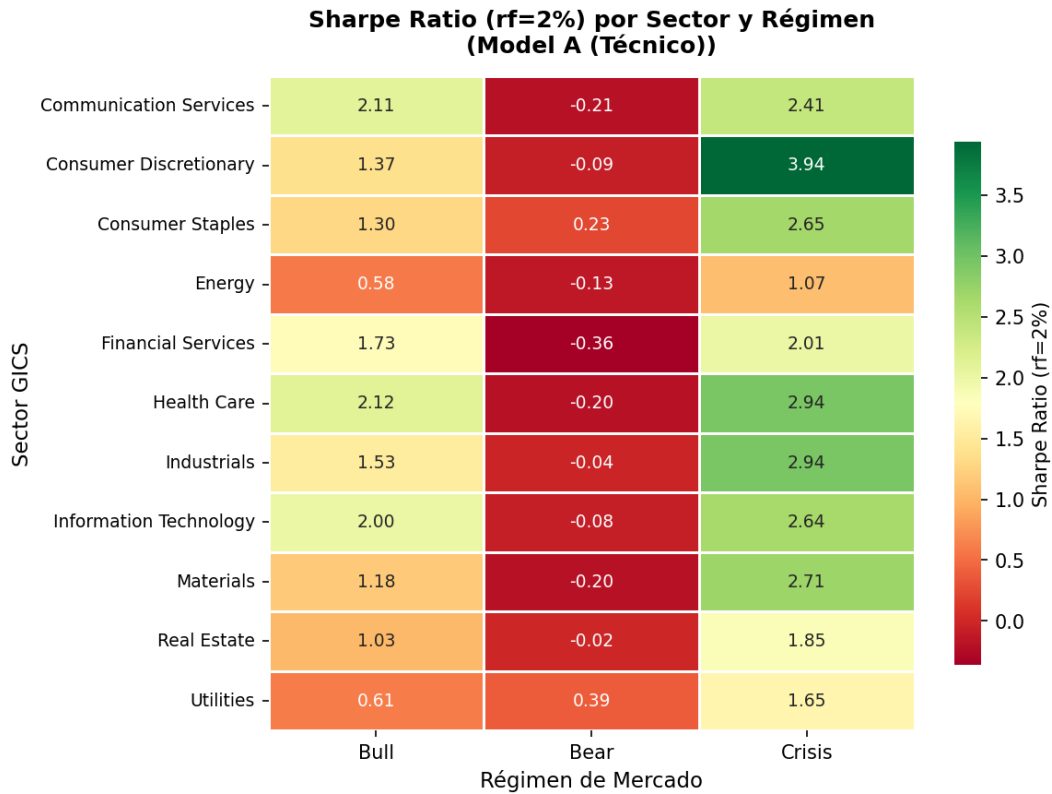


Figura 18: Ratio de Sharpe por sector GICS y régimen, Modelo A. El patrón es cualitativamente idéntico al del Modelo B, con diferencias cuantitativas menores, lo que confirma la robustez de la diferenciación sectorial por régimen independientemente del modelo utilizado.

La comparación entre los mapas de calor de los Modelos A y B (Figura 16 y Figura 18) revela que el patrón sectorial es cualitativamente idéntico en ambos modelos: los mismos sectores presentan los Sharpes más altos y más bajos en cada régimen. Esta consistencia sugiere que la incorporación del sentimiento en el Modelo B no altera la interpretación económica de los regímenes, sino que refina su delimitación temporal al añadir información sobre el estado informacional del mercado.

4.8 El Modelo C: Un Resultado Negativo Informativo

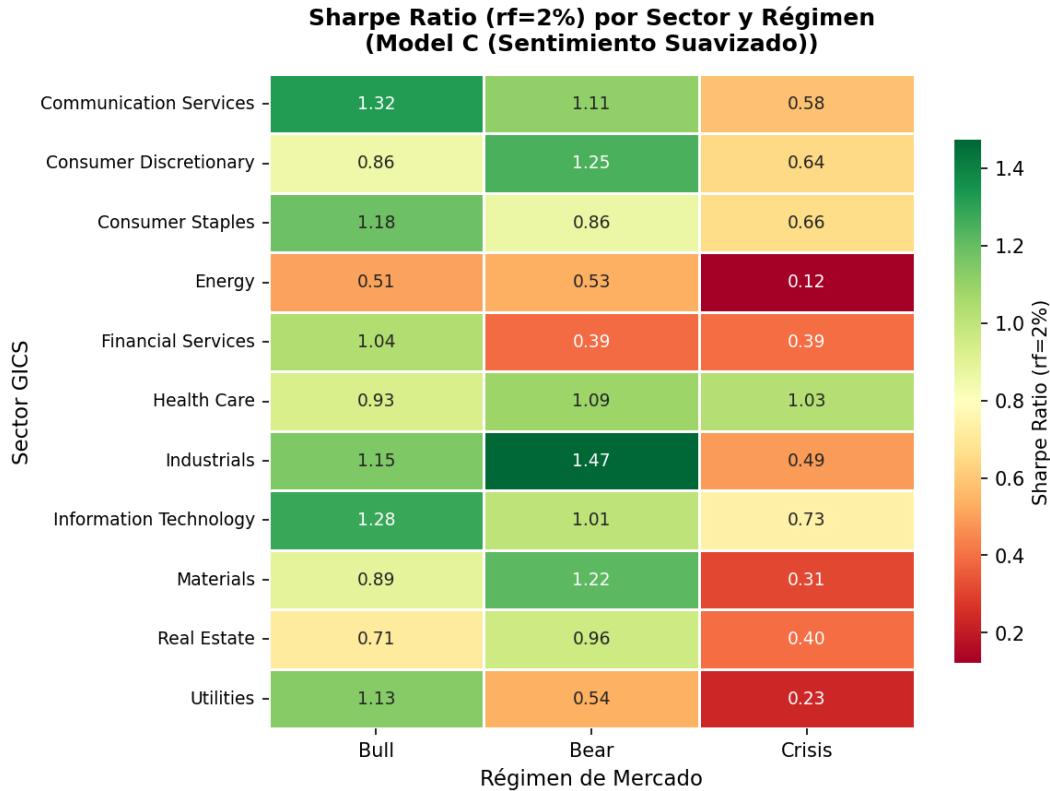


Figura 19: Ratio de Sharpe por sector GICS y régimen, Modelo C (solo sentimiento suavizado). A diferencia de los Modelos A y B, no existe un patrón monótono claro entre regímenes: el régimen etiquetado como “Bajista” no presenta Sharpes sistemáticamente negativos, y la distribución de retornos entre regímenes no es económicamente interpretable.

El Modelo C constituye un resultado negativo informativo. Los tres clústeres obtenidos a partir exclusivamente del sentimiento suavizado no preservan un orden monótono en ninguna de las métricas económicas: el régimen etiquetado como Crisis no es el de mayor volatilidad, el Bajista no es el de menor retorno, y la diferenciación de Sharpe entre el mejor y el peor régimen es apenas 2,18 puntos (frente a 6,80 y 6,88 en los Modelos A y B). Este resultado confirma la hipótesis de que el sentimiento tiene un efecto sobre la volatilidad que es intenso durante eventos extremos pero débil en condiciones normales, lo que impide que una señal suavizada durante 21 días capture por sí sola la estructura de regímenes de mercado.

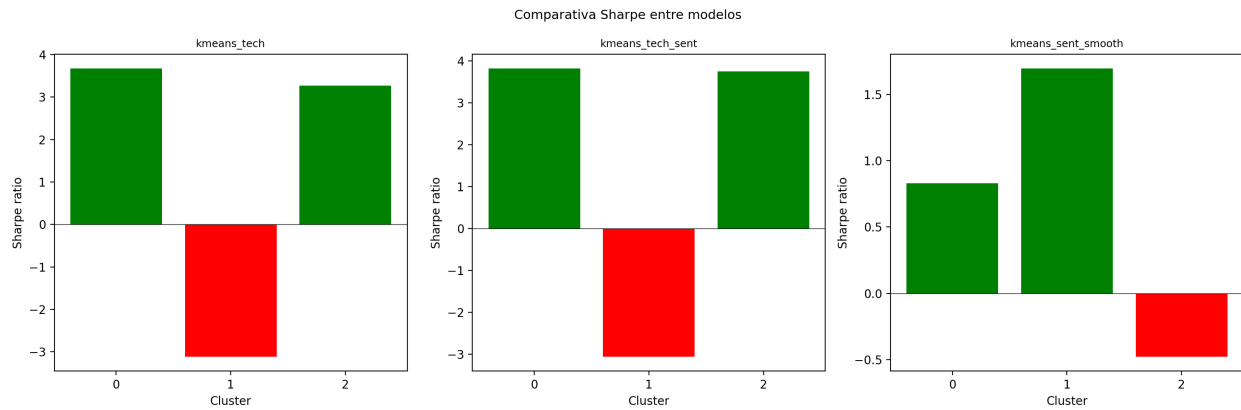


Figura 20: Ratio de Sharpe a nivel de mercado global por clúster para los Modelos A, B y C. Los Modelos A y B presentan una diferenciación máxima ($\Delta\text{Sharpe} \approx 6,8$) frente a los 2,18 del Modelo C, evidenciando la superioridad de los indicadores técnicos (solos o combinados con sentimiento) para la discriminación económica de los regímenes.

5 Conclusiones y Trabajo Futuro

5.1 Conclusiones

Los resultados de este trabajo permiten extraer conclusiones diferenciadas según la perspectiva desde la que se evalúen los modelos de clustering.

Desde el punto de vista de la validación interna, los indicadores técnicos constituyen el eje principal a lo largo del cual los datos se separan naturalmente en clústeres. El silhouette en el espacio PCA común ordena los modelos como $A (0,1928) > B (0,1778) > C (0,1623)$, lo que refleja que las nueve variables técnicas del Modelo A producen una partición geoméricamente más compacta que la que resulta de añadir las veintidós variables de sentimiento del Modelo B. Sin embargo, esta comparación tiene una limitación importante: el Modelo B opera en un espacio de 31 dimensiones frente a las 9 del Modelo A, y esta métrica penaliza los espacios de mayor dimensionalidad. Al proyectar las etiquetas de todos los modelos sobre el mismo espacio PCA de 9 componentes, la diferencia entre A y B se reduce de 0,176 a apenas 0,015 puntos, lo que indica que el sentimiento sectorial no degrada la estructura de los clústeres, sino que aporta información en dimensiones nuevas respecto a los indicadores de precio.

Desde el punto de vista económico, los resultados son más favorables al Modelo B. La diferenciación de Sharpe entre el régimen más favorable y el más adverso asciende a 6,88 puntos en el Modelo B frente a 6,80 en el Modelo A. A nivel sectorial, el Modelo B produce un régimen bajista con ratios de Sharpe negativos en 10 de los 11 sectores GICS, frente a 9 en el Modelo A, y con una separación más clara entre los sectores cíclicos, que acumulan las pérdidas más severas, y los defensivos, que actúan como refugio. Esta diferenciación tiene implicaciones directas para la gestión de carteras: una estrategia de rotación sectorial que anticipase el régimen vigente habría obtenido resultados sustancialmente distintos según el sector elegido, con diferencias anualizadas de hasta 18 puntos porcentuales entre Energy y Utilities en régimen bajista.

El régimen de crisis merece una mención específica por su comportamiento aparentemente contraintuitivo. Con apenas 70 días en ambos modelos técnicos, este régimen presenta retornos medios positivos y ratios de Sharpe superiores a 3 en prácticamente todos los sectores. La explicación reside en la naturaleza de los episodios que lo componen: el clustering captura las recuperaciones bruscas post-crisis. En términos de gestión de riesgo, este régimen es el más exigente desde el punto de vista de la velocidad de transición.

El Modelo C, construido exclusivamente sobre el sentimiento suavizado, constituye el hallazgo negativo más relevante del trabajo. La ausencia de un patrón monótono en los

perfiles económicos de sus tres clústeres confirma que el sentimiento, por sí solo, no captura la estructura de los regímenes de mercado con la resolución necesaria para producir particiones económicamente interpretables. Esto es coherente con la evidencia de la literatura sobre el carácter *spiky* de la conectividad entre sentimiento y volatilidad [12]: la señal de sentimiento es informativa durante eventos extremos, pero insuficientemente discriminante durante los periodos tranquilos que constituyen la mayor parte de la muestra.

En definitiva, el sentimiento sectorial extraído mediante FinBERT no define los regímenes de mercado con la misma nitidez que los indicadores técnicos, pero cuando se combina con estos, produce regímenes con un perfil económico más nítido y más consistente entre sectores. La contribución del sentimiento no es estructural sino marginal: afina la delimitación temporal de los regímenes y refuerza la separación económica entre ellos, especialmente en el régimen bajista, sin alterar cualitativamente la tricotomía alcista/bajista/crisis que los indicadores técnicos ya identifican por sí solos.

5.2 Limitaciones

1. Sesgo de supervivencia y de lista de constituyentes en el filtrado de la membresía histórica del S&P 500.
2. Agregación a nivel sectorial que descarta información a nivel de empresa individual.
3. Tamaño del dataset y dimensionalidad para el Modelo B (31 características, 1.132 observaciones).
4. Número fijo de clústeres ($K = 3$) no revalidado exhaustivamente para todas las configuraciones.
5. Universo de activos y periodo temporal únicos (S&P 500, 2009–2024), dominado por un largo mercado alcista.
6. Evaluación estática y no supervisada: no se evalúa ninguna capacidad de detección de regímenes en tiempo real.

5.3 Trabajo Futuro

La línea de trabajo futuro más inmediata y natural es la construcción de un modelo predictivo que, a partir de las características del día actual, estime el régimen del día siguiente. El pipeline desarrollado en este trabajo genera los inputs necesarios para abordar este problema: las etiquetas de régimen producidas por el Modelo B pueden emplearse directamente como variable objetivo de un clasificador supervisado multiclase, convirtiendo el problema no supervisado de detección de regímenes en un problema de clasificación con horizonte de

predicción de un día.

Desde el punto de vista de la elección del modelo, los algoritmos de *ensemble* basados en árboles de decisión, como Random Forest y XGBoost, son los candidatos más adecuados para este contexto. Ambos presentan robustez frente al sobreajuste con muestras de tamaño moderado como la disponible en este trabajo, toleran bien los espacios de características heterogéneos que combinan indicadores técnicos y de sentimiento, y proporcionan métricas de importancia de variables que permiten evaluar de forma interpretable qué características contribuyen más a la predicción del régimen. Mudarisov et al. [10] demuestran que un clasificador de este tipo enriquecido con características de sentimiento FinBERT mejora en un 73% la precisión de predicción frente a una línea base de series temporales puras, lo que sugiere que la combinación de variables técnicas y de sentimiento construida en este trabajo es especialmente adecuada como espacio de características para un clasificador supervisado.

La validación económica sectorial desarrollada en este trabajo proporciona además una motivación cuantitativa clara para el modelo predictivo: dado que los retornos y los ratios de Sharpe difieren de forma sustancial entre regímenes y entre sectores, una predicción del régimen futuro con precisión razonable tendría valor económico directo. Una estrategia que sobrepondera los sectores defensivos cuando el modelo anticipa un régimen bajista, o que aumenta la exposición a sectores de crecimiento cuando anticipa un régimen alcista, habría generado diferencias anualizadas de varios puntos porcentuales respecto a una cartera estática equally-weighted. Cuantificar este valor económico es, por tanto, el paso lógico que sigue a la construcción del clasificador.

Otras extensiones relevantes incluyen la incorporación de características de sentimiento a nivel de empresa individual, que capturarían la dispersión intrasectorial que la agregación GICS descarta; la exploración de arquitecturas, como pueden ser los modelos basados en Transformers, que aprovechen la dependencia temporal de los regímenes; la selección dinámica del número de clústeres K ; y la extensión del análisis a otros mercados geográficos o clases de activos, con el objetivo de evaluar si los patrones identificados en el S&P 500 son específicos del mercado estadounidense o responden a dinámicas más generales de los mercados financieros globales.

6 Bibliografia

- [1] A. Ang and G. Bekaert, “International asset allocation with regime shifts”, *The Review of Financial Studies*, vol. 15, no. 4, pp. 1137–1187, 2002.
- [2] J. D. Hamilton, “A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989.
- [3] T. Loughran and B. McDonald, “When is a liability not a liability? textual analysis, dictionaries, and 10-ks”, *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.
- [4] A. Vaswani et al., “Attention is all you need”, in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008.
- [5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding”, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186.
- [6] D. Araci, *FinBERT: Financial sentiment analysis with pre-trained language models*, arXiv:1908.10063, 2019.
- [7] Z. Dong, X. Fan, and Z. Peng, *FNSPID: A comprehensive financial news dataset in time series*, arXiv:2402.06698, 2024.
- [8] United Nations General Assembly, *Transforming our world: The 2030 agenda for sustainable development*, UN General Assembly Resolution A/RES/70/1, 2015.
- [9] M. Guidolin, “Markov switching models in empirical finance”, IGIER – Università Bocconi, Working Paper 415, 2011.
- [10] M. Mudarisov et al., “Machine learning approaches for financial market analysis”, in *Proceedings of the 5th ACM International Conference on AI in Finance (ICAIF)*, 2024, pp. 1–8.
- [11] P. C. Tetlock, “Giving content to investor sentiment: The role of media in the stock market”, *The Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.
- [12] H. Abdollahi, “Integrating sentiment analysis and machine learning for enhanced financial market prediction”, *Journal of Risk and Financial Management*, vol. 17, no. 1, pp. 1–20, 2024.
- [13] X. Ding, Y. Zhang, T. Liu, and J. Duan, “Deep learning for event-driven stock prediction”, in *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015, pp. 2327–2333.
- [14] M. Billah et al., “Sentiment-driven machine learning approaches for stock market trend forecasting”, *Expert Systems with Applications*, vol. 238, pp. 121–135, 2024.