**Análisis Detallado de Clústeres de Empresas para Predicción de Quiebra**

El análisis de clústeres (KMeans sobre componentes PCA) ha identificado tres grupos distintos de empresas dentro de nuestro conjunto de datos. Estos grupos muestran perfiles financieros y tasas de quiebra muy diferentes, lo que los convierte en una herramienta valiosa para contextualizar el riesgo.

**Resumen General de los Clústeres**

A continuación, se presenta el conteo de empresas y la proporción de quiebras para cada clúster:

|  |  |  |  |
| --- | --- | --- | --- |
| **Clúster** | **Conteo de Empresas** | **Proporción de Quiebras (Clase 1)** | **Observación Principal** |
| **0** | 2085 empresas | **0.14%** | **Bajo Riesgo / Saludables** |
| **1** | 4 empresas | **75.00%** | **¡Alto Riesgo Extremo!** |
| **2** | 4730 empresas | **4.52%** | **Riesgo Moderado / Promedio** |

*(Nota: Los porcentajes de quiebra se refieren a la proporción de empresas que quebraron dentro de ese clúster en el conjunto de datos de entrenamiento.)*

**Interpretación Detallada de Cada Clúster**

Para entender lo que significa cada clúster, analizamos la media de las características financieras (valores escalados) de las empresas que lo componen. Un valor positivo indica que la característica es superior a la media general del dataset, mientras que un valor negativo indica que es inferior.

**Clúster 0: El Grupo de Bajo Riesgo (Empresas Saludables)**

* **Conteo:** 2085 empresas (aproximadamente el 30.5% del total).
* **Tasa de Quiebra:** **0.14%** (Extremadamente baja).
* **Significado:** Este clúster representa a las **empresas más sólidas y financieramente estables** en el conjunto de datos. Son las empresas con menor probabilidad de quiebra.
* **Perfil Financiero Típico (Basado en Medias Escaladas):**
  + **Rentabilidad y Flujo de Caja (Valores Altos y Positivos):**
    - roac\_before\_interest\_and\_depreciation\_before\_interest (0.786)
    - roaa\_before\_interest\_and\_percent\_after\_tax (0.744)
    - operating\_gross\_margin (0.475)
    - cash\_flow\_rate (0.560)
    - total\_income\_total\_expense (0.045)
    - cash\_flow\_to\_total\_assets (0.506)
    - cfo\_to\_assets (0.704)
    - persistent\_eps\_in\_the\_last\_four\_seasons (0.690)
    - retained\_earnings\_to\_total\_assets (0.476)
    - net\_income\_to\_stockholders\_equity (0.160)
    - *Implicación:* Estas empresas son altamente rentables, generan mucho flujo de caja de sus operaciones y retienen una buena parte de sus ganancias.
  + **Solvencia y Baja Deuda (Valores Altos Positivos o Bajos Negativos):**
    - debt\_ratio\_percent (-0.646)
    - equity\_to\_liability (0.479)
    - borrowing\_dependency (-0.241)
    - current\_liability\_to\_assets (-0.421)
    - working\_capital\_to\_total\_assets (0.815)
    - cash\_total\_assets (0.827)
    - *Implicación:* Tienen muy poca deuda en relación con sus activos o patrimonio, una alta proporción de capital propio, y una excelente liquidez y capital de trabajo.
  + **Crecimiento (Valores Cercanos a Cero o Ligeramente Negativos):**
    - total\_asset\_growth\_rate (-0.078)
    - *Implicación:* El crecimiento puede no ser explosivo, pero esto es normal para empresas ya establecidas y estables.

**Clúster 2: El Grupo de Riesgo Moderado (Empresas Promedio)**

* **Conteo:** 4730 empresas (aproximadamente el 69.3% del total).
* **Tasa de Quiebra:** **4.52%** (Significativamente más alta que el Clúster 0, pero no extrema).
* **Significado:** Este es el clúster más grande y representa a la **mayoría de las empresas con un perfil financiero promedio o ligeramente por debajo del promedio**. Aunque su tasa de quiebra es moderada, contiene la mayor cantidad de quiebras en términos absolutos debido a su tamaño.
* **Perfil Financiero Típico (Basado en Medias Escaladas):**
  + **Rentabilidad y Flujo de Caja (Valores Negativos):**
    - roac\_before\_interest\_and\_depreciation\_before\_interest (-0.344)
    - roaa\_before\_interest\_and\_percent\_after\_tax (-0.325)
    - operating\_gross\_margin (-0.209)
    - cash\_flow\_rate (-0.246)
    - *Implicación:* Estas empresas son menos rentables y generan menos flujo de caja que la media, lo que indica una presión sobre sus operaciones.
  + **Solvencia y Deuda (Valores Positivos o Bajos Negativos):**
    - debt\_ratio\_percent (0.282)
    - equity\_to\_liability (-0.211)
    - borrowing\_dependency (0.079)
    - current\_liability\_to\_assets (0.183)
    - *Implicación:* Tienen una mayor proporción de deuda, menor capital propio en relación con la deuda, y una liquidez y capital de trabajo más ajustados en comparación con las empresas saludables.
  + **Crecimiento (Valores Cercanos a Cero o Ligeramente Positivos):**
    - total\_asset\_growth\_rate (0.035)
    - *Implicación:* Pueden mostrar un ligero crecimiento, pero no lo suficiente para compensar las debilidades en rentabilidad y solvencia.

**Clúster 1: El Grupo de Alto Riesgo Extremo (Empresas en Crisis)**

* **Conteo:** **4 empresas** (¡el clúster más pequeño, solo el 0.06% del total!).
* **Tasa de Quiebra:** **75.00%** (¡Extremadamente alta! 3 de 4 empresas quebraron).
* **Significado:** Este clúster es un **indicador de alarma crítico**. Representa a empresas en una situación financiera muy precaria, con una altísima probabilidad de quiebra.
* **Perfil Financiero Típico (Basado en Medias Escaladas):**
  + **Rentabilidad y Flujo de Caja (Valores Extremadamente Bajos y Negativos):**
    - roac\_before\_interest\_and\_depreciation\_before\_interest (-2.837)
    - roaa\_before\_interest\_and\_percent\_after\_tax (-3.115)
    - operating\_gross\_margin (-0.769)
    - cash\_flow\_rate (-0.437)
    - persistent\_eps\_in\_the\_last\_four\_seasons (-2.974)
    - retained\_earnings\_to\_total\_assets (-2.024)
    - *Implicación:* Sufren pérdidas masivas, tienen una capacidad nula o negativa para generar flujo de caja de sus operaciones, y sus ganancias retenidas son muy negativas, indicando pérdidas acumuladas significativas.
  + **Solvencia y Deuda (Valores Extremadamente Altos Positivos o Bajos Negativos):**
    - debt\_ratio\_percent (3.345)
    - borrowing\_dependency (31.904)
    - contingent\_liabilities\_net\_worth (21.119)
    - inventory\_and\_accounts\_receivable\_net\_value (21.395)
    - equity\_to\_liability (-0.742)
    - working\_capital\_to\_total\_assets (-2.002)
    - current\_liability\_to\_assets (2.340)
    - current\_liabilities\_equity (29.931)
    - *Implicación:* Tienen niveles de deuda insostenibles, una dependencia extrema del endeudamiento, problemas graves con el inventario y las cuentas por cobrar, y una solvencia prácticamente inexistente (muchos más pasivos que patrimonio neto, o patrimonio neto negativo). La liquidez es crítica.
  + **Crecimiento (Valores Muy Negativos):**
    - total\_asset\_growth\_rate (-0.487)
    - *Implicación:* La empresa está en contracción severa.

**Conclusión y Utilidad para el Negocio**

Este análisis de clústeres es fundamental para entender el **contexto** de las predicciones de tu modelo supervisado.

* **Clúster 0:** Empresas a las que se les puede otorgar crédito o invertir con alta confianza.
* **Clúster 2:** Empresas que requieren un monitoreo más cercano o condiciones más estrictas. Contienen la mayoría de las quiebras, pero de forma más dispersa.
* **Clúster 1:** ¡Estas son las **señales de alarma más urgentes**! Cualquier empresa que caiga en este clúster, incluso si el modelo supervisado no la clasifica directamente como "quiebra" (lo cual sería un Falso Negativo muy peligroso), debería ser objeto de una investigación inmediata y profunda.

La combinación del modelo predictivo (RandomForest\_ClassWeighted) y este análisis de clústeres proporciona una visión poderosa y multidimensional del riesgo de quiebra empresarial.