

Modelo de evaluación de riesgos en startups

Marcenaro y Boffelli

¿Qué es una start up?



¿Qué es una start up?

- Emprendimiento reciente con idea innovadora
- Potencialmente muy rentable en el futuro
- Busca inversores para hacer despegar su idea
- Bajos costos iniciales, poco personal
- Representa una inversión de riesgo

Algunos ejemplos



El objetivo del proyecto

Intentaremos acercarnos a una solución para la cuestión de cómo saber cuándo una startup es una buena inversión mediante un análisis de los datos encontrados y un proceso de Machine Learning.



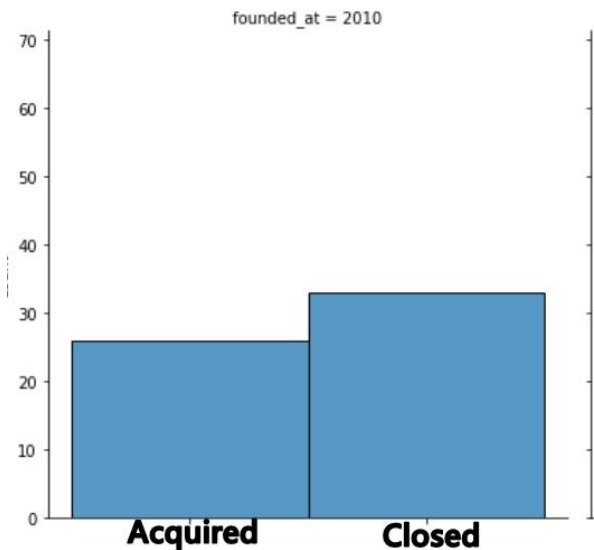
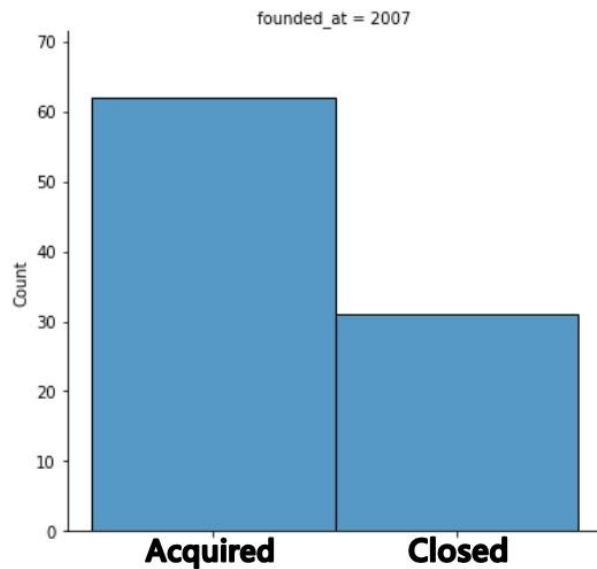
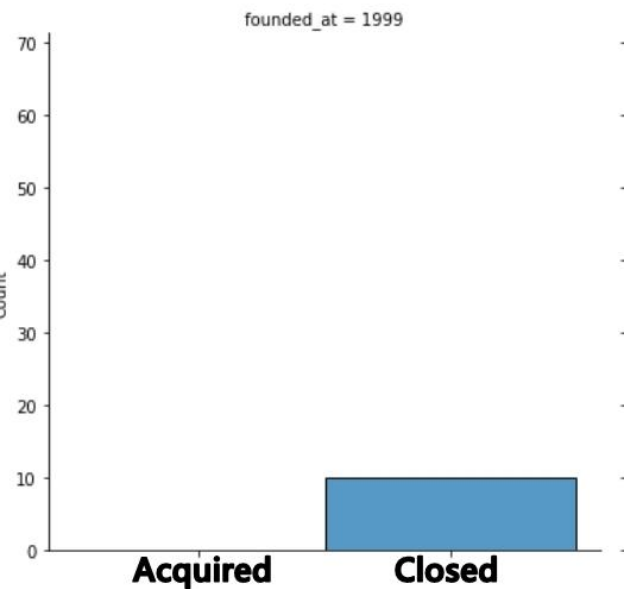
Fase I: EDA (Exploratory Data Analysis)

Se comienza a entender el dataset mientras se ubican datos faltantes o erróneos, realizando visualizaciones y distintos análisis y procesos.

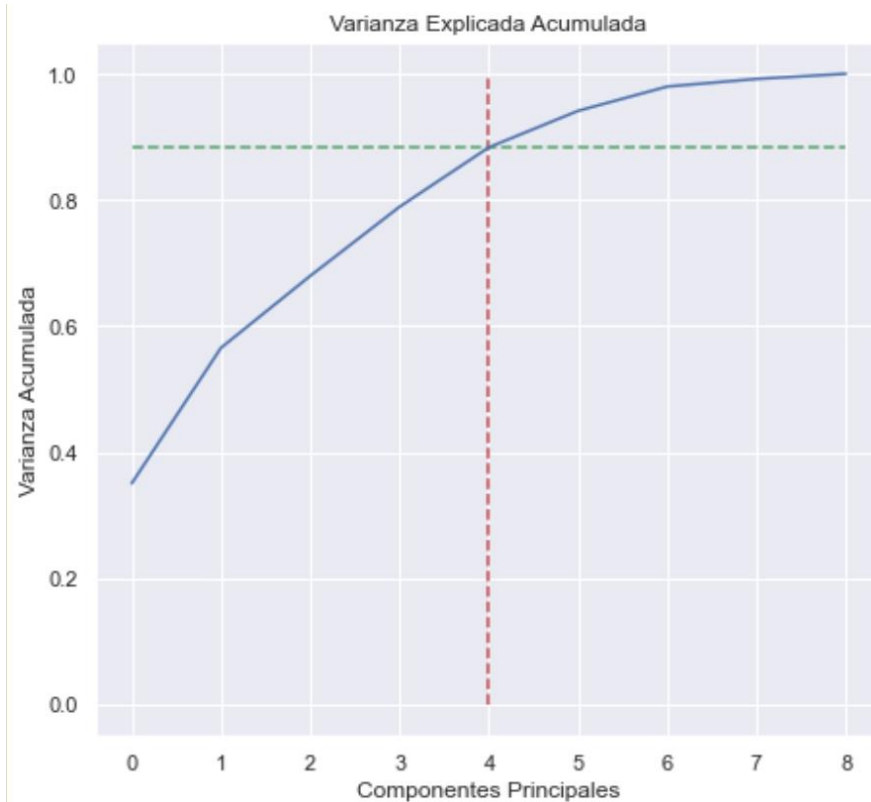
Algunos descubrimientos

- Datos equilibrados pero poca cantidad
- Eliminación de algunas columnas innecesarias
- Columnas con edades negativas
- Datos representativos de los procesos del mercado
- Se encontraron los 5 componentes numéricos que mejor explican la varianza

Algunos descubrimientos: correspondencia con comportamiento del mercado



Algunos descubrimientos: Componentes principales



Fase II: Búsqueda y aplicación del modelo

Debíamos aplicar un modelo de aprendizaje supervisado y de clasificación. Nuestras opciones fueron:

- Árbol de decisión
- Random Forest
- XG Boost
- Regresión logística

A continuación veremos los respectivos resultados a partir de las métricas de **accuracy score** y **matriz de confusión** para los modelos hipertuneados.

Accuracy Scores

Árbol de decisión

- 84% de aciertos sobre set de entrenamiento y 78% set de evaluación.

Random Forest *(Resultó ser el mejor para esta métrica)*

- 86% de aciertos sobre set de entrenamiento y 81% set de evaluación.

XG-Boost

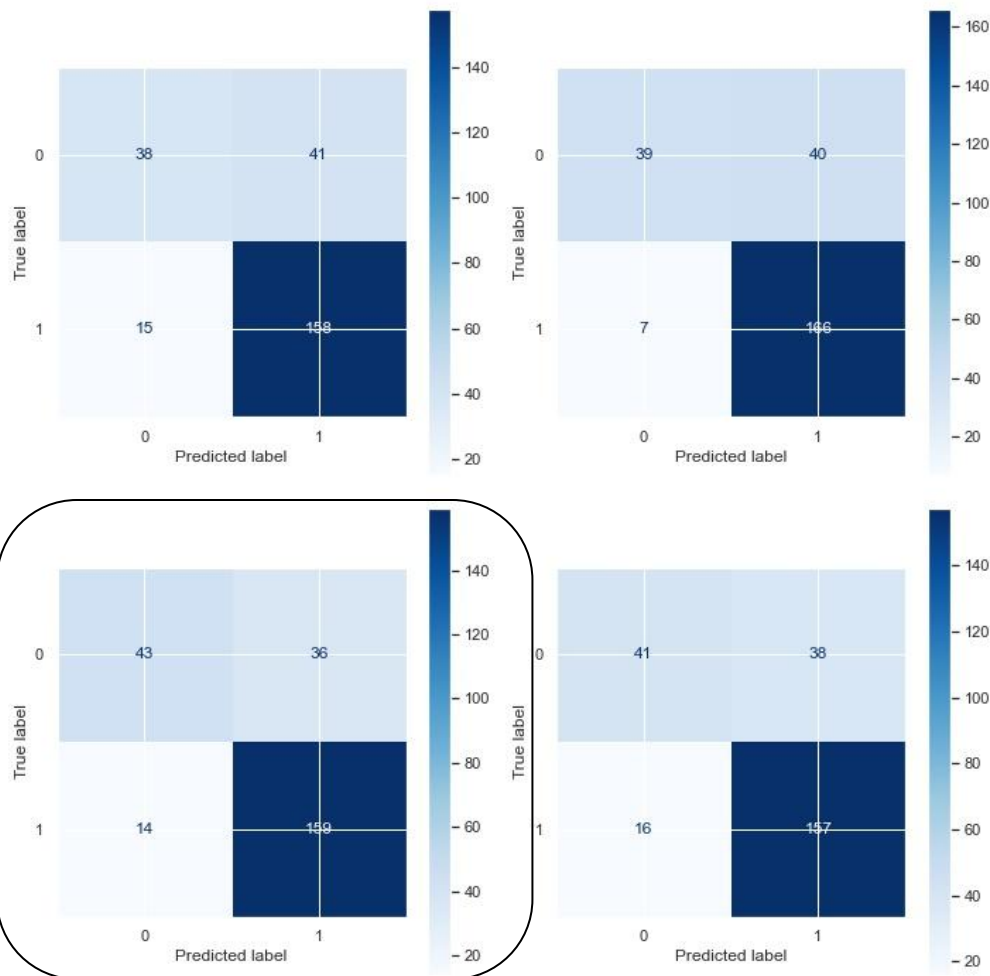
- 86% de aciertos sobre set de entrenamiento y 80% set de evaluación.

Regresión Logística

- 78% de aciertos sobre set de entrenamiento y 79% set de evaluación.

Matriz de Confusión

El modelo óptimo es aquel menos sensible a los **falsos positivos**. El modelo que mejor se adapta en este sentido es el de **XG-Boost**.



Conclusiones

- Las fluctuaciones de mercado deben ser tenidas en cuenta antes de invertir.
- El modelo puede ser tomado en cuenta por inversores o por quienes decidan continuar esta investigación.