Noviembre 03, 2022. Bogotá Colombia

**Universidad del Rosario – Maestría MACC**

**Presentado por Emilio Muñoz**

**Proyecto final Aprendizaje automático de maquina I**

**Visión general de una industria, negocio o problema.**

La industria del capital riesgo está compuesta por inversores financieros que destinan capital en nuevos modelos de negocio denominados Startups con tasas de éxito extremadamente inciertas. Esto hace que el retorno de las inversiones sea alto, pero también riesgoso. Por lo tanto, para gestionar el riesgo de las inversiones de riesgo, los fondos de capital siempre están buscando nuevas formas de reducir la incertidumbre y pronosticar el éxito de las nuevas empresas. Es por esto por lo que a través de la ciencia de datos podemos intentar predecir el éxito de un nuevo emprendimiento en base a los exitosos anteriores.

Este problema es relevante porque el ecosistema de emprendimiento es uno de los sectores más importantes en la economía actual del país. Es responsable de la creación de nuevos negocios capaces de generar miles de puestos de trabajo y también de desarrollar nuevas tecnologías y servicios que resuelvan problemas y dolores reales. Ayudar a reducir el riesgo involucrado en las inversiones podría llevar a los fondos de riesgo a asignar más capital y encontrar Startups más exitosas.

Por lo tanto, agregaría valor crear un modelo que contribuya a brindar sugerencias sobre la probabilidad de éxito de la industria de las startups colombianas, ya que es un mercado en crecimiento. Rappi es un claro ejemplo del impacto que tiene la industria de riesgo en la economía de un país. Ahora mismo es una de las empresas más valiosas del país y ha generado más de 150.000 empleos indirectos.

**Problema específico que se debe resolver.**

Primero, necesitamos entender el historial de éxito de las startups en Colombia con base en la información disponible en Pitchbook, Linkedin y Crunchbase. Con esta información, y una vez que identifiquemos las características clave que determinan el éxito de las startups, podríamos predecir la probabilidad de éxito o la valoración de una startup y dar recomendaciones de futuras inversiones.

**Para desarrollar el problema se definió la siguiente lista de pasos**

1. Carga y limpieza de datos.
2. EDA para entender los datos desde una perspectiva en conjunto e individual
3. Probar distintos modelos
   1. Regresiones lineales para predecir la valoración
   2. Redes neuronales para predecir la valoración
   3. Kprototypes, permite usar variables categóricas para formar clusters y así encontrar grupos de las startups. Adicional, gracias a la experiencia de negocio, se puede identificar cuáles son más propensos al éxito y así evaluar una nueva Startup dependiendo el grupo asignado por el algoritmo.
4. Diseñar un mecanismo para que se pueda evaluar una nueva startup y mejorar las oportunidades de éxito para el fondo que requiera este modelo.

**Resultados**

* Modelar la valoración mediante una regresión dio un r^2 de 0.72 después de utilizar Ridge. Lo que seria un modelo aceptable para empezar a evaluar la valoración de las Startups.
* La red neuronal planteada para predecir la valoración de una Startup presenta un pésimo desempeño y se sugiere utilizar métodos para optimizarla o cambiar su arquitectura.
* El modelo de Kprototypes es el que presento uno de los mejores resultados. En conjunto con el conocimiento de negocio se pudieron identificar 6 clusters relevantes y que son coherentes con los ciclos de vida de una startup. Los grupos identificados fueron los siguientes:
  + Validando producto (top)
  + Validando producto (Neutral)
  + Creciendo (top)
  + Creciendo (Luchando)
  + Mature (Pre IPO)
  + Mature (exit)
* Finalmente podemos decir que tanto la valoración como el grupo al que pertenece una nueva Startup que se este evaluando serian insumos ideales a la hora de tomar una decisión de inversión por parte de un fondo. Se podría hacer un backtesting si se consigue información del pasado de las startups evaluadas.

**Conclusiones**

Considero que este proyecto es un excelente primer acercamiento para tratar de utilizar datos para robustecer los procesos de identificación de oportunidades de inversión en un fondo de capital de riesgo. Adicional vemos como los distintos modelos se pueden complementar para tener incluso una versión mucho más integral sobre el posible éxito de una Startup. Los pasos siguientes podrían ser mejorar las fuentes de datos y optimizar los modelos, especialmente la red neuronal.