**Clase 3**

1. Implementar un [modelo para predecir la distribucion de rating](https://docs.google.com/presentation/d/19QWqaGuTCZgN8iMrbzdyfm7Xfw8jmk_coCpX_-C1qTg/edit#slide=id.g9e4e9bf97a_0_10) dada una película
   1. Evaluarlo con [métricas apropiadas a nuestro caso de negocio](https://docs.google.com/presentation/d/19QWqaGuTCZgN8iMrbzdyfm7Xfw8jmk_coCpX_-C1qTg/edit#slide=id.ga3cef763fd_0_74)
   2. Discutir: Vale la pena la complejidad que introduce este modelo?
2. Qué estrategias utilizarías para codificar las variables
   1. El género de una película
   2. Actores, directores, escritores
   3. Existe alguna otra relevante en el [dataset](https://www.imdb.com/interfaces/)?
   4. Tiene sentido complementar [feature based](https://docs.google.com/presentation/d/19QWqaGuTCZgN8iMrbzdyfm7Xfw8jmk_coCpX_-C1qTg/edit#slide=id.ga3cef763fd_0_105) con [embedding based](https://docs.google.com/presentation/d/19QWqaGuTCZgN8iMrbzdyfm7Xfw8jmk_coCpX_-C1qTg/edit#slide=id.ga3cef763fd_0_111)?
3. Analizar las implicancias de entrenar los embeddings para el crew para usarlas de features con
   1. Los datos crudos de imdb
   2. Los datos filtrados para que sean largometrajes posteriores a 1970
   3. Training + test set
   4. Solo training
4. Evaluar la siguiente hipótesis: El crew de una película es buen predictor del rating
5. Entrenar el un Word2Vec en algún escenario de los expuestos en el punto 2
   1. Alimentar con las representaciones de Word2Vec a un modelo para predecir el rating
   2. Evaluar si la representación aprendida por Word2Vec es útil para predecir el rating
6. Entrenar un Word2Vec en algún otro escenario del punto 2 y evaluarlo en el modelo. Evaluar los resultados en vista a
   1. Utilidad en relación al [producto que queremos construir](https://docs.google.com/presentation/d/19QWqaGuTCZgN8iMrbzdyfm7Xfw8jmk_coCpX_-C1qTg/edit#slide=id.ga3cef763fd_0_52)
   2. Overfitting / underfitting
   3. Bias / variance trade-off