# Práctica 9

# Procesamiento de Lenguaje Natural

Javier Pellejero Ortega & Zhaoyan Ni

Inteligencia Artificial

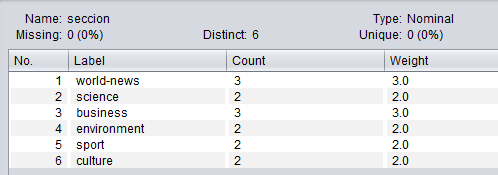
Grupo 11

Doble grado Matemáticas e ingeniería informática

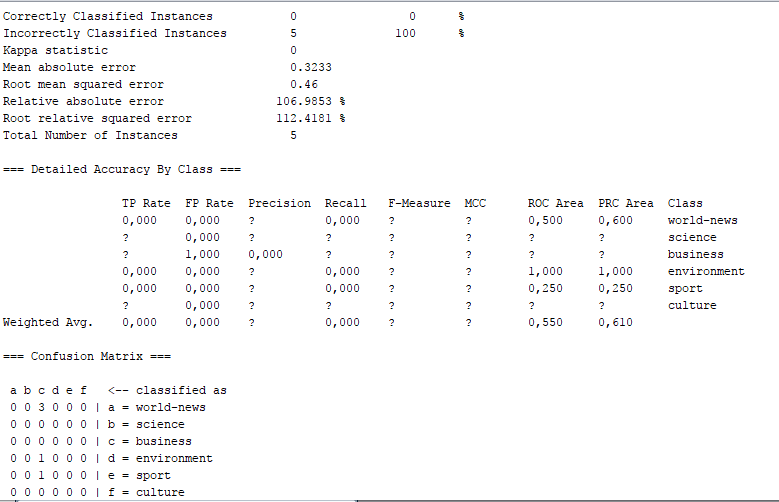
PARTE 2:

Para realizar esta parte, hemos buscado párrafos de noticias que están publicadas en periódicos digitales de 6 secciones diferentes: *World-new, Science, Business, Environment, Sport and Culture*.

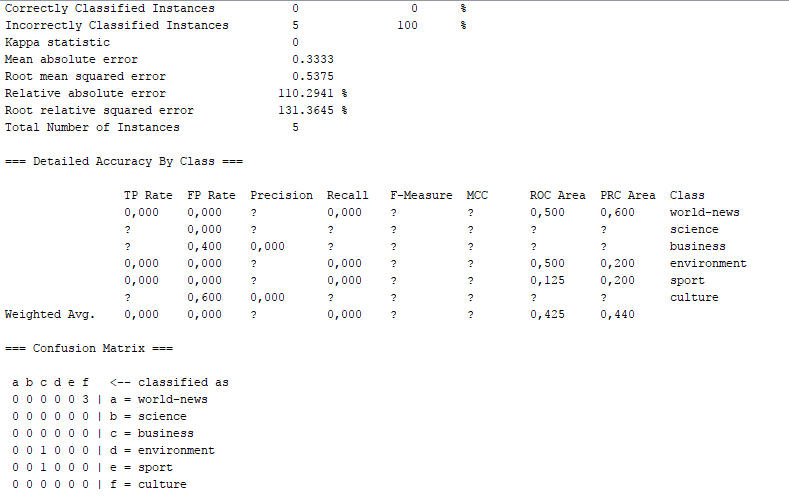
En el archivo .arff, están 14 instancias, de las cuales, 3 son de la clase *world-news*, 2 de la clase *science*, 3 de la clase *business*, 2 de la clase *environment*, 2 de la clase *sport* y 2 de la clase *culture*.



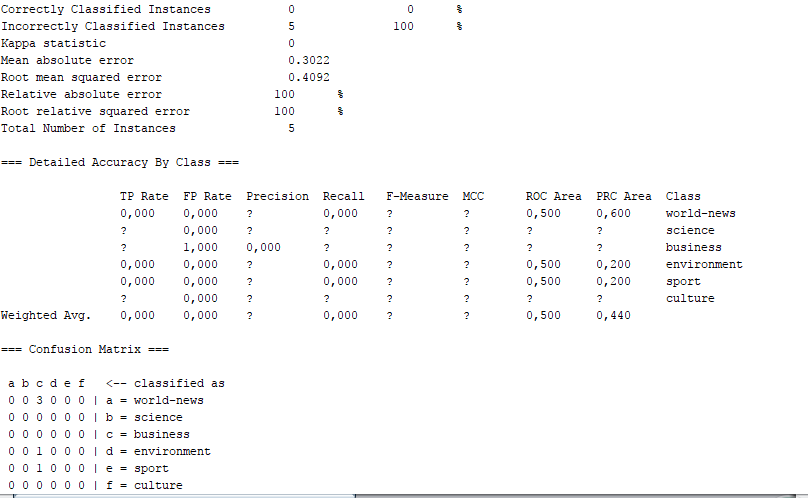
Hemos intentado clasificar estas 14 instancias con 5 clasificadores, en todos ellos, hemos utilizado dos tercios de los datos para entrenar un clasificador y un tercio de los datos sirve como el conjunto de validación. Los resultados obtenidos son siguientes:



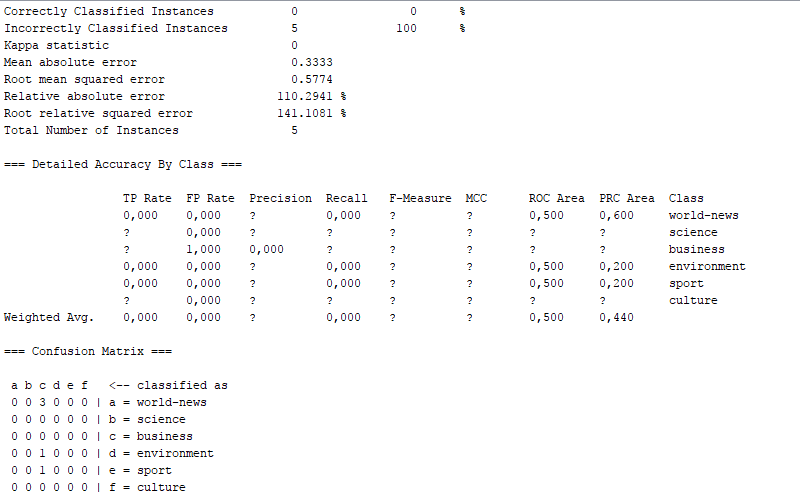
*Resultado obtenido con clasificador RandomForest(trees)*



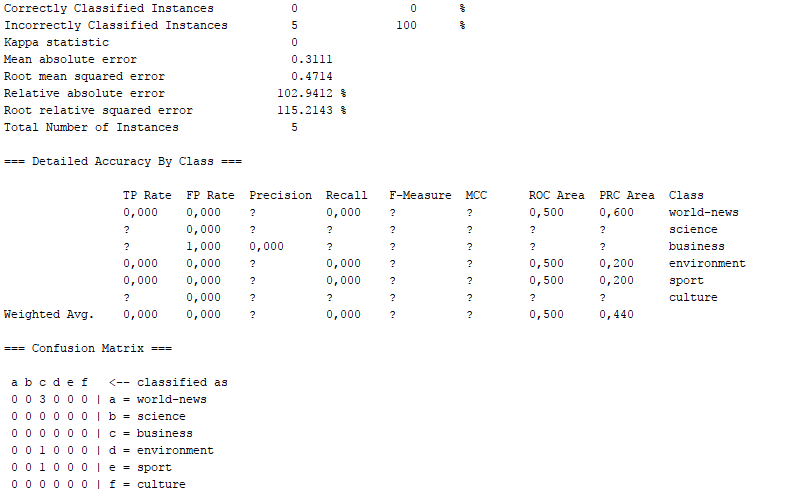
*Resultados obtenidos aplicando J48 (tree)*



*Resultados obtenidos aplicando ZeroR (rules)*



*Resultados obtenidos aplicando OneR (rules)*



*Resultados obtenidos aplicando lazy (IBK)*

Como podemos observar, en todos los resultados, no hemos clasificador bien ninguna instancia. Esto puede ser debido tenemos demasiados atributos comparando con las instancias, puesto que solo tenemos 14 instancias, y el número de atributos depende de los textos que hemos elegido porque hemos utilizado el filtro *StringtoWordVector* para convertir los textos en vectores de palabras. Además, la aparición de los nombres propios también puede dificultar la clasificación.

BIBLIOGRAFÍA:

https://www.theguardian.com/science/2018/jun/11/trials-begin-of-a-saliva-test-for-prostate-cancer

https://www.theguardian.com/business/2018/jun/11/ryanairs-uk-cabin-crew-to-be-represented-by-union-for-first-time-unite

https://www.theguardian.com/technology/2018/jun/11/bitcoin-price-cryptocurrency-hacked-south-korea-coincheck

https://www.theguardian.com/science/2018/jun/06/when-a-dinosaur-fossil-is-gone-its-gone-forever

https://www.theguardian.com/sport/2018/jun/11/wales-wayne-pivac-shortlist-to-succeed-warren-gatland-rugby-union

https://www.theguardian.com/sport/2018/jun/11/chris-froome-lizzie-deignan-doping-cases-reputations

https://www.theguardian.com/world/2018/jun/12/new-zealand-coalition-under-strain-as-jacinda-ardern-prepares-for-maternity-leave

https://www.theguardian.com/world/2018/jun/12/us-de-facto-embassy-in-taiwan-reopens-as-symbol-of-strength-of-ties

https://www.theguardian.com/politics/2018/jun/08/acerbic-and-firm-mary-wilson-remembered-fondly-after-death-at-102

https://www.theguardian.com/commentisfree/2018/jun/11/brexit-uk-fishermen-fishing-industry-quotas-uk-government

https://www.theguardian.com/games/2018/jun/11/e3-2018-bethesda-and-microsoft-unveil-fallout-elder-scrolls-vi-halo-and-gears-of-war

https://www.theguardian.com/business/2018/jun/11/poundworld-administration-jobs-high-street

https://www.theguardian.com/science/2018/jun/09/asteroid-mining-space-prospectors-precious-resources-fuelling-future-among-stars

https://www.theguardian.com/science/2018/jun/11/trials-begin-of-a-saliva-test-for-prostate-cancer

https://www.theguardian.com/world/2018/jun/12/ireland-blasphemy-referendum