

Proyecto #2

Aprendizaje de máquina: clasificadores

IC-6200 Inteligencia Artificial
Ingeniería en Computación
Instituto Tecnológico de Costa Rica
Prof. Diego Munguía

El objetivo es programar un algoritmo de aprendizaje de máquina y aplicarlo a un ejercicio de clasificación con datos reales.

- El algoritmo a implementar es el clasificador *Naïve Bayes*.
- La tarea de clasificación: analizar reseñas de un restaurante local y clasificarlas como buenas o malas.
- La plataforma de implementación es `node.js`.

Cabe notar que para este ejercicio la muestra no se presenta en forma tabulada, sino que cada observación se presenta en formato `JSON`, por ejemplo:

```
{
  "review": "Excelente ubicación para evitar la fila para el tren. El choripán
  "class": "good"
}
```

De esta forma, cada palabra de la reseña será un rasgo de la observación.

Se debe realizar un paso de limpieza de datos antes de entrenar a su algoritmo, se le sugiere considerar los siguientes aspectos para estandarizar la muestra:

- Eliminar *stop words*.
- Eliminar puntuación.
- Evitar la sensibilidad a mayúsculas, ej: “Excelente!” y “excelente” deberían ser el mismo rasgo.

- Eliminar acentos (debido a las faltas de ortografía de quienes escribieron las reseñas).

Debe entrenar y probar la predictabilidad de su algoritmo. En el `README.md` debe documentar los resultados de exactitud, precisión y reclamo de sus pruebas. Dado que la muestra es pequeña se le recomienda utilizar el método Monte Carlo de validación cruzada, https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada#Validaci.C3.B3n_cruzada_aleatoria (https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada#Validaci.C3.B3n_cruzada_aleatoria).

Para efectos de aplicación, su implementación debe ofrecer una función `classify(review)` que reciba una reseña como un string, y que retorne `true` si la reseña es buena o `false` en caso contrario.

El proyecto debe ser trabajado en parejas, fecha de entrega miércoles 1 de junio 2016 a más tardar a las 11:59 pm. El proyecto será trabajado en un repositorio privado de gitlab.