Marco Ferraro | B82957  
Tarea Weka | Inteligencia Artificial

Para este trabajo, se utilizó un data set de películas. Las películas son de origen de USA o de GERMANY. Esta clasificación va a ser el label. Los datos están en 2 documentos csv, los cuales vienen adjuntados en la entrega de esta tarea.

Para este trabajo, el set de datos se dividió en training y testing.

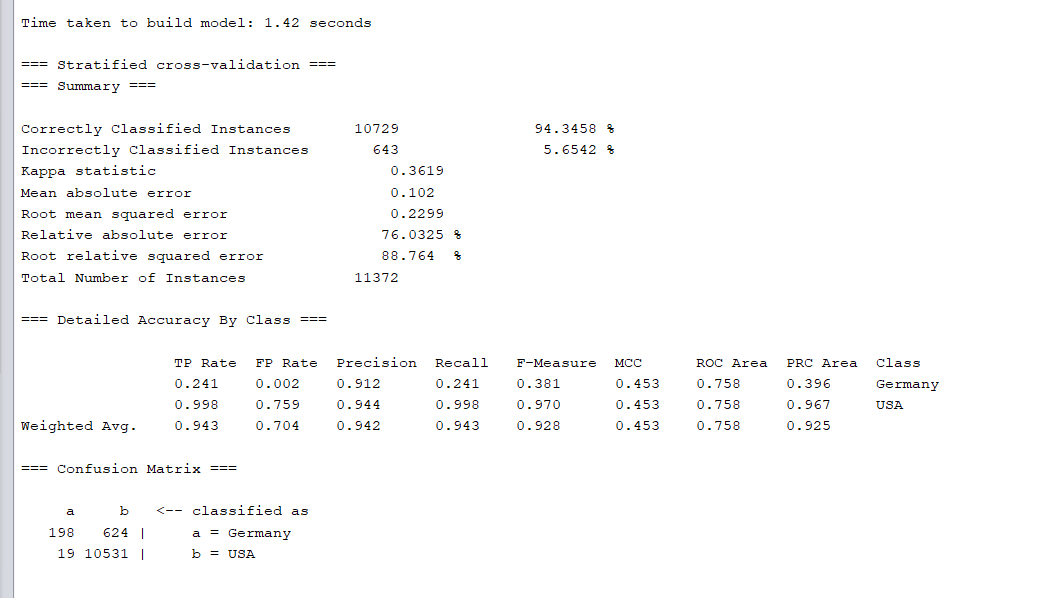


Ilustración 1 Resultados Modelo Training 5 Folds

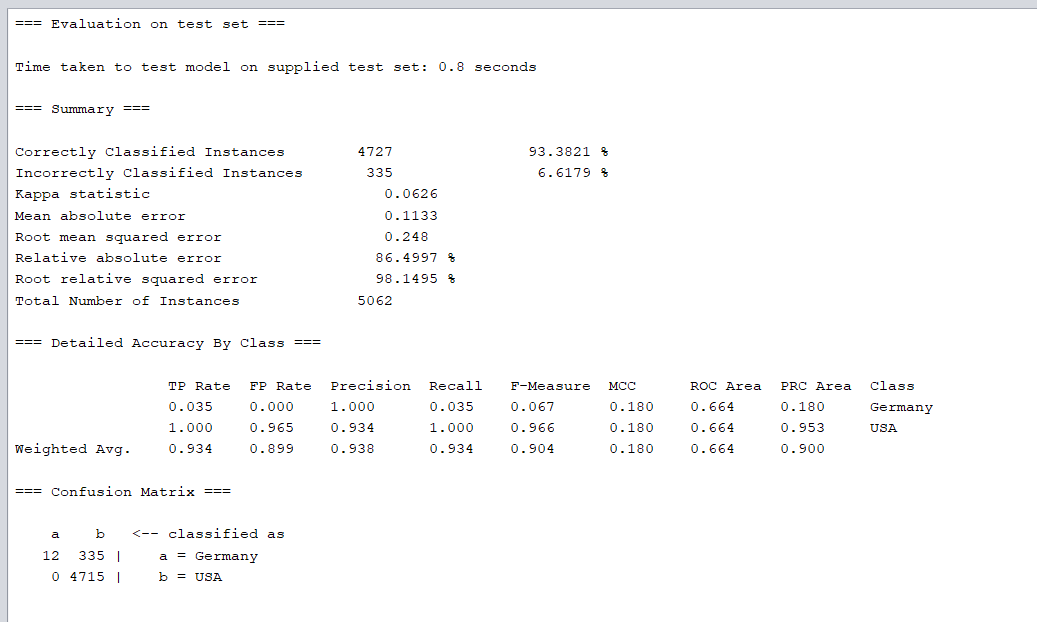


Ilustración 2 Evaluación en test set

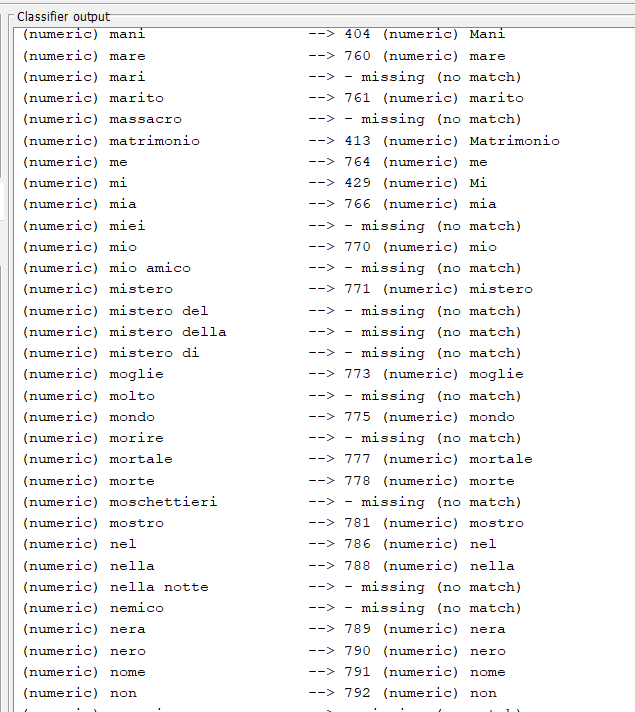


Ilustración 3 Output de clasificación con datos de prueba

Como podemos apreciar, en el set de datos de testing hay muchas palabras que el modelo no reconoce, ya que se reportan muchos misses. Se puede decir que esto es debido a la naturaleza de los datos, que el archivo de pruebas posee muchas palabras desconocidas para el archivo de entrenamiento y construcción del modelo.