## UNIVERSIDAD DE LA LAGUNA

### SECCIÓN DE FÍSICA DE LA FACULTAD DE CIENCIAS

# MEMORIA DE PRÁCTICAS EXTERNAS



Entidad: Instituto de Astrofísica de Canarias

Autor: Adrián Hernández Padrón

TUTORES: IGNACIO TRUJILLO CABRERA & SÉBASTIEN COMERÓN

Curso 2020/2021

# Índice

1.	Entidad colaboradora	2
2.	Descripción detallada de las tareas, trabajos desarrollados y departamentos de la entidad a los que ha estado asignado	2
3.	Relación de los problemas planteados y el procedimiento seguido para su	
	resolución	3
	3.1. Objetos que no se encontraron en GAIA	3
	3.2. Objetos que se encontraron en GAIA	4
	3.2.1. Metodología	4
	3.2.2. Categorías	6
	3.2.3. Clasificación final	11
	3.2.4. Conclusión	12
4.	Valoración de los conocimientos y competencias adquiridos en relación con los estudios universitarios	15
5.	Identificación de las aportaciones que, en materia de aprendizaje, han su- puesto las prácticas	15
6.	Autoevaluación de las prácticas y sugerencias de mejora	16
	6.1. Autoevaluación	16
	6.2. Sugerencias de mejora	16

#### 1. Entidad colaboradora

Estas prácticas se realizaron siguiendo la oferta propuesta por la comisión de prácticas externas del grado de Física de prácticas+TFG, en donde se le dio la oportunidad al alumnado que estaba desarrollando su TFG en el IAC/Departamento de astrofísica de realizar dichas prácticas en esta entidad.

Debido a las circustancias actuales estas prácticas se realizaron en remoto, aunque siempre se dio la oportunidad de acceder al IAC aunque nunca fue necesario puesto que tanto el trabajo como las reuniones con los tutores se realizaron en remoto.

El Instituto de Astrofísica de Canarias (IAC) es un organismo público de investigación español que gestiona dos de los mejores observatorios internacionales del mundo. Administrativamente, es un Consorcio Público, integrado por la Administración General del Estado Español, la Administración Pública de la Comunidad Autónoma de Canarias, la Universidad de La Laguna (ULL) y el Consejo Superior de Investigaciones Científicas (CSIC).

La misión principal del IAC es realizar y promover cualquier tipo de investigación astrofísica o relacionada con ella, así como desarrollar y transferir su tecnología; difundir los conocimientos astronómicos, colaborar en la enseñanza universitaria especializada de astronomía y astrofísica y formar y capacitar personal científico y técnico en todos los campos relacionados con la Astrofísica; administrar los centros, observatorios e instalaciones astronómicas ya existentes y los que en el futuro se creen o incorporen a su administración, así como las dependencias a su servicio; y fomentar las relaciones con la comunidad científica nacional e internacional.

### 2. Descripción detallada de las tareas, trabajos desarrollados y departamentos de la entidad a los que ha estado asignado

La tarea asignada fue trabajar sobre los outliers generados por el estudio AUTO-MATED UNSUPERVISED CLASSIFICATION OF THE SLOAN DIGITAL SKY SURVEY STELLAR SPECTRA USING k-MEANS CLUSTERING(J. Sánchez Almeida and C. Allende Prieto 2013 ApJ 763 50) . Este trabajo se centró en la clasificación, usando machine learning no supervisado, de 173.390 espectros estelares de entre 3800 a 9200 amstrongs pertenecientes a los programas del Sloan Digital Sky Survey(SDSS) SEGUE y SEGUE-2.

La técnica de machine learning no supervisado usada como la que se conoce como k-meaning clustering, esta técnica tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano. De manera que los 173.390 espectros iniciales se clasificaron en 25 grupos distintos.

Sin embargo, como resultado del algoritmo se obtuvieron 2212 espectros los cuales no encajaban en ninguna clasificación, estos espectros son los conocidos como outliers y fueron los que se propusieron para analizar y clasificar. En estos outliers se encontraban espectros de sistemas estelares, quasar con redshift elevado, galaxias, estrellas y espectros con una naturaleza extraña. El objetivo de esta clasificación espectral era identificar estos espectros de naturaleza extraña y separarlos para realizar un análisis mas profundo sobre estos.

Para realizar la clasificación de los espectros de estos outliers se planteo de la siguiente manera. La idea era buscar estos objetos en la base de datos de GAIA, concretamente en GAIA edr3, para realizar una comprobación de que objetos aparecían en esta base de datos así como para la obtención de algunos parámetros necesarios para su posterior clasificación.

# 3. Relación de los problemas planteados y el procedimiento seguido para su resolución

Se comenzó, por tanto, realizando esta búsqueda en GAIA de los objetos iniciales, para ello se utilizó el modulo de Astropy en Python. Esta búsqueda se hizo a través de VizieR para las coordenadas de cada objeto (RA\_ICRS, DE\_ICRS) en un radio de 3 arcosegundos, debido a que el estudio se realizó en 2013 se consideró que este rango sería apropiado para abarcar el movimiento de cualquier objeto desde entonces hasta ahora.

```
In [6]: #Take the ra and dec columns from the outliers
ra = np.array(outliers['ra'])
dec = np.array(outliers['dec'])

vizier_gaia_outliers = pd.DataFrame()
vizier_gaia_notmatch = []
print(ra[6], dec[6])

40.871126000000004 0.642712

In [10]: inicio = time.perf_counter() #time counter
for i in range(0, len(ra)):
    coord = SkyCoord(ra=ra[i], dec=dec[i], unit=(u.degree, u.degree), frame='icrs')
    radio = u.Quantity(3, u.arcsec)
    r = Vizier_(catalog="I/350/gaiaedr3").query_region(coord, radius=radio) #query to gaia
    try:
        vizier_gaia_query = r[0].to_pandas()
        except Indextror:
        vizier_gaia_outliers = vizier_gaia_outliers.append(vizier_gaia_query, ignore_index=True) #We save all the rows in a DataFrame
    final = time.perf_counter()
    print("Tiempo de ejecucion:", final - inicio)

* Itempo de ejecucion: 1067.432562
```

Figura 1: Código usado para acceder a GAIA edr3 a través de VizieR usando Python.

Con ello se obtuvo que, de los 2212 objetos que se tenían inicialmente, se obtuvo un resultado de que 2377 objetos aparecian en GAIA y 38 no aparecían en GAIA. Tras la limpieza (eliminar duplicados y filas vacías) se quedó en 2312 objetos que si aparecían y 38 que no aparecían en GAIA, se puede ver que los resultados obtenido no coinciden con los iniciales, esto es debido a que al estar buscando en la base de datos de GAIA entorno a un cierto rango definido, se daba la posibilidad que para un objeto del SDSS se encontraran en GAIA dos o mas objetos(ej: sistemas binarios).

#### 3.1. Objetos que no se encontraron en GAIA

De manera que ya se tenían los primeros pasos en esta clasificación y se propuso el primer análisis de estos objetos que no aparecían en GAIA. Para ello se observaron el espectro y la imagen de cada uno de los objetos, esto se hizo buscando cada una de las coordenadas en la base de datos del SDSS. De manera que de los 38 objetos iniciales, habían dos regiones vacías, 34 galaxias y dos objetos que parecían estrellas pero que no

tenían un espectro que poder interpretar, por lo que se concluyó en que ningún objeto de este grupo era digno de realizar un estudio a parte.

#### 3.2. Objetos que se encontraron en GAIA

El análisis y clasificación de estos objetos supuso la mayor parte del trabajo. Dado que el número de objetos con el que estábamos trabajando no era muy elevado, se decidió hacer una clasificación manual(revisando los espectros uno por uno) de estos objetos. Para ello, se decidió separar estos objetos bajos ciertos criterios:

- Paralaje > 3\*Paralaje\_error
- Movimiento propio > 3\*Movimiento propio\_error
- Restantes

#### 3.2.1. Metodología

Estos tres criterios no se aplicaron simultaneamente sobre el total sino que primero se aplicó el criterio del paralaje y después se aplico el del movimiento propio sobre los objetos que no entraban dentro del criterio del paralaje. De esta manera no habían objetos dentro de las tres clasificaciones repetidos. Para separar los datos obtenidos en GAIA en estos tres grupos se utilizó Python, usando el módulo de Pandas para trabajar con tablas en DataFrame y poder manejarlas de una manera intuitiva.

Figura 2: Ejemplo del código usado para aplicar el criterio del paralaje, como salida se generaban dos tablas; los objetos que si entraban dentro de un paralaje bien definido y los que no.

Una vez se tenían separado los objetos pertenecientes a cada una de los grupos, se exportaba únicamente las coordenadas RAJ2000 y DEJ2000 debido a que para analizar el espectro en el Skyserver (web del SDSS donde aparecía toda la información necesaria del objeto) solo necesitabamos conocer las coordenadas.

```
In [48]: stars_pm_gaia_coordinates= pd.DataFrame(stars_pm_gaia, columns=['RAJ2000','DEJ2000'])
others_plx_pm_gaia_coordinates= pd.DataFrame(others_plx_pm_gaia, columns=['RAJ2000','DEJ2000'])
stars_pm_gaia_coordinates.to_csv(r'c:\Users\Usuario\Documents\data\Tables\tars_pm_gaia_coordinates.csv', index = False)
others_plx_gaia_coordinates.to_csv(r'C:\Users\Usuario\Documents\data\Tables\others_plx_pm_gaia_coordinates.csv', index = False)
```

Figura 3: Ejemplo del código usado exportar las tablas con las coordenadas.

Con esto se procedió a realizar la clasificacion estelar de estos objetos analizando su espectro y su imagen en la base de datos SDSS, para ello se escribió un pequeño código en Bash de manera que se creara un archivo de texto en el que aparecieran las coordenadas del objeto seguido de su enlace al **Skyserver**.

Figura 4: Script usado para generar el fichero donde se realizaba la clasificacón. El código comentado corresponde a las obtención de las imágenes de cada objeto, se prescindió de el al final debido a que en la url estaba toda la información deseada.

```
(1)298.0691765785,77.93073799704
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=298.0691765785&dec=77.93073799704
(2)304.31485802755,77.72191921685
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=304.31485802755&dec=77.72191921685
(3)349.81806466443,53.74375172487
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=349.81806466443&dec=53.74375172487
(4)42.19472601343,-0.69346376085
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=42.19472601343&dec=-0.69346376085
(5)42.11858389974,1.15056645182
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=42.11858389974&dec=1.15056645182
(6)41.94189750859,0.41408739744
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=41.94189750859&dec=0.41408739744
```

Figura 5: Estructura del fichero generado donde se realizaba la clasificación.

Cuando se revisaba el objeto, se añadía una etiqueta de manera que el archivo inicial se modificaba de la siguiente manera:

```
(1)298.0691765785,77.93073799704 : star
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=298.0691765785&dec=77.93073799704
(2)304.31485802755,77.72191921685 : star
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=304.31485802755&dec=77.72191921685
(3)349.81806466443,53.74375172487 : star
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=349.81806466443&dec=53.74375172487
(4)42.19472601343,-0.69346376085 : star
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=42.19472601343&dec=-0.69346376085
(5)42.11858389974,1.15056645182 : star
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=42.11858389974&dec=1.15056645182
(6)41.94189750859,0.41408739744 : star
http://skyserver.sdss.org/dr16/en/tools/explore/Summary.aspx?ra=41.94189750859&dec=0.41408739744
```

Figura 6: Método usado para la clasificación.

El motivo de esto era agilizar el análisis, de esta manera se tenía que estar preocupando de meter cada objeto en cada una de las clasificaciones una vez revisado el espectro. Se repetía este proceso para cada una de los grupos y una vez finalizado para todos los objetos pertenecientes al grupo usando herramientas de Bash se guardaba cada uno de los objetos etiquetados en la carpeta correspondiente:

```
~/Practicas/cleaning > cat PM GAIA | grep -A 1 "star" | grep -v -E '\--' >> Estrellas
```

Figura 7: Comando escrito en Bash utilizado para separar los objetos del archivo original en los diferentes ficheros. Este comando se repetía para cada uno de los grupos.

Este método fue bastante útil debido a que, en primera instancia, se falló a la hora de filtrar los objetos por el criterio del movimiento propio. Nos percatamos de este error al observar la presencia de algunos quasar en la clasificación que presentaban movimiento propio. Sin embargo como ya se tenían los objetos clasificados con su correspondiente etiqueta, simplemente con obtener la tabla de objetos con movimiento propio correcta y repetir el comando en la figura 7 se pudo clasificar correctamente los objetos sin tener que repetir el análisis.

#### 3.2.2. Categorías

Las etiquetas usadas y por tanto las categorías finales de la clasificación fueron las siguientes:

- Estrellas: Estos fueron todos los objetos que mostraban espectros similares a los de la clasificación iniciales. En este grupo se encuentran:
  - Estrellas con espectros bien definidos: Estos espectros no encajaron en la clasificación de 2013 por algun motivo que desconocemos; nuestro pensamiento inicial es que estos espectros cuando se realizó la clasificación inicial eran espectros que no entraban en la clasificación pero tras el paso de los años estos espectros se han mejorado de tal manera que ahora si estarían dentro de esta.

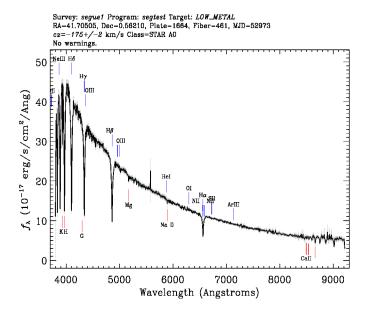


Figura 8: Ejemplo de estrella con un espectro bien definido.

 Estrellas con espectros ruidosos: Estos espectros suelen pertenecer a estrellas con una magnitud muy elevada, no encajan en la clasificación de 2013 debido a su ruido.

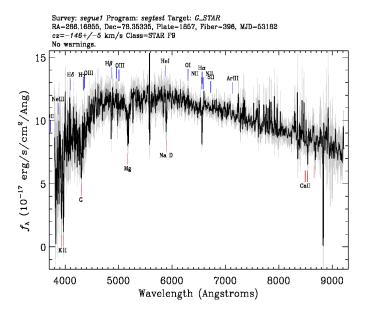


Figura 9: Ejemplo de estrella con un espectro con mucho ruido.

• Estrellas con fallos espectrales: Espectros que presentaban fallos en alguna parte de su espectro.

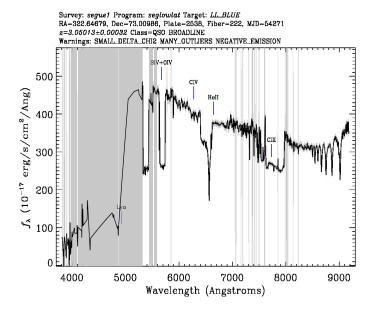


Figura 10: Ejemplo de fallo espectral.

• Quasars: Aunque los espectros seleccionados para la clasificación de 2013 eran de estrellas, la presencia de quasars en la clasificación se debe a un error en las etiquetas por parte del SDSS.

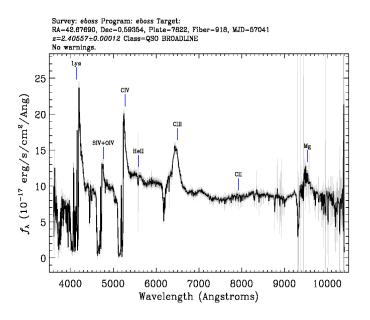


Figura 11: Ejemplo de un espectro de un quasar.

• Galaxias: De la misma manera que con los quásares ocurre con las galaxias.

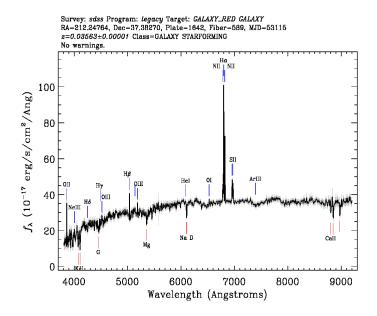


Figura 12: Ejemplo de un espectro de una galaxia.

■ Suma de espectros: Inicialmente esta categoría tenía el nombre de Estrellas binarias, esto se cambió al final, ya que al contrastar con un estudio de estrellas binarias en GAIA(Kareem El-Badry, Hans-Walter Rix, Tyler M. Heintz 2021) se comprobó que tan solo un 5 por ciento de los objetos clasificados como binarias estaban dentro de este estudio. La explicación de esto es que aunque estos objetos se veían en la imagen como binarias y su espectro mostraba la estructura de una suma de dos espectros, puede ser que no sea un sistema binario sino que los objetos se encuentren muy cerca en el plano de la imagen, sin embargo que una se encuentre mas profunda que otra. Esta suma de espectros podría entrar perfectamente dentro de la categoría de estrellas pero se decidió mantenerse separada dado el numero de objetos que se encontraron.

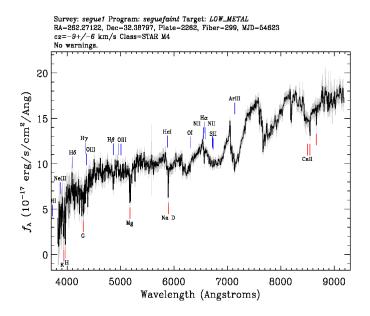


Figura 13: Ejemplo de un espectro de una estrella binaria.

■ Raros: Esta categoría era el objetivo de la clasificación. Objetos con espectros que no encajan en ninguna de las anteriones clasificaciones y por tanto requieren un análisis mas profundo.

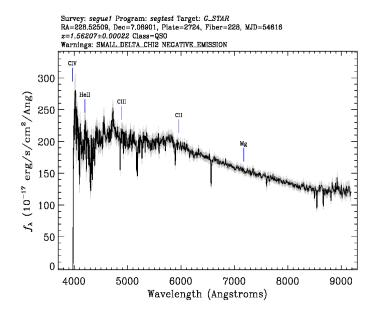


Figura 14: Ejemplo de un espectro raro.

• Objetos sin espectro: Como se dijo al comienzo, la busqueda en GAIA se hacia

entorno a un cierto rango, por lo que estos objetos sin espectro se deben a objetos que no pertenecen a la clasificación de 2013 pero que se encuentran muy próximos (en la imagen) a objetos que si se encuentran dentro de esta clasificación.

#### 3.2.3. Clasificación final

Objetos con paralaje: En este grupo se esperaba obtener un gran numero de estrellas y a una magnitud no muy elevada debido a que presentaban un paralaje bien definido.

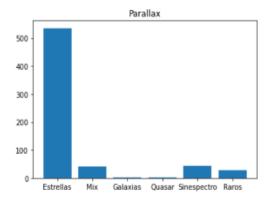


Figura 15: Histograma de los objetos con paralaje definido; estrellas(536), mezcla de espectros(40), objetos sin espectros(43), galaxias(2), quasars(1), raros(28).

Objetos con movimiento propio: En este grupo se esperaba obtener un gran numero de estrellas, estas estrellas como no tenían un paralaje bien definido presentaban en su mayoría un espectro muy ruidoso debido a la elevada magnitud a la que se encuentran.

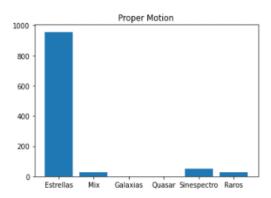


Figura 16: Histograma de los objetos con paralaje definido; estrellas(957),mezcla de espectros(26), objetos sin espectros(48), galaxias(0), quasars(0), raros(28).

 Resto: Dentro de este grupo se esperaba encontrar un gran numero de quasars, galaxias y algunas estrellas con espectros muy ruidosos las cuales no tenían un movimiento propio bien definido.

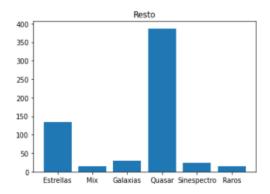


Figura 17: Histograma de los objetos restantes; estrellas(135),mezcla de espectros(14), objetos sin espectros(24), galaxias(29), quasars(387), raros(15).

Con esto ya se consiguió separar los objetos en los diferentes grupo, solo quedaría, por tanto, un análisis profundo de los objetos clasificados como raros para cada una de los grupos.

#### 3.2.4. Conclusión

Jorge Sanchez Almeida, uno de los autores del estudio sobre el que se estaba trabajando, estuvo siguiendo desde el principio el trabajo y fué el encargado de revisar estos objetos raros. De los 71 objetos que había clasificado como raros, Jorge consideró que 35 eran realmente objetos con espectros a tener en cuenta.

De todos estos objetos el más relevante y que se consideró que había que seguir en detalle fue el siguiente objeto (ra:45.65908545342, dec:37.40599576882) el cual se ve como una estrella con movimiento propio pero presenta un espectro de galaxia.

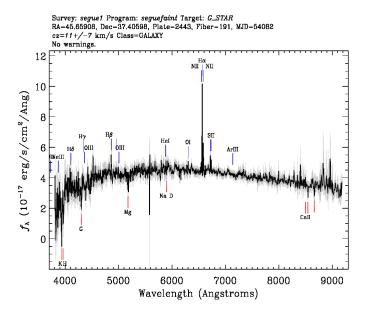


Figura 18: Extremely interesting object. Star-like, but with a galaxy spectrum. Clear Ha (+the two [NII]) and Hb, but no [OIII]. And it has proper motion measured!.

Otros ejemplos de los espectros de estos objetos que se consideraron como raros acompañados del la descripción de Jorge :

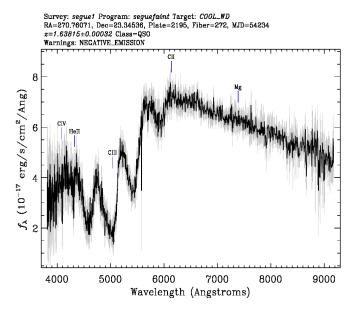


Figura 19: Trully interesting. Classified as quasar. Cannot be because it has proper motion. To be followed up.

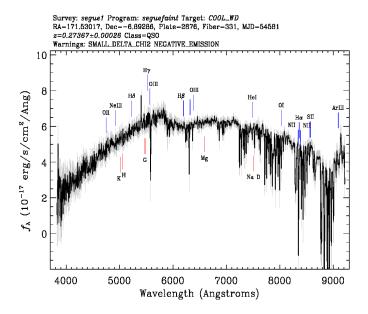


Figura 20: Looks like a miss-classified QSO. Very weird spectrum. Worth following up.

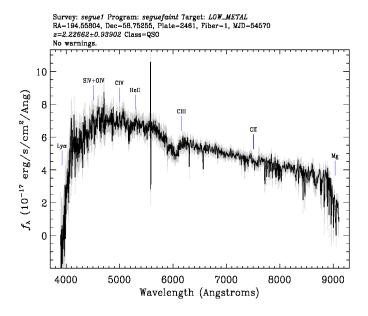


Figura 21: Weird stellar spectrum (stellar because the deep line at 4000 AA). Deserves a double check.

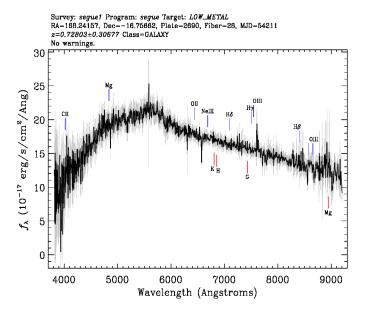


Figura 22: Very weird piece-wise spectrum. Rather than being at 6000 AA, the break happens at 5500 AA. There are plenty of uncorrected sky lines.

La lista completa de estos objetos raros se puede ver en este **pdf** en el que se encuentra las coordenadas del objeto, el enlace al Skyserver del SDSS y el comentario de Jorge sobre el objeto.

# 4. Valoración de los conocimientos y competencias adquiridos en relación con los estudios universitarios

En relación con los estudios universitarios se podrían destacar asignaturas en las que se tratan espectros y técnicas de computación como Técnicas Experimentales II, Computación Científica y Astrofísica y Cosmología. También debido a la oferta prácticas+TFG, es necesario mencionar los conocimientos que he aprendido durante el TFG puesto que este me ha dotado de conocimiento en el ámbito de bases de datos astronómicas asi como algunas herramientas de programación.

# 5. Identificación de las aportaciones que, en materia de aprendizaje, han supuesto las prácticas

En cuanto a las competencias adquiridas a lo largo de las prácticas externas, he conseguido adquirir un profundo conocimiento en cuanto a espectros, estos conocimientos no son solo en espectros estelares, sino también espectros de quasares y galaxias.

También para la realización de esta clasificación fueron necesarias herramientas de programación en Python y Bash, pese a que tenia algo de conocimiento en estos campos nunca habia tenido la experiencia de aplicarlos en un ejercicio real, creo que gracias a

las practicas externas he conseguido afianzar y familiarizarme con dichos conceptos en este ámbito.

### Autoevaluación de las prácticas y sugerencias de mejora

#### 6.1. Autoevaluación

He salido muy satisfecho de estas prácticas externas, he sentido desde el primer minuto mucha confianza por parte de mis tutores y mucha libertad a la hora de trabajar. Debido a esto he ido desarrollando estas prácticas de una manera autónoma en el sentido que he podido tomar el camino que he visto más óptimo y, sin embargo, en todo momento mis tutores han sabido guiarme de una manera profesional.

En cuanto a esto último, creo que la manera en la que he planteado el problema propuesto ha sido bastante acertada, el orden seguido y la forma en la que he que he ordenado y manejado los datos. Esto me ha permitido que la ejecución de los pasos durante el trabajo haya sido muy óptima de manera que me he podido dedicarle suficiente tiempo al análisis de cada uno de los espectros y no estar mas tiempo del necesario en el manejo de los datos y tablas. Siempre que iba mostrando resultados a los tutores estaban de acuerdo con lo presentado la mayoría de las veces y cuando se presentaba algún tipo de problema he sabido identificarlo y resolverlo rápidamente.

Todo esto me hace estar bastante contento del trabajo realizado y con el resultado final de la clasificación.

#### 6.2. Sugerencias de mejora

Como sugerencia de mejora en estas practicas creo que en el inicio se deberían haber centrado en adquirir conocimientos sobre los espectros en vez de empezar directamente a clasificarlos. Si que es verdad que conforme iba analizando espectros iba cogiendo soltura y conocimiento sobre estos, pero en la primera parte de la clasificación no tenía la certeza de que lo que estaba clasificando era correcto.