



Clasificación Automática de Galaxias en base a su Espectro

- ANEXO DE GESTIÓN -

Arvin Daswani Daswani
Pedro Martín Gómez
Silvia Peraza Delgado
Jose Luis Quintero García
Vicente Tetuani Sánchez

Índice

Entrevistas	3
Entrevista 1	3
Entrevista 2	4
Entrevista 3	5
Entrevista 4	8
Entrevista 5	11
Entrevista 6	15
Guión para las entrevistas	16
Guión general	16
Guión reunión 5 (02/07/2019)	16

1. Entrevistas

Resumen de las entrevistas realizadas:

Reunión	Fecha	Objeto
1	08/04/2019	Primera toma de contacto
2	29/04/2019	Revisar el alcance del proyecto
3	11/06/2019	Comprender el funcionamiento interno del IAC
4	19/06/2019	Conocer el grupo de trabajo de Big Data del IAC
5	02/07/2019	Evaluar los avances realizados en el tratamiento de los datos
6	26/08/2019	Presentación de los primeros resultados

1.1. Entrevista 1. Primera toma de contacto

Acta	
Fecha	08/04/19
Lugar	Instituto Astrofísico de Canarias
Asistentes	Jorge Sánchez Almeida (Astrofísico del IAC) Ana Monreal (Astrofísico del IAC) Arvin Daswani José Luis Quintero

Conocemos a Jorge Sánchez y Ana Monreal, astrofísicos del IAC. Nos presentamos y explicamos el objeto del máster. Nos interesa conocer aquellos problemas en los que les podamos ayudar (previamente hemos visto algunos trabajos donde se ha usado ML en el campo de la astrofísica).

Nos plantean que ellos mismos han realizado tiempo atrás clasificaciones de galaxias usando K-means. El objeto fundamental de estos estudios era la obtención de outliers en esta clasificación, dado que son estas galaxias raras las que plantean retos de investigación más interesantes.

Hablamos del SDSS, la base de datos que podría emplearse (básicamente, es un inventario de una zona del universo, que se va ampliando a través de varias campañas).

Este dataset es público y contiene varios tipos de datos relativos a distintos objetos. En este caso, lo que nos interesa serían los espectros de las galaxias, interesándose por la clasificación de las mismas según sus espectros.

Hablamos de distintos conceptos de astrofísica que nos anuncian un campo enormemente interesante y retador.

Nos pasan información sobre varios papers de referencia para ir familiarizándonos con los términos más propios de la astrofísica aplicada a la clasificación de galaxias.

Se da por finalizada la reunión emplazándonos en vernos próximamente para continuar avanzando.

1.2. Entrevista 2. Revisar el alcance del proyecto.

Acta	
Fecha	29/04/2019
Lugar	Instituto Astrofísico de Canarias
Asistentes	Jorge Sánchez Almeida (Astrofísico del IAC) Ana Monreal (Astrofísico del IAC) Silvia Peraza Vicente Tetuani José Luis Quintero

En esta ocasión detallamos algunos aspectos tras el estudio de los papers, fundamentalmente en lo relacionado con el alcance y objeto del proyecto.

Intentamos conseguir el programa del paper de Jorge Sánchez, basado en K-means, pero cuando éste se realizó aún no estaba extendido el uso de repositorios, por lo que el programa está perdido en su almacén de documentación. Este sistema ASK que Sánchez Almeida definió tenía problemas, dado que los clusters no quedaban bien definidos y los resultados eran demasiado aleatorios.

Nos queda claro que nos van a ayudar en este proyecto, dado que están muy interesados en ver las posibilidades actuales del big data, pero que tampoco quieren emplear demasiado tiempo en ello.

También detallamos sobre la base de datos SDSS y el acceso a los datos. Estos están contenidos con archivos en formato FITs, que requiere un tratamiento especial para su uso. Posteriormente a la entrevista, Ana nos envía un detallado tutorial para trabajar con FITs y enlaces a otros tutoriales.

También hablamos del tamaño de los datos y de algunos campos adicionales requeridos junto al espectro (como es el valor de redshift), necesario para corregir el propio espectro, dado que éste presenta desplazamiento en las amplitudes de ondas debido a la distancia de la galaxia y es necesario corregirlo.

1.3. Entrevista 3. Comprender el funcionamiento del IAC

Acta	
Fecha	11/06/2019
Lugar	Instituto Astrofísico de Canarias
Asistentes	Carlos Martín Galán (Gerente de Informática del IAC) Arvin Daswani Silvia Peraza Vicente Tetuani

Nos recibe Carlos Martín en su despacho y le explicamos el proyecto brevemente. Le pedimos información sobre el funcionamiento interno del IAC, así como su sistemas de financiación.

Financiación

Comenzamos con una explicación de Carlos sobre las distintas fuentes de financiación del IAC, repartidas a groso modo de la siguiente manera:

- ★ Presupuestos Generales del Estado (20% aproximadamente).
- ★ Financiación autonómica (pequeñas aportaciones).
- ★ Ayudas y subvenciones de proyectos de investigación, que realmente son la fuente principal de financiación del IAC (Planes de ayuda a la investigación Europeos/No europeos, Fondos FEDER, etc...)

Respecto a este último punto, nos explicó el flujo de la financiación de proyectos que se resume en los siguientes puntos:

- ★ Un científico tiene una idea sobre una investigación que quiere realizar.
- ★ Se pone en contacto con la entidad financiadora para pedir fondos.
- ★ Presenta un proyecto donde especifica los costes del mismo:
 - Personal
 - Infraestructura
 - Software
 - Formación
- ★ Una vez aprobado el proyecto, el dinero va a parar al IAC y no al investigador.
- ★ El IAC saca las licitaciones con proveedores externos y los procesos de selección para reclutar al personal adicional necesario, así como la dotación de recursos internos.
- ★ A la hora de contratar productos y servicios externos, el IAC se rige por la ley de contratación del estado con las siguientes pautas:
 - Contratos menores de 5K€ -> Contratación Directa
 - Contratos entre 5K€ y 15K€-> Negociado sin publicidad (invitación a 3 empresas).
 - Contratos mayores de 15K€ -> Concurso abierto.

Funcionamiento interno del IAC

Los investigadores están organizados por áreas de conocimiento. Entre ellos compiten por conseguir llevar a cabo sus proyectos, así como comparten conocimientos de forma muy espontánea mientras toman algo en la cafetería. Así es como llegan a detectar que tienen problemas comunes o tareas que se solapan, lo que Carlos apunta que sucede con frecuencia.

Brevemente nos habla sobre la existencia de la OTRI, Oficina de Transferencia de Resultados de Investigación para canalizar la relación entre las universidades y las empresas.

El IAC no cuenta con un departamento especializado que asuma la funciones de Big Data, y ni tiene previsto crearse, si bien es cierto que existe un grupo de trabajo formado por varios investigadores que se reúnen una vez a la semana, con la finalidad de tratar la resolución de ciertos problemas de Big Data. Recientemente han adquirido un ordenador de 120K €.

Le decimos que sería muy interesante para nosotros tener una entrevista con ellos para que nos den ideas de posibles proyectos. Carlos nombra a Marc Huertas, pero queda pendiente que nos pase los datos de contacto para cerrar otra entrevista.

Viabilidad de nuestro proyecto

Carlos comenta que ve interesante la creación de una empresa para prestar servicios de Big Data al IAC.

Respecto a la financiación de la supuesta compañía que se pretende simular en el proyecto, nos preguntó qué infraestructura pensábamos utilizar. Le respondemos que Cloud con Azure y nos dice que consideremos utilizar la red IRIS. Según comenta los investigadores podrían contratar dicha infraestructura, por lo que podría resultar más barata que Azure.

Además, el IAC cuenta con un nodo del supercomputador MareNostrum, ubicado en La Palma. El 50% de su uso es para investigadores del IAC y el otro 50% está destinado al supercomputador.

También le preguntamos por posibles KPIs para poder monitorizar el proyecto, diciéndole que habíamos pensado en los siguientes:

- ★ Horas de trabajo ahorradas a los investigadores en clasificación de galaxias.
- ★ Referencias al proyecto, White Papers e investigaciones.

Inicialmente no se le ocurrió ninguna más, pero posteriormente apuntó a la posibilidad de medir el nivel de implantación del proyecto, mediante la suma de las subvenciones obtenidas por el mismo, un valor económico fácilmente medible.

También comentó que durante el proceso de baremación de los proyectos, se valoran positivamente los siguientes aspectos:

- ★ Desarrollo del tejido empresarial local (contratación a terceros locales)

- ★ Igualdad de género en los proyectos
- ★ Cuestiones medioambientales

Competencia

Nos habla de una empresa americana que realizó trabajos con la NASA con oficina comercial en España y que contactó con ellos para realizar trabajos de Big Data. La empresa se llama Data Robot (<https://www.datarobot.com/>) y tienen un comercial en España (Felipe Ortin <https://www.linkedin.com/in/felipe-ort%C3%ADn-b2446980/>).

Le echamos un vistazo rápido a un video publicitario, pero parece tratarse de un software que tiene embebidos varios algoritmos y va mostrando el resultado de la ejecución de los mismos.

1.4. Entrevista 4. Conocer el grupo de trabajo de Big Data del IAC.

Acta	
Fecha	19/06/2019
Lugar	Instituto Astrofísico de Canarias
Asistentes	Marc Huertas (Astrofísico del IAC) Arvin Daswani Silvia Peraza Vicente Tetuani José Luis Quintero

Conocemos a Marc Huertas en las dependencias del IAC y le comentamos que Carlos Martín nos habló de él y del grupo de técnicas de Big Data que se había creado dentro del IAC.

Presentación

Nos presentamos el equipo de trabajo y explicamos brevemente el proyecto que vamos a realizar con el IAC. Le comentamos a Marc que contactamos con Jorge Sánchez Almeida y con Ana Monreal para ofrecernos para la realización de un proyecto y que ellos nos propusieron realizar una clasificación de galaxias en base al espectro.

Grupo especializado en Machine Learning, Deep Learning y Big Data

Marc Huertas, astrofísico de formación pero con mucha experiencia en temas de Ingeniería Informática, nos comenta que lleva trabajando con técnicas de Machine Learning desde el año 2006 y acaba de aterrizar en el IAC recientemente. Nos cuenta que, tras su incorporación en el equipo de investigadores, observó que dentro de la institución hay mucha demanda de técnicas de Machine Learning y de Deep Learning, pero que no existe nada dentro del instituto que se dedique a investigaciones con estas técnicas. Marc se refiere a documentación, recursos, equipos, etc... Nombra a Andrés Asensio Ramos, otro astrofísico del IAC que se dedica también a estos temas desde antes de su llegada. Decide entonces replicar la creación de un grupo de Big Data que desarrolló en su etapa en París.

Se crea entonces dentro del IAC un grupo multidisciplinar y transversal a toda la organización especializado en técnicas de Machine Learning, Deep Learning y Big Data en general. Según nos comenta, entre los astrofísicos hay muchísima demanda al respecto. Se trata de un grupo de alrededor de entre 10 y 20 personas (aunque en su momento se apuntaron más de 40) que se reúne cada 15 días para discutir cualquier temática relacionado con el ML aplicado a la astrofísica. En estas reuniones se tratan tanto temas generales, como por ejemplo el debate de algún estudio o investigación publicada, como particulares, como en el caso de investigadores que piden consejo en técnicas de ML o DL para sus propios estudios.

Evolución de algoritmos

Marc Huertas comenta que, en la actualidad, los algoritmos de Machine Learning y Deep Learning aplicados a la astrofísica se encuentran en una fase de transición. Los algoritmos actuales están bien para cuando se aplican entre 1 y 2 millones de objetos. No obstante, nos indica que es necesaria una renovación de los mismos ya que no van a servir para la cantidad de datos que estarán disponibles a corto o medio plazo (se prevé que crezcan x100 en 2 ó 3 años). Cita por ejemplo la misión espacial [EUCLID](#), de la que él forma parte.

Sobre la evolución de estas técnicas, Marc indica que en el año 2015 se escribió el primer paper que relacionaba técnicas de ML/DL con la astrofísica. El segundo artículo está firmado por él y entonces apenas se publicaba un artículo al mes o cada dos meses relacionado con este tema. En la actualidad, es raro no ver más de un artículo a la semana.

Con respecto al tipo de algoritmos, las técnicas han ido evolucionando y el Machine Learning ha quedado algo obsoleto en la astrofísica, dando paso a herramientas de Deep Learning. Según nos comenta, su uso más frecuente es el tratamiento de imágenes, clasificación y detección, sin embargo también es posible emplearlos sobre espectros. En la actualidad también se están empleando técnicas de regresión para realizar ciertas mediciones.

Externalización de servicios

Hoy en día no hay empresas que presten servicios de consultoría en materia de astrofísica. Son los mismos astrofísicos los que al encontrarse ante la necesidad de aplicar técnicas de ML o DL, indagan sobre la materia, recurriendo a otros astrofísicos con más experiencia o buscando códigos de ejemplo en Github, gracias a que cada vez son más los científicos que comparten sus desarrollos.

En esta línea, Marc nos cuenta que todas las semanas recibe consultas de compañeros acerca de qué herramientas utilizar en sus investigaciones o si algún problema se puede resolver mediante la aplicación de ML o DL. Lo habitual en estos casos es que el investigador desarrolle la solución por sus propios medios y posteriormente abandone el desarrollo del algoritmo una vez haya satisfecho sus necesidades.

Con respecto a la posible contratación de una consultora externa, Marc puntualiza dos aspectos que considera críticos:

- ★ La capacidad de entender los requisitos del problema sin consumir grandes cantidades de tiempo por parte del investigador.
- ★ Que los resultados obtenidos no sean de utilidad para el investigador.

Investigación propia y próximos pasos

Preguntamos a Marc sobre el tipo de trabajos que ha desarrollado y nos comenta que ha resuelto muchos temas de clasificación pero siempre morfológica, a través de imágenes. Con respecto a espectros ha realizado alguna colaboración puntual. Por otra parte, nos comenta que está actualmente muy interesado en las simulaciones. Es posible realizar desarrollos mediante Machine Learning o Deep Learning que sean capaces de, a partir de las observaciones y de toda la información disponible, intentar reconstruir toda la historia hasta el origen del universo.

Profundizando sobre algoritmos en concreto, Marc nos comenta que el Machine Learning clásico se ha quedado obsoleto. Funciona correctamente pero falta profundidad y por ello los investigadores no suelen encontrarlo útil. Actualmente los mejores resultados vienen de aplicar Deep Learning. Nos comenta Marc que podría ser interesante revisar los Variational Autoencoders (VAEs), que es una mejora del Random Forest utilizando redes convolucionales. También nos aconseja mirar técnicas de Active Learning, que son algoritmos capaces de identificar objetos problemáticos. Con esto del Active Learning Marc Huertas se refiere principalmente a descartar en las investigaciones los Outliers “obvios”, los que son producto de problemas de instrumentación (errores, calibración, etc...). También habla de la posibilidad de usar Generative Adversary Networks (GANs).

Por último, Marc nos indica que con espectros apenas ha trabajado y que le resulta un tema interesante, incluso propone el desarrollo de alguna máquina o software capaz de crear y simular espectros. También nos propone comparar los resultados que obtengamos, con el resultado del algoritmo Random Forest de Dalya Baron.

Nos indica que le mantengamos informado de nuestra evolución en este proyecto y se pone a nuestra disposición para cualquier duda o necesidad que pudiera surgir.

1.5. Entrevista 5. Evaluación de los avances alcanzados en el tratamiento de datos.

Acta	
Fecha	02/07/2019
Lugar	Instituto Astrofísico de Canarias
Asistentes	Jorge Sánchez-Almeida (Astrofísico del IAC) Ana Monreal (Astrofísico del IAC) Silvia Peraza Vicente Tetuani Pedro Martín

Entrevista de corte técnico para evaluar los avances realizados en la parte de tratamiento de datos y evolución de los mismos. Llevamos con nosotros una copia impresa del guión que se puede consultar en el siguiente apartado.

Ana trae impreso un documento en francés, (nos lo enviará por correo), donde se explica de forma clara y resumida qué es una galaxia, de qué está compuesta, tipos de galaxias, qué es un espectro, diferentes tipos de espectros (emisión, absorción).

También lleva consigo una copia impresa de un libro de ML aplicado a la astrofísica. No tenemos ocasión de revisarlo durante la entrevista, pero comenta que nos puede servir de apoyo.

Entre sus papeles también nos enseña una copia de un paper de Jorge del 2010, donde nos indica que debemos leernos el apartado 3.1. También queda pendiente que nos envíe este documento.

Verificación de los primeros pasos en el tratamiento de los datos

En base al nuestro guión impreso, se les presenta una muestra del dataset con un total de 535 galaxias, creado a partir de los archivo fits que nos indicaron en reuniones anteriores.

Comparando las gráficas y los espectros representados, validamos con ellos que la representación que muestra la web de SDSS no tiene aplicada el redshift, por lo que el desfase observado entre ambas es normal.

Se confirma pues que los primeros pasos en el tratamiento de los datos se han dado en la dirección correcta.

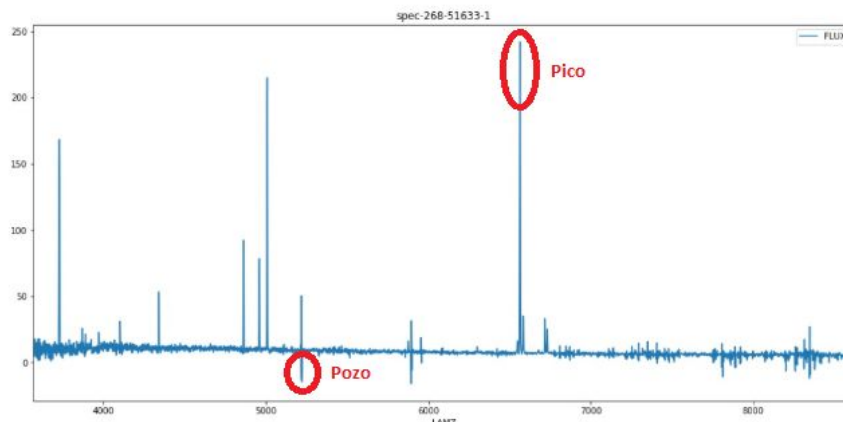
Resolución de dudas

Se plantean las siguientes dudas sobre los datos obtenidos:

- ★ Desplazamiento al rojo (redshift). Jorge nos explica que el universo está en expansión debido al Big Bang y los objetos que viajan más rápido, se separan de nosotros a mayor y por lo tanto están más lejos. Si observáramos las galaxias de cerca, los valores coincidirían con los contenidos en la BBDD, sin embargo, como la observación es desde La Tierra, hay que aplicar esta corrección.
- ★ Las longitudes de onda de Sloan son en el vacío, por lo que es posible nuestros resultados difieran un “poco” por la influencia de la atmósfera terrestre.
- ★ Líneas de emisión y absorción: Ana nos explica que las líneas vienen derivadas de cambios bruscos en el espectro y se representan como líneas muy oscuras (absorción) o muy claras (emisión) en el mismo.

Gracias a ellas se pueden identificar factores claves de la galaxia, como su edad, forma o composición respecto a elementos químicos. Nos confirman que gracias a ellas se pueden calcular el factor de corrección del redshift (Z).

Las líneas de emisión son los picos y las de absorción son los pozos.



Jorge nos explica que, como el flujo es la luz que llega, tiene que ser un valor positivo. Si nos encontramos con valores negativos se debe a fallos durante la observación, por ejemplo a causa de la atmósfera de La Tierra (imposible de evitar ya que observamos desde abajo). Sin embargo, nos indica que los algoritmos que hagamos tienen que ser inmunes a este ruido.

- ★ Los elementos Sky Flux y Best Fits en las gráficas de SDSS nos muestra el espectro teniendo en cuenta la luz recibida, propias del entorno de la tierra (luna, contaminación lumínica), mientras que el segundo es una aproximación media de los valores de la gráfica. En cualquier caso no se utilizan ninguna de las dos medidas.
- ★ Respecto a los elementos que nos interesan para clasificar las galaxias como “raras” nos indican algunos aspectos de los espectros que pueden resultar interesantes:
 - Galaxias solapadas: se debe que estamos observando una galaxia que tiene una delante, por lo que el espectro tiene líneas de ambas galaxias.
 - Columnas anchas: el hecho de encontrar líneas de emisión anchas puede indicar galaxias con mucho movimiento interno.
 - Doble pico o asimétrica: líneas con doble cumbre o que no representan una forma gaussiana simétrica.

- Forma de “fantasma invertido”: tres picos donde las líneas laterales son más altas que la central.
 - Espectros altamente rojos: aquellos donde la línea principal (se llama línea principal la tendencia del espectro filtrando los picos) tiende a ascender a medida que nos acercamos al rojo (mayor longitud de onda).
 - Alta excitación (Helio 2)
 - Falsos positivos: hay que prestar atención a los espectros que se identifican como outliers pero que en realidad son espectros defectuosos por mediciones erróneas.
- ★ Ana nos comenta que en la BBDD podemos encontrar objetos catalogados como “GALAXY” erróneamente, ya que estas etiquetas están puestas por ellos (se refiere a Sloan).

Siguientes pasos

Una vez confirmado que el cálculo del dataset es correcto, para continuar con el tratamiento de los datos y que se puedan aplicar los algoritmos, hay que realizar los siguientes pasos:

- ★ Normalizar el muestreo. Para poder comparar las galaxias, hay que normalizar las muestras en el eje de las abscisas. Para ello tomaremos las frecuencias del primer dataset y realizaremos las mediciones en ellas para todos los espectros utilizando interpolación.
- Ana nos comenta que por ejemplo podríamos usar un rango 3.500 a 9.000, separados en 0,4 Ångstrom.
- ★ Crear un dataset, donde cada columna sea una de estas frecuencias (por ejemplo, los logaritmos) y cada fila sea cada un espectro.
 - ★ Para identificar un espectro usaremos la combinación: Fiber + MJD + Plate.
 - ★ Ana comenta que en total podremos encontrar con unas 2.000 columnas. Si no podemos procesar toda la información, existen zonas de espectros más interesantes que otras para su estudio, por lo que se podría acortar.
 - ★ Ana comenta también que no vamos a usar el patrón estelar (Template).

- ★ Normalizar Flux. Es importante normalizar la intensidad de la luz, porque sino nuestros algoritmo clasificarían los más luminosos y los menos luminosos (ver el artículo de 2010 de Jorge).

Para ello se realizará la mediana dentro de una ventana de longitudes de onda para cada espectro y se normalizarán las mediciones en este sentido. A efectos comparativos, esta ventana será la misma que utilizó Sánchez Almeida en su estudio.

1.6. Entrevista 6. Presentación de los primeros resultados.

Acta	
Fecha	26/08/2019
Lugar	Instituto Astrofísico de Canarias
Asistentes	Jorge Sánchez-Almeida (Astrofísico del IAC) Ana Monreal (Astrofísico del IAC) Pedro Martín Arvin Daswani Silvia Peraza José Luis Quintero

Consideramos esta entrevista cómo la entrega del sprint 1 para el Cliente. Dada la situación creada con los clusters y las dificultades en su obtención, esta primera entrega consistirá en la presentación de la arquitectura y la muestra de algunos outliers resultantes.

Durante la entrevista se aprecia la diferencia de expectativas entre los dos representantes con los que hablamos: Jorge estaba mucho más interesado en la segmentación (posiblemente porque este es el trabajo que había realizado con anterioridad), mientras que a Ana le interesan más los outliers. Esta diferencia de expectativas complica la presentación, dado que deriva casi toda la reunión a las dificultades que hemos tenido para la obtención de los clusters.

Hablamos sobre la conveniencia o no de normalizar los datos como lo hemos hecho y del tratamiento de la aleatoriedad en los resultados. A Jorge le parece adecuado que en los clusters resultantes se mantengan tan solo un 70% de los miembros, cosa que a nosotros en principio nos parecía demasiado poco (considerábamos que estos resultados parecían más un "convenio" que resultados naturales). Con esto en mente, podría ser más sencillo presentar resultados adecuados de cada al Cliente para próximas entregas.

En cuanto a los outliers, les parecen bastante prometedores, sobre todo teniendo en cuenta que solo hemos tratado una muestra de 3000 galaxias. Muestran bastante insistencia en que tomemos una muestra mucho más amplia (al menos unas 20000) para tener resultados más significativos. Tomamos nota de estas expectativas para próximas entregas.

En estos outliers, resulta alguno especialmente interesante, cómo uno donde las líneas de emisión son especialmente anchas.

2. Guión para las entrevistas

Previa a la realización de las entrevistas, hemos elaborado un guión con las preguntas que consideramos más relevantes para poder comprender el negocio, el problema y valorar la viabilidad de nuestra solución:

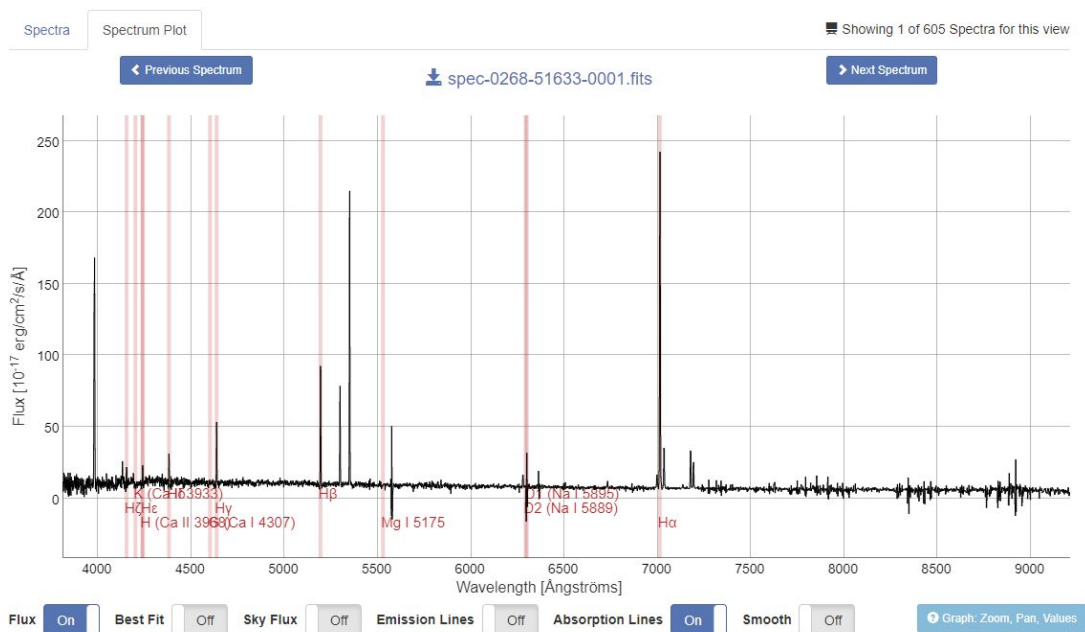
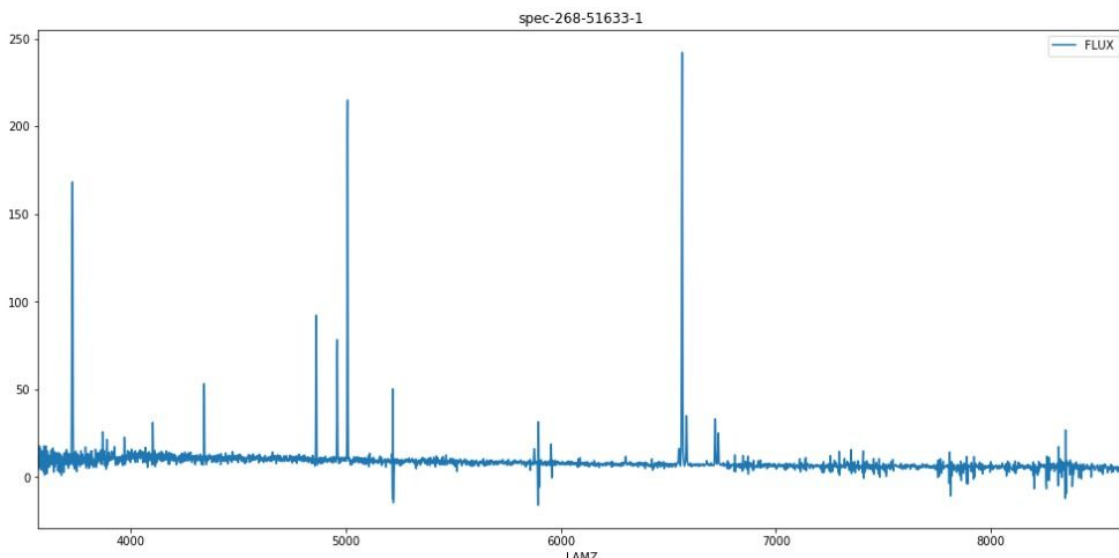
2.1. Guión general

Número	Objeto
1	¿Cómo se diseñan y eligen los proyectos en el IAC?
2	¿Y en otros Institutos de Investigación?
3	¿Cómo se financian los proyectos?
4	¿Quién aprueba la financiación?
5	¿Cómo entran las subcontrataciones?
6	¿Qué papel jugaría Galassify en los proyectos? ¿En cuántos?
7	¿Cómo eligen hoy? ¿Manualmente? ¿Quién hace ese trabajo?
8	¿Han usado algún algoritmo de ML antes? ¿Cuántas veces?
9	¿Qué se hace con la enorme cantidad de datos que sale de los observatorios?
10	¿Cómo colaboran entre sí los científicos? ¿Cómo compiten?
11	¿Cuál es el presupuesto medio de un proyecto del IAC? ¿Y la parte de subcontratación?
12	¿Estarían dispuestos a pagar 80.000€ por Galassify?
13	¿Qué papel juega el prestigio en la aprobación de proyectos?
14	¿Cómo se obtiene prestigio?

2.2. Guión reunión 5 (02/07/2019)

Para esta reunión en concreto, y dado que su contenido era de validación técnica de resultados, elaboramos un guión que permitiera validar los resultados obtenidos hasta ese momento.

1. Los espectros que vienen representados en los Fits, coinciden con los que se pueden consultar en la web sdss.org. Sin embargo en estos espectros parece no estar corregido el redshift. ¿Es así?



- Absorption

- Ca(H) = 3933.7A
- Ca(K) = 3968.5A
- G-band = 4304.4A
- Mg = 5175.3A
- Na = 5894.0 A

- Emission

- [OII] = 3727.3A
- H δ = 4102.8A
- H γ = 4340.0A
- H β = 4861.3A
- [OIII] = 4959.0A
- [OIII] = 5006.8A
- H α = 6562.8A
- S₂ = 6716.0A

2. ¿Qué significa en la gráfica las medidas Best Fit y Sky Flux?

3. Aspecto del Dataset para realizar las pruebas

	FIBER	FLUX	LAM	LAMZ	LOGLAM	MJD	Plate	Z
0	592	3.788501	3815.926059	2852.359438	3.5816	52026	523	0.337814
1	592	2.896744	3816.806003	2853.017186	3.5817	52026	523	0.337814
2	592	0.340875	3817.684054	2853.673519	3.5818	52026	523	0.337814
3	592	1.086873	3818.562307	2854.330003	3.5819	52026	523	0.337814
4	592	5.743586	3819.442859	2854.988205	3.5820	52026	523	0.337814
5	592	1.677638	3820.321517	2855.644991	3.5821	52026	523	0.337814
6	592	-2.625701	3821.202475	2856.303497	3.5822	52026	523	0.337814
7	592	-2.027008	3822.081537	2856.960586	3.5823	52026	523	0.337814
8	592	-0.916980	3822.962901	2857.619394	3.5824	52026	523	0.337814
9	592	2.416374	3823.842369	2858.276786	3.5825	52026	523	0.337814
10	592	6.944658	3824.724138	2858.935898	3.5826	52026	523	0.337814
11	592	-0.521999	3825.604011	2859.593593	3.5827	52026	523	0.337814
12	592	6.659282	3826.484086	2860.251438	3.5828	52026	523	0.337814
13	592	1.728912	3827.366465	2860.911006	3.5829	52026	523	0.337814
14	592	1.299521	3828.246946	2861.569155	3.5830	52026	523	0.337814
15	592	3.765015	3829.129731	2862.229026	3.5831	52026	523	0.337814
16	592	4.049691	3830.010617	2862.887479	3.5832	52026	523	0.337814
17	592	4.486362	3830.893809	2863.547654	3.5833	52026	523	0.337814
18	592	4.167349	3831.775101	2864.206409	3.5834	52026	523	0.337814
19	592	0.212614	3832.656596	2864.865316	3.5835	52026	523	0.337814

4. Líneas de absorción y líneas de emisión. ¿Su principal utilidad es la de calcular Z ? ¿Cómo se identifican?
5. ¿Qué zonas del espectro son las más interesantes?
6. ¿Cómo se reconoce un outlier por su espectro? Ejemplos.
7. Ejemplo de estudio de interés y qué tipo de espectros buscan.
8. Bibliografía interesante (artículos preferiblemente) donde se hable de los espectros y de los outliers.