# INTRODUCCION

Buenos días, mi nombre es Pablo Santidrian y voy a presentar mi proyecto de final de grado llamado Jellyfish Forecast.

Lo primero es explicar las motivaciones y objetivos del proyecto. En los últimos años, las apariciones de medusas han ido en ascenso en muchas partes del mundo. Esto causa grandes impactos económicos ya que, si la gente no va a las playas, el consumo en los establecimientos de los alrededores se verá mermado. Pero también causa impactos en la salud de las personas pues la picadura de algunas especies de medusas puede provocar grandes heridas o incluso la muerte como es el caso de la que trataremos hoy aquí.

La especie que se trata de predecir es la Physalia physalis, o más conocida como la fragata o carabela portuguesa. Esta contiene un veneno muy potente que puede llegar a producir una parada cardiaca.

Por este motivo, es de gran interés el estudio de sus floraciones y el conseguir una herramienta capaz de predecirlas.

Para esto se marcaron una serie de objetivos al principio de proyecto:

Recopilar y tratar una serie de datos necesarios para el entrenamiento del modelo predictivo

Aprender el uso de técnicas de aprendizaje automático.

Crear una aplicación web capaz de mostrar las predicciones.

# TRABAJOS RELACIONADOS

En este aspecto, ya existen ciertos trabajos que tratan el mismo tema desde enfoques diferentes.

Por ejemplo, existen registros de avistamientos en los que cualquiera puedes introducir donde has visto cierta especie de medusa, junto a su tamaño o la densidad de medusas que existe. Es el caso de Perseus, que es un proyecto europeo que estudia el impacto humano en los mares del sur de Europa.

<http://www.perseus-net.eu/en/jellyfish_map/index.html>

El siguiente paso, es que realizan aplicaciones como infomedusa. Esta cuenta con unas serie de observadores locales como pueden ser los socorristas, policías o personal sanitario que aportan información actualizada sobre el estado de las playas.

<https://play.google.com/store/apps/details?id=es.infomedusa&hl=es>

Por último, existen las que realizan predicciones en función que condiciones climatológicas u oceánicas. En este grupo podemos encontrar a NOAA, que se trata de una administración nacional de EEUU que tiene un apartado experimental que predice la probabilidad de encontrar medusas, pero no su abundancia.

[https://ocean.weather.gov/Loops/ocean\_guidance.php?model=Sea\_Net tles&area=Prob&plot=prob&day=0&loop=1](https://ocean.weather.gov/Loops/ocean_guidance.php?model=Sea_Net%09tles&area=Prob&plot=prob&day=0&loop=1)

# TRABAJO

Un siguiente paso a estos proyectos es el de predecir las apariciones de medusas, pero también la cantidad de ellas. Este es el objetivo de nuestro proyecto. Se trata de realizar una predicción del número de medusas que pueden aparecer en las costas de Chile en función de las condiciones oceánicas. A parte de esto, se ha realizado una aplicación web para representar estas y que puedan ser consultadas.

Para conseguir esto necesitábamos datos de los avistamientos de medusas y los datos oceánicos. Los datos de avistamientos nos los proporcionó Antonio Canepa, que ha sido profesor en esta universidad. Estos datos se trataban de un histórico de los reportes de avistamientos realizados por los capitanes de los puertos o el ministerio de salud en la mayoría de los casos. La mayor parte de estos registros se trataban de medusas varadas en la costa.

FOTO DEL EXCEL

La segunda parte de los datos las conseguimos gracias a COPERNICUS. Esta es una organización europea que recopila datos climatológicos y oceánicos por todo el mundo gracias a una serie de satélites. Tienen habilitado un servidor FTP del que se pueden descargar estos paquetes de datos.

Una vez recogidos los datos, era necesario su tratamiento pues contenían información de todo el mundo y una serie de variables que no nos son de utilidad. Para ello se realizó un crop de las coordenadas que abarcaban todo el país de Chile y se eliminaron las variables que no nos eran necesarias. Este paso probablemente en un futuro no sea necesario pues existe una API de la que descargar los datos indicando las coordenadas mínimas y máximas, así como las variables que nos interesan. Al principio del proyecto se probó su uso, pero no funcionaba del todo bien. Tardaba demasiado en devolver los datos de la petición y daba algunos errores a la hora de la descarga por lo que me decante por el uso del FTP.

Una vez hecho esto, se unieron los registros de avistamientos con los datos oceánicos correspondientes. Para esto se realizaron diferentes pruebas para ver cuales ofrecían mejores resultados.

FOTO CUADRANTES

Por parte, se probó con diferente numero de cuadrantes desde el punto del avistamiento. una También se probó con un desfase de fechas respecto del día del registro. El resultado de esto fue que la mejor combinación que se encontró era la de 2 celdas y dos días de desfase.

Finalmente se procedió al entrenamiento del modelo, para esto se utilizó la biblioteca Scikit-Learn de aprendizaje automático en Python. Se probaron diferentes algoritmos de minería de los cuales finalmente el que mejor resultados devolvió fue el de árboles de decisión con el que obtuvo una puntuación de 0,25 utilizando como parámetro de puntuación el coeficiente de determinación o R2.

FOTO MODELOS

A la hora de entrenar estos modelos se vio que los datos iniciales se podían considerar una serie temporal ya que son registros recogidos a lo largo de los años.

FOTO SERIE

Por ello a la hora de entrenar los modelos se utilizó una técnica utilizada en estas series temporales. Es similar a la validación cruzada con la diferencia de que no se utilizar registros del futuro para predecir el pasado.

FOTO TIMESPLITS

Una vez realizado el modelo, se introdujo en una aplicación web creada para este proyecto. La web es bastante sencilla. Cuenta con tres pantallas, una de presentación, una con enlaces de contacto y la página en la que se muestran las predicciones. Esta se mostrará a continuación en un video de muestra.

Además, se han realizado unas pruebas automatizadas para comprobar el buen funcionamiento de la aplicación con Selenium.

FOTO WEB

# HERRAMIENTAS

Para la realización del proyecto se ha utilizado principalmente el lenguaje Python. En el tratamiento de los datos y entrenamiento del modelo utilicé jupyter notebook sobre todo por la facilidad de ejecutar fragmentos sueltos de código y poder consultar los dataframes y dataset de manera más fácil.

Para la ejecución de los entrenamientos y para la descarga de los datos desde el FTP, se me facilitó el acceso remoto a un equipo de cómputo de la universidad ya que ambos suponían dejar el equipo funcionando durante bastante tiempo. Para poder dejar ejecutando los scripts, aunque la conexión estuviese cerrada se utilizó tmux que se trata de una herramienta que nos permite lanzar terminales que corren en segundo plano por lo que, aunque la conexión se cerrase por ejemplo porque que se había desconectado la vpn que era algo relativamente habitual el progreso no se perdía.

En la parte de la aplicación web, se utilizo la biblioteca de Python Flask que nos permite crear una aplicación web. Con esto se ha desplegado a través de la plataforma Heroku. Esta es una herramienta que tiene una parte gratuita que cubre las necesidades de nuestra aplicación. Además, nos ofrece la posibilidad de auto desplegarse cada vez que se realiza un push en nuestro repositorio. Por este motivo es por lo que mi proyecto está dividido en dos repositorios. Para el auto despliegue es necesario que los archivos estén en la raíz del repositorio y no en una subcarpeta.

A parte de esto, se han utilizado múltiples bibliotecas de Python para poder trabajar con las estructuras de datos como pueden ser pandas, xarray, netcdf que era necesario para leer los archivos descargados desde copernicus o folium para cargar los mapas en la aplicación web.

# CONCLUSIONES

Para acabar con la presentación, las conclusiones del proyecto. Se puede decir que el proyecto a cumplido con los objetivos y los requisitos marcados. Desde un principio se sabía que sería complicado conseguir una buena predicción por parte del modelo. Se han realizado muchas pruebas y aun así las predicciones realizadas por el mismo no se acercan mucho a la realidad, aunque hacer funcionar correctamente a un modelo de estas características probablemente lleve suficiente tiempo como para un trabajo de final de master.

A pesar de esto he aprendido sobre diferentes técnicas de aprendizaje automático, así como los tratamientos de datos previos que son necesarios. También he conseguido desarrollar una aplicación web que permite consultar la predicción realizada por el modelo y mostrarla en una serie de gráficos.

Muchas gracias.