



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería
Informática**

**Jellyfish Forecast
Documentación Técnica**



Presentado por Pablo Santidrian Tudanca
en Universidad de Burgos — 19 de junio
de 2020

Tutor: José Francisco Díez Pastor y Álgvar
Arnaiz González

Índice general

Índice general	I
Índice de figuras	III
Índice de tablas	IV
Apéndice A Plan de Proyecto Software	1
A.1. Introducción	1
A.2. Planificación temporal	1
A.3. Estudio de viabilidad	13
Apéndice B Especificación de Requisitos	15
B.1. Introducción	15
B.2. Objetivos generales	15
B.3. Catálogo de requisitos	15
B.4. Especificación de requisitos	16
Apéndice C Especificación de diseño	21
C.1. Introducción	21
C.2. Diseño de datos	21
C.3. Diseño procedimental	21
C.4. Diseño arquitectónico	21
Apéndice D Documentación técnica de programación	23
D.1. Introducción	23
D.2. Estructura de directorios	23
D.3. Manual del programador	23

D.4. Compilación, instalación y ejecución del proyecto	23
D.5. Pruebas del sistema	23
Apéndice E Documentación de usuario	25
E.1. Introducción	25
E.2. Requisitos de usuarios	25
E.3. Instalación	25
E.4. Manual del usuario	25
Bibliografía	27

Índice de figuras

A.1. <i>Burndown chart</i> del Sprint 1	3
A.2. <i>Burndown chart</i> del Sprint 2	4
A.3. <i>Burndown chart</i> del Sprint 3	5
A.4. <i>Burndown chart</i> del Sprint 4	6
A.5. <i>Burndown chart</i> del Sprint 5	7
A.6. <i>Burndown chart</i> del Sprint 6	8
A.7. <i>Burndown chart</i> del Sprint 7	9
A.8. <i>Burndown chart</i> del Sprint 8	10
A.9. <i>Burndown chart</i> del Sprint 9	11
A.10. <i>Burndown chart</i> del Sprint 10	12
B.1. Diagrama de casos de uso	17

Índice de tablas

A.1. Equivalencias <i>Story Points</i> y tiempo estimado	2
A.2. Costes de personal	13
A.3. Licencias de bibliotecas y herramientas utilizadas	14
B.1. Caso de uso 1: Visualizar predicción	17
B.2. Caso de uso 2: Exportar resultados	18
B.3. Caso de uso 3: Visualización del mapa	18
B.4. Caso de uso 4: Consulta predicción	19
B.5. Caso de uso 4.1: Consulta modelo predictivo	19
B.6. Caso de uso 4.2: Selección de la fecha a consultar	20

Apéndice A

Plan de Proyecto Software

A.1. Introducción

En el siguiente apéndice se realizará una estimación de tiempo, trabajo y dinero que será necesario invertir para la realización de este proyecto.

En una primer apartado se tratará sobre la evolución temporal que ha tenido el proyecto, indicando los principales cambios y objetivos de cada *Sprint*, así como una estimación del tiempo que se estimaba, sería necesario invertir antes de la realización del mismo.

En la segunda parte se centra en la viabilidad del proyecto tanto en el ámbito económico (estimación de costes y beneficios) como en el legal (licencias software, etc.).

A.2. Planificación temporal

En este apartado se procederá a explicar con detalle cuál ha sido el resultado de la planificación del proyecto. Esta planificación se ha realizado utilizando una metodología ágil basada en *sprints* de una duración de una o dos semanas en función de las necesidades y el tiempo disponible debido a otras cargas de trabajo diferentes a este proyecto.

En estos *sprints* se van marcando ciertos objetivos que serán revisados junto a los tutores en las reuniones al final de los mismos. Los objetivos del siguiente *sprint* serán marcados durante dichas reuniones.

Para el control de tiempos se ha utilizado la herramienta ZenHub siendo la valoración de los *Story Points* la siguiente:

Story Points	Estimación temporal
1	1 hora
2	1,5 horas
3	2 horas
4	2,5 horas
5	3 horas
6	3,5 horas
7	4 horas
8	6 horas
9	9 horas

Tabla A.1: Equivalencias *Story Points* y tiempo estimado

Sprint 1 (29/01/2020 - 05/02/2020)

En esta primera reunión se marcó el comienzo del proyecto. Ya se había hablado anteriormente con uno de los tutores (José Francisco) del interés sobre el proyecto propuesto del que también formaba parte de los tutores Álar Arnaiz.

Al ser la primera reunión se habló de las herramientas que se iban a utilizar así como acordar los primeros objetivos de este *sprint*:

- Crear el repositorio.
- Añadir la plantilla de L^AT_EX a la documentación.
- Crear cuenta en la plataforma *Copernicus*.
- Investigar el funcionamiento básico de las librerías a utilizar.
- Leer una serie de papers que me proporcionaron sobre las medusas.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Se estimó unas 10 horas de trabajo de las que finalmente se invirtieron 8 horas quedando sin terminar una *issue*.

Figura A.1: *Burndown chart* del Sprint 1

Sprint 2 (13/02/2020 - 28/02/2020)

En la segunda reunión se comentó la existencia de una API para la descarga de los datos meteorológicos como alternativa a la descarga de una gran cantidad de datos a través del FTP.

Por otro lado, se me proporcionaron apuntes de la asignatura de minería de datos para su lectura y aprendizaje.

Por último, los tutores me recomendaron iniciar la documentación del plan de proyecto de los *Sprints* que se fuesen sucediendo para no acumular trabajo y se pudiera olvidar detalles del mismos.

Los objetivos fueron los siguientes:

- Realizar script para la descarga de los datos.
- Comenzar a documentar el plan del proyecto.
- Lectura de apuntes y papers.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Se estimaron unas 8 horas de trabajo de las que finalmente se invirtieron 9 horas.

Figura A.2: *Burndown chart* del Sprint 2

Sprint 3 (28/02/2020 - 17/03/2020)

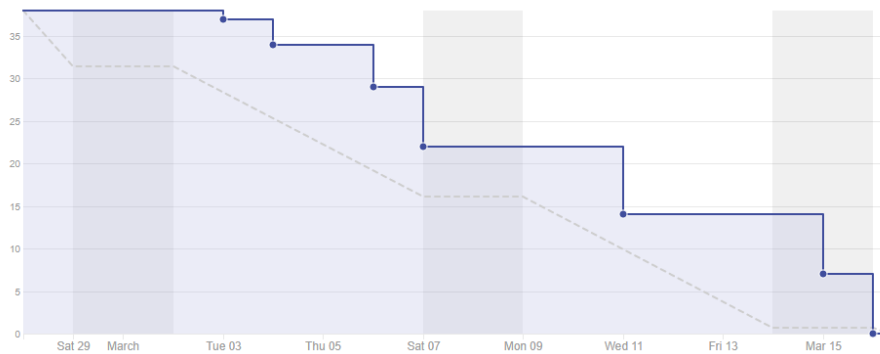
En esta tercera reunión hablamos sobre la descarga de los datos, de las dos opciones posibles nos quedamos con la descarga por FTP por ser más fiable. Además este *Sprint* se centrará en su mayor parte en documentación como las herramientas a utilizar o cuestiones teóricas sobre las medusas aparte de comenzar a desarrollar la base de la aplicación web.

Se marcaron los siguientes objetivos:

- Comienzo desarrollo web.
- Elaboración de parte de la documentación teórica.
- Descarga de los datos en un equipo de cómputo habilitado en la Universidad.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Se estimó unas 19 horas y finalmente se realizaron 24. La causa del desvío de horas principalmente fueron: el comienzo del desarrollo web por el desconocimiento previo y el comentar el código creado para la descarga de los datos necesarios pues se corrigieron errores y se mejoró la salida por pantalla con una barra de descarga más visual.

Figura A.3: *Burndown chart* del Sprint 3

Sprint 4 (17/03/2020 - 30/03/2020)

La cuarta reunión se hizo mediante *Skype* con José Francisco debido a la cuarentena por el coronavirus. Se habló sobre la necesidad de utilizar la VPN de la universidad por este mismo motivo para poder tener acceso a la maquina remota.

Por otra parte, se mostró el avance de la web acordando que el siguiente paso debería ser la introducción de los mapas para lo que se comentaron varias bibliotecas de las que se podía hacer uso.

Los objetivos que se marcaron fueron:

- Continuación desarrollo web.
- Introducción de mapas en la aplicación web.
- Conexión a la VPN de la universidad para conseguir dejar los datos descargándose aunque la sesión esté cerrada.
- Continuación de la documentación.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Figura A.4: *Burndown chart* del Sprint 4

Se estimaron 15 hora y media y finalmente se realizaron 16.

Sprint 5 (30/03/2020 - 14/4/2020)

Esta reunión se realizó también de manera remota por *Skype* con ambos tutores. Se mostró la implementación de los mapas en la aplicación web así como varias mejoras en la interfaz de la misma. También avances realizados en la memoria y ciertas mejoras propuestas por los tutores.

Sobre la web, a pesar de haber empezado su desarrollo, se recomendó el realizar unos bocetos de la misma para definir la estructura a seguir.

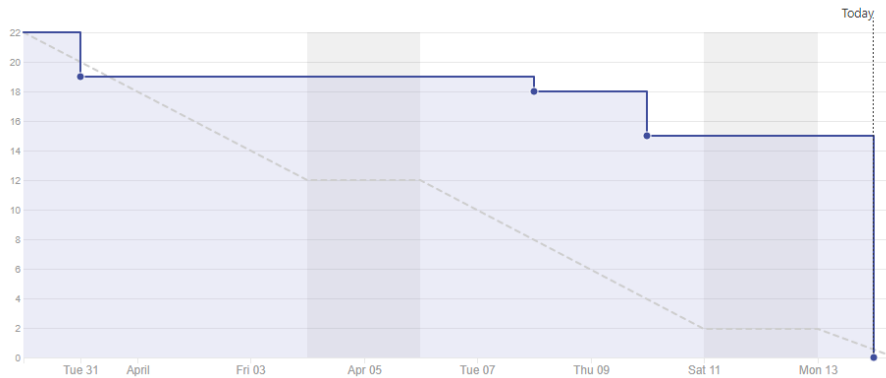
Por otro lado, una vez descargados los datos oceánicos el siguiente paso será generar una estructura de datos para poder entrenar al modelo.

Por último, se acordó que el siguiente paso en la realización de la memoria debía ser la finalización de los objetivos del proyecto y definir los requisitos.

Los objetivos que se marcaron fueron los siguientes:

- Generar estructura de datos.
- Definir objetivos del proyecto.
- Definir requisitos.
- Definir aspectos relevantes.
- Realizar bocetos aplicación web.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Figura A.5: *Burndown chart* del Sprint 5

Se estimaron 12,5 horas y finalmente se realizaron 18,5.

Sprint 6 (15/4/2020 - 24/4/2020)

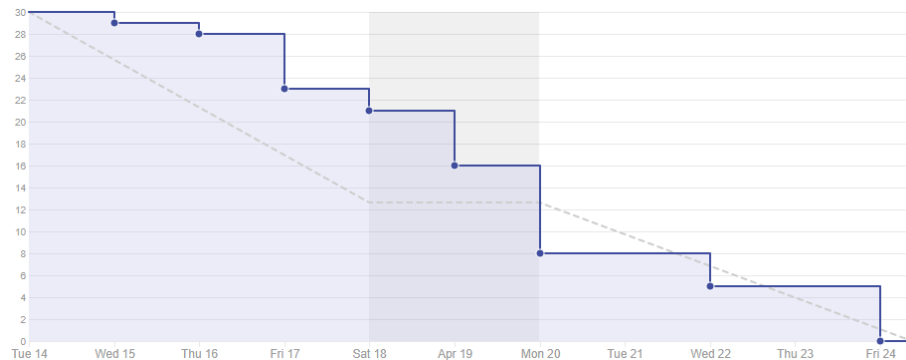
En esta reunión se mostró los avances realizados en la generación de la estructura de datos que posteriormente se utilizará en la creación del modelo de predicción. Se comentaron diferentes problemas encontrados como la existencia de valores nulos en los datos de origen así como la mejor forma de localizar los cuadrantes contiguos a las playas.

También se realizaron cambios en la memoria y anexos.

Para este sprint se marcaron como objetivo:

- Añadir mas cuadrantes a la estructura de datos mencionada así como asegurar la selección de los cuadrantes contiguos a las playa.
- Redacción de casos de uso.
- Investigar el despliegue de la página web.
- Añadir diferentes partes documentación.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Figura A.6: *Burndown chart* del Sprint 6

En un principio se estimaron 16 horas y media y finalmente se realizaron 21 horas.

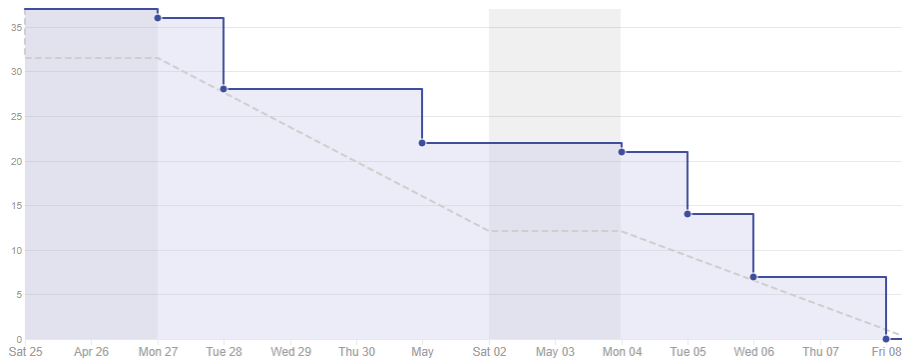
Sprint 7 (25/4/2020 - 8/5/2020)

La temática de la reunión de este sprint fue similar a la anterior. Se habían solucionado los problemas con los valores nulos y añadido más cuadrantes a la estructura de datos, esto dio lugar a algunos problemas por la orientación de las playas. Por este motivo se revisarán las playas para ver si es posible reducir el número de las mismas y así hacer una predicción más precisa. En relación a esto, se empezará con la realización del modelo predictivo aprendiendo la utilización de la biblioteca *Scikit-Learn*. Por último, se mostró el auto despliegue de la aplicación web con *heroku* desde un repositorio a parte del original.

Para este sprint se marcaron como principales objetivos:

- Mejoras en la interfaz de la aplicación web.
- Revisar los datos obtenidos de las playas para reducir el dataset inicial si fuese necesario.
- Aprender la utilización de *scikitLearn*.
- Corrección y ampliación de anexos y memoria.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Figura A.7: *Burndown chart* del Sprint 7

En un principio se estimaron 20 horas y finalmente se realizaron 27.

Sprint 8 (12/5/2020 - 2/6/2020)

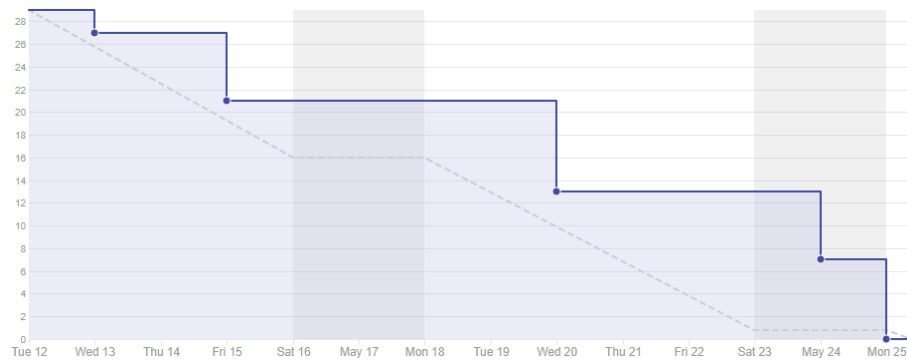
Esta reunión se basó principalmente en la realización del modelo. Se comentaron algunas correcciones en la estructura de los datos (valores nulos, desfase temporal en la organización de los avistamientos).

Por otra parte, hablamos de diferentes algoritmos de minería de datos que se podría utilizar para realizar pruebas y con los que se obtienen mejores resultados así como otras transformaciones que realizar a la estructura de datos para lograr mejores resultado como la normalización de los datos o dar más importancia a unas clases que a otras.

Para este sprint se marcaron como principales objetivos:

- Arreglos en la estructura de datos.
- Investigar los diferentes algoritmos.
- Probar algoritmos de minería de datos.
- Documentar las pruebas y teoría de minería de datos.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Figura A.8: *Burndown chart* del Sprint 8

En un principio se estimaron 12 horas y finalmente se realizaron 20.

Sprint 9 (3/6/2020 - 10/6/2020)

La reunión estuvo basada en los algoritmos de predicción a utilizar. Se mostraron los diferentes algoritmos probados con algunos resultados y se propuso por parte de los tutores la prueba de alguno más.

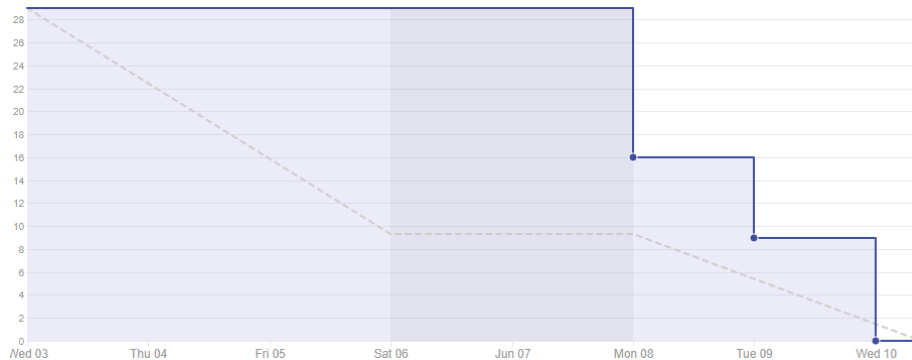
Aparte de esto, hablamos sobre la necesidad de tratar los datos como series temporales debido a la naturaleza de los mismos. También se obtuvo un nuevo conjunto de datos con un mayor número de lecturas de las que se tenía hasta el momento.

Para este sprint se marcaron como principales objetivos:

- Seguir realizando pruebas con diferentes algoritmos de minería.
- Investigar sobre las series temporales.
- Tratar nuevo Excel.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Gráfico

Figura A.9: *Burndown chart* del Sprint 9

En un principio se estimaron 14 horas y finalmente se realizaron 22 hora y media.

Sprint 10 (10/6/2020 - 17/6/2020)

La temática de la reunión fue la misma que las anteriores. Las pruebas realizadas aplicando las series temporales a los modelos no aportaron buenos resultados por lo que se decidió realizar estas con validación cruzada y viendo que algoritmo arroja mejores resultados, utilizarlo finalmente con las series temporales.

En cuanto a la página web, se mostró la introducción de algunas funcionalidades viendo que existían bugs que subsanar.

Para este sprint se marcaron como principales objetivo:

- Pruebas de algoritmos con validación cruzada.
- Arreglar fallos e introducir gráfico en la página web.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

Figura A.10: *Burndown chart* del Sprint 10

En un principio se estimaron 18 horas y finalmente se realizaron 41.

Sprint 11 (17/6/2020 - 24/6/2020)

En primer lugar, se trató las ultimas pruebas realizadas con los modelos. Se obtuvo un resultado aceptable por lo que será el modelo a introducir finalmente en la pagina web.

Por otra parte, se mostró los avances en la aplicación web que estaba casi terminada por completo. Se propuso por parte de los tutores la introducción de un nuevo gráfico mostrando un historial de los avistamientos en la playa seleccionada. En cuanto a la forma de introducir el modelo, se barajaron varias formas. Finalmente se subirán una serie de los datos de origen con los que pueda trabajar el modelo.

Para este sprint se marcaron como principales objetivo:

- Finalizar la aplicación web.
- Dejar finalizada la documentación a falta de la revisión final por parte de los tutores.

Las *issues* para este *Sprint* se pueden ver [aquí](#).

En un principio se estimaron XX horas y finalmente se realizaron XX.

Resumen

A.3. Estudio de viabilidad

Viabilidad económica

Costes

Coste software

Todas las herramientas software utilizadas son de uso gratuito por lo que el coste de este apartado es cero.

Coste hardware

El coste de dispositivos es muy bajo. Unicamente se ha utilizado un ordenador personal valorado en 800 euros. Suponiendo una amortización de 5 años, el coste amortizado seria:

$$\frac{800}{12 * 5} * 5 = 66,67\text{€} \quad (\text{A.1})$$

Alpha?

Coste de personal

Se considera que el proyecto se ha realizado por un desarrollador trabajando 6 horas diarias durante 5 meses:

Concepto	Coste
Salario bruto del trabajador	1200 €
Contingencias comunes (23,6 %)	283,2 €
Desempleo (5,5 %)	66 €
Fogasa (0,2 %)	2,4 €
Formación profesional (0,6 %)	7,2 €
Coste total mensual	1558,80 €

Tabla A.2: Costes de personal

El coste del empleado seria de 1558,80 euros mensuales. Si aplicamos esto por los 5 meses de trabajo el coste total del personal ascendería a 7794€.

Beneficios

El proyecto esta realizado para ser disfrutado de manera gratuita y sin publicidad, por lo que a corto plazo no se obtendrían beneficios. Se podría plantear la posibilidad de agregar algunas funcionalidades extras a futuro, siendo estas de pago a través de algún tipo de suscripción mensual.

Viabilidad legal

Para la realización del proyecto se han utilizado multitud de biblioteca de Python de dominio público. A continuación se expondrán las principales herramientas utilizadas.

Tabla A.3: Licencias de bibliotecas y herramientas utilizadas

Librería	Versión	Descripción	Licencia
VsCode	1.46.1	Editor de código.	MIT
Jupyter Notebook	6.0.3	Aplicación para el desarrollo de código en múltiples lenguajes.	BSD
Scikit-Learn	0.22.1	Biblioteca para aprendizaje automático en Python.	BSD
Flask	1.1.1	Biblioteca para crear aplicaciones web en Python.	BSD
Jinja2	2.11.1	Motor para integrar Python con documentos HTML.	BSD
Folium	0.11.0	Biblioteca para generar mapas en Python	MIT
tmux	3.1b	Multiplexador de terminales.	ISC
Bootstrap	3.3.7.1	Biblioteca para desarrollar aplicaciones web responsive.	MIT
JQuery	3.2.1	Biblioteca para optimizar JavaScript.	MIT
Morris.js	0.5.1	Biblioteca para crear gráficos en Javascript y HTML.	MIT

Apéndice B

Especificación de Requisitos

B.1. Introducción

En este apéndice se recogerán los diferentes objetivos del proyecto con sus correspondientes requisitos funcionales y no funcionales que marcan el desarrollo de este proyecto.

B.2. Objetivos generales

Este proyecto tiene como objetivo el desarrollo de un modelo que nos permita predecir la presencia de medusas en las costas de Chile en función de las condiciones marítimas.

El modelo resultante se utilizará en un aplicación web con la que poder consultar la predicción de la fecha requerida ayudando a su visualización mediante representaciones gráficas.

B.3. Catálogo de requisitos

Requisitos funcionales

RF-1 Obtención de los datos: Se debe ser capaz de descargar los datos necesarios de manera automática a través de un servidor FTP.

RF-2 Filtrado de los datos: Los datos descargados han de ser tratados, descartando las zonas geográficas distintas al lugar de estudio así como las variables ambientales que no sean de utilidad.

RF-3 Cruce de datos: Los datos filtrados se han de cruzar con los de avistamientos, obteniendo una estructura de los avistamientos con su respectiva fecha, localización y variables marítimas.

RF-4 Consultar modelo: La aplicación debe ser capaz de realizar una consulta al modelo predictivo a partir de los datos introducidos por el usuario.

RF-5 Mostrar mapa: La aplicación debe ser capaz de mostrar un mapa con el que poder interactuar.

RF-6 Introduccion de datos para su visualización: El usuario debe poder seleccionar una serie de datos con los que realizar las consultas.

RF-6.1 Elección de fechas: Se debe poder seleccionar una fecha de la que obtener información.

RF-6.2 Elección de playa: Se debe poder elegir una playa de la que obtener información.

RF-6.3 Visualización de resultados: El usuario debe ser capaz de visualizar los resultados de una playa en la fecha especificada.

RF-7 Exportación de resultados: Se deben poder descargar los resultados de la consulta.

Requisitos no funcionales

RNF-1 Rendimiento: La aplicación debe tener buenos tiempos de respuesta.

RNF-2 Usabilidad: La aplicación debe ser intuitiva, de manera que al usuario no le suponga un esfuerzo su uso.

RNF-3 Diseño *responsive*: Se debe garantizar una correcta visualización en diferentes dispositivos de distintas dimensiones.

RNF-4 Internacionalización La aplicación debe disponer de varios idiomas.

B.4. Especificación de requisitos

Actores

Solo existe un tipo de actor, aquel usuario que consulta las predicciones.

Diagrama de casos de uso

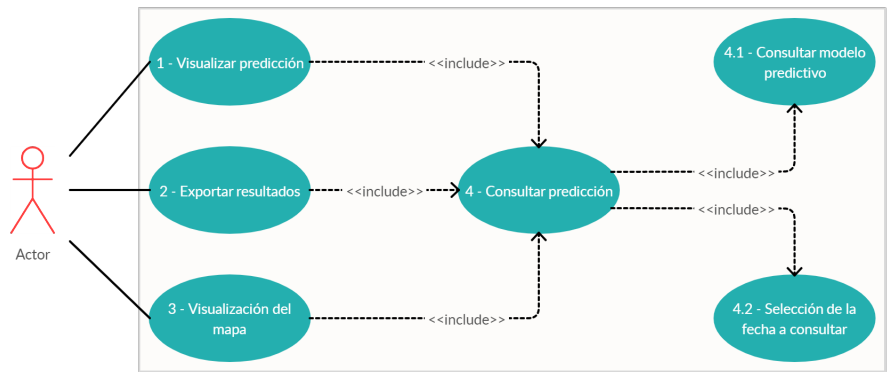


Figura B.1: Diagrama de casos de uso

Casos de uso

CU-1: Visualizar predicción		
Descripción	Permite al usuario visualizar toda la información relativa a la consulta realizada.	
Pre-condiciones	La fecha introducida, es una fecha válida.	
	El nombre de la playa introducida, es una nombre válido.	
Requisitos	RF-5, RF-6	
Secuencia normal	Paso	Acción
	1	El usuario debe acceder a la pestaña «Mapas».
	2	El usuario introduce una fecha.
	3	El usuario podría elegir una playa en particular si lo desea o hacer una búsqueda general.
	4	El usuario pulsa el botón de búsqueda.
Excepciones	5	Se muestra por pantalla el resultado de la consulta.
	1	Fecha introducida no válida.
Frecuencia	2	El nombre de la playa no es válido.
Importancia	Alta	

Tabla B.1: Caso de uso 1: Visualizar predicción

CU-2: Exportar resultados		
Descripción	Permite al usuario descargar el resultado de la consulta. La fecha introducida, es una fecha válida.	
Pre-condiciones	El nombre de la playa introducida, es una nombre válido.	
Requisitos	RF-6, RF-7	
	Paso	Acción
Secuencia normal	1	El usuario debe acceder a la pestaña «Mapas».
	2	El usuario introduce una fecha.
	3	El usuario podría elegir una playa en particular si lo desea o hacer una búsqueda general.
	4	El usuario pulsa el botón de búsqueda.
	5	Se muestra por pantalla el resultado de la consulta.
	6	El usuario pulsa el botón de descarga.
	7	El archivo se descarga en el dispositivo.
Excepciones	1	Fecha introducida no válida.
	2	El nombre de la playa no es válido.
Frecuencia	Baja	
Importancia	Alta	

Tabla B.2: Caso de uso 2: Exportar resultados

CU-3: Visualización del mapa		
Descripción	Permite al usuario visualizar un mapa sobre el que se superpondrán los datos	
Pre-condiciones	-	
Requisitos	RF-5	
	Paso	Acción
Secuencia normal	1	El usuario debe acceder a la pestaña «Mapas».
	2	Se muestra la ventana de consultas en la que aparece el mapa.
Excepciones	-	
Frecuencia	Alta	
Importancia	Alta	

Tabla B.3: Caso de uso 3: Visualización del mapa

CU-4: Consultar predicción		
Descripción	Permite al usuario realizar una consulta.	
Pre-condiciones	La fecha introducida, es una fecha válida.	
	El nombre de la playa introducida, es una nombre válido.	
Requisitos	RF-4	
Secuencia normal	Paso	Acción
	1	El usuario debe acceder a la pestaña «Mapas».
	2	El usuario introduce una fecha.
	3	El usuario podría elegir una playa en particular si lo desea o hacer una búsqueda general.
	4	El usuario pulsa el botón de búsqueda.
Excepciones	1	Fecha introducida no válida.
	2	Las coordenadas no existen o no son válidas.
Frecuencia	Alta	
Importancia	Alta	

Tabla B.4: Caso de uso 4: Consulta predicción

CU-4.1: Consultar modelo predictivo		
Descripción	La aplicación web realiza una consulta al modelo predictivo con los parámetros marcados por el usuario.	
Pre-condiciones	-	
Requisitos	RF-4	
Secuencia normal	Paso	Acción
	1	La aplicación web realiza una llamada al modelo predictivo.
	2	El modelo devuelve el resultado de la consulta.
	3	La aplicación web imprime por pantalla dichos resultados.
Excepciones	1	Fecha introducida no válida.
	2	Las coordenadas no existen o no son válidas.
Frecuencia	Alta	
Importancia	Alta	

Tabla B.5: Caso de uso 4.1: Consulta modelo predictivo

CU-4.2: Selección de la fecha la consultar		
Descripción	El usuario elige una fecha en la que realizar a consulta.	
Pre-condiciones	-	
Requisitos	RF-4	
	Paso	Acción
Secuencia normal	1	El usuario debe acceder a la pestaña "Mapas".
	2	El usuario introduce una fecha.
Excepciones	1	Fecha introducida no válida.
Frecuencia	Alta	
Importancia	Alta	

Tabla B.6: Caso de uso 4.2: Selección de la fecha a consultar

Apéndice C

Especificación de diseño

- C.1. Introducción
- C.2. Diseño de datos
- C.3. Diseño procedimental
- C.4. Diseño arquitectónico

Apéndice D

Documentación técnica de programación

- D.1. Introducción
- D.2. Estructura de directorios
- D.3. Manual del programador
- D.4. Compilación, instalación y ejecución
del proyecto
- D.5. Pruebas del sistema

Apéndice E

Documentación de usuario

- E.1. Introducción
- E.2. Requisitos de usuarios
- E.3. Instalación
- E.4. Manual del usuario

Bibliografía
