**Capítulo 3 - Procesamiento de datos para predicción de oleaje**

Básicamente, la información que utilizamos para entrenar a los clasificadores consiste en dos grandes grupos de datos. El primero es un histórico de datos del modelo WW3, situados en cuatro coordenadas alrededor de la isla Oahu, Hawaii a unos cuantos kilómetros de la costa, de la cual disponemos de diferentes lecturas del comportamiento del océano (periodo de ola, altura de ola, dirección de ola, dirección del viento y velocidad del viento) cada 3 horas durante los 365 dias del año desde 1997 hasta la fecha. El segundo grupo es una base de datos de observaciones visuales de la altura de las olas de diferentes playas alrededor de toda la costa de Oahu desde el año 1987 hasta el año 2004 inclusive. Estas observaciones realizadas por un experto en el area nos brinda por cada dia del año la altura de la ola más grande que arribo a la costa ese dia. Además estas observaciones sabemos (ya que la fuente lo especifica) que fueron realizadas durante el periodo de luz solar de dia en esa playa.

A partir de estos dos grupos de datos, comenzamos a implementar diferentes estrategias de filtrado y utilizacion de los mismos hasta encontrar una que optimice los aciertos en la prediccion de las olas en las diferentes costas de la isla.

**Datos de entrada:**

Modelo WW3

* Altura de ola (metros).
* Período de ola (segundos).
* Dirección de ola (grados).
* Dirección del viento (coordenadas cartesianas).
* Velocidad del viento (metros por segundo).

Observación visual

* Altura de la ola (HSF, escala hawaiana en pies). Esta es una medida adoptada por los surfistas hawaianos, fue convertida en una medida más representativa para cualquier persona, que indica la altura en metros del frente de la ola desde la base de la misma a la cresta (Through to crest scale).

Dado que las lecturas de WW3 se actualizan cada 3 horas, tanto durante el día como la noche, y debido a que las observaciones visuales se han realizado en horarios en que se cuenta con luz solar, el primer filtro aplicado a la información que ofrece el modelo WW3 fue eliminar todas las lecturas que no correspondían con un horario de luz natural en Oahu.

Por otra parte, las observaciones realizadas nos indican la altura de la ola más grande que arribo a la costa cada día del año, es decir que disponemos de una lectura diaria en HSF, lo cual fue convertida a metros (1 HSF = 2 trough-to-crest aproximadamente, 1 pie = 0,3048 metros) por cuestiones de practicidad, por esto el segundo filtro aplicado a las lecturas de WW3 fue dejar únicamente la lectura diaria con la mayor altura de ola. De esta forma ambos conjuntos de datos cuentan con una lectura diaria que indica la ola más grande captada y durante una hora en Oahu en la que hay luz solar.

Los filtros anteriormente descriptos fueron los que finalmente se utilizaron como filtrado básico de los datos de entrada en todas las evaluaciones que se describirán en el capítulo siguiente. Adicionalmente se puede dar el caso que alguna de las pruebas desarrolladas a continuación también le aplique algún filtro especifico a los datos de entrada. Otros filtros generales que también tuvimos en cuenta para realizar pruebas (descartados por no mejorar los resultados) fueron por ejemplo, utilizar las lecturas diarias del WW3 en el que la altura de la ola se acerca más al promedio de altura de olas de cada día, filtrar las lecturas cuya dirección de ola no estaba dirigida a la costa a evaluar, agrupar las lecturas en dos grupos (dos rangos de meses) que distinguen la temporada de olas grandes de la temporada de olas pequeñas, etc. .