

UNIVERSIDAD DE ALCALÁ

**Escuela Politécnica Superior**

Máster Universitario en Ingeniería de  
Telecomunicación



Proyecto Fin de Máster

**DESARROLLO DE SOLUCIONES DE  
INTELIGENCIA ARTIFICIAL PARA  
DATACIÓN DE ESTRELLAS**

Autor: Jarmi Recio Martínez  
Director: Roberto Javier López Sastre

**TRIBUNAL:**

*Presidente: D. Juan Antonio Martínez Rojas*

*Vocal 1º: D. Óscar Rodríguez Polo*

*Vocal 2º: D. Roberto Javier López Sastre*

**CALIFICACIÓN:**..... **FECHA:**.....



# Desarrollo de soluciones de inteligencia artificial para datación de estrellas

Jarmi Recio Martínez.

28 de julio de 2021



*A mi familia y amigos.*



# Agradecimientos

A Roberto, por darme la oportunidad de aportar mi granito de arena en la que ha sido mi primera publicación.





# Índice general

<b>Agradecimientos</b>	<b>V</b>
<b>Resumen</b>	<b>XVII</b>
<b>Abstract</b>	<b>XIX</b>
<b>Resumen Extendido</b>	<b>XXI</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Trabajos relacionados</b>	<b>5</b>
2.1. Girocronología . . . . .	5
2.2. Modelos de IA para datación estelar . . . . .	6
<b>3. Base de datos</b>	<b>9</b>
3.1. Entendiendo la base de datos . . . . .	11
<b>4. Modelos de Inteligencia Artificial para datación estelar</b>	<b>15</b>
4.1. Modelos . . . . .	15
4.2. Escenarios de evaluación . . . . .	18
4.3. Resultados . . . . .	23
4.3.1. Configuración experimental . . . . .	24
Implementación . . . . .	24
Métricas de evaluación . . . . .	24
Gráficas . . . . .	24
4.3.2. Benchmark A: Problema de datación estelar . . . . .	24
Stacking . . . . .	25
Neural Network . . . . .	27
Gaussian Process . . . . .	27
kNN . . . . .	27
Random Forest . . . . .	28
Support Vector Regression . . . . .	28
Decision Tree Regression . . . . .	29

Linear Regressor . . . . .	29
Bayesian Regression . . . . .	29
4.3.3. Benchmark B: Escenario de generalización . . . . .	29
4.3.3.1. Benchmark B1 . . . . .	31
Neural Network . . . . .	31
4.3.3.2. Benchmark B2 . . . . .	32
Bayesian Regression . . . . .	33
kNN . . . . .	34
Random Forest . . . . .	35
Support Vector Regression . . . . .	35
Decision Tree Regression . . . . .	35
Linear Regressor . . . . .	35
Bayesian Regression . . . . .	35
Gaussian Process . . . . .	35
kNN . . . . .	36
Random Forest . . . . .	36
Support Vector Regression . . . . .	37
Decision Tree Regression . . . . .	37
Linear Regressor . . . . .	37
Neural Network . . . . .	38
Gaussian Process . . . . .	38
4.3.4. Benchmark C: Rendimiento sobre la muestra de control . . . . .	39
Gaussian Process . . . . .	40
Stacking . . . . .	42
Neural Network . . . . .	42
kNN . . . . .	42
Random Forest . . . . .	42
Support Vector Regression . . . . .	43
Decision Tree Regression . . . . .	43
Linear Regressor . . . . .	44
Bayesian Regression . . . . .	44
4.3.5. Conclusiones de la comparativa . . . . .	45
<b>5. Meta-learning para estimación de edades estelares . . . . .</b>	<b>47</b>
5.1. Introducción . . . . .	47
5.2. Model-Agnostic Meta-Learning Algorithm . . . . .	48
5.3. MAML en regresión . . . . .	49
5.4. MAML para datación estelar . . . . .	52
5.4.1. Inyección de los datos . . . . .	52
5.4.2. Proceso de aprendizaje . . . . .	55
5.4.3. Resultados obtenidos . . . . .	56

<i>ÍNDICE GENERAL</i>	IX
<b>6. Conclusiones</b>	<b>59</b>
6.1. Modelos clásicos de IA para datación de estrellas . . . . .	59
6.2. Meta-Learning aplicado al problema de la estimación de la edad de las estrellas	60
6.3. Conclusiones personales . . . . .	61
<b>Bibliografía</b>	<b>63</b>



# Lista de figuras

1.	Benchmark A: Rendimiento de los modelos en función del MAE. En esta configuración, las dos mejores aproximaciones son Neural Network (nnet) y Gaussian Process (gp), y por lo tanto, su Stacking. . . . .	XXIII
2.	Benchmark B1: Rendimiento de los modelos en función del MAE. . . . .	XXIV
3.	Benchmark B2: Rendimiento de los modelos en función del MAE. . . . .	XXIV
4.	Benchmark C: Rendimiento de los modelos en función del MAE. . . . .	XXV
3.1.	Diagrama HR mostrando la muestra MS. . . . .	11
3.2.	$M - P_{\text{rot}}$ para las estrellas elegidas. . . . .	12
3.3.	$M - P_{\text{rot}}$ para las estrellas elegidas. La edad se muestra en escala de color. . . .	12
3.4.	$Edad - P_{\text{rot}}$ para las estrellas elegidas. Las estrellas astrosísmicas y de cúmulos se muestran en colores diferentes. También se representa la regresión lineal obtenida usando estos dos grupos. . . . .	13
4.1.	Arquitectura de la red neuronal implementada. Se han usado 3 capas ocultas, con 50 unidades en cada una seguida de una ReLU. La salida no tiene función de activación, para realizar directamente la estimación de la edad. . . . .	18
4.2.	Benchmark A: Distribución de las estrellas que forman el subconjunto de entrenamiento. . . . .	19
4.3.	Benchmark A: Distribución de las estrellas que forman el subconjunto de prueba. . . .	19
4.4.	Benchmark B1: Distribución de las estrellas que forman el subconjunto de entrenamiento. . . . .	20
4.5.	Benchmark B1: Distribución de las estrellas que forman el subconjunto de prueba. . . .	20
4.6.	Benchmark B2: Distribución de las estrellas que forman el subconjunto de entrenamiento. . . . .	21
4.7.	Benchmark B2: Distribución de las estrellas que forman el subconjunto de prueba. . . .	22
4.8.	Benchmark C: Distribución de las estrellas que forman el subconjunto de entrenamiento. . . . .	23
4.9.	Benchmark C: Distribución de las estrellas que forman el subconjunto de prueba. . . .	23
4.10.	Benchmark A: Rendimiento de los modelos en función del MAE. En esta configuración, las dos mejores aproximaciones son Neural Network (nnet) y Gaussian Process (gp), y por lo tanto, su Stacking. . . . .	25

4.11. Benchmark A: Rendimiento para el modelo Stacking. . . . .	26
4.12. Benchmark A: Predicción detallada para el modelo de Stacking. Se observa la edad real (en azul), la predicción de los modelos (en naranja), y el error correspondiente (en rojo). . . . .	26
4.13. Benchmark A: MAE en función del rango de edad del conjunto de prueba. . . . .	27
4.14. Modelo Neural Network . . . . .	28
4.15. Modelo Gaussian Process . . . . .	28
4.16. Modelo k-Nearest Neighbor . . . . .	28
4.17. Modelo Random Forest . . . . .	28
4.22. Benchmark B1: Rendimiento de los modelos en función del MAE. . . . .	29
4.18. Modelo Support Vector Regression . . . . .	30
4.19. Modelo Decision Tree Regressor . . . . .	30
4.20. Modelo Linear Regressor . . . . .	30
4.21. Modelo Bayesian Regression . . . . .	30
4.23. Benchmark B2: Rendimiento de los modelos en función del MAE. . . . .	30
4.24. Benchmark B1: Rendimiento para el modelo Neural Network. . . . .	31
4.25. Benchmark B1: Predicción detallada para el modelo de Neural Network. Se observa la edad real (en azul), la predicción de los modelos (en naranja), y el error correspondiente (en rojo). . . . .	32
4.26. Benchmark B1: MAE en función del rango de edad del conjunto de prueba. . . . .	32
4.27. Benchmark B2: Rendimiento para el modelo Bayesian Regression. . . . .	33
4.28. Benchmark B2: Predicción detallada para el modelo de Bayesian Regression. Se observa la edad real (en azul), la predicción de los modelos (en naranja), y el error correspondiente (en rojo). . . . .	34
4.29. Benchmark B2: MAE en función del rango de edad del conjunto de prueba. . . . .	34
4.30. Modelo k-Nearest Neighbor . . . . .	36
4.31. Modelo Random Forest . . . . .	36
4.32. Modelo Support Vector Regression . . . . .	36
4.33. Modelo Decision Tree Regressor . . . . .	36
4.34. Modelo Linear Regressor . . . . .	37
4.35. Modelo Bayesian Regression . . . . .	37
4.36. Modelo Gaussian Process . . . . .	37
4.37. Modelo k-Nearest Neighbor . . . . .	38
4.38. Modelo Random Forest . . . . .	38
4.39. Modelo Support Vector Regression . . . . .	38
4.40. Modelo Decision Tree Regression . . . . .	38
4.41. Modelo Linear Regressor . . . . .	39
4.42. Modelo Neural Network . . . . .	39
4.43. Modelo Gaussian Process . . . . .	39
4.44. Benchmark C: Rendimiento de los modelos en función del MAE. . . . .	40

4.45. Benchmark C: Rendimiento para el modelo Gaussian Process. El punto amarillo se corresponde con la representación de la estimación de la edad del Sol. . . . .	41
4.46. Benchmark C: Predicción detallada para el modelo de Gaussian Process. Se observa la edad real (en azul), la predicción de los modelos (en naranja), y el error correspondiente (en rojo).El punto amarillo se corresponde con la representación de la estimación de la edad del Sol. . . . .	41
4.47. Benchmark C: MAE en función del rango de edad del conjunto de prueba. . . .	42
4.48. Modelo Stacking . . . . .	43
4.49. Modelo Neural Network . . . . .	43
4.50. Modelo k-Nearest Neighbor . . . . .	43
4.51. Modelo Random Forest . . . . .	43
4.52. Modelo Support Vector Regression . . . . .	44
4.53. Modelo Decision Tree Regression . . . . .	44
4.54. Modelo Linear Regressor . . . . .	44
4.55. Modelo Bayesian Regression . . . . .	44
5.1. Resultados del estudio de regresión sobre las ondas sinusoidales con $K = 5$ . . .	51
5.2. Resultados del estudio de regresión sobre las ondas sinusoidales con $K = 10$ . . .	51
5.3. Resultados del estudio de regresión sobre las ondas sinusoidales con $K = 20$ . . .	52
5.4. Esquema de la inyección de datos durante la fase de entrenamiento. . . . .	54
5.5. Esquema de la inyección de datos durante la fase de prueba. . . . .	54
5.6. Esquema del funcionamiento interno del modelo durante el proceso de entrenamiento. . . . .	55
5.7. Esquema del funcionamiento interno del modelo durante el proceso de prueba. . .	56
5.8. Comparación del MSE para las distintas configuraciones de la red neuronal . .	57





# Lista de tablas

1.	Edad estimada para el Sol en Gyr. El método más preciso es rf. . . . .	XXV
3.1.	Características proporcionadas en la base de datos para cada estrella. . . . .	9
4.1.	Edad estimada para el Sol en Gyr. El método más preciso es rf. . . . .	40
4.2.	Precisión de todos los métodos en los diferentes benchmarks. La precisión se mide como el porcentaje de estimaciones de edad que caen dentro del margen de confianza asociado a cada estrella. . . . .	45
5.1.	MSE para las diferentes arquitecturas y cada una de las actualizaciones de gradiente. . . . .	57



# Resumen

En astronomía, la edad es una de las propiedades estelares más difíciles de medir y la girocronología es una de las técnicas más prometedoras para esta tarea. Consiste en datar estrellas usando su periodo de rotación y relaciones lineales empíricas con otras propiedades estelares observadas. Sin embargo, estas aproximaciones no permiten reproducir todos los datos observados, lo que resulta en posibles desviaciones en la estimación de la edad.

En este contexto, se propone explorar el problema de la datación estelar utilizando la girocronología desde la perspectiva de la Inteligencia Artificial (IA). Para hacerlo, se presenta un estudio exhaustivo de evaluación comparativa de los modelos de IA en regresión de última generación entrenados y probados para la datación estelar mediante girocronología. También se explora en este trabajo cómo aplicar modelos de meta-learning para el problema de la datación estelar con pocos datos. Los resultados de ambos estudios son bastante prometedores, y abren una nueva línea de investigación en el campo de la estimación de la edad de las estrellas mediante IA.

**Palabras clave:** inteligencia artificial, aprendizaje automático, datación estelar, girocronología, astronomía.



# Abstract

In astronomy, age is one of the most difficult stellar properties to measure, and gyrochronology is one of the most promising techniques for this task. It consists of dating stars using their period of rotation and empirical linear relationships with other observed stellar properties. However, these approximations do not allow to reproduce all the observed data, which results in possible deviations in the age estimate.

In this context, it is proposed to explore the problem of stellar dating using gyrochronology from the perspective of AI. To do so, we present a comprehensive benchmarking study of state-of-the-art regression Artificial Intelligence models trained and tested for gyrochronology stellar dating. How to apply meta-learning models to the stellar dating problem with little data is also explored in this paper. The results of both studies are quite promising, and open a new line of research in the field of estimating the age of stars using AI.

**Keywords:** artificial intelligence, machine learning, stellar dating, gyrochronology, astronomy.



# Resumen Extendido

La edad de las estrellas es una de las propiedades más difíciles de medir, ya que no puede observarse de forma directa, y a su vez, es algo fundamental para profundizar en otros problemas astrofísicos, como la búsqueda de vida fuera del sistema solar. Entre las técnicas disponibles para abordar esta tarea se encuentra la girocronología, que permite estimar las edades estelares mediante el periodo de rotación de las estrellas.

Las estrellas nacen con una determinada velocidad de rotación y sufren una reducción de esta velocidad por el proceso de frenado magnético. Se sabe que cuanto mayor sea la rotación, mayor será su frenado, y por lo tanto, todas las estrellas tienden a converger a una velocidad de rotación similar para una edad y masa concreta. Es la girocronología la técnica que explota este hecho.

Esta prometedora técnica presenta una gran ventaja frente a otras, que consiste en la facilidad para obtener entradas de observación a analizar. Aunque, por otro lado, existen una serie de incertidumbres que la convierten en una técnica de estimación de edades estelares poco precisa, además de presentar limitaciones en cuanto al rango de masa de las estrellas sobre las que se puede aplicar esta técnica.

Teniendo esto en cuenta, se propone explorar el problema de la datación estelar utilizando la girocronología desde la perspectiva de la Inteligencia Artificial. La idea consiste en reemplazar las combinaciones lineales y las técnicas tradicionales que se emplean en la girocronología convencional por una aproximación de regresión de aprendizaje automático.

Hasta el momento, existen pocos trabajos que aborden la datación de estrellas mediante técnicas de Inteligencia Artificial. Por ejemplo, en [53], se presenta por primera vez una Red Neuronal Bayesiana aplicada a la estimación de la edad de las estrellas.

En este trabajo se propone un enfoque diferente al que han seguido otros autores en sus estudios, más detallados en la Sección 2, ya que los modelos anteriores abordan el problema desde la perspectiva bayesiana. En cambio, aquí se propone un problema de regresión puro, donde los diferentes modelos estudiados se enfrentan a la estimación de un valor particular para la edad y se evalúan en consecuencia.

La base de datos principal de este trabajo se compone de 1464 estrellas. Este conjunto de datos contiene varias características de cada una de las estrellas, tomando especial importancia: la temperatura efectiva  $T_{\text{eff}}$ , el logaritmo de la gravedad superficial  $\log g$ , la masa estelar  $M$ ,

el radio estelar  $R$ , la metalicidad  $[\text{Fe}/\text{H}]$ , el periodo de rotación  $P_{\text{rot}}$  y la edad a estimar. Estas edades precisas provienen de la astrosismología y de la pertenencia a cúmulos. Las 4 edades principales de los cúmulos son: 0.79, 1, 2.5 y 4.2 Gyrs.

Para seleccionar las estrellas que alimentarán los modelos en la fase de entrenamiento, validación y test, se realiza un filtrado previo sobre la base de datos comentada. En ella se seleccionan las estrellas FGK (estrellas de clases espectrales medias) y con periodo de rotación inferior a 50 días. El motivo de este filtrado se explica mediante la física que hay detrás de la girocronología, ya que ésta se ve limitada en un rango de masa de entre 0.7 y 2, y los periodos de rotación superiores a 50 días presentan dificultades a la hora de explicarse con la estructura estelar actual.

Es necesario añadir que esta muestra exhibe dos sesgos importantes: 1) la muestra perteneciente a la astrosismología está sesgada hacia las estrellas masivas y viejas; y 2) la muestra basada en cúmulos está sesgada hacia edades más jóvenes.

Los modelos que se han alimentado con esta muestra están detallados en la Sección 4.1 y son: Linear Regressor, Bayesian Regression, Decision Tree Regressor, Random Forest Regressor, Support Vector Regression, Gaussian Process, kNN, Neural Network y Stacking, siendo este último un método de ensamblado de modelos formado por Neural Network y Gaussian Process en el nivel 0 y por Linear Regressor en el nivel 1. La Red Neuronal empleada es un perceptron multicapa, está formada por 3 capas ocultas *fully-connected* con 50 unidades en cada una de ellas.

El estudio de rendimiento de los modelos y su correspondiente comparativa se lleva a cabo mediante tres *benchmarks*. En estos, se proponen tres escenarios de evaluación con el fin de evaluar diferentes capacidades mediante el error medio absoluto (MAE) de los modelos. El MAE permite cuantificar la precisión de las estimaciones de edad comparando el valor predicho frente al observado.

**Problema de datación estelar (Benchmark A)** Configuración para evaluar los modelos de regresión bajo un esquema clásico de división de datos en entrenamiento y prueba, siguiendo una distribución de 80 %/20 % respectivamente. En este escenario, los modelos que mejor rendimiento muestran son Neural Network, Gaussian Process y su Stacking, presentando un MAE inferior a 0.46 Gyrs, tal y como se puede apreciar en la Figura 1.



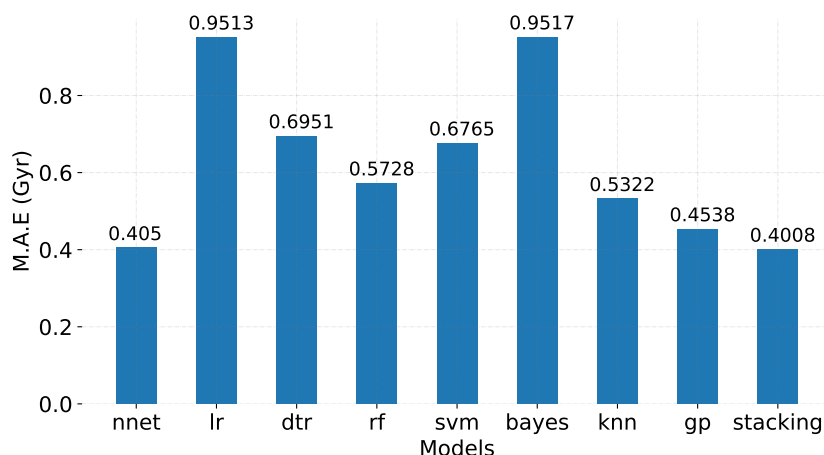


Figura 1: Benchmark A: Rendimiento de los modelos en función del MAE. En esta configuración, las dos mejores aproximaciones son Neural Network (nnet) y Gaussian Process (gp), y por lo tanto, su Stacking.

**Capacidad de generalización (Benchmark B)** Este Benchmark se compone de dos configuraciones diferentes: 1) Benchmark B1, donde se evalúa la capacidad de los modelos para estimar edades de estrellas cuyo rango está fuera del de las estrellas vistas durante el entrenamiento; y 2) Benchmark B2, en el que se evalúa la capacidad de generalización entrenando los modelos con las estrellas pertenecientes a cúmulos y probando con el resto de estrellas. Los resultados de los modelos para estos dos escenarios muestran peor MAE que para el Benchmark A, concluyendo que todos los modelos presentan dificultades a la hora de generalizar. Para la configuración del Benchmark B1, el modelo que mejor resultado presenta es el de Neural Network, mientras que para el Benchmark B2 es uno de los modelos que peor rendimiento muestra, siendo Bayesian Regression el que mejor resultado devuelve en este escenario. Ver Figura 2 y Figura 3 para visualizar todos los rendimientos.

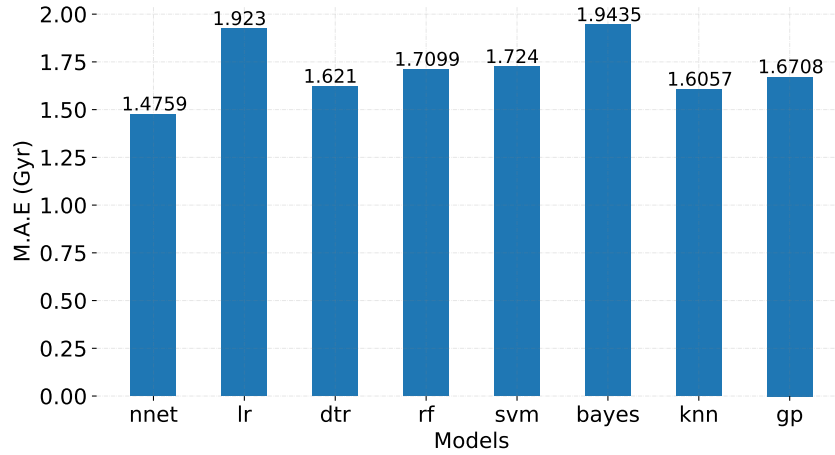


Figura 2: Benchmark B1: Rendimiento de los modelos en función del MAE.

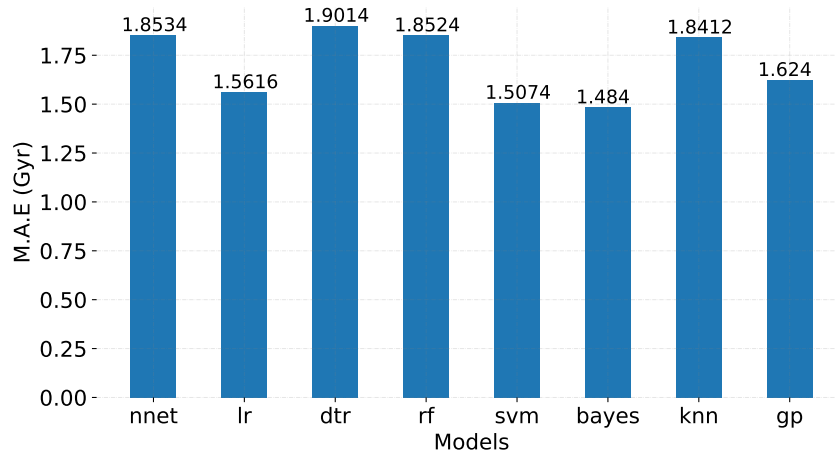


Figura 3: Benchmark B2: Rendimiento de los modelos en función del MAE.

**Regresión sobre una muestra de control (Benchmark C)** En este último escenario se evalúan los modelos sobre una muestra de control de 32 estrellas, extraídas del conjunto de datos filtrado y desconocidas en todos los entrenamientos realizados. Entre estas 32 estrellas se encuentra el Sol, haciéndose énfasis en la estimación de su edad. En esta ocasión, los tres mejores modelos son Gaussian Process, Neural Network y su Stacking. La Figura 4 muestra el MAE para cada modelo. En cuanto a la estimación de la edad del Sol, tal y como se puede observar en la Tabla 1, el modelo que más se aproxima a los 4.6 Gyrs es Random Forest Regressor, estimando una edad de 4.76 Gyrs.

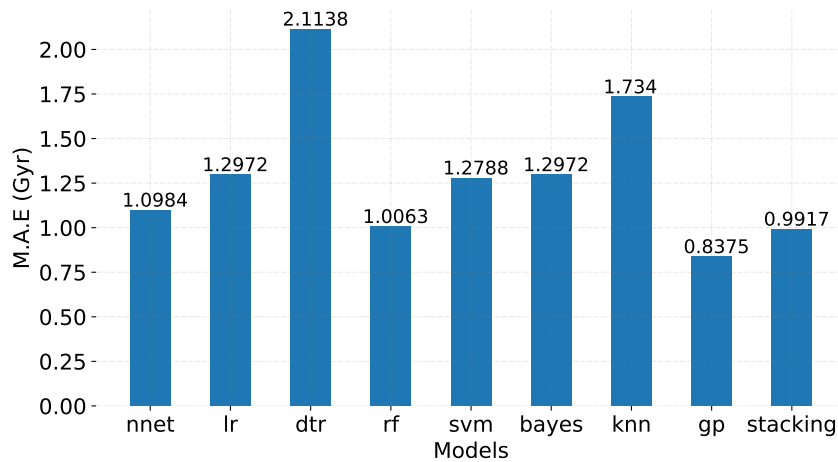


Figura 4: Benchmark C: Rendimiento de los modelos en función del MAE.

Modelos	nnet	lr	dtr	rf	svm	bayes	knn	gp	stacking
Sol (4.6 Gyr)	4.02	3.92	4.49	4.76	3.81	3.91	4.49	3.88	4.02

Tabla 1: Edad estimada para el Sol en Gyr. El método más preciso es rf.

Se obtienen tres conclusiones claras tras el estudio: 1) el modelo de Neural Network proporciona buenos resultados en 3 de los 4 escenarios evaluados, pudiendo asignarle el título simbólico de ganador; 2) todos los modelos tienden a subestimar, por lo que se cree que sería interesante tratar de corregir este sesgo en futuras líneas de investigación; y 3) el estudio revela resultados prometedores, obteniendo MAEs inferiores a 0.5 Gyrs, pudiendo considerarse como gran avance en el campo de la datación de estrellas.

Además del *benchmarking* de los modelos clásicos de Inteligencia Artificial, se realiza un estudio sobre la aplicación del algoritmo MAML (*Model-Agnostic Meta-Learning*) [20] para el problema de estimación de edades estelares. Este algoritmo se denomina agnóstico del modelo porque es compatible con cualquier algoritmo entrenado con descenso de gradiente. Y se caracteriza por el metaaprendizaje porque se entrena en varias tareas de aprendizaje con el objetivo de que sea capaz de resolver nuevas tareas con una pequeña cantidad de muestras.

La motivación principal para usar este tipo de algoritmos se basa en su capacidad de aprendizaje con pocas muestras, esto es útil cuando se trabaja con conjuntos de datos reducidos. Este es uno de los aspectos principales por los que aplicar este tipo de algoritmos sobre el problema de datación estelar. Además de que se parte de un conjunto de datos relativamente reducido, con el paso del tiempo se obtendrán de conjuntos de datos muy reducidos, provenientes de diferentes fuentes y misiones espaciales. Este tipo de algoritmos facilitan el uso de estos conjuntos reducidos para reforzar los modelos probados con estrellas anteriores.

Durante la fase de metaentrenamiento, el modelo se entrena para que sea capaz de adaptarse a un número muy elevado de tareas, todas con pocas muestras. Esto se consigue mediante la configuración de aprendizaje tipo *K-Shot*, donde el modelo se entrena para que aprenda nuevas tareas usando  $K$  muestras de la observación inicial y la retroalimentación proporcionada por la función de pérdidas para cada tarea. El ajuste del modelo sigue una regla de aprendizaje basada en gradientes, con el objetivo de aprender rápidamente sobre nuevas tareas extraídas de la distribución total sin llegar al sobreajuste.

El funcionamiento detallado del algoritmo y la inyección de los datos relativos a las características de las estrellas se explican en la Sección 5.4. Para evaluar los resultados obtenidos de este experimento, se emplea la misma métrica que se utiliza en el trabajo original [20], el error cuadrático medio entre la edad estelar estimada y la edad estelar real. En este cálculo se tienen en cuenta los errores obtenidos en todos los pasos de actualización de gradiente, por lo que, el error cuadrático medio final se obtienen haciendo la media de todos estos. Esta evaluación se realiza sobre distintas arquitecturas que pueden componer la red neuronal principal del modelo. Originalmente, se desarrolla con una red neuronal de 2 capas ocultas y 40 unidades cada una, pero los mejores resultados obtenidos se consiguen con una arquitectura de 3 capas ocultas y 20 unidades cada una, obteniendo un  $MSE = 0.01$  en sus últimas actualizaciones de gradiente.

Tras analizar los métodos clásicos y el algoritmo MAML se concluye que los resultados obtenidos en ambos experimentos son muy prometedores para el campo de la estimación de edades estelares. En el análisis de modelos clásicos destaca el modelo de la red neuronal, que funciona notablemente en 3 de los 4 escenarios propuestos, aunque se demuestre que todos los modelos tienden a subestimar. Respecto al algoritmo MAML, destaca una configuración de red de 3 capas ocultas con 20 unidades cada una, lo que permite obtener valores de MSE cercanos a 0 en la última actualización de gradiente. Además, se cree conveniente el uso de otros algoritmos de metaaprendizaje que permitan comparar sus resultados con los obtenidos por el algoritmo MAML.

# Capítulo 1

## Introducción

Según expresa Soderblom en [60], se pueden medir muchas de las propiedades físicas claves de las estrellas. La masa se puede determinar a partir de la física básica, o también se puede inferir a partir de su tipo espectral, particularmente para la secuencia principal. La composición de las estrellas, no es fácil de determinar pero, se puede seguir un proceso bien comprendido en el que se conocen sus limitaciones. Sin embargo, la edad, que es el tercer determinante clave del estado físico de una estrella, no se puede medir directamente, solo se puede estimar. Únicamente se tiene la edad precisa y exacta de una estrella, el Sol. Pero esta edad no ha sido revelada por la propia estrella, sino que, se consigue estimar gracias a la posibilidad de estudiar en un laboratorio la materia que compone el sistema solar, estudio que no se puede hacer para ninguna otra estrella.

En algunas ocasiones, es posible determinar las edades si se colocan ciertos eventos en la secuencia correcta y se es capaz de comprender la duración relativa de cada una de las fases. Pero, normalmente se necesita unir observaciones de objetos o sucesos independientes para alcanzar la comprensión de un proceso. A continuación, se añade una serie de ejemplos que remarcan la utilidad de conocer la edad de ciertos objetos y su colocación temporal respecto al resto de sucesos:

- La formación y evolución de los discos protoplanetarios parece ocurrir en los primeros 100 Myr de la vida de una estrella, y los discos de escombros se forman más tarde. En la actualidad, apenas se puede limitar esta escala temporal con los métodos disponibles, y para comprender la física de los procesos implicados, es necesario observar las diferencias en los procesos de formación y conocer la escala temporal entre ellos, además del de las estrellas implicadas.
- Se han descubierto muchos planetas gigantes gaseosos alrededor de estrellas de tipo solar cercanas. Para obtener una mejor comprensión de, por ejemplo, la dinámica de tales sistemas y la migración hacia el interior de los planetas, se requieren edades precisas.
- Se podría comprender mejor el comportamiento y el futuro de nuestro Sol si pudiéramos crear una verdadera cohorte basada en la masa, la composición y la edad. En la actuali-

dad, las comparaciones del Sol con otras estrellas a menudo se basan en las propiedades estelares, el proceso es autorreferencial.

- Para comprender completamente la historia de formación, enriquecimiento y dinámica de la Galaxia, se necesita poder asignar edades a las estrellas individuales. Las estrellas de baja masa viven lo suficiente para estar presentes en todas las épocas de formación estelar, lo que las hace atractivas para estos fines.
- La búsqueda de vida y cualquier comprensión obtenida al observar signos de vida más allá de nuestro sistema solar tiene una profunda necesidad de medir las edades estelares si se espera obtener información sobre la evolución biológica.

La comprensión completa de una gran variedad de fenómenos astrofísicos significa que, al menos, se conoce la secuencia correcta de eventos y sus correspondientes escalas de tiempos. Muchos de estos fenómenos están relacionados con estrellas individuales o conjuntos de estrellas cercanas, por lo que se necesitan formas confiables de llegar a las edades de estos objetos.

Partiendo de que la edad es la propiedad más difícil de obtener, ya que es la única que no se puede observar *directamente* en ningún caso y debe inferirse utilizando diversos métodos, Soderblom presentó dos estudios que describen la mayoría de estas técnicas [59, 60].

Dentro del grupo de aproximaciones para la datación estelar, una de las técnicas más prometedoras es la girocronología o datación de estrellas mediante su periodo de rotación. Las estrellas nacen con un rango de velocidades de rotación por diferentes razones físicas, ver [6]. En este trabajo, Barnes introduce la rotación como la característica observable más fundamental de las estrellas, además de ser una de las que más precisión admiten en su medición. El trabajo de Barnes se enfoca en la interpretación de datos relativos al periodo rotacional. Las observaciones de las estrellas de cúmulos abiertos sugieren que las estrellas en rotación se encuentran principalmente en dos secuencias. Estas secuencias y el número de estrellas que las componen, evolucionan sistemáticamente con la edad de los cúmulos.

Para entender mejor el estudio de Barnes, es necesario explicar que las isócronas son curvas en el diagrama de Hertzsprung-Russell (HR) que representan una población de estrellas de la misma edad. El diagrama de Hertzsprung-Russell, por definición, es un gráfico de dispersión de estrellas que representa la magnitud absoluta de una estrella en luz visible frente a su tipo espectral.

El proceso llamado frenado magnético reduce la velocidad de rotación de la estrella con el tiempo [54]. Desde finales de los 60's [10, 28, 40, 41, 64], se sabe que el frenado rotacional, o ley de frenado, es una función del periodo rotacional. Es decir, cuanto mayor sea la rotación, mayor será el frenado. Por lo tanto, con el tiempo, todas las estrellas tienden a converger a una velocidad de rotación similar para una edad y masa determinadas. Este efecto fue demostrado empíricamente por primera vez por Skumanich [58]. La girocronología es la técnica de datación estelar que explota este hecho [9, 59]. Tradicionalmente, la girocronología se ha aplicado ajustando una expresión lineal que contiene la edad estelar, su periodo de rotación y algunas aproximaciones de su masa. En trabajos recientes, Angus y otros [2] [4] han propuesto nuevas

relaciones lineales para la girocronología combinadas con otras técnicas de datación, como el ajuste de isócronas o la cinemática estelar, en un marco bayesiano para una estimación de edad más precisa.

En [4], por ejemplo, se demuestra que la combinación de la girocronología con el ajuste de isócronas mejora la precisión de las estimaciones de edad para estrellas enanas FGK sobre un único ajuste de isócronas. Además, se determina que la incorporación de los periodos de rotación al modelo de evolución estelar que emplean, ayuda a mejorar la precisión del resto de parámetros estelares. Por otro lado, también se deduce que la incertidumbre y los valores atípicos de los periodos de rotación se traducen en estimaciones inexactas por parte de la girocronología. Como conclusiones generales: 1) la forma óptima de determinar la edad de las estrellas implica combinar todos los observables disponibles que estén relacionados con la edad, y 2) ya que la girocronología funciona bien para estrellas enanas FGK y el ajuste de isócronas funciona bien para subgigantes y estrellas calientes, la combinación de ambos métodos da como resultado edades consistentemente precisas.

Sin embargo, debido a la dificultad para recolectar muestras de datos que sean lo suficientemente grandes para alimentar modelos de aprendizaje automático y que sean lo suficientemente precisas, no se ha intentado abordar el problema de la estimación de la edad de las estrellas como un ejercicio de regresión utilizando técnicas de Inteligencia Artificial. En este trabajo, se explora esta novedosa perspectiva, y las principales contribuciones son las siguientes.

En primer lugar, se ha recopilado un conjunto de datos empleados para determinar edades estelares mediante girocronología. En general, el conocimiento aportado sobre la girocronología se basa principalmente en los cúmulos estelares, los únicos sistemas con una gran cantidad de estrellas y una determinación precisa de la edad. Para este trabajo se ha construido una muestra completa y precisa con estrellas provenientes de diversos cúmulos, pero también provenientes de la astrosismología [1, 42, 61]. La aparición de la astrosismología como una herramienta eficaz para datar estrellas ha permitido añadir una gran cantidad de estrellas de campo a la muestra. Todas las estrellas tienen una temperatura efectiva precisa ( $T_{\text{eff}}$ ), metalicidad ( $[\text{Fe}/\text{H}]$ ), logaritmo de la gravedad superficial ( $\log g$ ), luminosidad ( $L$ ), masa ( $M$ ), radio ( $R$ ), edad y periodo de rotación ( $P_{\text{rot}}$ ). La Sección 3 detalla la construcción de la muestra de datos, analizando fortalezas y debilidades. Supone un reto cuando se quiere estimar las edades estelares utilizando la girocronología, principalmente porque aún no se comprende completamente la física y las dependencias físicas del fenómeno de frenado rotacional. Sin embargo, se cree que este es un problema interesante para que las técnicas de aprendizaje automático ofrezcan lo mejor de sí mismas.

En segundo lugar, se presenta por primera vez un estudio comparativo de modelos de regresión de Inteligencia Artificial de última generación utilizando el conjunto de datos comentado (para conocer en detalle los casos que forman el estudio comparativo, consultar la Sección 4.2). Se incluye en el análisis un conjunto heterogéneo y completo de enfoques, que cubren: kNN clásico y regresión lineal o bayesiana; modelos combinados (bosque aleatorio, staking); proceso gaussiano; y redes neuronales. Se proponen hasta tres *benchmarks* diferentes, donde se comparan todos estos modelos para entender su precisión, robustez y capacidad de generalización.

En tercer lugar, se estudia el algoritmo MAML (Model Agnostic Meta Learning) para tratar de solucionar el problema de la estimación de edades estelares. Este algoritmo se denomina agnóstico del modelo, siendo compatible con cualquier modelo entrenado con descenso de gradiente, y se basa en el metaaprendizaje para conseguir la resolución de nuevas tareas usando un número muy reducido de muestras. Este algoritmo permitirá el uso de conjuntos de datos muy reducidos provenientes de futuros estudios y futuras misiones que aporten información de nuevas estrellas. El análisis detallado del algoritmo y su aplicación se encuentra en la Sección 5.2.

Finalmente, en la Sección 4.3, se revelan los resultados de los experimentos llevados a cabo con los modelos clásicos, donde algunos modelos logran un MAE de  $< 0,5$  Gyr en estimación de edades estelares, lo que puede considerarse un avance significativo en el campo. Además, para fomentar una mayor investigación sobre el problema, se publica la muestra de datos, los códigos para reproducir los experimentos y todos los protocolos de evaluación diseñados. Por otro lado, en la Subsección 5.4.3 se detallan los resultados obtenidos con el algoritmo MAML. Aquí se puede observar un valor de error cuadrático medio muy próximo a 0 para las configuraciones más complejas de la red neuronal principal del modelo.



# Capítulo 2

## Trabajos relacionados

### 2.1. Girocronología

El beneficio principal de la girocronología está relacionado con la facilidad para obtener las entradas de observación requeridas: representaciones fotométricas o espectroscópicas de la masa estelar y el periodo de rotación. Por otro lado, presenta una serie de incertidumbres que la convierten en una técnica de datación poco precisa, además de tener un rango de aplicación relativamente estrecho en términos de masa. En [6] y [7], se publica por primera vez una ecuación empírica que permite estimar la edad estelar en función de su periodo de rotación y su índice de color. Esto puede considerarse como el origen de la girocronología como técnica práctica. Esta relación empírica ha sido revisada posteriormente, en [33], [8] y [1], utilizando una muestra de datos astrosísmicos en la última de estas.

Pero en la vida real no suele ser tan sencillo. El frenado rotacional está ligado a la existencia de un campo magnético estelar bipolar. El mejor candidato para la dinamo que genera este campo es una zona convectiva exterior bien desarrollada. Por lo tanto, solo las estrellas con masas inferiores al límite de Kraft [30], alrededor de 1.3 M, pueden reducir su velocidad de rotación con el tiempo, siguiendo las leyes de ruptura conocidas. Pero, [67] mostró que las estrellas A (estrellas blancas de la secuencia principal) también tienen velocidades de rotación que evolucionan con el tiempo, esto puede deberse a que en estas estrellas también pueden encontrarse fuertes campos magnéticos, o al menos sus firmas [5], y por lo tanto, por encima del límite de Kraft, el frenado rotacional puede seguir existiendo. En este rango de masas, por encima de 1.3 M, el mecanismo exacto de frenado es casi desconocido.

Para las estrellas jóvenes, la girocronología solo puede aplicarse a cúmulos de estrellas o grupos en movimiento [18]. Para las estrellas individuales, la incertidumbre asociada solo permite una declaración general sobre si es o no es joven. Por otro lado, [61] encontró, comparando un conjunto de 21 estrellas datadas usando astrosismología con modelos de frenado rotacional [62], que a partir de cierto valor del número de Rossby, el frenado rotacional parece detenerse y la velocidad rotacional permanece en un estado estacionario, lo que hace que la girocronología tradicional sea casi inútil. En cualquier caso, este cambio en el régimen del frenado rotacional

aún está en debate, con trabajos que lo confirman [24], [29], [42], [43] y [63], y algunos otros que concluyen que no se observa, especialmente en el caso de gemelos solares en términos de masa y metalicidad [9] y [32].

Para profundizar sobre las últimas conclusiones obtenidas en relación a los periodos de rotación se revisa la última de las publicaciones comentada en el párrafo anterior. En [24], se miden los periodos de rotación de más de 8000 estrellas K2 de secuencia principal. En este estudio se detecta la brecha en la distribución del periodo de rotación y se muestra que es poco probable que se trate de una peculiaridad, ya que aparece en todas las campañas de K2. Tras revisar diferentes explicaciones y argumentos sobre la brecha, se deduce que lo más probable es que la brecha resulte de un giro estancado en la secuencia de rotación rápida para estrellas de baja masa, seguido de una rápida evolución a través de la brecha hacia la secuencia de rotación lenta.

## 2.2. Modelos de IA para datación estelar

Se pueden encontrar pocos trabajos que aborden el problema de la estimación de edades estelares mediante Inteligencia Artificial [3], [4], [19] y [53]. Sander y Das presentan en [53] un catálogo con distancias, masas y edades de 3 millones de estrellas de la segunda publicación de datos de Gaia [15]. Las edades se estiman siguiendo un ajuste bayesiano a la estructura estelar y los modelos de evolución, para caracterizar sus funciones de densidad de probabilidad. Angus y otros [3], [4] proponen una combinación de las relaciones lineales para girocronología con isócronas ajustadas bajo un marco de estimación bayesiano. Su enfoque utiliza un proceso de cadena de Markov Monte Carlo para inferir las edades. Es en [53] donde se encuentra por primera vez una Red Neuronal Bayesiana aplicada al problema de estimación de la edad. Técnicamente, Sander y Das [53] entrenan un perceptrón multicapa, con una sola capa oculta de 10 neuronas, para generar distribuciones posteriores predictivas para la masa, edad y distancia de las estrellas. De esta manera, su modelo puede reemplazar la dependencia de la técnica de las isócronas.

Las diferencias de todos estos trabajos con las desarrolladas en este proyecto son las siguientes. Todos los modelos anteriores abordan el problema desde una perspectiva bayesiana, centrándose en las predicciones de distribuciones posteriores de las edades. En cambio, aquí se propone un problema de regresión puro, donde los modelos se enfrentan a la estimación de un valor particular para la edad y se evalúan en consecuencia. Además, por este motivo, la comparación directa con estos trabajos anteriores no es tarea fácil.

Además de los trabajos mencionados, como prueba de lo prometedor que supone la aplicación de técnicas de Inteligencia Artificial para la datación estelar, durante la fase final de este proyecto se han encontrado trabajos recientes en los que se aplica aprendizaje automático para la predicción de edades estelares. En [35], se presenta un modelo formado por dos redes neuronales convolucionales en cascada. Este método de aprendizaje profundo utiliza Gaia DR2 y fotometría 2MASS para predecir las edades estelares de las estrellas previas a la secuencia prin-

cial. En [25], se presenta un estudio de cúmulos abiertos utilizando un telescopio de imágenes ultravioleta a bordo del ASTROSAT y Gaia EDR3. En este trabajo se combinan parámetros astrométricos, fotométricos y sistemáticos para entrenar un algoritmo de aprendizaje automático junto con un modelo de mezcla gaussiana para la determinación de la pertenencia al cúmulo.



# Capítulo 3

## Base de datos

En este capítulo detallamos la base de datos con edades estelares que va a ser utilizada para validar los modelos de IA para estimación de edad estelar. En concreto se trata de una muestra de 1464 estrellas con edades precisas procedentes de la astrosismología o la pertenencia a cúmulos. La muestra astrosísmica consta de 312 entradas para las cuales se han inferido observables estelares fundamentales precisos (temperatura efectiva, logaritmo de la gravedad superficial, masa, radio, contenido estelar de hierro sobre hidrógeno) a partir de una combinación de observaciones fotométricas, espectroscópicas y astrosismología. La contribución más significativa proviene de [55], con 224 entradas. Otras se obtuvieron de [16],[21],[56] y [57].

Característica	Unidad	Descripción
$i.d$	-	ID de la estrella
$T_{\text{eff}}$	K	Temperatura estelar efectiva
$\log g$	dex	Logaritmo de la gravedad superficial
$M$	Masas Solares	Masa estelar
$R$	Radios Solares	Radio estelar
$[\text{Fe}/\text{H}]$	dex	Contenido de Hierro sobre Hidrógeno
$P_{\text{rot}}$	Días	Periodo de rotación estelar
Edad	Gyr	Edad estelar

Tabla 3.1: Características proporcionadas en la base de datos para cada estrella.

En términos de periodo de rotación, de las 312 entradas obtenidas de la astrosismología, se tomaron 293 periodos de [21]. Los periodos restantes se tomaron de [16], [34], [36], [37], [47] y [51]. García y otros [21] analizaron la tasa de rotación de la superficie en el subconjunto de estrellas de tipo solar Kepler, estudiadas en [21]. Los mismos métodos de análisis implementados en [21] se adoptan en [16]. En [47], por otro lado, Nielsen et al. calcularon un periodograma Lomb-Scargle (LS). Eligieron como periodo de rotación el valor mediano de todos los picos registrados de potencia máxima medidos en varios trimestres de los datos de Kepler. En [51] también se usa el periodograma LS, pero restringiendo el análisis exclusivamente a los datos de cadencia larga del tercer trimestre de Kepler.

Se complementa la muestra con estudios de cúmulos realizados por la misión Kepler/K2.

Se recopilan un total de 1152 entradas tomadas de [9], [23], [38], [39] y [50]. En [9] y [23], los autores estudiaron el antiguo cúmulo M67, analizando los datos de las curvas de luz Kepler/K2 Campaign 5. M67 es un objetivo interesante para la girocronología, ya que tiene aproximadamente la misma edad y comparte una composición química similar a la del Sol. M67 también es el grupo más antiguo de la muestra. Barner y otros [9], derivaron los periodos de rotación de superficie utilizando una combinación de cuatro métodos: minimización de la dispersión de fase, longitud mínima de la cuerda, método de detección de la señal del periodo bayesiano y función de autocorrelación. En [23], se utilizó el método de análisis de periodograma de Lomb-Scargle. La edad del cúmulo se estableció por [9] y su concordancia con las edades derivadas de la cromosfera [22] y la isócrona [11]. Meibom y otros en [38] estudiaron NGC 6811 [26] y [39], y reportaron periodos para estrellas en NGC 6819 [27]. Estos dos grupos cierran la brecha de edad entre Praesepe y M67. Estos autores emplearon el método de periodograma LS para obtener los periodos de rotación. Además, para todos los periodos de rotación reportados, examinaron visualmente el periodograma y las curvas de luz, y también verificaron los periodos de forma independiente utilizando el algoritmo CLEAN. Se observó a Praesepe durante la Campaña 5 de Kepler/K2. En [50], los autores identificaron los periodos de rotación de superficie aplicando el periodograma Lomb-Scargle. Tomaron el periodo correspondiente al pico más fuerte del periodograma como periodo de rotación (con algunas excepciones). El estudio produjo periodos para más del 80 % de todas las curvas de luz de Praesepe. La estimación de la edad de este grupo ha sido objeto de cierto debate, con el valor más reciente establecido en  $760 \pm 60$  Myr por [12].

Para trabajar con la mayor exactitud y precisión, se han completado la muestra de los cúmulos con masas y radios derivados con un modelo de Random Forest de aprendizaje automático del paquete R `empiricalRelationsMR` [44].

La muestra es, por lo tanto, una mezcla de cuatro cúmulos con edades de 0.79, 1, 2.5 y 4.2 Gyr que reúnen un total de 1152 estrellas, con masas y radios estimados mediante aprendizaje automático y datos empíricos. Por otro lado, se tienen 312 estrellas con masas, radios y edades determinadas mediante astrosismología y periodos de rotación precisos. Finalmente, se han seleccionado aquellas estrellas FGK y de secuencia principal con periodos de rotación inferiores a 50 días. La razón de este filtrado es que la física detrás de la girocronología ocurre principalmente en estos tipos estelares y, por lo tanto, está limitada a un rango de masas de 0.7 a 2. Además, los periodos de rotación superiores a 50 días difícilmente pueden explicarse por la estructura estelar actual y los modelos de evolución.

Se van a apartar 32 estrellas no agrupadas (incluido el Sol) datadas utilizando astrosismología con fines de prueba. Por lo tanto, se trabaja con una muestra final formada por 397 estrellas en cúmulos más 240 estrellas estudiadas mediante astrosismología, es decir, un total de 637 estrellas.

Desafortunadamente, esta muestra exhibe dos sesgos importantes: 1) la muestra astrosísmica está sesgada hacia las estrellas masivas y ancianas; y 2) la muestra basada en cúmulos está cuantificada por edad y sesgada hacia edades más jóvenes. En los experimentos se muestra que, a pesar de los sesgos, las técnicas de aprendizaje automático pueden extraer información

confiable del conjunto de datos para estimar edades estelares. Con el tiempo, especialmente con las misiones espaciales actuales y futuras, estos sesgos se irán mitigando progresivamente, con la consecuente mejora de las estimaciones.

### 3.1. Entendiendo la base de datos

En la Figura 3.1, se muestra la posición en el diagrama de Hertzsprung-Russell de todas las estrellas MS (secuencia principal) seleccionadas. También se muestra cuáles son los miembros de un cúmulo y cuáles se han caracterizado mediante astrosismología. Aquí se puede ver que las estrellas astrosísmicas cubren la zona más masiva y/o evolucionada del diagrama HR, y las estrellas del cúmulo son más jóvenes y también cubren la región de baja masa.

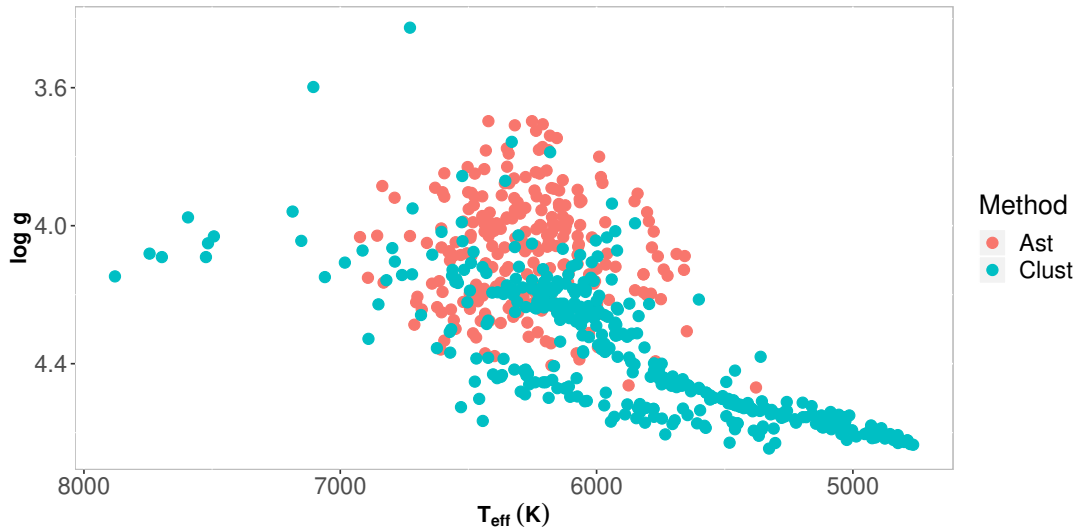


Figura 3.1: Diagrama HR mostrando la muestra MS.

La idea clásica de la girocronología consiste en estimar las edades estelares utilizando cualquier indicador de la masa estelar y el periodo rotacional estelar. Como se tiene una estimación de masa para toda la muestra de entrenamiento, se puede representar directamente  $M$  vs.  $P_{\text{rot}}$  (Fig. 3.2), evitando usar otros indicadores. En este gráfico, también se diferencian las estrellas astrosísmicas de los cúmulos. No hay diferencias claras entre ellos, excepto el sesgo ya mencionado de la muestra astrosísmica hacia estrellas más masivas. El límite de Kraft está claro alrededor de  $1,2M_{\odot}$ .

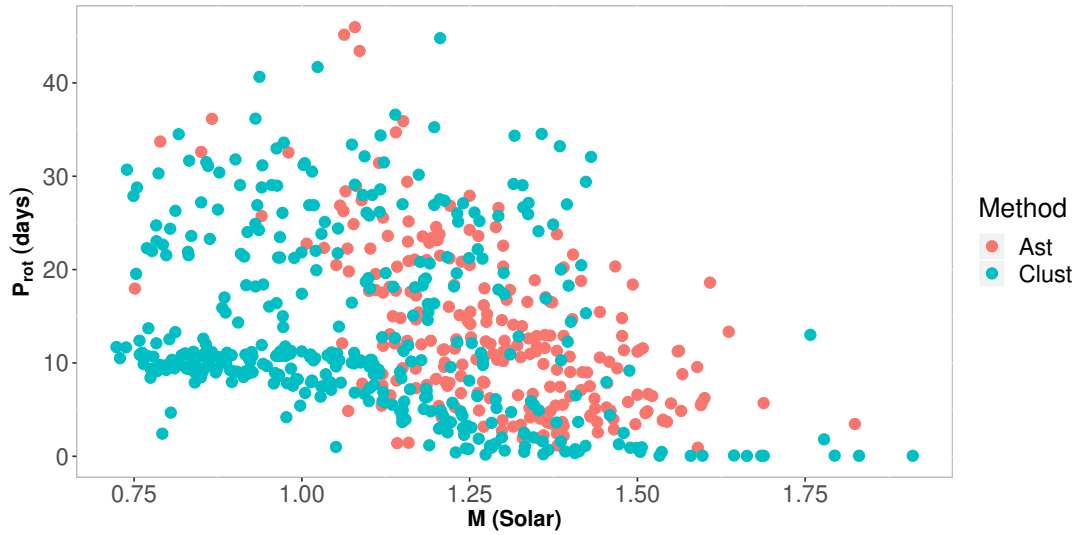


Figura 3.2:  $M - P_{\text{rot}}$  para las estrellas elegidas.

Si se añade la información de la edad estelar a la gráfica  $M$  vs.  $P_{\text{rot}}$ , se obtiene la Figura 3.3. Aquí se puede observar que, con una gran dispersión, cuanto mayor es  $P_{\text{rot}}$ , mayor es la edad. Es destacable que también es cierto para estrellas masivas, por encima del límite de Kraft. Es decir, incluso en ausencia de una envolvente convectiva exterior desarrollada, también se produce la reducción de la velocidad de rotación con el tiempo.

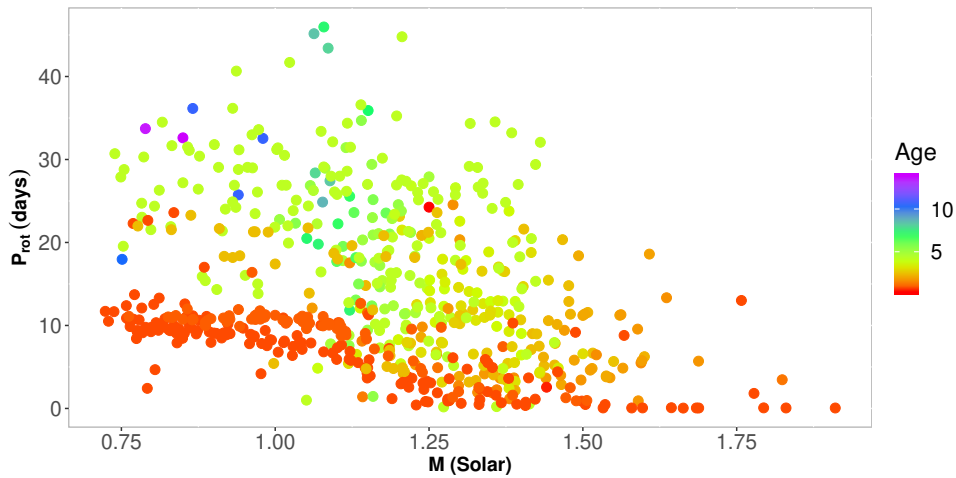


Figura 3.3:  $M - P_{\text{rot}}$  para las estrellas elegidas. La edad se muestra en escala de color.

Si se representa la edad frente a su periodo de rotación, se obtiene la Figura 3.4. En general, se confirma que a mayor edad, mayor periodo de rotación, con una gran dispersión, principalmente para estrellas en cúmulos debido a la dependencia de masa del frenado de rotación. También se ha ajustado una regresión lineal a esta relación, diferenciando entre las estrellas



astrosísmicas y los cúmulos. Aquí se puede observar uno de los sesgos de la muestra. Estas relaciones lineales tienen la misma tendencia pero apenas son diferentes para cada subgrupo. En cualquier caso, la dispersión es realmente grande y no se puede asegurar que estas regresiones se puedan utilizar para estimar la edad. Esta es la razón por la que se pasa a métodos de análisis basados en Inteligencia Artificial para generar modelos adecuados para estimaciones de edades estelares.

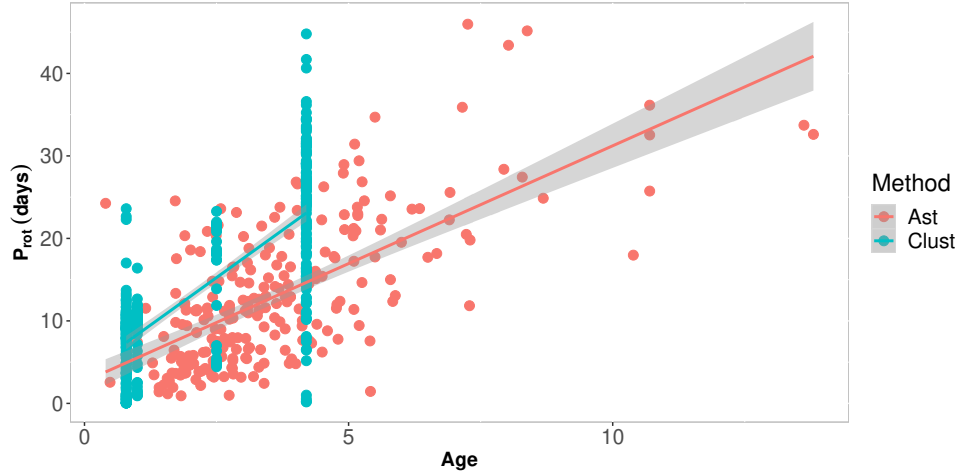


Figura 3.4:  $\text{Edad} - P_{\text{rot}}$  para las estrellas elegidas. Las estrellas astrosísmicas y de cúmulos se muestran en colores diferentes. También se representa la regresión lineal obtenida usando estos dos grupos.



## Capítulo 4

# Modelos de Inteligencia Artificial para datación estelar

Este apartado contiene, en primer lugar, la descripción de los modelos clásicos de IA que forman parte del análisis comparativo, detallando cuales son los hiperparámetros que se optimizan para llegar a los resultados obtenidos. En segundo lugar, se detallan los 3 *benchmarks* que se plantean para evaluar el rendimiento de todos los modelos, incluyendo gráficas que ilustran las edades que componen los conjuntos de entrenamiento y prueba. Por último, se detallan los resultados obtenidos en cada uno de los 3 *benchmarks*, adjuntando diferentes gráficas que ilustran el rendimiento de todos los modelos en los escenarios de interpolación, control y el escenario base.

### 4.1. Modelos

**Linear Regressor** Primero se entrena un modelo de datación de estrellas asumiendo que la edad es combinación lineal de las características proporcionadas en la muestra de datos. Sea  $\hat{a}$  el valor de la edad predecida, entonces

$$\hat{a}(w, x) = w_0 + w_1x_1 + \dots + w_Dx_D, \quad (4.1)$$

donde  $x$  y  $w$  codifican el vector de características de entrada y el coeficiente del regresor lineal, respectivamente. Se sigue el enfoque de mínimos cuadrados para el aprendizaje del modelo, resolviendo un problema de la forma:  $\min_w ||Xw - a||_2^2$ .  $X$  es una matriz que contiene todos los vectores de entrenamiento y  $a$  es el vector de edades.

Durante la fase de aprendizaje del modelo, se lleva a cabo una optimización de su hiperparámetro interno *normalize* a través de un proceso de *grid search* con validación cruzada. De forma predeterminada se encuentra desactivado, pero al activarlo, los regresores  $X$  se normalizan antes de la regresión restando la media y dividiendo por la norma l2.

**Bayesian Regression** Se implementa un regresor Bayesian Ridge para estimar un modelo

probabilístico del problema de datación estelar con la forma:

$$p(a|X, w, \alpha) = \mathcal{N}(a|Xw, \alpha) , \quad (4.2)$$

donde se asume la edad salida  $a$  como una distribución gaussiana sobre  $Xw$ .  $\alpha$  se estima directamente a partir de los datos que se tratan como una variable aleatoria. El parámetro de regularización se ajusta a los datos disponibles, introduciendo sobre los hiperparámetros del modelo,  $w$ , el siguiente gaussiano esférico  $p(w|\lambda) = \mathcal{N}(w|0, \lambda^{-1}\mathbf{I}_D)$ .

Para este regresor, se optimiza con validación cruzada el hiperparámetro asociado al número máximo de iteraciones, que tomará los valores 100, 200, 300, 400, 500.

**Decision Tree Regressor** Se usa un modelo de árbol de decisión [14]. Los árboles de decisión son un método de aprendizaje no paramétrico cuyo objetivo es crear un modelo capaz de predecir una variable objetivo siguiendo reglas de decisión simples inferidas de las características de los datos. Destacan por ser un método fácil de entender e implementar, pero, también por convertirse en modelos demasiado complejos con dificultad para generalizar. Durante el aprendizaje, el modelo optimiza sus hiperparámetros internos a través de un proceso de *grid search* con validación cruzada. Específicamente, se ajustan: la estrategia para elegir la división en cada nodo del árbol (mejor o aleatorio); el número mínimo de muestras requeridas para estar en un nodo hoja (5, 10, 50, 100); y la función para medir la calidad de la división (error cuadrático medio, error cuadrático medio con la puntuación de mejora de Friedman para posibles divisiones, error medio absoluto con reducción de la desviación de Poisson para encontrar divisiones).

**Random Forest Regressor** Este modelo de regresión se implementa siguiendo el trabajo de Breiman [13], donde los árboles de decisión se construyen a partir de las muestras de datos extraídas con reemplazo del conjunto de entrenamiento. Los árboles de decisión individuales suelen tender al sobreajuste. Los bosques aleatorios logran reducir la varianza del modelo al combinar diversos árboles, aunque suele ser a costa de un ligero sesgo. La aleatoriedad inyectada en los bosques produce árboles de decisión con errores de predicción desacoplados. Al tomarse el promedio de las predicciones, algunos errores pueden anularse. De nuevo se usa un proceso de *grid search* más validación cruzada para ajustar los siguientes hiperparámetros: número de árboles (5, 10, 50, 100); número mínimo de muestras requeridas para estar en un nodo hoja (5, 10, 50, 100); y la función para medir la calidad de la división (error cuadrático medio o error medio absoluto).

**de Support Vector Regression** Las máquinas de vectores soporte son un modelo que representa los puntos de la muestra en el espacio, separando una clase de otra mediante hiperplanos, tratando de mantener la máxima distancia posible con las muestras. El objetivo de este modelo consiste en encontrar el hiperplano que separe ambas clases de forma óptima. Una de las grandes ventajas de este modelo es que presenta una alta efectividad en espacios de gran dimensión, además, permite especificar diferentes funciones de kernel para la función de decisión. En el estudio se incluye este modelo de regresión, basado en la implementación de LibSVM

para regresión, [17], siguiendo la aproximación  $\epsilon$ -SVR. Se emplea el kernel RBF, y se realiza una *grid search* con validación cruzada para ajustar el parámetro  $C$  (1, 10, 100, 1000).

**Gaussian Process** Se entrena un regresor de proceso gaussiano [49], donde se supone que la media es constante y cero. Los procesos gaussianos son un método genérico de aprendizaje supervisado. Algunas de sus ventajas principales se basan en que su predicción es probabilística, lo que le permite calcular intervalos de confianza y reajustar la predicción en base a estos. Además, al igual que las máquinas de vectores soporte, también permite especificar diferentes kernels. Este método también presenta algunas desventajas, como su pérdida de eficiencia en espacios de gran dimensión. Se usa un kernel que es la suma de los siguientes tres kernels:

- RBF:  $k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2l^2}\right)$ , donde  $d(\cdot, \cdot)$  es la distancia euclídea y  $l$  es el parámetro de escala de longitud. Este es un kernel estacionario, las siglas RBF provienen de *Radial-basis function* aunque también se conoce como kernel exponencial al cuadrado.
- Producto escalar:  $k(x_i, x_j) = \sigma_0^2 + x_i \cdot x_j$ . Es un kernel no estacionario que se obtiene de la regresión lineal usando  $N(0, 1)$  sobre los coeficientes de  $x_i (i = 1, \dots, N)$  y  $N(0, \sigma_0^2)$  para el sesgo.
- White kernel:  $k(x_i, x_j) = \text{noise\_level}$  if  $x_i == x_j$  else 0. La aplicación clave del White kernel es formar parte de un kernel suma donde se explica el ruido de la señal como independiente e idénticamente distribuido siguiendo una función normal.

**k-Nearest Neighbors** También se incluye en la comparativa un regresor kNN. El principio de este modelo consiste en encontrar el número predefinido de muestras de entrenamiento más cercanas en distancia al nuevo punto y predecir la etiqueta a partir de ellas. A pesar de su simplicidad, este modelo ha tenido éxito en gran cantidad de problemas de clasificación y regresión. Para este problema, se ajusta el parámetro  $k$  (1, 5, 10, 15, 20, 50) y el tipo de función de ponderación que se usa para escalar la contribución de cada vecino. Se exploran dos tipos de funciones de ponderación: a) uniforme, donde todos los puntos en cada vecindario se ponderan por igual; y b) distancia, que pondera los puntos por la inversa de su distancia a un punto de consulta.

**Neural Network** La arquitectura implementada se muestra en la figura 4.1. Ésta consiste en un perceptrón multicapa. El modelo incluye una capa de entrada, un conjunto de 3 capas ocultas *fully-connected* con 50 unidades en cada una, y la capa de salida encargada de la regresión de las edades de las estrellas. Cada unidad oculta va seguida de una función de activación ReLU. Se usa como función de pérdidas el error cuadrático medio,  $Loss(\hat{a}, a, W) = \frac{1}{2} \|\hat{a} - a\|_2^2 + \frac{\alpha}{2} \|W\|_2^2$ , donde  $W$  codifica los pesos de la red, y  $\alpha$  es el parámetro de regularización. La técnica de propagación inversa [31] se utiliza para el aprendizaje del modelo con el optimizador SGD [52]. Durante el aprendizaje, se fija  $\alpha = 0,01$ , y se usa una política de tasa de aprendizaje adaptativa, comenzando con un valor 0,1.

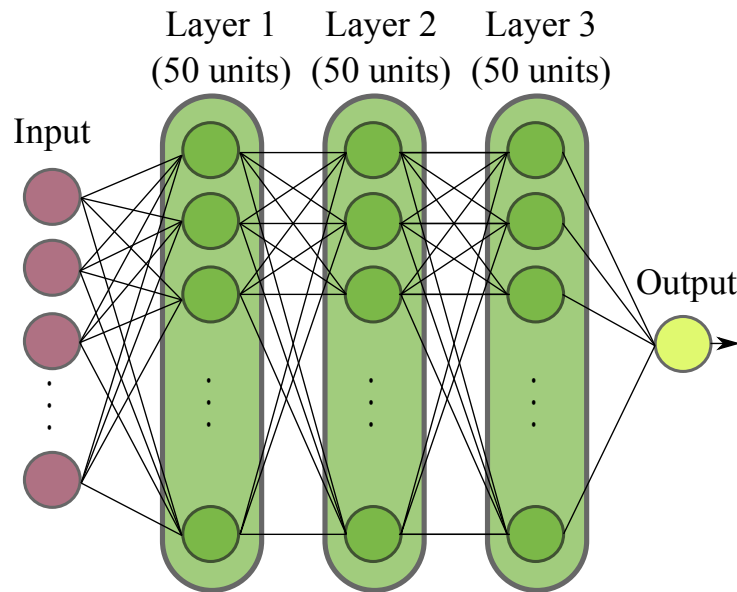


Figura 4.1: Arquitectura de la red neuronal implementada. Se han usado 3 capas ocultas, con 50 unidades en cada una seguida de una ReLU. La salida no tiene función de activación, para realizar directamente la estimación de la edad.

**Stacking** Finalmente, se emplea un método de ensamblado en aprendizaje automático conocido como generalización apilada o stacking [65]. El modelo consiste en apilar la salida de los estimadores individuales y usar un regresor para calcular la predicción final. Esto permite tener en cuenta el potencial de cada estimador individual, utilizando su salida como entrada de un estimador final. Por tanto, las predicciones de cada regresor individual (nivel 0) se apilan y se introducen en un estimador final (nivel 1), que calcula la predicción de la edad. Para este proyecto, en el nivel 0, se integran: Neural Network y Gaussian Process. Para la última capa (nivel 1), se incorpora un Linear Regressor. Durante el entrenamiento, y para generalizar y evitar sobreajuste, el estimador de nivel 1 se entrena con muestras externas (tomadas del conjunto de entrenamiento), siguiendo una metodología de validación cruzada.

## 4.2. Escenarios de evaluación

La base de datos, detallada en la Sección 3, ofrece más de 600 estrellas, con edades precisas, donde se propone construir los siguientes escenarios de evaluación.

**Problema de datación estelar (Benchmark A)** En esta configuración, se propone evaluar los diferentes modelos de regresión siguiendo un esquema clásico de división de datos de entrenamiento/prueba. A partir de la distribución de la muestra de datos, se define un conjunto de entrenamiento y un conjunto de pruebas, donde el 80 % y el 20 % de las estrellas se han incluido al azar, respectivamente. En la Figura 4.2 se detalla la distribución de las estrellas en el

conjunto de entrenamiento y en la Figura 4.3 la del conjunto de prueba. Tras la observación de ambas figuras se puede concluir: 1) los límites de edad son similares para ambos conjuntos y 2) predominan las estrellas pertenecientes a los cúmulos de 0.7 y 4.2 Gyr.

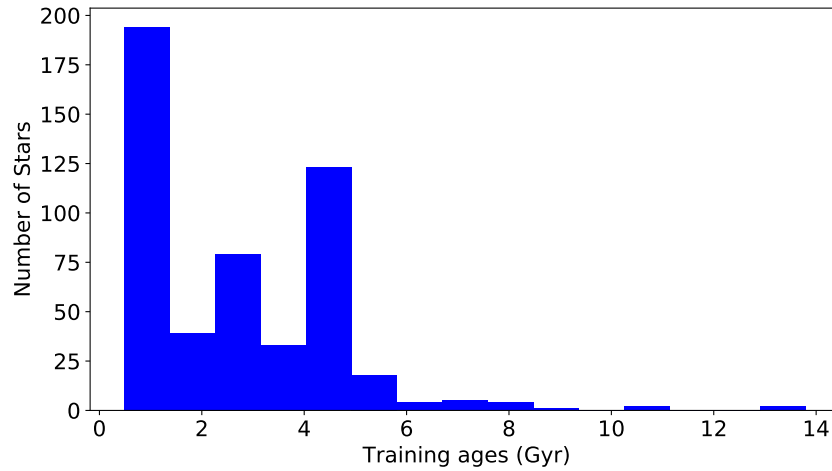


Figura 4.2: Benchmark A: Distribución de las estrellas que forman el subconjunto de entrenamiento.

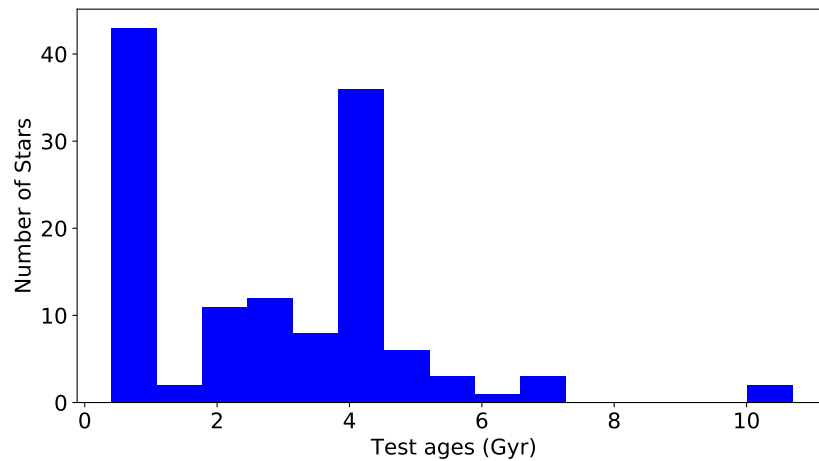


Figura 4.3: Benchmark A: Distribución de las estrellas que forman el subconjunto de prueba.

**Capacidad de generalización (Benchmark B)** En primer lugar, se proporciona un estudio de generalización para los modelos, que consiste en realizar el entrenamiento sobre estrellas *jóvenes*, para posteriormente evaluar su rendimiento en estrellas *ancianas*. Este escenario, denominado Benchmark B1, es interesante para evaluar la capacidad de los modelos de Inteligencia

## 20CAPÍTULO 4. MODELOS DE INTELIGENCIA ARTIFICIAL PARA DATACIÓN ESTELAR

Artificial para trabajar con estrellas cuyo rango está fuera del conjunto de entrenamiento. Los rangos de edad de entrenamiento y prueba son  $[0,4, 4,2]$  y  $[4,2, 13,8]$  Gyr, respectivamente. En la Figura 4.4 se detalla la distribución de las estrellas en el conjunto de entrenamiento y en la Figura 4.5 la del conjunto de prueba.

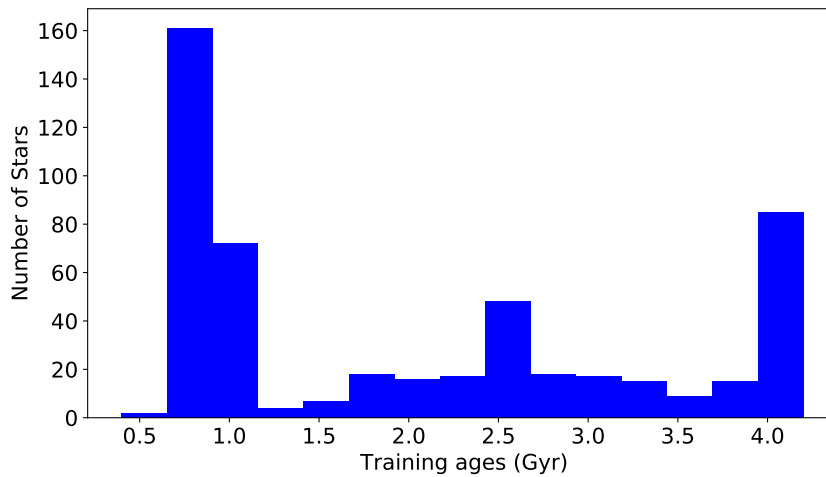


Figura 4.4: Benchmark B1: Distribución de las estrellas que forman el subconjunto de entrenamiento.

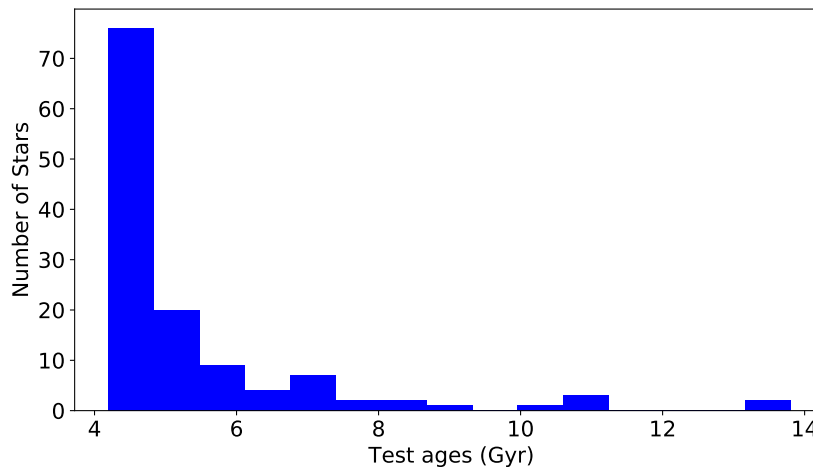


Figura 4.5: Benchmark B1: Distribución de las estrellas que forman el subconjunto de prueba.

También se propone un segundo escenario para evaluar la capacidad de generalización de los modelos cuando se entrenan solo con estrellas pertenecientes a cúmulos. En la práctica, la girocronología se utiliza para datar estrellas individuales que pueden tener cualquier edad. Las



estrellas de los cúmulos suelen estar datadas gracias a su pertenencia a los propios cúmulos. En este Benchmark B2, se propone el entrenamiento de los modelos utilizando únicamente estrellas pertenecientes a cúmulos. El resto de estrellas se incluyen en el conjunto de prueba. En total, este escenario recoge 397 y 240 muestras para entrenamiento y prueba, respectivamente.

En la Figura 4.6 se detalla la distribución de las estrellas que forman el conjunto de entrenamiento, aquí se observa la concentración de estas en las 4 edades principales de cada uno de los cúmulos (0.79, 1, 2.5 y 4.2 Gyr). Por otro lado, la Figura 4.7 detalla la distribución del conjunto de prueba. En esta gráfica se observa una mayor dispersión sobre el rango total de edades, aunque sigue apareciendo una mayor concentración de estrellas en la edad de 2.4 Gyr.

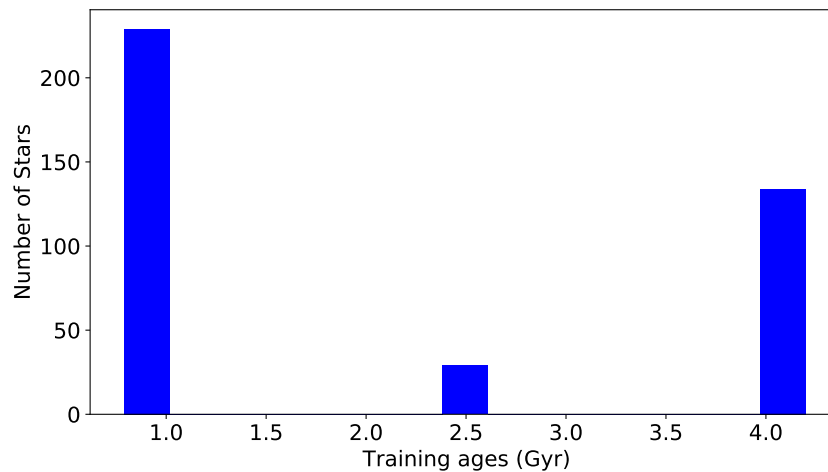


Figura 4.6: Benchmark B2: Distribución de las estrellas que forman el subconjunto de entrenamiento.

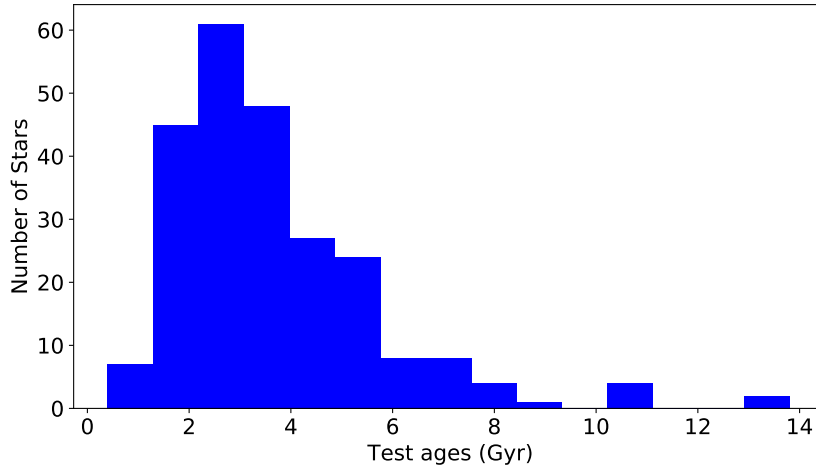


Figura 4.7: Benchmark B2: Distribución de las estrellas que forman el subconjunto de prueba.

**Regresión para datación estelar sobre una muestra de control (Benchmark C)** Aquí se propone examinar el rendimiento de todos los modelos sobre una muestra de datos de control compuesta por estrellas que no pertenecen a ningún cúmulo, y con una distribución de edades más realista/generalista. Específicamente, en este Benchmark C, se evalúan los modelos entrenados con toda la muestra de datos sobre un conjunto de 32 estrellas nuevas, no pertenecientes a cúmulos, incluyendo el Sol. Este conjunto independiente de control se ha obtenido de las mismas fuentes basadas en astrosismología utilizadas para recopilar la información de la muestra descrita en la Sección 3. Se han reservado 32 estrellas de estas fuentes, antes de construir la base de datos. El rango de edad de este conjunto va de 1,2 a 10,1 Gyr. Este rango se superpone con el rango de edad utilizado en la muestra de entrenamiento ( $[0,4, 13,8]$  Gyr), e incluye algunas edades nuevas no observadas. Se presta especial atención a la precisión de los modelos para la estimación de la edad del Sol (4.6 Gyr). En la Figura 4.8 se detalla la distribución de las estrellas en el conjunto de entrenamiento y en la Figura 4.9 la del conjunto de prueba.

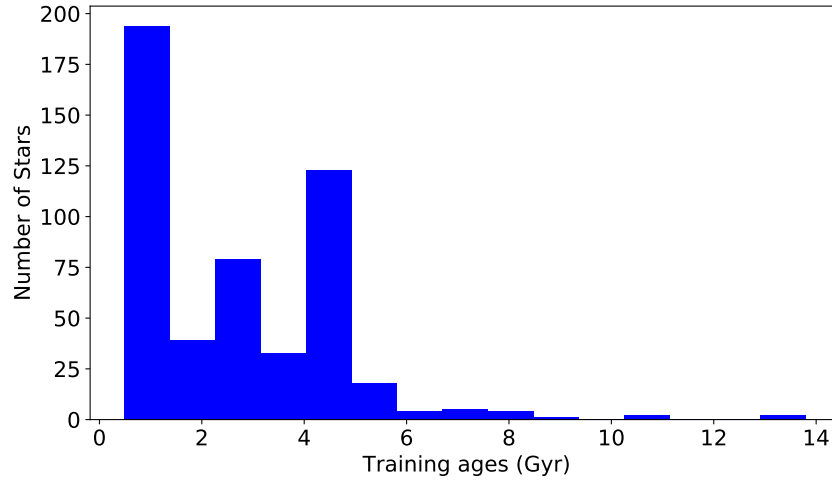


Figura 4.8: Benchmark C: Distribución de las estrellas que forman el subconjunto de entrenamiento.

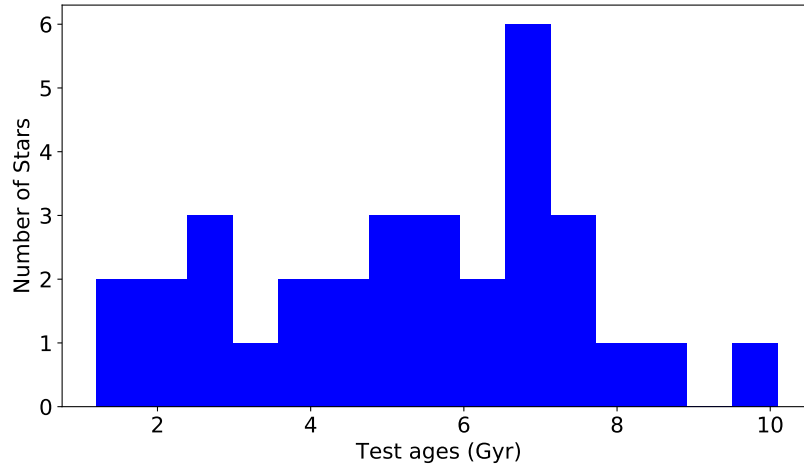


Figura 4.9: Benchmark C: Distribución de las estrellas que forman el subconjunto de prueba.

## 4.3. Resultados

A continuación, se presenta la configuración experimental y se informa de los resultados obtenidos en la evaluación de los modelos detallados en la Sección 4.1. Este *benchmarking* se realiza a través de los 3 escenarios descritos: el efecto de los diferentes modelos (Benchmark A); la capacidad de generalización (Benchmark B); y el rendimiento sobre una muestra de control (Benchmark C).

### 4.3.1. Configuración experimental

**Implementación** Para poder realizar una comparación entre modelos, se han creado todos en Python, usando la librería scikit-learn [48]. Se utilizan las siguientes siglas para identificar los modelos implementados: Neural Network (nnnet), Linear Regressor (lr), Decision Tree Regressor (dtr), Random Forest Regressor (rf), Support Vector Regression (svm), Bayesian Regression (bayes), k-Nearest Neighbors (kNN), Gaussian Process (gp) y Stacking (stacking).

**Métricas de evaluación** La métrica de evaluación principal utilizada es el error medio absoluto,  $MAE = \frac{\sum_{i=1}^N |a_i - \hat{a}_i|}{N}$ , donde  $a_i$  y  $\hat{a}_i$ , son la edad proporcionada por el conjunto de datos y la edad estimada por un modelo de regresión, respectivamente. Dado que el conjunto de datos proporciona información sobre la precisión de la edad de cada estrella, en forma de límites de error, también se propone utilizar como métrica de evaluación de la precisión el porcentaje de predicciones de la edad de las estrellas que se encuentran dentro del intervalo de confianza proporcionado por el propio conjunto de datos.

**Gráficas** Antes de abordar cada uno de los *benchmarks*, es necesario introducir brevemente las gráficas que permiten profundizar en el rendimiento de los modelos. En primer lugar, se muestra un gráfico de barras que recoge un resumen con el MAE de cada uno de los modelos. A continuación y para cada uno de los modelos, se observan 3 gráficas diferentes. La primera de las gráficas permite observar las estimaciones que caen dentro del margen de error. La segunda gráfica muestra detalladamente la predicción, incluyendo el error obtenido para cada uno de los casos. La tercera gráfica muestra un histograma que recoge el MAE en función de la edad de las estrellas del conjunto de prueba.

### 4.3.2. Benchmark A: Problema de datación estelar

La Figura 4.10 muestra el rendimiento de todos los modelos, permitiendo la comparativa del MAE de cada modelo. En este escenario es interesante observar que se tienen 2 modelos, más su Stacking, que presentan los mejores resultados, estableciendo un margen de rendimiento con respecto a los demás. Son la Red Neuronal y el Proceso Gaussiano. Su Stacking reduce ligeramente el mejor MAE, de la Red Neuronal, de 0,405 a 0,400.

Analizando los 6 modelos restantes, se observa que, la mayoría de los modelos presentan dificultades en la estimación de la edad de las estrellas más viejas. También es relevante el resultado proporcionado por un modelo tan simple como un kNN (ver Figura 4.15 y Figura 4.15), con un MAE de solo 0.53 Gyr. El éxito de este modelo se explica mediante los cúmulos que forman la base de datos, donde se encuentran muchos elementos concentrados en 2 valores de edad muy específicos, lo que permite que modelos como kNN hagan estimaciones muy precisas sobre las estrellas de prueba con estas edades. Este hecho se corrobora en la Tabla 4.2, donde se observa que la mejor precisión la logra el modelo kNN, seguido por la Red Neuronal.

En general, a excepción de lr y bayes, todos los modelos ofrecen un MAE por debajo de 0.7 Gyr, lo que supone un gran avance en el campo de la datación estelar.

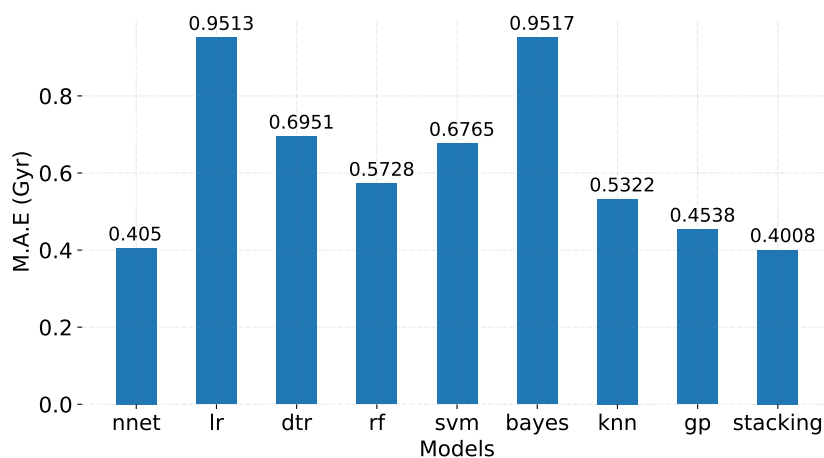


Figura 4.10: Benchmark A: Rendimiento de los modelos en función del MAE. En esta configuración, las dos mejores aproximaciones son Neural Network (nnet) y Gaussian Process (gp), y por lo tanto, su Stacking.

**Stacking** A continuación, se muestran las 3 gráficas que detallan el rendimiento del modelo Stacking, el cual ha presentado el menor MAE en este benchmark, 0.400 Gyr. En la Figura 4.11 se observa que se tienen estimaciones dentro del margen de confianza para todo el espectro de edades, aunque, también se observa una acumulación de fallos en la estimación de las edades cercanas a 0.7 y 4.2 Gyr. En la Figura 4.12 se aprecia que los errores más pronunciados se dan para edades inferiores a 1, concretamente, las edades pertenecientes al cúmulo de 0.7 Gyr. Además, en la Figura 4.13 se observa una diferencia del error medio entre edades de 0.45 Gyr, lo que supone una diferencia relativamente alta.

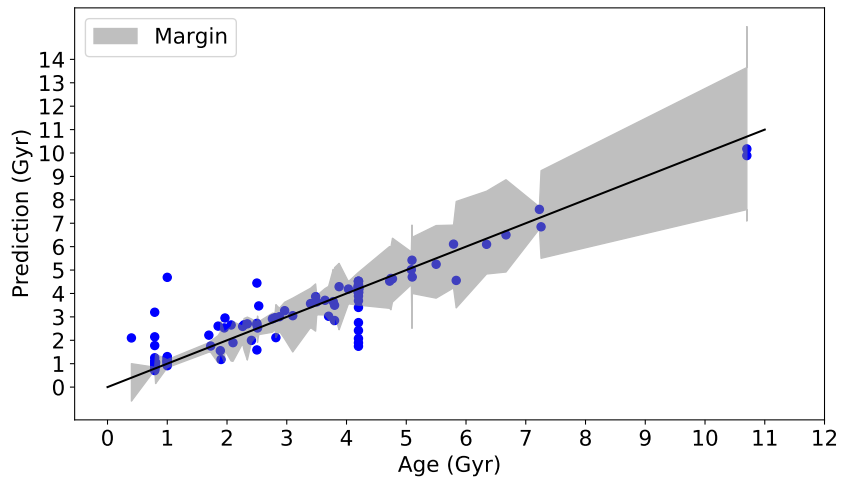


Figura 4.11: Benchmark A: Rendimiento para el modelo Stacking.

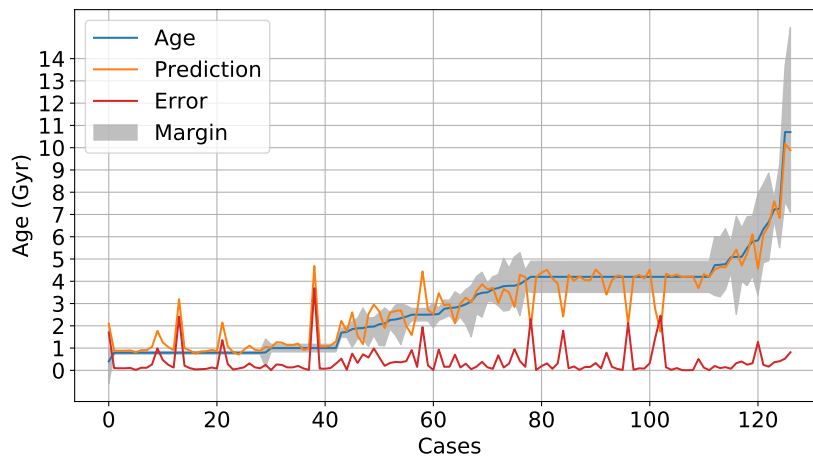


Figura 4.12: Benchmark A: Predicción detallada para el modelo de Stacking. Se observa la edad real (en azul), la predicción de los modelos (en naranja), y el error correspondiente (en rojo).

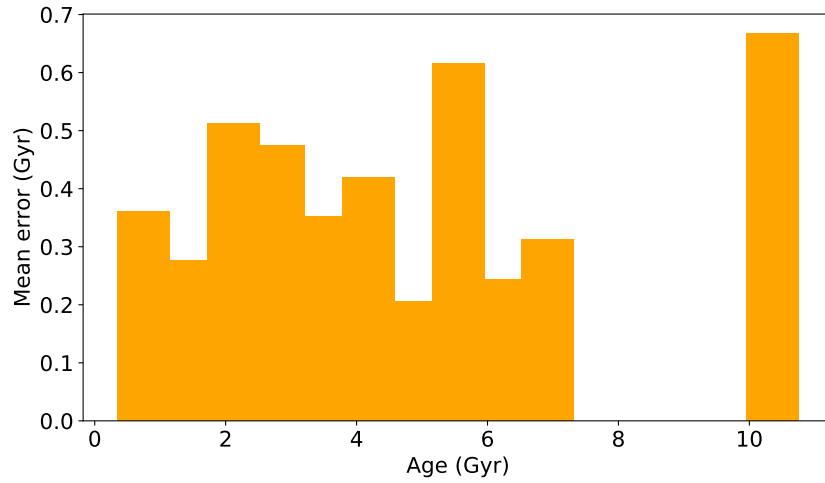


Figura 4.13: Benchmark A: MAE en función del rango de edad del conjunto de prueba.

En las Figuras 4.14, 4.15, 4.16 y 4.16 se adjuntan las gráficas que detallan el rendimiento de los modelos Neural Network, Gaussian Process, kNN y Random Forest. En sus correspondientes gráficas se puede apreciar:

**Neural Network** La red neuronal es el segundo mejor modelo, después del stacking, con un MAE de 0.405 Gyr. Al igual que con el modelo del Stacking, en las gráficas de la red neuronal se aprecia cierta dificultad para colocar todas las estrellas los cúmulos de 0.7 y 4.2 Gyr dentro del margen de confianza. Por otro lado, el error medio entre edades varia desde 0.2 a 0.8 Gyr, aumentando la diferencia del modelo de Stacking.

**Gaussian Process** El tercer mejor modelo es el Proceso Gaussiano, con un MAE de 0.453 Gyr. A diferencia de los dos modelos anteriores, aquí se empiezan a observar dificultades para estimar edades entre 0.4 y 4.2 Gyr, obteniendo un mayor número de puntos fuera del margen de confianza. Además, la segunda gráfica, relativa al error para cada edad, muestra errores de valor constante pero con mayor presencia a lo largo de todo el conjunto de edades. También destaca, en la tercera figura, la similitud que muestra el error medio para edades de hasta 7 Gyr, que es 4 veces menor que el error medio para las edades de más de 10 Gyr.

**kNN** Las gráficas del modelo kNN, el cual ha presentado una MAE de 0.532 Gyr, muestran una mayor dispersión de las estrellas estimadas, incluyendo menos estrellas dentro del intervalo de confianza. Por primera vez se observa una mala estimación para las edades más altas. Además, la figura de errores, muestra picos más pronunciados que los vistos hasta el momento, lo que indica una degradación en el rendimiento de los resultados obtenidos.

## 28CAPÍTULO 4. MODELOS DE INTELIGENCIA ARTIFICIAL PARA DATACIÓN ESTELAR

**Random Forest** El último modelo de este bloque es el de Random Forest, con un MAE de 0.572 Gyr. En este caso se observa una dispersión relativamente controlada, respecto a lo visto anteriormente, pero se siguen presentando dificultades para estimar las edades de 0.7, 1 y 4.2 Gyr. En la segunda figura se observan errores de valor constante para todo el espectro de edades, aunque, el mayor error medio se sigue teniendo en las edades superiores a 10 Gyr.

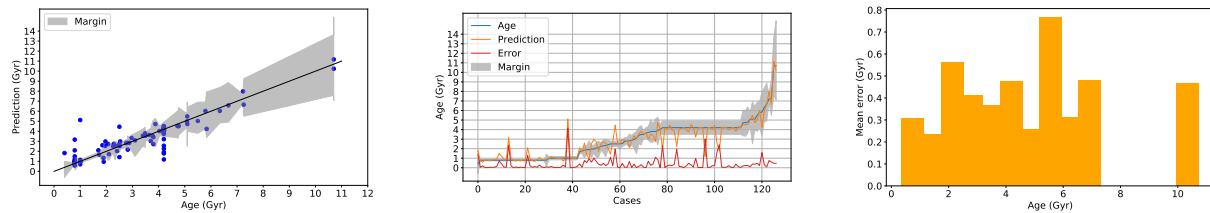


Figura 4.14: Modelo Neural Network

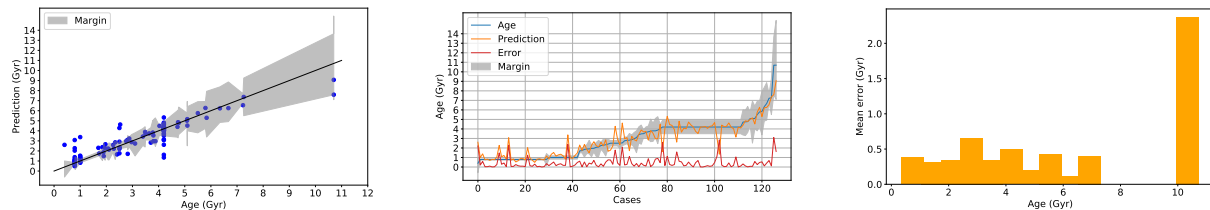


Figura 4.15: Modelo Gaussian Process

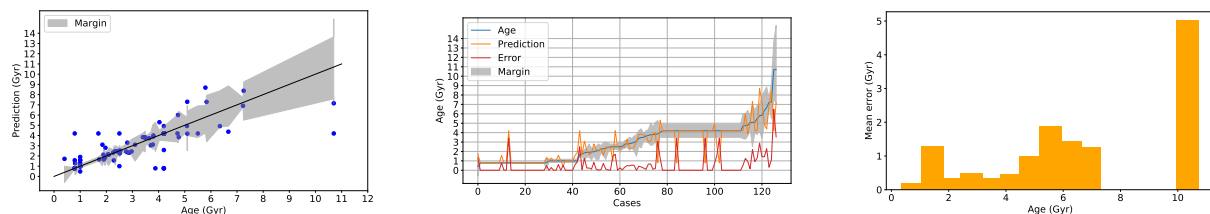


Figura 4.16: Modelo k-Nearest Neighbor

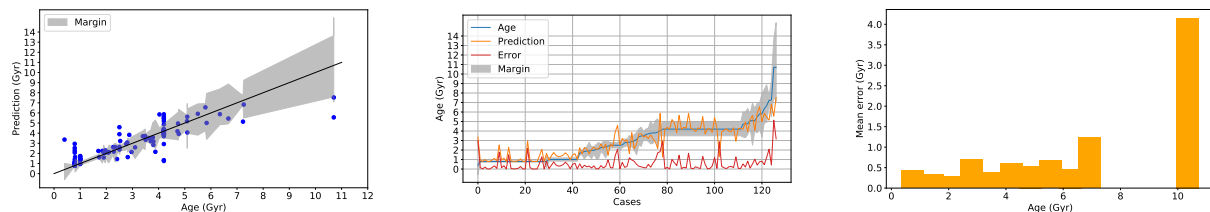


Figura 4.17: Modelo Random Forest

En las Figuras 4.18, 4.19, 4.20 y 4.21 se adjuntan las gráficas que detallan el rendimiento de los modelos Support Vector Regression, Decision Tree Regressor, Linear Regressor y Bayesian Regression. En sus correspondientes gráficas se puede apreciar:

**Support Vector Regression** El modelo Support Vector Regression ha presentado un MAE de 0.676 Gyr. Se siguen presentando dificultades para estimar correctamente todas las estrellas de



los cúmulos de 0.7, y sobre todo, 4.2 Gyr. También se aprecia cierta correlación entre la edad y el error medio obtenido en las estimaciones.

**Decision Tree Regression** Para el modelo Decision Tree Regression, con un MAE de 0.695 Gyr, se visualiza una nube de estimaciones muy dispersa, proporcionando pocas estrellas dentro del intervalo de confianza. Además de presentar problemas en la estimación de los cúmulos, muestra dificultades en la estimación de las edades medias, algo que no se había producido con demasiada intensidad en los modelos anteriores.

**Linear Regressor** El modelo Linear Regressor, con un MAE de 0.951 Gyr, sigue mostrando problemas de estimación para las estrellas pertenecientes a los cúmulos. Además, la figura relativa al error muestra un error constante para todo el espectro de edad, saliéndose continuamente del margen de error asociado a las estrellas.

**Bayesian Regression** Por último, el modelo Bayesian Regression. Este muestra los peores resultados junto con el modelo Linear Regressor, ambos con un MAE de 0.951 Gyr. Tal y como se ha visto en el modelo anterior, las figuras muestran un error constante para todas las edades del conjunto de prueba. No se encuentran picos de error excesivamente elevados, pero la constante presencia de errores de estimación lo convierten en el peor modelo de este *benchmark*.

### 4.3.3. Benchmark B: Escenario de generalización

Aquí se analiza la capacidad de generalización de todos los modelos para el problema de datación estelar. Las Figuras 4.22 y 4.23 muestran el rendimiento en los escenarios de evaluación B1 y B2, respectivamente.

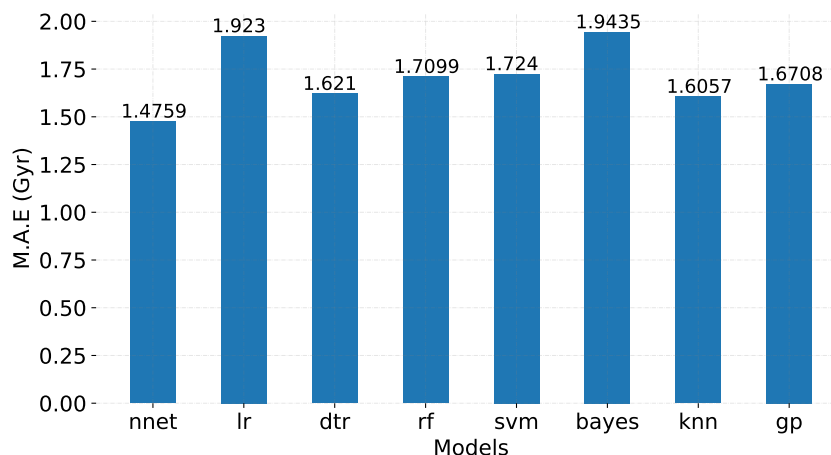


Figura 4.22: Benchmark B1: Rendimiento de los modelos en función del MAE.

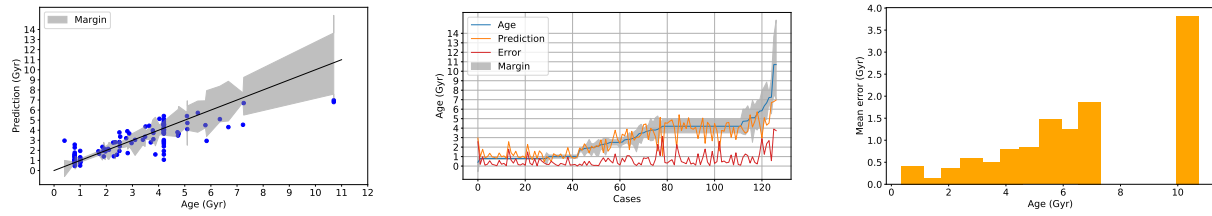


Figura 4.18: Modelo Support Vector Regression

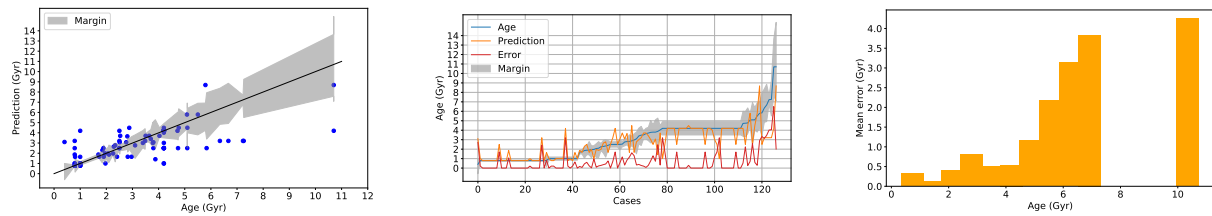


Figura 4.19: Modelo Decision Tree Regressor

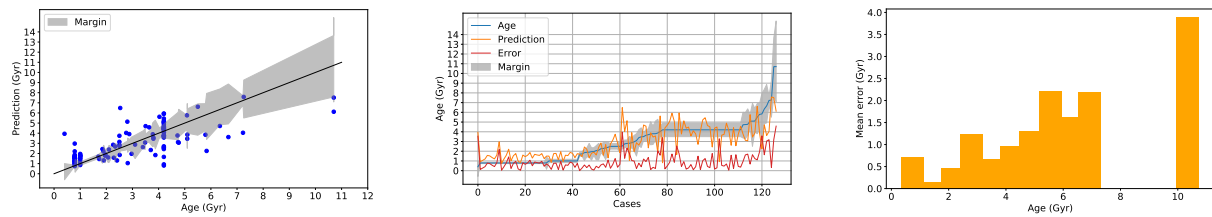


Figura 4.20: Modelo Linear Regressor

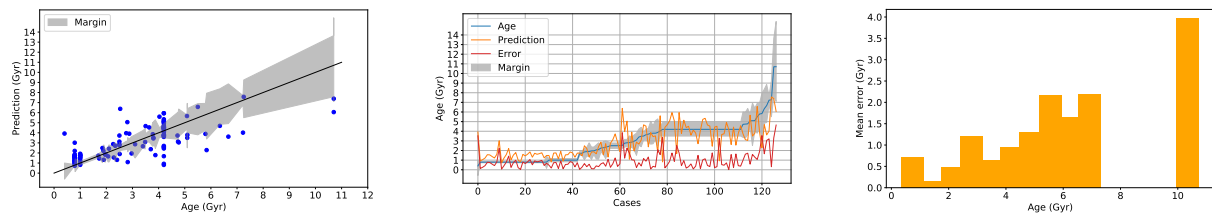


Figura 4.21: Modelo Bayesian Regression

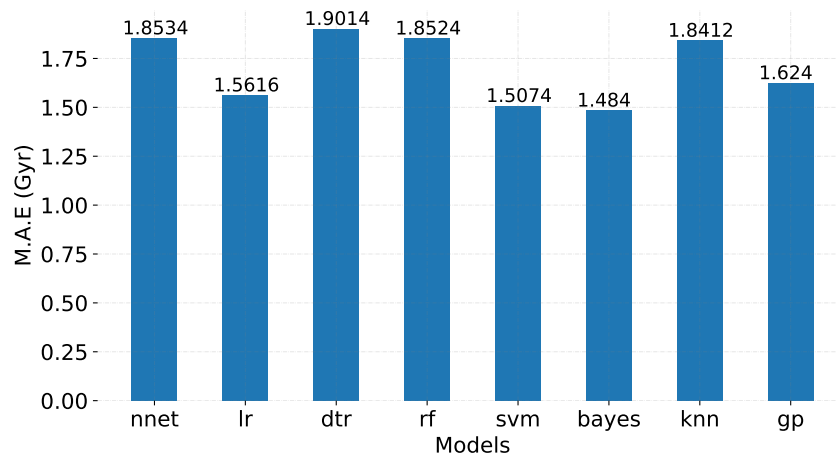


Figura 4.23: Benchmark B2: Rendimiento de los modelos en función del MAE.

### 4.3.3.1. Benchmark B1

El Benchmark B1 se centra en analizar la precisión de los modelos estimando edades de estrellas más antiguas a las que han visto durante el entrenamiento. Obviamente, el error promedio de los modelos ha aumentado. El mejor rendimiento se observa en la Red Neuronal, con un MAE de 1,47 Gyr. Las estimaciones detalladas se pueden ver en la Figura 4.24 y Figura 4.25, donde se puede observar cómo la Red Neuronal es capaz de generalizar hasta 6 Gyr, asignando estimaciones de edad que en su mayoría se encuentran dentro del margen de confianza. La Tabla 4.2 confirma este hecho, donde la Red Neuronal también exhibe la precisión más alta para el Benchmark B1.

**Neural Network** A continuación, se muestran las 3 gráficas que detallan el rendimiento del modelo Neural Network, el cual presenta el menor MAE de este *benchmark*, 1.475 Gyr. En la Figura 4.24 se ve claramente que el modelo tiende a subestimar las edades estelares, consiguiendo únicamente incluir dentro del margen de error las edades más bajas. Esto es lógico ya que a partir de 4.2 Gyr son todas edades desconocidas para el modelo. En la Figura 4.26 se aprecia el aumento del error medio con el aumento de la edad a estimar. En la Figura 4.25 se aprecia cómo la predicción sigue la tendencia de una línea recta que oscila sobre el mayor valor que formaba parte del conjunto de entrenamiento.

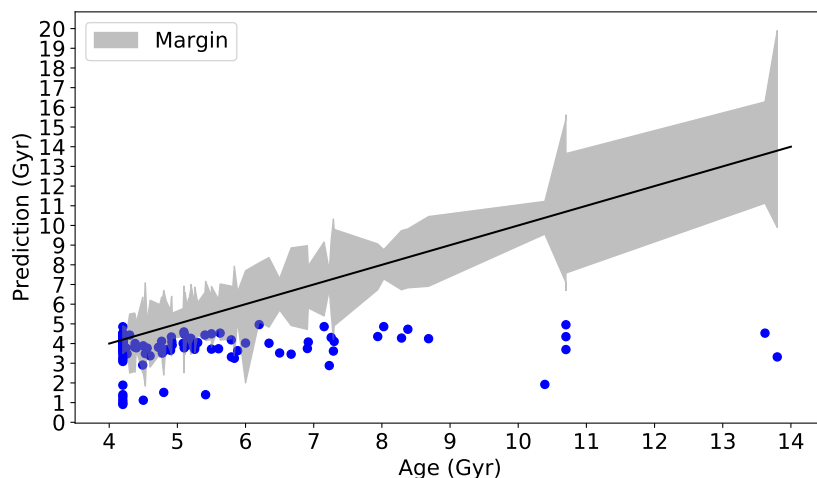


Figura 4.24: Benchmark B1: Rendimiento para el modelo Neural Network.

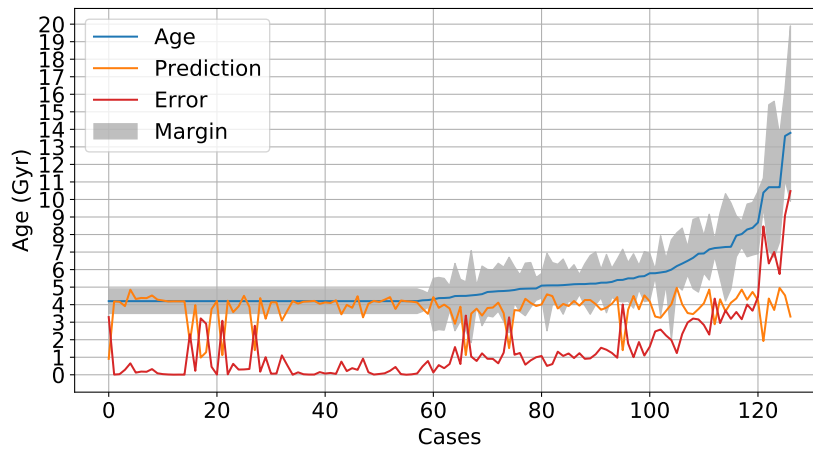


Figura 4.25: Benchmark B1: Predicción detallada para el modelo de Neural Network. Se observa la edad real (en azul), la predicción de los modelos (en naranja), y el error correspondiente (en rojo).

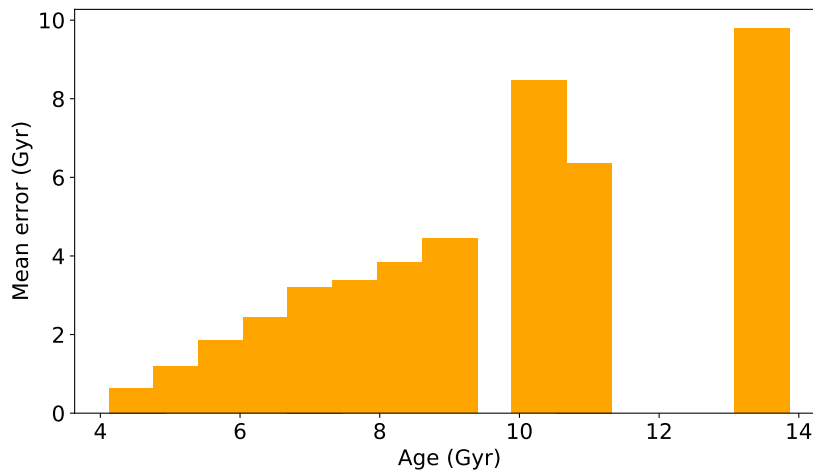


Figura 4.26: Benchmark B1: MAE en función del rango de edad del conjunto de prueba.

#### 4.3.3.2. Benchmark B2

La historia es completamente diferente en el Benchmark B2. La Red Neuronal no es capaz de ser la mejor generalizando. Su error asciende hasta 1,85 Gyr, siendo superada por Linear Regressor, Random Forest, Support Vector Regression, k-Nearest Neighbors, Gaussian Process y Bayesian Regression. Este último resulta ser el mejor modelo en este escenario, con un MAE de 1.48 Gyr. De la Figura 4.27 y la Figura 4.28 se llega a la conclusión de que un regresor

bayesiano es capaz de proporcionar edades precisas en el rango de 1 Gyr a 5 Gyr, es decir, el rango cubierto por la muestra de entrenamiento en este caso. En términos de precisión, la Tabla 4.2 revela que el modelo SVR es el modelo ganador, seguido de cerca por el Proceso Gaussiano y el regresor lineal. El modelo bayesiano, aunque tiene el MAE más bajo, presenta una precisión del 29.71 %.

**Bayesian Regression** A continuación, se muestran las 3 gráficas que detallan el rendimiento del modelo Bayesian Regression, el cual presenta el menor MAE de este *benchmark*, 1.484 Gyr. Las figuras vuelven a mostrar que el modelo tiende a subestimar, algo normal si se tiene en cuenta que la edad más alta conocida en el entrenamiento es 4.2 Gyr. De la Figura 4.27 se observa que la mayoría de las estimaciones caen fuera del margen de error asociado a las estrellas. En la Figura 4.28 se observa que el estimador tiene con facilidad a edades cercanas a 4.2 y no alcanza valores superiores a 6 Gyr. En la línea de todo lo comentado, la Figura 4.29 muestra que el error medio aumenta según aumentan las edades estelares a estimar.

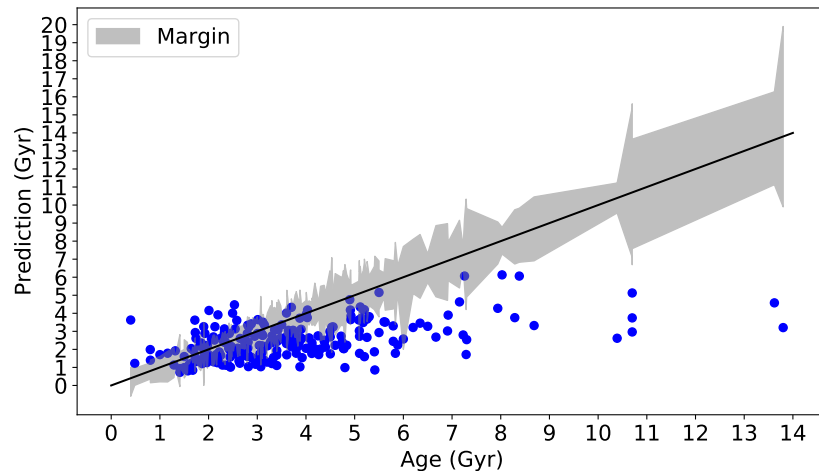


Figura 4.27: Benchmark B2: Rendimiento para el modelo Bayesian Regression.

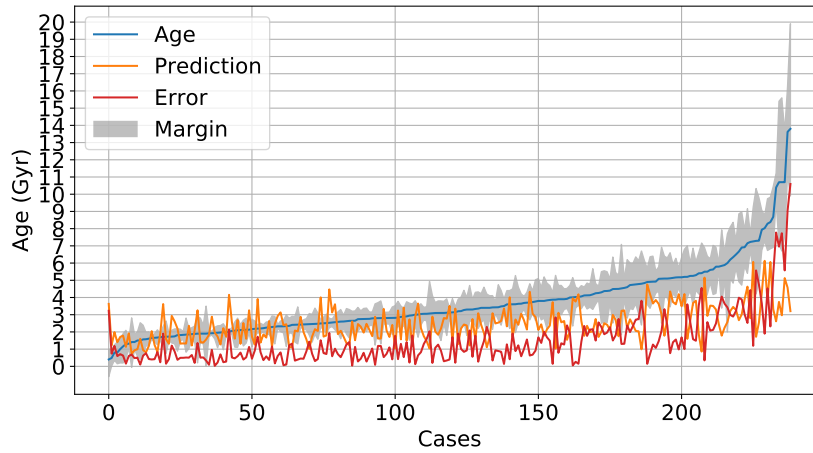


Figura 4.28: Benchmark B2: Predicción detallada para el modelo de Bayesian Regression. Se observa la edad real (en azul), la predicción de los modelos (en naranja), y el error correspondiente (en rojo).

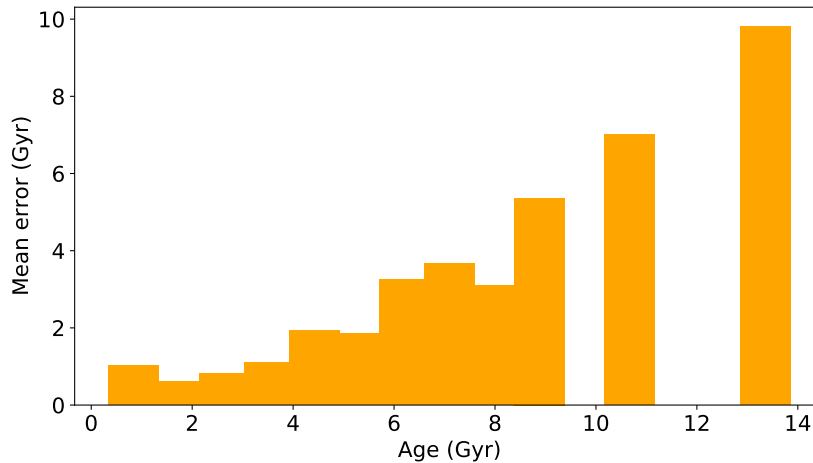


Figura 4.29: Benchmark B2: MAE en función del rango de edad del conjunto de prueba.

A continuación, se detalla el resto de modelos del Benchmark B1, en las Figuras 4.30, 4.31, 4.32 y 4.33 se adjuntan las gráficas correspondientes al rendimiento de los modelos kNN, Random Forest, Support Vector Regression y Decision Tree Regression. En sus correspondientes gráficas se puede apreciar:

**kNN** El modelo kNN presenta un MAE de 1.605 Gyr. En las figuras se observa que solo se realizan estimaciones dentro del margen de error para edades inferiores a 5 Gyr. Además, la

segunda figura muestra que el modelo es incapaz de generar estimaciones por encima de 4.2 Gyr, la edad máxima que conoce del conjunto de entrenamiento.

**Random Forest** En el modelo Random Forest, con un MAE de 1.709 Gyr, se observa un rendimiento similar al visto en el modelo anterior. No consigue generar edades superiores a 4.2 Gyr y se observa un aumento del error medio en estimación según aumenta la edad a estimar.

**Support Vector Regression** El modelo Support Vector Regression, con un MAE de 1.724 Gyr, presenta diferencias respecto a los dos modelos anteriores. Este modelo sí es capaz de generar edades superiores a los 4.2 Gyr. Aún así, se mantiene el problema de subestimación y solo se consiguen estimaciones dentro del margen de error para edades inferiores a 7 Gyr.

**Decision Tree Regression** El modelo Decision Tree Regression ha presentado un MAE de 1.621 Gyr. La primera figura ilustra los problemas que tiene el modelo para generar edades superiores a 4.2 Gyr. De hecho, parece que el modelo solo conoce estrellas de 4.2 y 3.2 Gyr, pues, tiende a esos valores durante todo el espectro de edades. Tal y como se ha visto hasta el momento, el error medio de estimación aumenta con la edad real a estimar.

En las Figuras 4.34, 4.35 y 4.36 se adjuntan las gráficas que detallan el rendimiento de los modelos Linear Regressor, Bayesian Regression y Gaussian Process para el Benchmark B1. En sus correspondientes gráficas se puede apreciar:

**Linear Regressor** El modelo Linear Regressor muestra el segundo peor rendimiento de este *benchmark* con un MAE de 1.923 Gyr. La primera figura muestra los problemas que tiene el modelo para estimar todo el espectro de edades, incluidas las estrellas de 4.2 Gyr que ya conoce y para las que estima edades entre 1 y 5 Gyr. El modelo consigue generar edades de hasta 6 Gyr, pero no consigue introducir demasiadas estimaciones dentro de los márgenes de error.

**Bayesian Regression** El modelo Bayesian Regression ha presentado los peores resultados del *benchmark* con un MAE de 1.943 Gyr. En la primera figura se observa algo similar al modelos Linear Regressor, pues, presenta problemas para estimar las estrellas de 4.2 Gyr, edad que conoce del conjunto de entrenamiento. Además, se sigue observando la subestimación para el resto de edades. La segunda figura muestra en detalle el error constante que mantiene el modelo para todo el espectro de edades, consiguiendo pocas estimaciones dentro del margen de error.

**Gaussian Process** Por último, el modelo Gaussian Process, con un MAE de 1.670 Gyr. Aquí se siguen observando para estimar la edad de 4.2 Gyr, pero no se llega al nivel de los dos modelos anteriores. Este modelo sigue la tendencia de todos los que hemos visto hasta el momento, no es capaz de estimar edades superiores a 4.2 Gyr aunque consigue introducir alguna estimación dentro del margen de error. Evidentemente, el error medio aumenta con las edades reales a estimar.

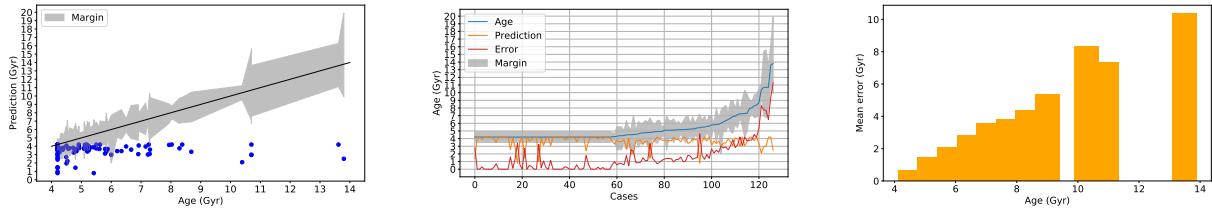


Figura 4.30: Modelo k-Nearest Neighbor

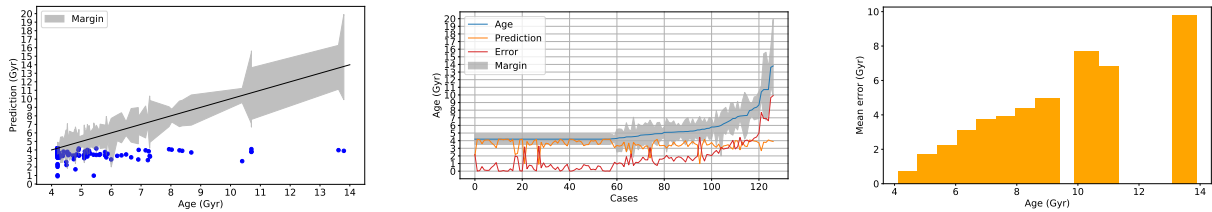


Figura 4.31: Modelo Random Forest

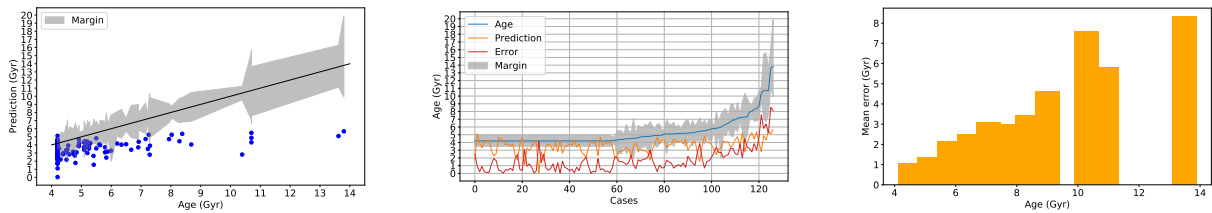


Figura 4.32: Modelo Support Vector Regression

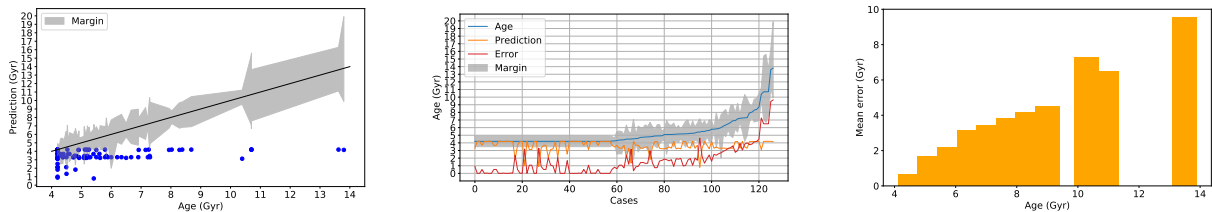


Figura 4.33: Modelo Decision Tree Regressor

A continuación, se detalla el resto de modelos del Benchmark B2, en las Figuras 4.37, 4.38, 4.39 y 4.40 se adjuntan las gráficas correspondientes al rendimiento de los modelos kNN, Random Forest, Support Vector Regression y Decision Tree Regression. En sus correspondientes gráficas se puede apreciar:

**kNN** El modelo kNN ha presentado un MAE de 1.841 Gyr. En la primera imagen se observan con mucha claridad los problemas que muestra el regresor lineal para estimar edades que no sean las de los cúmulos. La segunda figura confirma que solo se estiman valores de 0.79, 1, 2.5 y 4.2 Gyr.

**Random Forest** El modelo Random Forest, con un MAE de 1.852 Gyr, muestra un rendimiento similar al kNN. En este caso la nube de puntos de la primera figura representa una



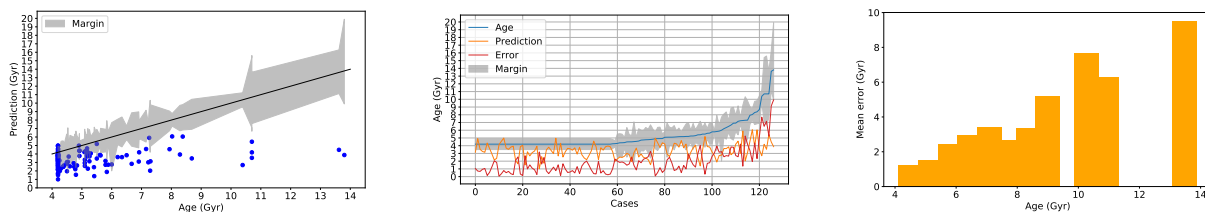


Figura 4.34: Modelo Linear Regressor

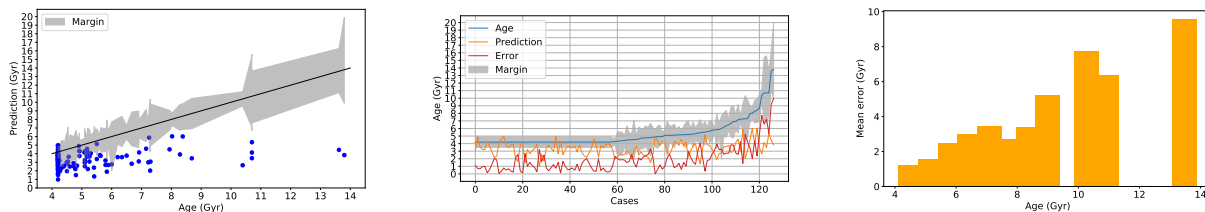


Figura 4.35: Modelo Bayesian Regression

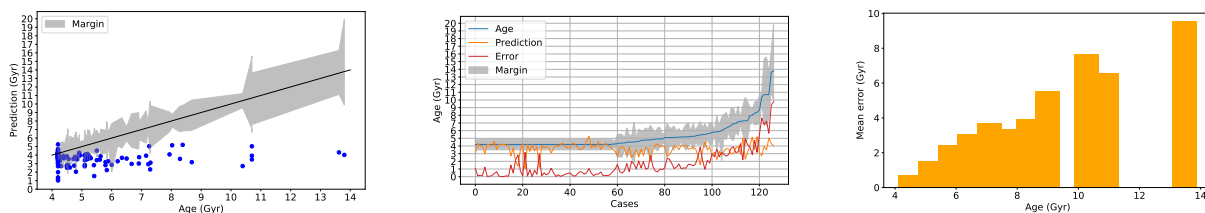


Figura 4.36: Modelo Gaussian Process

serie de estimaciones menos ajustadas a las edades de los cúmulos, aunque se sigue apreciando gran población de estimaciones en esas edades. Como parece lógico, al aumentar la edad real a estimar, aumenta el error medio de la estimación.

**Support Vector Regression** El modelo Support Vector Regression, con un MAE de 1.507 Gyr, muestra una nube de estimaciones que tiende a buscar la edad de las estrellas que no conoce, mostrando un rendimiento más natural que los dos anteriores. Aunque se consiguen estimaciones cercanas a los 6 Gyr, el modelo muestra dificultad para meter en el margen de error las estimaciones de estrellas de más de 4 Gyr.

**Decision Tree Regression** El modelo Decision Tree Regression, con un MAE de 1.901 Gyr, muestra un comportamiento similar al kNN, en que se ve con claridad su tendencia a estimar la edad de los cúmulos para todas las estrellas del conjunto de test.

En las Figuras 4.41, 4.42 y 4.43 se adjuntan las gráficas que detallan el rendimiento de los modelos Linear Regressor, Neural Network y Gaussian Process para el Benchmark B2. En sus correspondientes gráficas se puede apreciar:

**Linear Regressor** El modelo Linear Regressor, con un MAE de 1.561 Gyr, muestra un rendimiento similar a la máquina de vectores soporte. Tiene una gran nube de estimaciones que

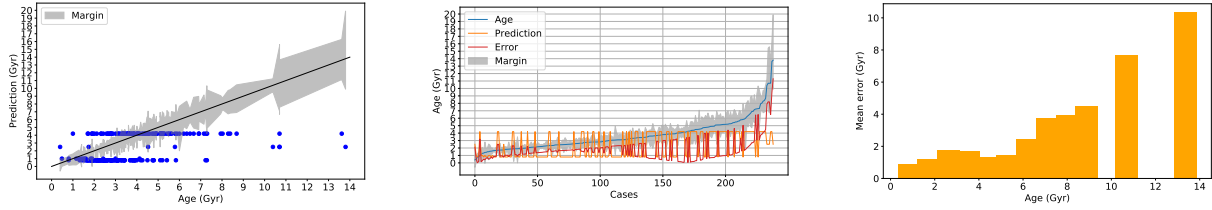


Figura 4.37: Modelo k-Nearest Neighbor

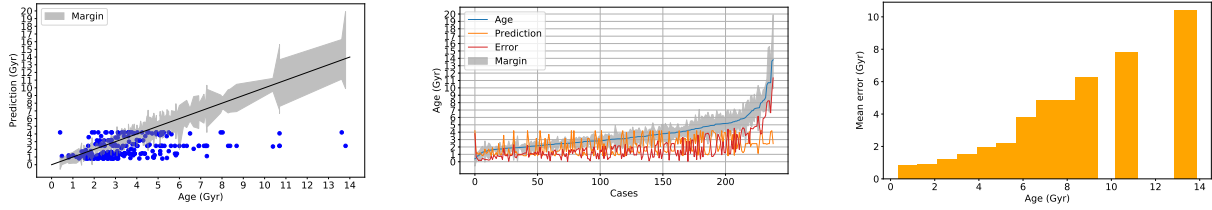


Figura 4.38: Modelo Random Forest

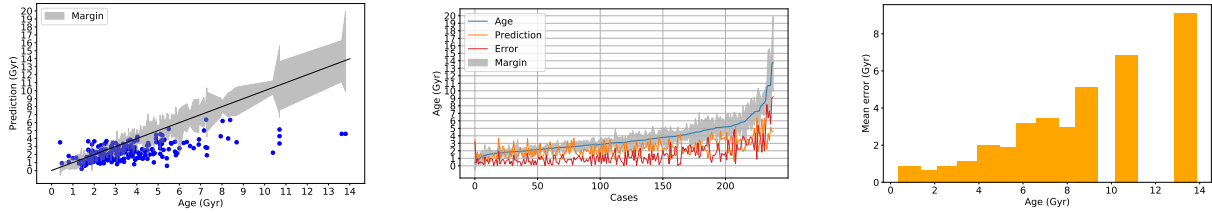
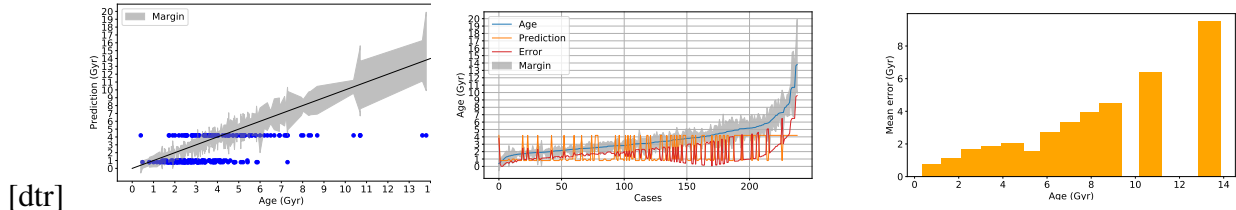


Figura 4.39: Modelo Support Vector Regression



[dtr]

Figura 4.40: Modelo Decision Tree Regression

caen dentro del margen de error para estrellas de hasta 4 Gyr, pero, para edades mayores presenta problemas en la estimación. La segunda figura muestra que se produce un error contante en la estimación de todo el espectro de edades.

**Neural Network** La red neuronal, a diferencia del resto de escenarios, es uno de los modelos que peor rendimiento presenta junto con Decision Tree Regression, mostrando un MAE de 1.853 Gyr. En la primera figura se observa que las estimaciones de todas las edades van de 0.7 a 4.2 Gyr, siendo estas las edades inferior y superior de los cúmulos.

**Gaussian Process** El último modelo de este *benchmark* es Gaussian Process, el cual ha presentado un MAE de 1.624 Gyr. En la primera figura se puede observar un comportamiento ya conocido, pues, hasta las estrellas de 4 Gyr, consigue estimaciones aproximadas. Pero, cuan-

do la edad a estimar pasa de 4 Gyr, se vuelve a demostrar la incapacidad para extrapolar y la tendencia a subestimar edades.

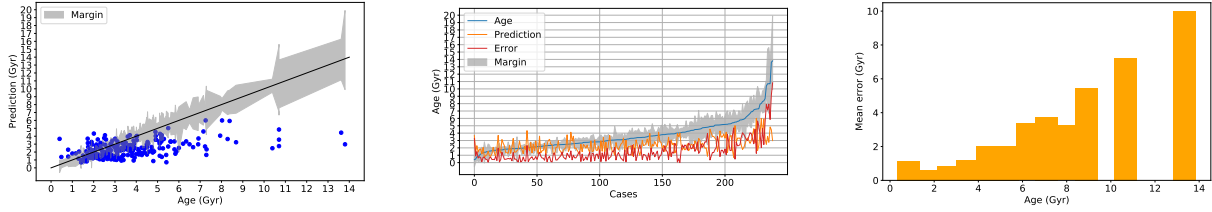


Figura 4.41: Modelo Linear Regressor

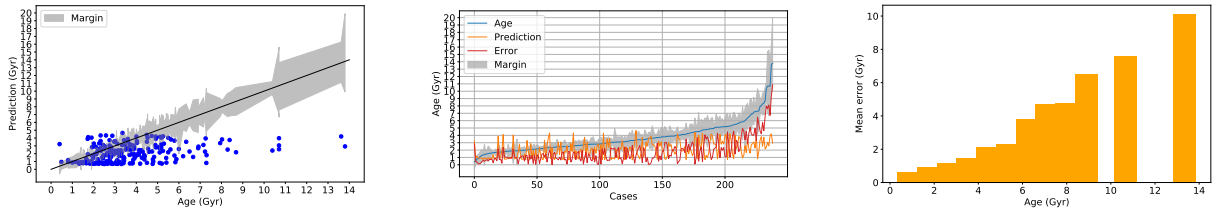


Figura 4.42: Modelo Neural Network

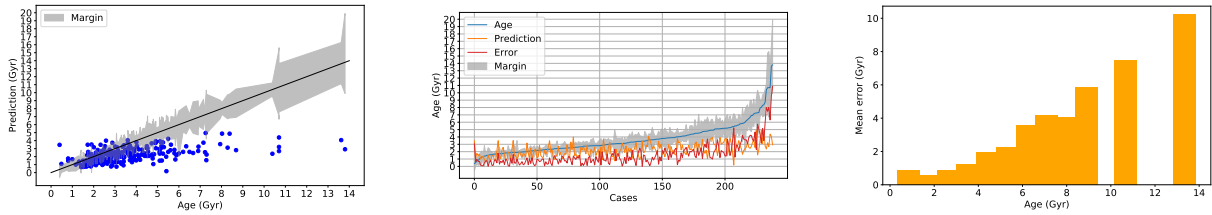


Figura 4.43: Modelo Gaussian Process

#### 4.3.4. Benchmark C: Rendimiento sobre la muestra de control

Se muestra el MAE y la precisión de cada método para el *benchmark* de control en la Figura 4.44 y en la Tabla 4.2, respectivamente. En esta ocasión, los tres mejores métodos son el Proceso Gaussiano, la Red Neuronal y su Stacking. Entre estos tres, destaca el Proceso Gaussiano. En la Figura 4.45 y Figura 4.46 se pueden inspeccionar las predicciones de las edades estelares usando este método frente a las edades de referencia para las estrellas de ese set. Proceso Gaussiano es el modelo con mayor precisión: 40,62 % de las predicciones caen dentro del margen de confianza del conjunto de datos. Para el Sol, Proceso Gaussiano predice una edad de 3,88 Gys frente a una edad aceptada de 4,6 Gyr, subestimándola. De hecho, el modelo tiende a subestimar ligeramente la mayoría de las edades. Se observa este comportamiento en todos lo modelos salvo en kNN y dtr.

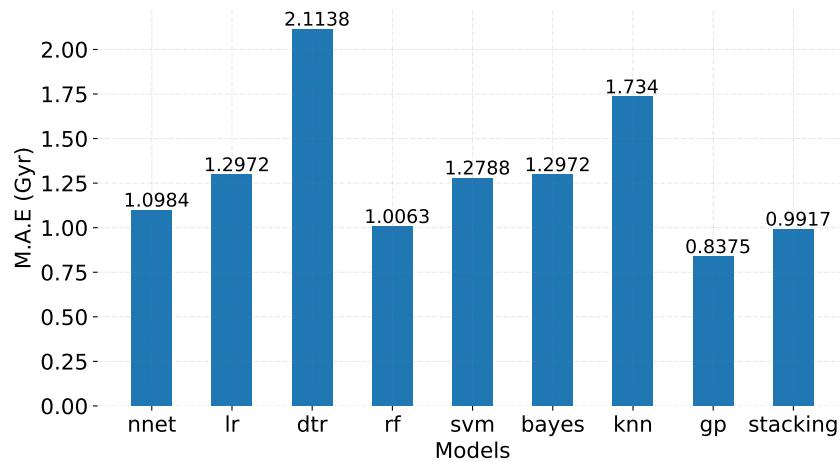


Figura 4.44: Benchmark C: Rendimiento de los modelos en función del MAE.

Respecto a la precisión en las estimaciones para la edad del Sol, 4,6 Gyr, se proporciona en la Tabla 4.1 la edad específica obtenida en cada uno de los modelos. Todos los modelos subestiman, y rf obtiene la predicción más cercana, seguido de dtr y kNN.

Modelos	nnet	lr	dtr	rf	svm	bayes	knn	gp	stacking
Sol (4.6 Gyr)	4.02	3.92	4.49	4.76	3.81	3.91	4.49	3.88	4.02

Tabla 4.1: Edad estimada para el Sol en Gyr. El método más preciso es rf.

**Gaussian Process** A continuación, se muestran las 3 gráficas que detallan el rendimiento del modelo Gaussian Process. El mejor modelo en este *benchmark*, con un MAE de 0.837 Gyr. En las dos primeras figuras se puede observar la precisión de las estimaciones, ya que, se podría decir que oscilan en torno a las edades reales. El modelo encuentra dificultad en estimar edades cercanas a los 7 Gyr, así como la edad del Sol que no se encuentra en el margen de error permitido.

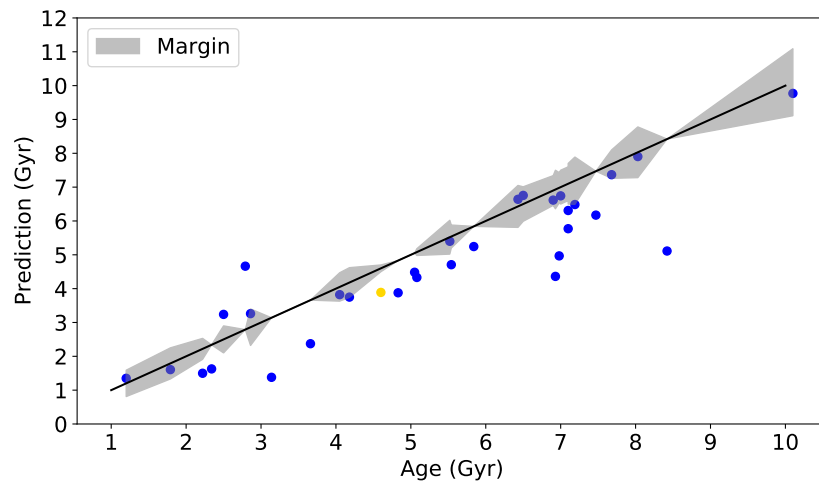


Figura 4.45: Benchmark C: Rendimiento para el modelo Gaussian Process. El punto amarillo se corresponde con la representación de la estimación de la edad del Sol.

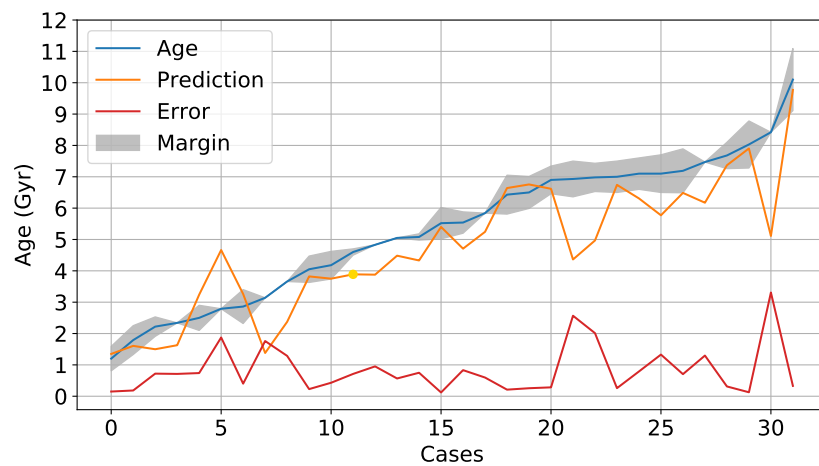


Figura 4.46: Benchmark C: Predicción detallada para el modelo de Gaussian Process. Se observa la edad real (en azul), la predicción de los modelos (en naranja), y el error correspondiente (en rojo). El punto amarillo se corresponde con la representación de la estimación de la edad del Sol.

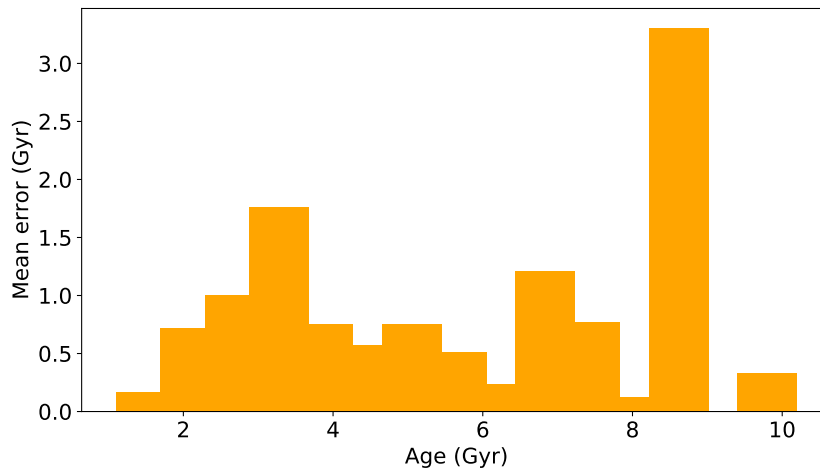


Figura 4.47: Benchmark C: MAE en función del rango de edad del conjunto de prueba.

A continuación, se detalla el resto de modelos del Benchmark C, en las Figuras 4.48, 4.49, 4.50 y 4.51 se adjuntan las gráficas correspondientes al rendimiento de los modelos Stacking, Neural Network, kNN y Random Forest. En sus correspondientes gráficas se puede apreciar:

**Stacking** El modelo Stacking ha presentado un MAE de 0.991 Gyr. En la primera gráfica se vuelve a observa cierta dificultad para estimar estrellas cercanas a los 7 Gyr, y además, se aprecia mayor problema en la estimación de las estrellas más jóvenes. Este modelo sigue sin estimar el Sol dentro del margen de error. En este caso, se obtiene mayor error medio en la estimación de las estrellas más viejas.

**Neural Network** La red neuronal, con un MAE de 1.098 Gyr, muestra un rendimiento similar al de los modelos anteriores. Presenta problemas en la estimación de las estrellas mayores de 7 Gyr y no consigue estimar la edad del Sol dentro del margen de error. Las dos primeras gráficas ilustran cierta subestimación en las estimaciones de todo el espectro de edades.

**kNN** El modelo kNN, con un MAE de 1.734 Gyr, presenta, a simple vista, mayor dispersión en la nube de edades estimadas. En las dos primeras figuras se observan edades superiores a 10 y 12 Gyr para estrellas de 7 Gyr. Por otro lado, se tiene una mejor estimación de la edad del Sol que con los modelos anteriores.

**Random Forest** El modelo Random Forest, con un MAE de 1.006 Gyr, también presenta una buena estimación de la edad del Sol. Además, las dos primeras figuras muestran la oscilación de las estimaciones en torno a las edades reales, aunque la mayoría no lo suficientemente buenas como para entrar en el margen de error.

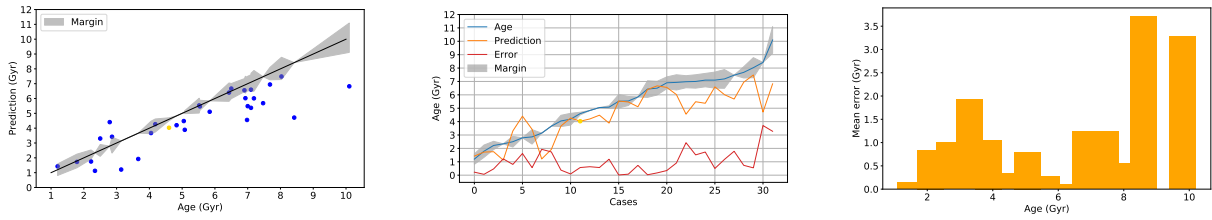


Figura 4.48: Modelo Stacking

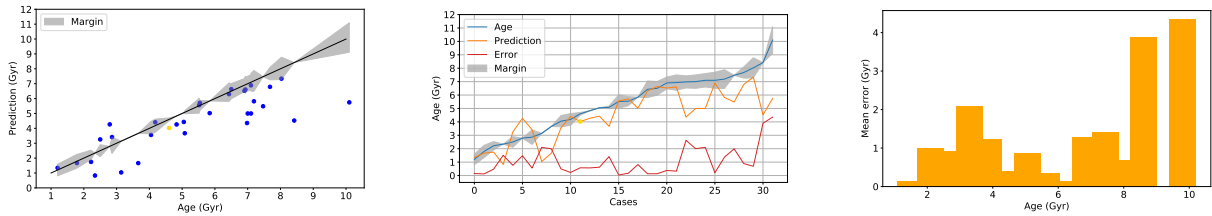


Figura 4.49: Modelo Neural Network

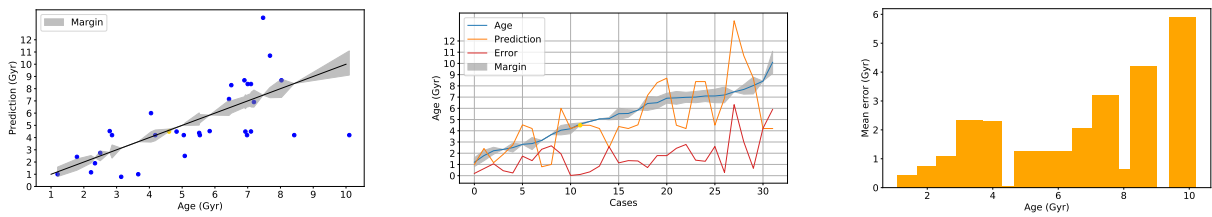


Figura 4.50: Modelo k-Nearest Neighbor

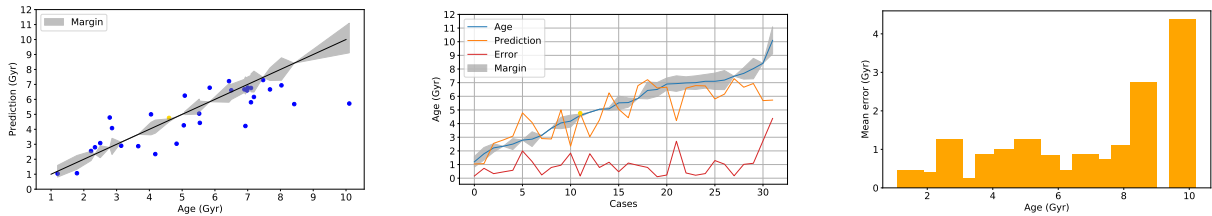


Figura 4.51: Modelo Random Forest

En la Figura ?? se adjuntan las gráficas que detallan el rendimiento de los modelos Support Vector Regression, Decision Tree Regression, Linear Regressor y Bayesian Regression para el Benchmark C. En sus correspondientes gráficas se puede apreciar:

**Support Vector Regression** El modelo de máquina de vectores soporte ha presentado un MAE de 1.278 Gyr. Las dos primeras figuras ilustran la tendencia del modelo para subestimar las edades de las estrellas a partir de los 3 Gyr. Además, en la tercera gráfica, se observa mayor error medio para la estimación de las estrellas más viejas.

**Decision Tree Regression** El modelo Decision Tree Regression, con un MAE de 2.113 Gyr, muestra el peor rendimiento del *benchmark*. Las dos primeras figuras muestran que la mayoría de las estimaciones están muy lejos de entrar dentro del margen de error asociado a las edades.

Por contra, la estimación de la edad del Sol es una de las pocas edades que sí se encuentra dentro del margen de error.

**Linear Regressor** El modelo Linear Regressor ha presentado un MAE de 1.297 Gyr. Sus dos primeras gráficas muestran subestimación en las estimaciones a partir de los 3 Gyr. Además, la mayor parte de estimaciones se encuentran fuera del margen de error, incluida la edad del Sol. La figura tres muestra un aumento del error medio con el aumento de edad a estimar.

**Bayesian Regression** Por último, se tiene el modelo Bayesian Regression, el cual ha presentado un MAE de 1.297 Gyr. Su comportamiento es similar al del modelo de regresión lineal, tomando un valor similar y una estimación de edades muy parecida.

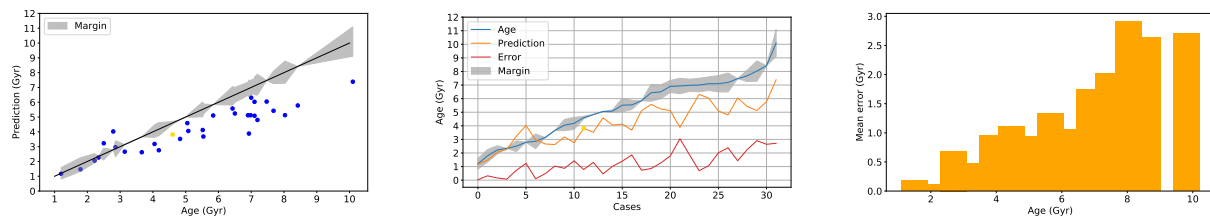


Figura 4.52: Modelo Support Vector Regression

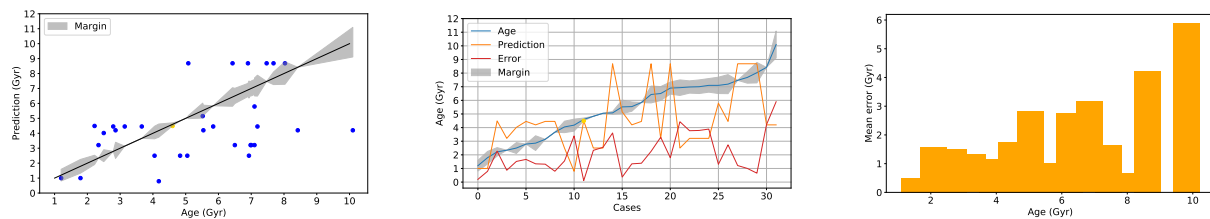


Figura 4.53: Modelo Decision Tree Regression

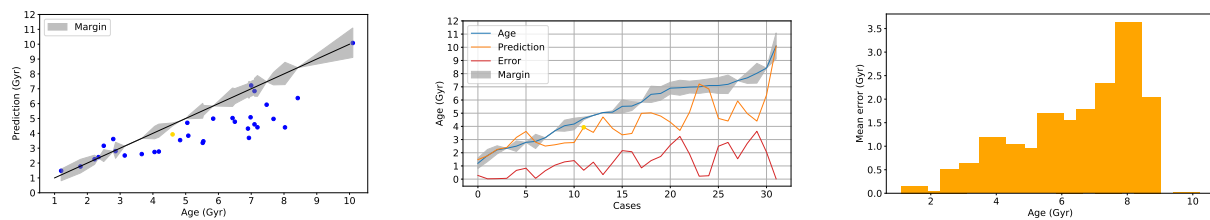


Figura 4.54: Modelo Linear Regressor

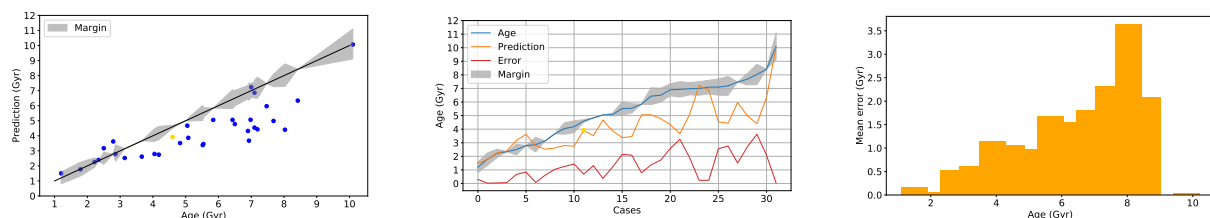


Figura 4.55: Modelo Bayesian Regression



### 4.3.5. Conclusiones de la comparativa

Comparando los resultados de los Benchmark A (Figura 4.10) y C (Figura 4.44), se puede concluir que:

1. Los modelos ganadores son comunes Neural Network, Gaussian Process y su Stacking.
2. k-Nearest Neighbors y Decision Tree Regression sufren una degradación considerable de su rendimiento, ya que dependen en gran medida de la distribución de datos utilizada durante su entrenamiento.
3. Bayesian Regression y Linear Regressor son similares, mostrando un incremento similar en sus correspondientes MAEs, pero al mismo tiempo ofreciendo cierta estabilidad entre escenarios.

Al observar la Tabla 4.2 después de conocer el rendimiento de todos los modelos, además de lo comentado en cada uno de los benchmarks, se llega a la conclusión de que: a) salvo para el Benchmark B2, nnet presenta una precisión alta en comparación con el resto de modelos, y b), salvo para el Benchmark B2, lr y bayes presentan una precisión muy inferior a la conseguida por el modelo de mejor rendimiento.

Benchmark	nnet	lr	dtr	rf	svm	bayes	knn	gp	stacking
<b>A</b>	73.23	35.43	63.78	60.63	46.46	32.28	<b>77.95</b>	62.20	62.99
<b>B1</b>	<b>59.84</b>	29.92	51.18	48.03	37.79	28.35	55.91	51.97	n/a
<b>B2</b>	28.45	30.96	21.34	25.94	<b>33.05</b>	29.71	25.52	30.96	n/a
<b>C</b>	34.37	21.87	9.37	21.87	12.50	21.87	15.62	<b>40.62</b>	34.37

Tabla 4.2: Precisión de todos los métodos en los diferentes benchmarks. La precisión se mide como el porcentaje de estimaciones de edad que caen dentro del margen de confianza asociado a cada estrella.



# Capítulo 5

## Meta-learning para estimación de edades estelares

### 5.1. Introducción

En este trabajo, además de realizar un estudio comparativo sobre los modelos tradicionales de Inteligencia Artificial para regresión de edades estelares, se propone abordar por primera vez, el aplicar técnicas de meta-learning a este problema. El meta-learning se define como el proceso de aprender a aprender con pocas muestras, o explicado de un modo mas técnico, sería el modo de entrenamiento de un modelo en varias tareas de aprendizaje, de modo que el modelo sea capaz de resolver nuevas tareas utilizando una pequeña cantidad de muestras.

En particular, en este proyecto se estudia la aplicación del algoritmo MAML (model-agnostic meta-learning) [20], para el problema de la estimación de edades estelares. Este algoritmo se denomina agnóstico del modelo porque es compatible con cualquier modelo entrenado con descenso de gradiente y aplicable a una variedad de diferentes problemas de aprendizaje, incluyendo clasificación, regresión y aprendizaje por refuerzo.

Para llevar a cabo este estudio, se parte de [20], donde se propone un algoritmo de meta-learning y agnóstico del modelo con enfoque en redes neuronales profundas. La idea clave de este método es entrenar los parámetros iniciales del modelo de forma que tenga un rendimiento máximo en una nueva tarea después de que los parámetros se hayan actualizado a través de uno o más pasos de gradiente calculados con una pequeña cantidad de datos de esa nueva tarea.

La contribución principal de este trabajo fin de máster reside en un algoritmo de modelado simple e independiente de las tareas para el meta-learning sobre datación estelar, capaz de conseguir un aprendizaje rápido en una nueva tarea a través de una pequeña cantidad de actualizaciones de gradiente. Es decir, vamos a emplear y evaluar el rendimiento de MAML en un escenario de meta-learning para la regresión de la edad de las estrellas.

## 5.2. Model-Agnostic Meta-Learning Algorithm

En esta sección se comienza describiendo los detalles de MAML [20]. Durante el meta-learning, el modelo se entrena para poder adaptarse a un número grande o infinito de tareas. A continuación, se presenta una aproximación genérica de una tarea, permitiendo abordar una variedad de problemas de aprendizaje, desde la clasificación hasta el aprendizaje por refuerzo.

Cada tarea consiste en una función de pérdidas, una distribución de las observaciones iniciales, una distribución de transición y una longitud de episodio. El modelo genera muestras con la longitud del episodio y la función de pérdidas proporciona retroalimentación específica para cada tarea. En el escenario de meta-learning, se tiene en cuenta una distribución de todas las tareas a las que se quiere adaptar el modelo. En la configuración del aprendizaje *K-shot*, el modelo se entrena para que aprenda nuevas tareas de la distribución total, usando únicamente  $K$  muestras de la observación inicial y la retroalimentación proporcionada por la función de pérdidas para cada tarea.

Durante el metaentrenamiento, se muestrea una tarea proveniente de la distribución total, el modelo se entrena con  $K$  muestras y se retroalimenta con la correspondiente función de pérdidas de cada tarea, y después se prueba sobre nuevas muestras de la misma tarea. Entonces, el modelo mejora considerando cómo varía el error de prueba sobre nuevos datos, provenientes de las observaciones iniciales, respecto a los parámetros.

Para el desarrollo del algoritmo se presenta un enfoque explícito para el problema. Ya que el modelo se ajustará utilizando una regla de aprendizaje basada en gradientes, el objetivo es hacer aprender a un modelo de tal manera que esta regla de aprendizaje basada en gradientes pueda hacer un avance rápido en nuevas tareas extraídas de la distribución total sin sobreajuste. Por tanto, el objetivo es encontrar los parámetros del modelo que sean sensibles a los cambios de las tareas, de modo que pequeños cambios en los parámetros producirán grandes mejoras en la función de pérdidas de cualquiera tarea.

A continuación, se detalla paso a paso del algoritmo MAML. Inicialmente se requiere la distribución de tareas a la que se quiere adaptar el modelo, además del tamaños de los pasos de actualización de hiperparámetros. Entrando en detalle en cada paso del algoritmo, en primer lugar, se inicializa de forma aleatoria el vector de parámetros con el que se parametriza el modelo. El bucle principal del algoritmo comienza con un muestreo de tareas sobre la distribución inicial. Para cada una de estas tareas se evalúa la función de pérdidas respecto a las  $K$  muestras seleccionadas. A continuación, se realiza la adaptación de parámetros mediante el descenso de gradiente. Una vez se ha finalizado con el conjunto de tareas, se actualiza el vector de parámetros que parametriza el modelo.

**Data:**  $p(T)$ : distribución de tareas

**Data:**  $\alpha, \beta$  : tamaños de los pasos de actualización de hiperparámetros

Inicialización aleatoria de  $\theta$ ;

**while not done do**

    Muestreo de un *batch* de tareas  $T_i \sim p(T)$ ;

**for all**  $T_i$  **do**

        Evaluar  $\nabla_{\theta} L_{T_i}(f_{\theta})$  respecto a las  $K$  muestras;

        Realizar la adaptación de parámetros mediante el descenso de gradiente:  $\theta'_i = \theta -$

$\alpha \nabla_{\theta} L_{T_i}(f_{\theta})$ ;

**end**

    Actualizar  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{T_i \sim p(T)} L_{T_i}(f_{\theta'_i})$

**end**

**Algorithm 1:** Model-Agnostic Meta-Learning

### 5.3. MAML en regresión

Para estudiar el rendimiento de un algoritmo MAML sobre el problema de regresión de estimación de edades estelares, se parte del desarrollo realizado por Chelsea Finn, entre otros [20].

En su trabajo se presenta una evaluación experimental con el objetivo de responder a tres cuestiones: 1) ¿Será capaz el algoritmo MAML de reproducir un aprendizaje rápido sobre las nuevas tareas?, 2) ¿Podrá ser usado este algoritmo para implementar metaentrenamiento en diferentes dominios, incluyendo regresión, clasificación y aprendizaje por refuerzo? y 3) ¿Puede el modelo aprendido continuar mejorando su rendimiento con actualizaciones de gradiente adicionales?

Los investigadores proponen tres escenarios de aplicación para tratar de responder a estas preguntas: 1) un problema simple de regresión donde se ilustran los principios básicos del algoritmo MAML, 2) un problema de reconocimiento de imágenes sobre los datasets Omniglot y MiniImagenet y 3) un problema de aprendizaje por refuerzo en el que se construyen varios conjuntos de tareas basadas en entornos de control continuo.

Conociendo los 3 escenarios expuestos en esta investigación y con el fin de abordar el problema de regresión de datación de estrellas, se toma como referencia el escenario de regresión simple. En este escenario, cada tarea implica una regresión de la entrada a la salida de una onda sinusoidal, donde la amplitud y la fase de la senoide varían en cada tareas.

Para la generación de los senos se necesitan conocer los valores que pueden tomar la amplitud y la fase de los senos. Una vez definido esto, se parte de un vector equiespaciado con tantas muestras como puntos se quiere que tenga el seno. Sobre este vector se aplica la función seno con su correspondiente valor de fase y se multiplica por un valor amplitud. Así se obtienen tantos vectores senos como vectores se han definido para el tamaño de *batch*.

En la fase de entrenamiento, se generan 25 vectores, con los 20 puntos correspondientes al seno, por cada una de las 70000 iteraciones de metaentrenamiento. Son 25 vectores porque

el tamaño de *batch* es igual a 25. Después, en la etapa de prueba, se producen 600 iteraciones en las que se genera 1 vector de 20 puntos para cada iteración. En esta ocasión, el tamaño de *batch* es igual a 1. Por lo tanto, el algoritmo tiene como entrada 25 vectores de 20 elementos por cada una de las 70000 iteraciones del metaentrenamiento y 1 vector de 20 elementos por cada una de las 600 iteraciones de prueba. De los 20 puntos que componen el seno, 10 se emplean íntegramente en el proceso de descenso de gradiente, por eso se denomina *10-shot*.

La función de pérdidas se corresponde con el error cuadrático medio entre la predicción y el valor real. El modelo regresor se basa en una red neuronal con 2 capas ocultas de tamaño 40 cada una, con función de activación ReLU. Para el entrenamiento del MAML, se utiliza un valor de actualización de gradiente igual a 10 con un tamaño fijo de *step* de 0.01 y se emplea Adam como metaoptimizador.

Para evaluar su rendimiento, se ajusta un solo modelo con un número variable las muestras empleadas para el correspondiente *K-shot* y se compara con dos *baselines*: 1) preentrenando sobre todas las tareas, y 2), usando un oráculo que reciba la amplitud y fase real de cada seno.

Las Figuras 5.1, 5.2 y 5.3 ilustran los resultados de esta comparativa para diferentes valores de *K*. Aquí se representa el MSE en función del valor de la iteración de actualización del gradiente, mostrando mejor resultado en las últimas actualizaciones de gradiente. Salvo para el oráculo, que ya conoce los valores de amplitud y fase, por lo que su valor de MSE es igual a 0 todo el tiempo.

En estas gráficas se puede observar la diferencia de rendimiento que se obtiene al comparar el algoritmo MAML frente al modelo preentrenado. En el mejor de los rendimientos obtenidos, el modelo preentrenado presenta un MSE de 1.5, mientras que el MAML presenta un MSE próximo a 0. Por otro lado, el oráculo presenta mejores resultados que el MAML para las primeras iteraciones de la actualización del gradiente, ya que toma un valor perpetuo igual a 0, pero el MAML llega a tomar valores comparables con la configuración *10-shot* e incluso similares con la configuración *20-shot*, con valores muy próximos a 0.

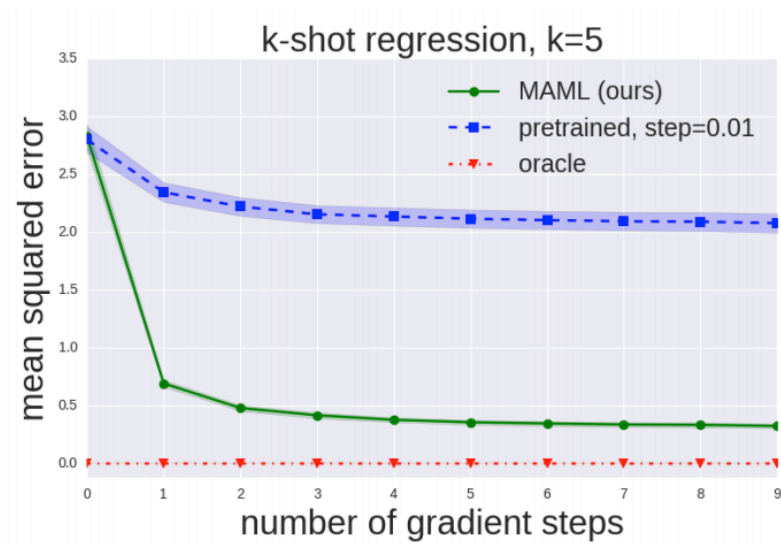


Figura 5.1: Resultados del estudio de regresión sobre las ondas sinusoidales con  $K = 5$ .

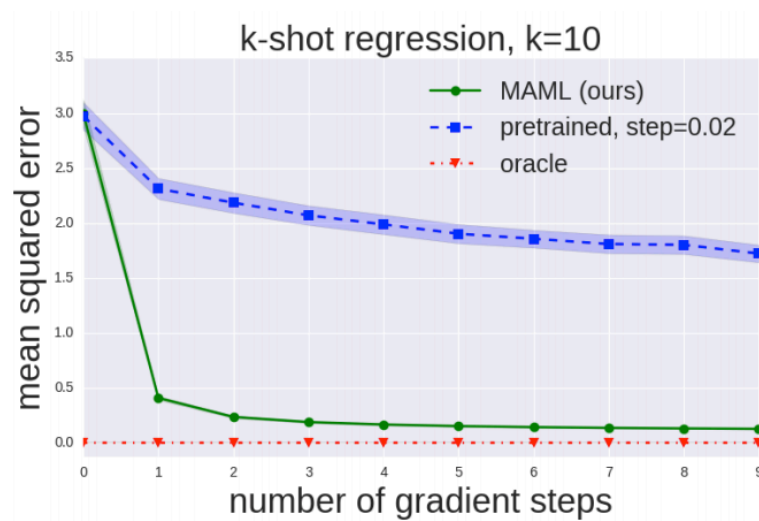


Figura 5.2: Resultados del estudio de regresión sobre las ondas sinusoidales con  $K = 10$ .

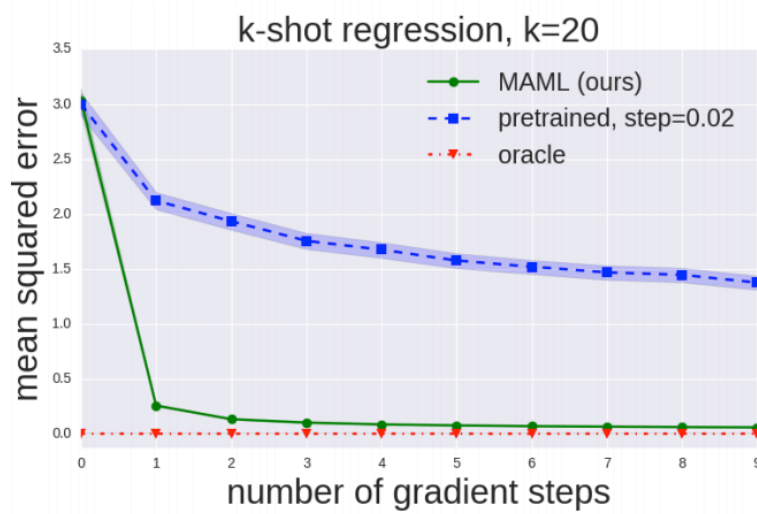


Figura 5.3: Resultados del estudio de regresión sobre las ondas sinusoidales con  $K = 20$ .

## 5.4. MAML para datación estelar

En esta sección, se detalla la aplicación del algoritmo MAML sobre el problema de la estimación de edades estelares. En primer lugar, se analiza la inyección de los datos en el algoritmo, teniendo en cuenta que las características de las estrellas que se emplean para predecir la edad son las mismas que se han utilizado en el Benchmarking principal de este proyecto, las cuales se detallan en la Sección 3. A continuación, se profundiza en el propio funcionamiento del algoritmo, donde se implementan los procesos comentados en el apartado anterior. Y finalmente, se analizan los resultados obtenidos por el algoritmo comparando diferentes configuraciones para la arquitectura de la red neuronal.

### 5.4.1. Inyección de los datos

Como se ha comentado en la introducción de este apartado, el set de datos elegido para estimar la edad estelar es el mismo que se detalla en la Sección 3. Este set se compone de 8 características asociadas a cada una de las estrellas que forman el dataset: temperatura estelar efectiva, logaritmo de la gravedad superficial, masa estelar, radio estelar, contenido de hierro sobre hidrógeno, luminosidad, periodo de rotación y la edad de la estrella.

Este dataset es idéntico al que se ha detallado en la Sección 3, ya que se parte de un conjunto de datos inicial sobre el que es necesario aplicar una serie de filtros asociados a la física detrás de la girocronología. Las características que se emplean para estimar la edad de las estrellas



son: masa estelar, radio estelar, temperatura efectiva, luminosidad, metalicidad, logaritmo de la gravedad superficial y periodo de rotación.

Tras es filtrado, el conjunto de datos cuenta con un número relativamente reducido de muestras en comparación con los experimentos de regresión estudiados en [20]. Por tanto, ya que la el conjunto de datos completo incluye valores de error por exceso y defecto para cada una de las características, se aumenta el tamaño del dataset siguiendo una distribución normal. Esta normal tiene como rango inferior el valor de cada característica menos su error por defecto y como rango superior, el valor de la característica más el error por exceso. Además, la función de la distribución normal tiene un parámetro asociado al número de muestras que se generan a partir de las características y sus valores de error asociado, este parámetro toma un valor igual a 20. Este proceso permite pasar de un conjunto de datos de 637 estrellas a un conjunto de datos de más de 13000 estrellas.

En la práctica esto se traduce en 10600 vectores de 7 elementos, es decir, sus características, para el proceso de metaentrenamiento y 2651 vectores, con sus correspondientes 7 elementos, para el proceso de prueba. Es necesario tener en cuenta que el tamaño del *batch* en el proceso de entrenamiento es igual a 20, por lo que las iteraciones totales de metaentrenamiento se obtienen de distribuir los 10600 vectores entre 20. Por otro lado, el tamaño de *batch* del proceso de prueba es igual a 1, por lo que, no será necesario hacer un reparto de los vectores en cada iteración.

La división del conjunto total de estrellas en entrenamiento y prueba tiene una relación 80/20 respectivamente. Además, las edades se ordenan de forma creciente, por lo que, el conjunto de entrenamiento contiene estrellas de hasta 4.2 Gyrs de edad, mientras que el conjunto de prueba comprende edades de entre 4.2 Gyrs y 18.8 Gyrs.

Durante la fase de entrenamiento del modelo, al algoritmo se le inyectan conjuntos de 20 estrellas, o lo que es lo mismo, el tamaño del *batch* es igual a 20, tal y como se ilustra en la Figura 5.4. De estas 20 muestras, 10 se emplean en el proceso de actualización de gradiente, es por esto que el experimento se cataloga como *few-shot* con  $K = 10$ , es decir, *10-shot*. Para la fase de prueba el tamaño del *batch* es igual a 1, por lo que, en cada iteración del *test* se inyecta una estrella. En la Figura 5.5 se observa este proceso de forma gráfica.

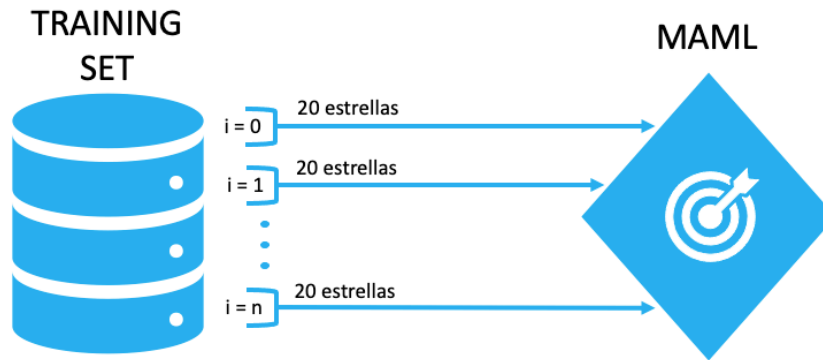


Figura 5.4: Esquema de la inyección de datos durante la fase de entrenamiento.

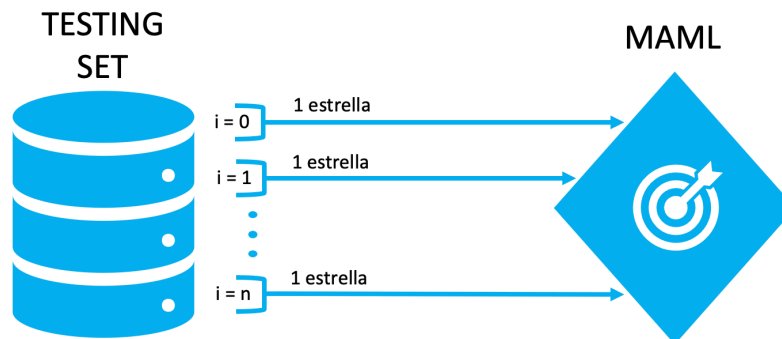


Figura 5.5: Esquema de la inyección de datos durante la fase de prueba.

### 5.4.2. Proceso de aprendizaje

Según se ha comentado en la sección anterior, el algoritmo procesa *batches* de 20 estrellas durante el proceso de entrenamiento, 10 de estas estrellas pasan a ser directamente *inputs* de la red neuronal del modelo y le proporcionan información del ajuste de pesos a las otras 10 estrellas que participarán en el proceso de actualización de gradiente. Esas 10 estrellas restantes, pasan por un proceso de actualización de gradiente en el que suceden 10 pasos, es decir, se producen 10 actualizaciones del gradiente hasta llegar a la estimación óptima. En cada una de las actualizaciones se produce una corrección de los pesos del modelo y se obtiene una medida de error asociada a la estimación de la edad por parte de la red neuronal con dichos pesos. La Figura 5.6 ilustra el funcionamiento descrito.

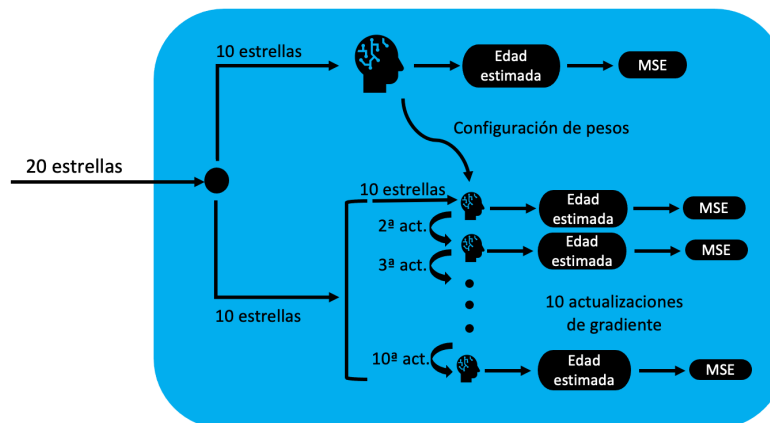


Figura 5.6: Esquema del funcionamiento interno del modelo durante el proceso de entrenamiento.

En el proceso de pruebas se sigue el mismo procedimiento, pero a diferencia de la fase de entrenamiento, el tamaño de *batch* es igual a 1. Esto permite, en cada iteración del *test*, obtener la edad estimada para cada estrella y las pérdidas asociadas en cada uno de los 10 pasos de actualización de gradiente. Además, se obtiene la estimación y las pérdidas de la red neuronal sin la corrección de sus pesos, es decir, sin actualización del gradiente. La Figura 5.7 ilustra el funcionamiento descrito.

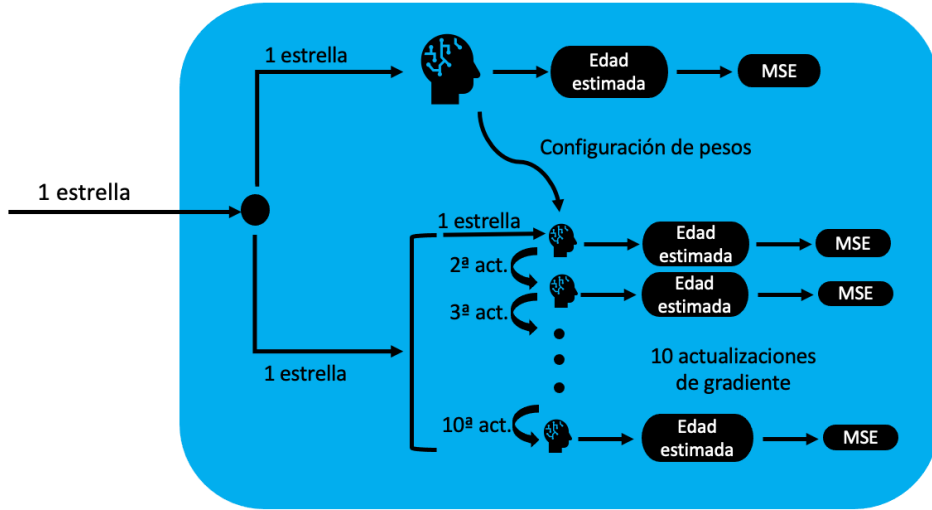


Figura 5.7: Esquema del funcionamiento interno del modelo durante el proceso de prueba.

### 5.4.3. Resultados obtenidos

Para evaluar los resultados obtenidos en este experimento, se trabaja con la misma métrica que se utiliza en el artículo original, el error cuadrático medio. Esta medida se obtiene aplicando la función de pérdidas (MSE) sobre la estimación del modelo y la edad real de la estrella.

Al finalizar la fase de prueba, se obtiene un vector con 11 posiciones para cada estrella del conjunto de *test*, la primera posición del vector se corresponde con el error obtenido sin descenso de gradiente y las 10 posiciones restantes con el error en cada una de las actualizaciones de gradiente.

Con el fin de encontrar la mejor arquitectura para la red neuronal, se realizan diferentes simulaciones variando la composición en cuanto al número de capas ocultas del modelo y el número de unidades de cada capa. La arquitectura publicada en el artículo original esta compuesta por 2 capas ocultas de 40 unidades cada una, además, se propone: 1) 2 capas ocultas de 20 unidades cada una, 2) 3 capas ocultas de 20 unidades cada una y 3) 1 capa oculta con 40 unidades. La Figura 5.8 ilustra los resultados obtenidos al comparar las diferentes arquitecturas.

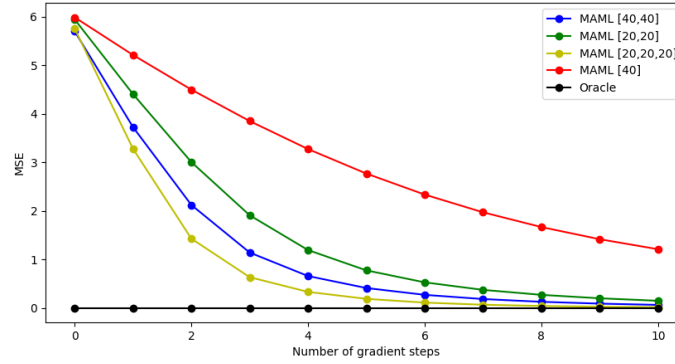


Figura 5.8: Comparación del MSE para las distintas configuraciones de la red neuronal

Tras analizar estos resultados, se concluye que el modelo que mejores resultados presenta es la red neuronal cuya arquitectura está formada por 3 capas ocultas y 20 unidades en cada una. Desde la primera iteración del gradiente, muestra un MSE menor al de todas las demás arquitecturas.

Por otro lado, la arquitectura que muestra los peores resultados de experimento es la red neuronal formada por una sola capa oculta de 40 unidades. Se observa un rendimiento considerablemente peor desde la primera hasta la última iteración de actualización de gradiente.

Sin tener en cuenta esta última configuración, las 3 arquitecturas restantes se aproximan, en las últimas actualizaciones de gradiente, a los resultados obtenidos por el oráculo ( $MSE = 0$ ). En la Tabla 5.1 se detallan los valores exactos.

Arquitectura	Act. 0	Act. 1	Act. 2	Act. 3	Act. 4	Act. 5	Act. 6	Act. 7	Act. 8	Act. 9	Act. 10
[40 40]	5.69	3.72	2.11	1.13	0.65	0.40	0.26	0.18	0.12	0.08	0.06
[20 20]	5.94	4.40	3.00	1.90	1.19	0.77	0.52	0.37	0.26	0.19	0.14
[20 20 20]	5.75	3.27	1.42	0.63	0.32	0.18	0.10	0.06	0.03	0.02	0.01
[40]	5.98	5.21	4.49	3.84	3.27	2.76	2.33	1.96	1.66	1.41	1.20

Tabla 5.1: MSE para las diferentes arquitecturas y cada una de las actualizaciones de gradiente.



# Capítulo 6

## Conclusiones

### 6.1. Modelos clásicos de IA para datación de estrellas

En la primera parte de este trabajo se ha presentado un análisis exhaustivo comparando el estado del arte en modelos en regresión de Inteligencia Artificial entrenados y probados para solucionar el problema de datación estelar. En base a los resultados obtenidos en este estudio, se informa de los siguientes hallazgos sobre el rendimiento y la capacidad de generalización de los modelos de regresión para la estimación de edades estelares.

En primer lugar, el estudio demuestra que un modelo de datación estelar basado en una Red Neuronal como la desarrollada en este proyecto, proporciona resultados notables en términos de compensación entre generalización y precisión. Los experimentos revelan que el rendimiento del modelo nnet es bueno en los Benchmarks A y C, donde se prueba sobre un conjunto de estrellas aleatorias de la base de datos y sobre un conjunto de control, respectivamente, pero también presenta buenos resultados en el Benchmark B1, donde se pone a prueba su capacidad de generalizar sobre estrellas de edades desconocidas. Es un modelo simple, un MLP con 3 capas ocultas, que podría extenderse (usando capas 1D CNN) posiblemente ofreciendo mejores resultados. Se deja la puerta abierta para futuras investigaciones. Si los profesionales del campo pretenden buscar la solución que proporcione el menor error, el mensaje es: debe utilizarse un modelo de stacking que combine procesos gaussianos y redes neuronales, esta combinación presenta los mejores rendimientos del Benchmark A y los segundos mejores del Benchmark C, además de presentar buenas precisiones en ambos escenarios de generalización, aunque no las mejores. Finalmente, se recomienda una regresión bayesiana si los datos de entrenamiento disponibles pertenecen principalmente a cúmulos de estrellas.

En segundo lugar, es importante destacar que *todos* los modelos tienden a subestimar en sus estimaciones, y no solo en el Benchmark B1, donde sería natural debido a que los modelos se entrenan con estrellas jóvenes y desconocen un amplio rango de edades. Por lo tanto, se ve una interesante línea de trabajo futura en relación a la reducción de este sesgo.

En tercer lugar, el estudio revela resultados prometedores para el problema de datación estelar. Un error de  $< 0,5$  Gyr se puede considerar como un gran avance en el campo.

Por último, destacar que este trabajo ha servido de apoyo para la publicación de un artículo para el *1st Workshop on AI for Space*. El artículo [45], titulado *AI for dating stars: a benchmarking study for gyrochronology*, se basa en la aplicación de modelos regresivos de aprendizaje automático para reemplazar las combinaciones lineales que se utilizan en el campo de la estimación de edades estelares. Los modelos de IA empleados y la forma de compararlos siguen el guión expuesto en el Capítulo 4 donde se realiza un *benchmarking* de los escenarios vistos en la Sección 4.2 con los modelos detallados en la Sección 4.1. Las conclusiones a las que se llegan en el artículo son idénticas a las obtenidas en este proyecto.

## 6.2. Meta-Learning aplicado al problema de la estimación de la edad de las estrellas

En la segunda parte del trabajo, se estudia el algoritmo MAML (Meta-Agnostic Meta-Learning) para resolver el problema de la regresión de las edades estelares en un planteamiento típico de meta-learning, o aprender a aprender con pocas muestras. Sobre este problema, se obtienen diferentes resultados asociados a diferentes arquitecturas desarrolladas para el modelo. Los peores resultados van asociados a la arquitectura más simple probada, con una sola capa oculta y 40 células, que muestra un  $MSE = 1.20$  en su última actualización de gradiente. En contraposición, los mejores resultados se puede observar en el modelo que cuenta con 3 capas ocultas de 20 unidades cada una, con un  $MSE = 0.01$ . Es importante destacar que las configuraciones más complejas consiguen resultados de MSE muy próximos a cero en la última actualización de gradiente. En cualquier caso, lo más relevante es que es la primera vez que se prueba una aproximación basada en meta-learning para abordar el problema de la datación estelar, y los resultados obtenidos son muy prometedores.

Respecto a la comparación de estos resultados con los obtenidos en el *benchmarking* anterior, resulta inviable realizar una comparación directa debido a la métrica empleada para la función de pérdidas. Durante el primer experimento con los modelos clásicos se ha empleado el error medio absoluto, mientras que en este análisis se utiliza el error cuadrático medio. Es por esto que, como próximos pasos, sería interesante realizar un estudio sobre otros modelos y algoritmos que ofrezcan resultados comparables.

Tras lo comentado, se establecen como próximos pasos a seguir el análisis de otros algoritmos enfocados en el aprendizaje del modelo mediante un número reducido de muestras. Reptile [46] es uno de estos algoritmos, matemáticamente es similar al algoritmo MAML estudiado en este proyecto ya que muestrea recursivamente una tarea, realizando un descenso de gradiente y actualizando los parámetros iniciales. Aún así, resultaría interesante una comparación de los rendimientos de ambos algoritmos. Otro estudio a tener en cuenta es el desarrollado en [66], en este caso el algoritmo se denomina MLTI (Meta-Learning Task Interpolation). En este análisis se generan tareas adicionales tras muestrear aleatoriamente un par de tareas e interpolar características y sus correspondientes etiquetas.



## 6.3. Conclusiones personales

Como reflexión personal, considero que el desarrollo de este proyecto me ha permitido adquirir conocimientos generales y específicos sobre el mundo de la inteligencia artificial y la datación de estrellas. Hasta el momento en el que empecé el TFM, no sabía absolutamente nada de las técnicas de datación estelar ni de su importancia para la evolución del ser humano y de la ciencia. Gracias a la transferencia de conocimiento por parte de Andrés y Roberto en estos campos, he podido aprender lo requerido para enfrentarme a este proyecto. Esto me ha permitido alcanzar dos importantes hitos para: 1) obtener unos resultados lo suficientemente satisfactorios como para poder catalogarse como prometedores dentro del ámbito de la datación estelar y 2) aportar mi granito de arena en la elaboración de un artículo [45] basado, en parte, en este TFM y publicado en el *1st Workshop on AI for Space, CVPR, 2021*. Todo el código para reproducir los resultados del artículo lo hemos liberado públicamente <sup>1</sup>.

Por otro lado, este proyecto también me ha ayudado a conocer un poco mejor la dinámica que podría seguirse en un proyecto de investigación. Acostumbrado al mundo de la empresa privada en el que todo se hace de prisa y con una calidad cuestionable, el tener que dedicarle horas a leer artículos y probar líneas de código distintas me permite adquirir nuevas competencias.

---

<sup>1</sup>El código está disponible en el siguiente repositorio: <https://github.com/gramuah/ai4datingstars>



# Bibliografía

- [1] Ruth Angus, Suzanne Aigrain, Daniel Foreman-Mackey, and Amy McQuillan. Calibrating gyrochronology using Kepler asteroseismic targets. , 450(2):1787–1798, June 2015. 3, 5
- [2] Ruth Angus, Angus Beane, Adrian M. Price-Whelan, Elisabeth Newton, Jason L. Curtis, Travis Berger, Jennifer van Saders, Rocio Kiman, Daniel Foreman-Mackey, Yuxi Lucy Lu, Lauren Anderson, and Jacqueline K. Faherty. Exploring the Evolution of Stellar Rotation Using Galactic Kinematics. , 160(2):90, August 2020. 2
- [3] Ruth Angus, Timothy D. Morton, and Daniel. Foreman-Mackey. stardate: Combining dating methods for better stellar ages. *Journal of Open Source Software*, 41(4):1469, 2019. 6
- [4] Ruth Angus, Timothy D. Morton, Daniel Foreman-Mackey, Jennifer van Saders, Jason Curtis, Stephen R. Kane, Megan Bedell, Rocio Kiman, David W. Hogg, and John Brewer. Toward Precise Stellar Ages: Combining Isochrone Fitting with Empirical Gyrochronology. , 158(5):173, November 2019. 2, 3, 6
- [5] L. A. Balona. Starspots on A stars. , 467(2):1830–1837, May 2017. 5
- [6] Sydney A. Barnes. On the Rotational Evolution of Solar- and Late-Type Stars, Its Magnetic Origins, and the Possibility of Stellar Gyrochronology. , 586(1):464–479, March 2003. 2, 5
- [7] Sydney A. Barnes. Ages for Illustrative Field Stars Using Gyrochronology: Viability, Limitations, and Errors. , 669(2):1167–1189, November 2007. 5
- [8] Sydney A. Barnes. A Simple Nonlinear Model for the Rotation of Main-sequence Cool Stars. I. Introduction, Implications for Gyrochronology, and Color-Period Diagrams. , 722(1):222–234, October 2010. 5
- [9] Sydney A. Barnes, Joerg Weingrill, Dario Fritzewski, Klaus G. Strassmeier, and Imants Platais. Rotation Periods for Cool Stars in the 4 Gyr old Open Cluster M67, The Solar-Stellar Connection, and the Applicability of Gyrochronology to at least Solar Age. , 823(1):16, May 2016. 2, 6, 10

- [10] J. W. Belcher and K. B. MacGregor. Magnetic acceleration of winds from solar-type stars. , 210:498–507, December 1976. **2**
- [11] A. Bellini, L. R. Bedin, B. Pichardo, E. Moreno, C. Allen, G. Piotto, and J. Anderson. Absolute proper motion of the Galactic open cluster M 67. , 513:A51, April 2010. **10**
- [12] Timothy D. Brandt and Chelsea X. Huang. The Age and Age Spread of the Praesepe and Hyades Clusters: a Consistent,  $\sim 800$  Myr Picture from Rotating Stellar Models. , 807(1):24, July 2015. **10**
- [13] L. Breiman. Random forests. *Machine Learning*, 5:5–32, 2001. **16**
- [14] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984. **16**
- [15] A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, L. Eyer, and et al. Gaia data release 2. *Astronomy Astrophysics*, 616:A1, Aug 2018. **6**
- [16] T. Ceillier, J. van Saders, R. A. García, T. S. Metcalfe, O. Creevey, S. Mathis, S. Mathur, M. H. Pinsonneault, D. Salabert, and J. Tayar. Rotation periods and seismic ages of KOIs - comparison with stars without detected planets from Kepler observations. , 456(1):119–125, February 2016. **9**
- [17] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. **17**
- [18] Jason L. Curtis, Marcel A. Agüeros, Eric E. Mamajek, Jason T. Wright, and Jeffrey D. Cummings. TESS Reveals that the Nearby Pisces-Eridanus Stellar Stream is only 120 Myr Old. , 158(2):77, August 2019. **5**
- [19] Payel Das and Jason L Sanders. MADE: a spectroscopic mass, age, and distance estimator for red giant stars with Bayesian machine learning. *Monthly Notices of the Royal Astronomical Society*, 484(1):294–304, 10 2018. **6**
- [20] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks, 2017. **xxv, xxvi, 47, 48, 49, 53**
- [21] R. A. García, T. Ceillier, D. Salabert, S. Mathur, J. L. van Saders, M. Pinsonneault, J. Ballot, P. G. Beck, S. Bloemen, T. L. Campante, G. R. Davies, Jr. do Nascimento, J. D., S. Mathis, T. S. Metcalfe, M. B. Nielsen, J. C. Suárez, W. J. Chaplin, A. Jiménez, and C. Karoff. Rotation and magnetism of Kepler pulsating solar-like stars. Towards asteroseismically calibrated age-rotation relations. , 572:A34, December 2014. **9**

- [22] Mark S. Giampapa, Jeffrey C. Hall, Richard R. Radick, and Sallie L. Baliunas. A Survey of Chromospheric Activity in the Solar-Type Stars in the Open Cluster M67. , 651(1):444–461, November 2006. 10
- [23] Guillermo Gonzalez. Variability among stars in the M 67 field from Kepler/K2-Campaign-5 light curves. , 459(1):1060–1068, June 2016. 10
- [24] Tyler A. Gordon, James R. A. Davenport, Ruth Angus, Daniel Foreman-Mackey, Eric Agol, Kevin R. Covey, Marcel Agüeros, and David Kipping. Stellar Rotation in the K2 Sample: Evidence for Broken Spindown. *arXiv e-prints*, page arXiv:2101.07886, January 2021. 6
- [25] Vikrant V Jadhav, Clara M Pennock, Annapurni Subramaniam, Ram Sagar, and Prasanta Kumar Nayak. Uocs – iii. uvit catalogue of open clusters with machine learning-based membership using gaia edr3 astrometry. *Monthly Notices of the Royal Astronomical Society*, 503(1):236–253, Jan 2021. 7
- [26] Kenneth Janes, Sydney A. Barnes, Søren Meibom, and Sadia Hoq. NGC 6811: An Intermediate-age Cluster in the Kepler Field. , 145(1):7, January 2013. 10
- [27] R. D. Jeffries, Tim Naylor, N. J. Mayne, Cameron P. M. Bell, and S. P. Littlefair. A lithium depletion boundary age of 22 Myr for NGC 1960. , 434(3):2438–2450, September 2013. 10
- [28] Steven D. Kawaler. Angular Momentum Loss in Low-Mass Stars. , 333:236, October 1988. 2
- [29] Leonid Kitchatinov and Alexander Nepomnyashchikh. How supercritical are stellar dynamos, or why do old main-sequence dwarfs not obey gyrochronology? , 470(3):3124–3130, September 2017. 6
- [30] Robert P. Kraft. Studies of Stellar Rotation. V. The Dependence of Rotation on Age among Solar-Type Stars. , 150:551, November 1967. 5
- [31] Y. LeCun, L. Bottou, G. Orr, and K.-R. Müller. *Efficient BackProp*, pages 9–48. Springer Berlin Heidelberg, 2012. 17
- [32] Diego Lorenzo-Oliveira, Jorge Meléndez, Jhon Yana Galarza, Geisa Ponte, Leonardo A. dos Santos, Lorenzo Spina, Megan Bedell, Iván Ramírez, Jacob L. Bean, and Martin Asplund. Constraining the evolution of stellar rotation using solar twins. , 485(1):L68–L72, May 2019. 6
- [33] Eric E. Mamajek and Lynne A. Hillenbrand. Improved Age Estimation for Solar-Type Dwarfs Using Activity-Rotation Diagnostics. , 687(2):1264–1293, November 2008. 5

- [34] Tsevi Mazeh, Hagai B. Perets, Amy McQuillan, and Eyal S. Goldstein. Photometric Amplitude Distribution of Stellar Rotation of KOIs—Indication for Spin-Orbit Alignment of Cool Stars and High Obliquity for Hot Stars. , 801(1):3, March 2015. 9
- [35] A. McBride, R. Lingg, M. Kounkel, B. Hutchinson, K. Covey, E. Howard, and R. Olney. Predicting Young Stellar Ages with Machine Learning. In *American Astronomical Society Meeting Abstracts #236*, volume 236 of *American Astronomical Society Meeting Abstracts*, page 331.04, June 2020. 6
- [36] A. McQuillan, T. Mazeh, and S. Aigrain. Stellar Rotation Periods of the Kepler Objects of Interest: A Dearth of Close-in Planets around Fast Rotators. , 775(1):L11, September 2013. 9
- [37] A. McQuillan, T. Mazeh, and S. Aigrain. Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample. , 211(2):24, April 2014. 9
- [38] Søren Meibom, Sydney A. Barnes, David W. Latham, Natalie Batalha, William J. Borucki, David G. Koch, Gibor Basri, Lucianne M. Walkowicz, Kenneth A. Janes, Jon Jenkins, Jeffrey Van Cleve, Michael R. Haas, Stephen T. Bryson, Andrea K. Dupree, Gabor Furesz, Andrew H. Szentgyorgyi, Lars A. Buchhave, Bruce D. Clarke, Joseph D. Twicken, and Elisa V. Quintana. The Kepler Cluster Study: Stellar Rotation in NGC 6811. , 733(1):L9, May 2011. 10
- [39] Søren Meibom, Sydney A. Barnes, Imants Platais, Ronald L. Gilliland, David W. Latham, and Robert D. Mathieu. A spin-down clock for cool stars from observations of a 2.5-billion-year-old cluster. , 517(7536):589–591, January 2015. 10
- [40] L. Mestel. Magnetic braking by a stellar wind-I. , 138:359, January 1968. 2
- [41] L. Mestel and H. C. Spruit. On magnetic braking of late-type stars. , 226:57–66, May 1987. 2
- [42] Travis S. Metcalfe and Ricky Egeland. Understanding the Limitations of Gyrochronology for Old Field Stars. , 871(1):39, January 2019. 3, 6
- [43] Travis S. Metcalfe and Jennifer van Saders. Magnetic Evolution and the Disappearance of Sun-Like Activity Cycles. , 292(9):126, September 2017. 6
- [44] Andy Moya, Federico Zuccarino, William J. Chaplin, and Guy R. Davies. Empirical Relations for the Accurate Estimation of Stellar Masses and Radii. , 237(2):21, August 2018. 10
- [45] López-Sastre Roberto Moya Andrés and Recio-Martínez Jarmi. Ai for dating stars: a benchmarking study for gyrochronology, 2021. 60, 61

- [46] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms, 2018. 60
- [47] M. B. Nielsen, L. Gizon, H. Schunker, and C. Karoff. Rotation periods of 12 000 main-sequence Kepler stars: Dependence on stellar spectral type and comparison with  $v \sin i$  observations. , 557:L10, September 2013. 9
- [48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 24
- [49] Carl Eduard Rasmussen and Christopher K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. 17
- [50] L. M. Rebull, J. R. Stauffer, L. A. Hillenbrand, A. M. Cody, J. Bouvier, D. R. Soderblom, M. Pinsonneault, and L. Hebb. Rotation of Late-type Stars in Praesepe with K2. , 839(2):92, April 2017. 10
- [51] Timo Reinhold, Ansgar Reiners, and Gibor Basri. Rotation and differential rotation of active Kepler stars. , 560:A4, December 2013. 9
- [52] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22:400–407, 1951. 17
- [53] Jason L Sanders and Payel Das. Isochrone ages for  $\sim 3$  million stars with the second Gaia data release. *Monthly Notices of the Royal Astronomical Society*, 481(3):4093–4110, 2018. **xxi**, 6
- [54] E. Schatzman. A theory of the role of magnetic activity during star formation. *Annales d’Astrophysique*, 25:18, February 1962. 2
- [55] Aldo Serenelli, Jennifer Johnson, Daniel Huber, Marc Pinsonneault, Warrick H. Ball, Jamie Tayar, Victor Silva Aguirre, Sarbani Basu, Nicholas Troup, Saskia Hekker, Thomas Kallinger, Dennis Stello, Guy R. Davies, Mikkel N. Lund, Savita Mathur, Benoit Mosser, Keivan G. Stassun, William J. Chaplin, Yvonne Elsworth, Rafael A. García, Rasmus Handberg, Jon Holtzman, Fred Hearty, D. A. García-Hernández, Patrick Gaulme, and Olga Zamora. The First APOKASC Catalog of Kepler Dwarf and Subgiant Stars. , 233(2):23, December 2017. 9
- [56] V. Silva Aguirre, G. R. Davies, S. Basu, J. Christensen-Dalsgaard, O. Creevey, T. S. Metcalfe, T. R. Bedding, L. Casagrande, R. Handberg, M. N. Lund, P. E. Nissen, W. J. Chaplin, D. Huber, A. M. Serenelli, D. Stello, V. Van Eylen, T. L. Campante, Y. Elsworth, R. L. Gilliland, S. Hekker, C. Karoff, S. D. Kawaler, H. Kjeldsen, and M. S. Lundkvist. Ages and fundamental properties of Kepler exoplanet host stars from asteroseismology. , 452(2):2127–2148, September 2015. 9

- [57] Víctor Silva Aguirre, Mikkel N. Lund, H. M. Antia, Warrick H. Ball, Sarbani Basu, Jørgen Christensen-Dalsgaard, Yveline Lebreton, Daniel R. Reese, Kuldeep Verma, Luca Casagrande, Anders B. Justesen, Jakob R. Mosumgaard, William J. Chaplin, Timothy R. Bedding, Guy R. Davies, Rasmus Handberg, Günter Houdek, Daniel Huber, Hans Kjeldsen, David W. Latham, Timothy R. White, Hugo R. Coelho, Andrea Miglio, and Ben Rendle. Standing on the Shoulders of Dwarfs: the Kepler Asteroseismic LEGACY Sample. II. Radii, Masses, and Ages. , 835(2):173, February 2017. 9
- [58] A. Skumanich. Time Scales for Ca II Emission Decay, Rotational Braking, and Lithium Depletion. , 171:565, February 1972. 2
- [59] David Soderblom. Astrophysics: Stellar clocks. , 517(7536):557–558, January 2015. 2
- [60] David R. Soderblom. The Ages of Stars. , 48:581–629, September 2010. 1, 2
- [61] Jennifer L. van Saders, Tugdual Ceillier, Travis S. Metcalfe, Victor Silva Aguirre, Marc H. Pinsonneault, Rafael A. García, Savita Mathur, and Guy R. Davies. Weakened magnetic braking as the origin of anomalously rapid rotation in old field stars. , 529(7585):181–184, January 2016. 3, 5
- [62] Jennifer L. van Saders and Marc H. Pinsonneault. Fast Star, Slow Star; Old Star, Young Star: Subgiant Rotation as a Population and Stellar Physics Diagnostic. , 776(2):67, October 2013. 5
- [63] Jennifer L. van Saders, Marc H. Pinsonneault, and Mauro Barbieri. Forward Modeling of the Kepler Stellar Rotation Period Distribution: Interpreting Periods from Mixed and Biased Stellar Populations. , 872(2):128, February 2019. 6
- [64] Edmund J. Weber and Jr. Davis, Leverett. The Angular Momentum of the Solar Wind. , 148:217–227, April 1967. 2
- [65] David H. Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–259, 1992. 18
- [66] Huaxiu Yao, Linjun Zhang, and Chelsea Finn. Meta-learning with fewer tasks through task interpolation, 2021. 60
- [67] J. Zorec and F. Royer. Rotational velocities of A-type stars. IV. Evolution of rotational velocities. , 537:A120, January 2012. 5