

Modelos de Clasificación de Fases de HEA

Diabb Zegpi D.

Metodología

El producto del proceso de ciencia de datos es un modelo predictivo para las fases de las aleaciones de alta entropía (HEA), y las etapas que constituyen tal proceso son la extracción, manipulación, transformación y visualización de datos, modelización (aprendizaje automático) y comunicación de resultados (Wickham 2014). Este proceso es una adaptación moderna del proceso KDD (Fayyad, Piatetsky-Shapiro, y Smyth 1996), que utiliza principios de desarrollo de software *tidy*, adoptados por la comunidad de científicos de datos que utilizan el lenguaje de programación R. La Figura 1 ilustra el proceso de ciencia de datos, desde la etapa de obtención de datos (**Import**) hasta la comunicación de resultados (**Communicate**), incluyendo el ciclo de entrenamiento y optimización de modelos (ver Wickham y Grolemund 2016, cap. 2).

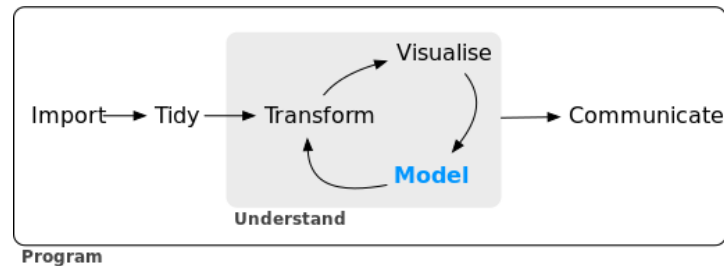


Figura 1: El proceso de ciencia de datos descrito por Wickham y Grolemund (2016).

Software

El software empleado en el análisis es de código abierto y está basado en el lenguaje de programación R (Team et al. 2013). R fue diseñado especialmente para análisis de datos y modelización (Kuhn y Silge 2022, cap. 1). Las estructuras de datos, gramática y filosofía de diseño del código usado en este trabajo son proveídas por la colección de paquetes para manipulación y visualización de datos denominada *tidyverse* (Wickham et al. 2019), y la colección de paquetes de aprendizaje automático llamada *tidymodels* (Kuhn y Wickham 2020). Las versiones del software empleado se exponen en la Tabla 1.

Tabla 1: Software y versiones.

Software	Versión
R	4.2.1
tidyverse	1.3.2
tidymodels	1.0.0

Referencias

- Fayyad, Usama, Gregory Piatetsky-Shapiro, y Padhraic Smyth. 1996. «From data mining to knowledge discovery in databases». *AI magazine* 17 (3): 37-37.
- Kuhn, Max, y Julia Silge. 2022. *Tidy Modeling with R*. " O'Reilly Media, Inc."
- Kuhn, Max, y Hadley Wickham. 2020. «Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles». *Boston, MA, USA.*[(accessed on 10 December 2020)].
- Team, R Core et al. 2013. «R: A language and environment for statistical computing».
- Wickham, Hadley. 2014. «Tidy Data». *Journal of Statistical Software* 59 (10): 1-23. <https://doi.org/10.18637/jss.v059.i10>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. «Welcome to the Tidyverse». *Journal of open source software* 4 (43): 1686.
- Wickham, Hadley, y Garrett Golemund. 2016. *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc."