

Si en entrenamiento tenemos un atributo que tiene por ejemplo un 30% de datos perdidos, y recuperamos información, y en el test tenemos ese mismo atributo con todos los datos perdidos, ¿que se suele hacer?

La misma estrategia que se haya seguido en entrenamiento se tiene que seguir en test.

Por ejemplo, si se ha remplazado por la media calculada sobre el conjunto de entrenamiento, cuando te llega un dato perdido en test, hay que remplazarlo también por la media (pero calculada sobre el conjunto de entrenamiento).

En el caso de sustituir por medias lo suyo es hacer la sustitución teniendo en cuenta la clase a la que pertenece el patrón al cual se le imputa el valor perdido. Por ejemplo, para patrones de la clase A, reemplazar por la media de ese atributo de todos los patrones que sean de clase A, y así sucesivamente con el resto de las clases.