

Rúbrica para la evaluación del trabajo de aprendizaje de modelos de decisión

Preprocesado del conjunto de datos (50 puntos)

- Trabajar con varias versiones del conjunto de datos:
 - Construcción razonada de varias versiones del conjunto de datos.
 - Evaluación del resto del trabajo en varias versiones del conjunto de datos.
 - Análisis comparativo de estas versiones en base a los modelos evaluados.

Hasta 16 puntos, dependiendo del número de versiones realizadas del conjunto de datos (factor multiplicativo, una versión 1x, dos o tres versiones 2x y cuatro o más versiones 3x), su uso a lo largo de todo el trabajo (1 punto por su uso en cada tipo de modelo: knn, árboles, modelos lineales y redes neuronales, hasta 4 puntos) y el análisis comparativo de estas (juicio de valor: sumando extra hasta 4 puntos).

- Análisis del conjunto de datos:
 - Estudio de valores ausentes.
 - Estudio de parámetros estadísticos: media, mediana, moda.
 - Análisis del tipo de características: numéricas-discretos, numéricas-continuos, categóricas.
 - Estudio de los valores de clasificación.

Hasta 4 puntos, uno por cada apartado.

- Incorporación del tratamiento de valores ausentes a las versiones del conjunto de datos.
 - Imputación de datos: relleno de valores ausentes.
 - Eliminación de datos con valores ausentes.
 - Eliminación de características.

Hasta 3 puntos, uno por cada apartado.

- Incorporación de la modificación de características a las versiones del conjunto de datos.
 - Label Encoding
 - One Hot Encoding
 - Discretizado
 - Normalización
 - Estandarización

Hasta 15 puntos dependiendo de la cantidad de métodos usados. Un único método 1 punto, dos métodos 3 puntos, tres métodos 6 puntos, cuatro métodos 10 puntos, los cinco métodos 15 puntos.

- Incorporación de la reducción del número de características a las versiones del conjunto de datos.
 - Reducción de características basada en modelos
 - Reducción recursiva de características
 - Análisis de componentes principales

Hasta 6 puntos dependiendo de la cantidad de métodos usados. Un único método 1 punto, dos métodos 3 puntos, tres métodos 6 puntos.

- Incorporación de la modificación del conjunto de ejemplos a las versiones del conjunto de datos.
 - Eliminación de muestras
 - Over-sampling
 - Under-sampling

Hasta 6 puntos dependiendo de la cantidad de métodos usados. Un único método 1 punto, dos métodos 3 puntos, tres métodos 6 puntos.

Construcción de modelos de decisión (46 puntos)

- Aprendizaje basado en instancias:

- Knn
- Ajuste de parámetros
- Documentación y análisis razonado de los resultados

Hasta 6 puntos. Construcción de modelos (1 punto), ajuste de parámetros (1 punto), documentación y análisis razonado de los resultados (juicio de valor: hasta 4 puntos).

- Árboles de decisión:

- Grado de dispersión
- Poda a priori
- Poda a posteriori
- Documentación y análisis razonado de los resultados

Hasta 10 puntos. Construcción de modelos (1 punto), análisis comparativo modificando el grado de dispersión (1 punto), poda a priori (dependiendo del número de factores considerados, 1 punto por un factor, 2 puntos por dos o tres factores, 3 puntos por cuatro o más factores), poda a posteriori (1 punto), documentación y análisis razonado de los resultados (juicio de valor: hasta 4 puntos).

- Ensemble learning con árboles:

- Random forest
- Gradient boosting
- Ajuste de parámetros
- Documentación y análisis razonado de los resultados

Hasta 10 puntos. Construcción de modelos (una técnica 1 punto, las dos técnicas 3 puntos), ajuste de parámetros (factor multiplicativo 2x por la puntuación de los modelos en los que se usa), documentación y análisis razonado de los resultados (juicio de valor: hasta 4 puntos).

- Modelos lineales:

- Regresión logística
- Máquinas de vectores soporte
- Ajuste de parámetros
- Documentación y análisis razonado de los resultados

Hasta 10 puntos. Construcción de modelos (una técnica 1 punto, las dos técnicas 3 puntos), ajuste de parámetros (factor multiplicativo 2x por la puntuación de los modelos en los que se usa), documentación y análisis razonado de los resultados (juicio de valor: hasta 4 puntos).

- Redes neuronales:

- Redes neuronales multicapa en scikit-learn
- Redes neuronales multicapa en keras
- Ajuste de parámetros
- Documentación y análisis razonado de los resultados

Hasta 10 puntos. Construcción de modelos (una técnica 1 punto, las dos técnicas 3 puntos), ajuste de parámetros (factor multiplicativo 2x por la puntuación de los modelos en los que se usa), documentación y análisis razonado de los resultados (juicio de valor: hasta 4 puntos).

Evaluación de modelos (45 puntos)

- Construcción equilibrada de modelos
 - Ajuste de hiperparámetros
 - Validación cruzada

Hasta 15 puntos. Para cada tipo de modelo (knn, árboles de decisión, regresión logística, máquinas de vectores soporte, redes neuronales) se valorará con 1 punto el aplicar alguna de las técnicas anteriores y 3 puntos aplicarlas las dos.

- Evaluación del rendimiento:
 - Matriz de confusión
 - Informe de clasificación
 - Documentación y análisis razonado de los resultados

Hasta 20 puntos. Para cada tipo de modelo (knn, árboles de decisión, regresión logística, máquinas de vectores soporte, redes neuronales) se valorará con 1 punto el aplicar alguna de las técnicas anteriores y 3 puntos aplicarlas las dos. Análisis razonado de los resultados (juicio de valor: hasta 5 puntos)

- Curvas de rendimiento:
 - Curva PR
 - Curva ROC
 - Documentación y análisis razonado de los resultados

Hasta 20 puntos. Para cada tipo de modelo (knn, árboles de decisión, regresión logística, máquinas de vectores soporte, redes neuronales) se valorará con 1 punto el aplicar alguna de las técnicas anteriores y 3 puntos aplicarlas las dos. Análisis razonado de los resultados (juicio de valor: hasta 5 puntos)

Presentación y defensa (59 puntos)

- Organización de la hoja de trabajo.
Juicio de valor: **hasta 9 puntos.**
- Documentación general de la hoja de trabajo.
Juicio de valor: **hasta 15 puntos.**
- Evaluación de la hoja de trabajo.
Juicio de valor: **hasta 5 puntos.**
- Presentación razonada de los resultados.
Juicio de valor: **hasta 10 puntos.**
- Respuesta a cuestiones planteadas durante la defensa.
Juicio de valor: **hasta 20 puntos.**

La valoración será la cantidad necesaria para alcanzar un total de 200 puntos, en los casos de valoración máxima de los apartados anteriores. En caso de que alguno de los subapartados anteriores no sea aplicable al conjunto de datos considerado (por ejemplo tratamiento de valores ausentes) se incrementará la puntuación de esta parte en los dos últimos apartados (presentación razonada y respuesta a cuestiones).