Práctica Final

# Rúbrica General

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item** | **Muy bien** | **Bien** | **Insuficiente** | **Muy deficiente** |
| **Identificación del problema** | El tipo de problema a resolver se identifica correctamente y resuelve un problema concreto y bien definido. | El tipo de problema a resolver se identifica correctamente y la variable a predecir está definida pero vagamente justificada. | El tipo de problema a abordar no está claro o la variable a predecir no es muy interesante para el dataset seleccionado. | El tipo de problema no queda nada claro ni la variable a predecir tampoco. |
| **Análisis del problema** | El análisis del dataset es adecuado y se justifica porque la variable objetivo está correlacionada con el resto de características.  Se han usado todos los plots necesarios para la justificación. | El análisis del dataset es adecuado y se justifica porque la variable objetivo está correlacionada con el resto de características.  Se echa en falta algún plot necesario para la justificación. | El análisis del dataset no es adecuado o no se justifica correctamente porque la variable objetivo está correlacionada con el resto de características. | No se justifica porque la variable objetivo está relacionada con el resto de características. |
| **Limpieza del dataset** | Se eliminan todos los NaN del dataset sin perder información relevante infiriendo de la forma más precisa posible. | Se eliminan todos los Nan del dataset usando las técnicas adecuadas sin perder información | Se eliminan todos los NaN del dataset perdiendo demasiada información. | No se eliminan todos los NaN. |
| **Características nuevas** | Se han extraído características nuevas cuando era necesario y se ha comprobado que ayudan a mejorar la precisión final. | Se han extraído características nuevas cuando era necesario pero no han sido correctamente justificadas. | Se han extraído características nuevas poco útiles o faltan algunas muy evidentes | No se han extraído características nuevas. |
| **Análisis del dataset** | Se usan técnicas que explican la varianza de los datos y se comprueba que características no aportan nada | Se usan técnicas que explican la varianza de los datos pero no se detectan las características que no aportan nada | Se usan técnicas para explicar la varianza de los datos, pero la justificación de los resultados es errónea. | No se usan técnicas que expliquen la varianza de los datos |
| **Seleccion de caracteristicas** | Se ha justificado la selección de características para predecir la variable objetivo y se ha comprobado usando algún algoritmo de feature selection diferente a PCA. | Se ha justificado la selección de características para predecir la variable objetivo y se ha comprobado usando únicamente PCA. | La justificación de la selección de las características para predecir la variable objetivo es vaga o errónea. | No se justifica la selección de las características. |
| **Reducción de dimensionalidad (PCA y MDS)** | Se ha aplicado un método de reducción de dimensionalidad y se analiza cuánta varianza puede llegar a explicar cada componente (PCA).  Además se comprueba los efectos de transformar los datos sobre un modelo de ML comparando respecto los datos originales. | Se ha aplicado un método de reducción de dimensionalidad pero no se analiza cuánta varianza puede llegar a explicar cada componente O no se comprueba los efectos de transformar los datos sobre un modelo de ML comparando respecto los datos originales. | Se ha aplicado un método de reducción de dimensionalidad pero el análisis de varianza es erróneo o no queda claro que los efectos sean los mencionados. | No se ha aplicado un método de reducción de dimensionalidad. |
| **Modelos de ML** | Se ha probado más de un modelo de ML y son adecuados para resolver el problema.  Se han ajustado los hyperparameters correctamente.  Ha tenido en cuenta usar CV y se ha validado con un juego de datos de test una vez obtenido el modelo final.  No hay signos de overfitting ni underfitting. | Se ha probado más de un modelo de ML y son adecuados para resolver el problema.  Ha tenido en cuenta usar CV y se ha validado con un juego de datos de test una vez obtenido el modelo final. | Únicamente se ha probado un único método de clasificación o no se ha hecho CV o varios de los métodos empleados no son adecuados para el problema. | No se ha empleado un modelo de ML o el modelo empleado no es adecuado para el problema. |
| **Análisis de las predicciones obtenidas** | Se justifica correctamente el accuracy obtenido gracias a todo el trabajo previo | Se justifica correctamente el accuracy obtenido pero algún detalle es erróneo | No Se justifica correctamente el accuracy obtenido | No se justifica el accuracy obtenido. |
| **Clustering** | En el caso de un problema de clasificación se han analizado los posibles clusters que formaban el problema y si se corresponden con el número de etiquetas propuesto por el dataset.  Se ha utilizado alguna métrica como el coste o calinski para justificar el número de clusters obtenidos. | En el caso de un problema de clasificación se han analizado los posibles clusters que formaban el problema.  Se ha utilizado alguna métrica como el coste o calinski para justificar el número de clusters obtenidos. | Se ha aplicado una técnica de clustering pero el análisis no es correcto. | No se ha aplicado una técnica de clustering. |
| **SVM (modelo obligatorio)** | Se han probado los kernels lineal, polinómico, RBF y se ha ejecutado un grid search para encontrar la configuración de parámetros que mejor accuracy proporciona. | Se han probado los kernels lineal, polinómico, RBF y se ha buscado una configuración de parámetros que mejore los resultados. | Alguno de los kernels no se ha probado y se han dejado los parámetros por defecto. | No se ha usado SVM. |
| **Neural Network (modelo opcional)** | Se ha creado una arquitectura que permite con los parámetros de loss, optimización y selección de la función de activación de la última capa obtener una accuracy similar o mejor a los modelos de ML tradicionales y están bien justificados. | Se ha creado una arquitectura que permite con los parámetros de loss, optimización y selección de la función de activación de la última capa obtener una accuracy similar o mejor a los modelos de ML tradicionales. | Se ha utilizado una Neural Network pero faltan muchos elementos por justificar o los resultados son mucho peores. | No se ha usado una Neural Network. |

Evaluación

|  |  |  |
| --- | --- | --- |
| Parte | Porcentaje | Criterios de calidad |
| Análisis previos dataset | 20% | Por qué se ha elegido este dataset y por qué es relevante |
| Limpieza dataset | 20% | Los datos han sido limpiados, normalizados y definido las variables correctamente |
| Elección de modelo | 10% | El modelo elegido permite modelar los resultados a predecir |
| Modelos aplicados | 20% | Se han aplicado correctamente los modelos y funciones |
| Predicciones | 20% | Los resultados tienen sentido, aportan valor y están bien representados |
| Evaluación / Métricas | 10% | Los resultados superan unos criterios de calidad y tests para garantizar que son correctos |