Resultados empíricos Comunidad Valenciana

* **Preprocesamiento de los datos.**

Inicialmente el conjunto de datos consta de 97 observaciones, de las cuales 15 son etiquetadas como eficientes según el modelo DEA-BCC, y los 82 restantes son etiquetadas como ineficientes. Esto hace que el dataset tenga un 15,46% de observaciones etiquetadas como eficientes y 84,53% como ineficientes.

Se propone un balanceo mediante la generación de unidades sintéticas. El dataset final para entrenar los modelos consta de 261 observaciones de las cuales 82 son eficientes y 179 tienen etiqueta ineficiente (31,42%-68,58%).

* **Modelos a entrenar (validación cruzada con 5 folds).**

SVM con kernel polinómica. El conjunto de hiperparámetros a tunear es:

"degree" = (1, 2, 3, 4, 5),

"scale" = (0.001, 0.1, 1, 10, 100),

"C" = (0.001, 0.1, 1, 10, 100)

*Hiperparámetros seleccionados: degree = 1, scale = 100 y C = 0.1*.

NN con una capa oculta. El conjunto de hiperparámetros a tunear es:

"size" = (1, 5, 10, 20),

"decay" = (0, 0.1, 0.01, 0.001, 0,0001)

*Hiperparámetros seleccionados: size = 5 y decay = 0.1*.

* **Métricas de rendimiento.**

**SVM**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sensitivity | Specificity | Pos Pred Value | Neg Pred Value | Precision | Recall |
| 0,72 | 0,97 | 0,92 | 0,88 | 0,92 | 0,72 |
|  |  |  |  |  |  |
| F1 | Prevalence | Detection Rate | Detection Prevalence | Balanced Accuracy |
| 0,81 | 0,31 | 0,23 | 0,25 | 0,85 |

**NN**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sensitivity | Specificity | Pos Pred Value | Neg Pred Value | Precision | Recall |
| 0,89 | 0,96 | 0,90 | 0,95 | 0,90 | 0,89 |
|  |  |  |  |  |  |
| F1 | Prevalence | Detection Rate | Detection Prevalence | Balanced Accuracy |
| 0,90 | 0,31 | 0,28 | 0,31 | 0,92 |

* Importancia de las variables

|  |  |  |
| --- | --- | --- |
| Variables | SVM | NN |
| **Inputs** | **0.089** | **0.244** |
| Total assets | 0.034 | 0.155 |
| Employees | 0 | 0.019 |
| Fixed assets | 0.007 | 0.051 |
| Personal expenses | 0.048 | 0.019 |
|  |  |  |
| **Output** | **0.911** | **0.755** |
| Operating income | 0.911 | 0.755 |

* Dataset de outputs según probabilidad deseada