Modelos y resultados

Predicción del ingreso

En esta sección la predicción de la clasificación del hogar en pobre o no se realiza mediante una primera predicción del ingreso por individuo . Este ingreso se agrupa por hogar , se ajusta por la cantidad de personas que hay en la unidad de gasto y por un arriendo estimado cuando no se está pagando arriendo; luego, se compara la línea de pobreza –estimada de acuerdo con la valoración de la canasta de bienes necesaria para satisfacer ciertas necesidades–, como se muestra a continuación:

La predicción del ingreso se realiza utilizando diferentes modelos y especificaciones. Respecto a modelos, se utilizaron modelos de regularización, de *random forest* y de redes; en cuanto a especificaciones se utilizaron las siguientes variantes:

1. Especificación 1: , siendo X todas las variables seleccionadas y descritas en la sección de datos, que son características de los individuos.
2. Especificación 2: ,, siendo todas las variables características de los individuos distintas de las descritas explícitamente en la ecuación.
3. Especificación 3: log(, siendo X todas las variables seleccionadas y descritas en la sección de datos, que son características de los individuos.
4. Especificación 4: , siendo todas las variables características de los individuos distintas de las descritas explícitamente en la ecuación.
5. Especificación 35 , siendo todas las variables características de los individuos distintas de las descritas explícitamente en la ecuación.

La evaluación de los modelos se realizó mediante el cálculo del *F1-score* en el conjunto de validación, a partir de una matriz de confusión que se construía con la predicción y con los valores reales del conjunto en cuestión. En la siguiente gráfica se muestran los mejores resultados del *F1-score* para los modelos evaluados:

Figura . Mejores resultados F1- score de predicción del parámetro de Pobre con la predicción del ingreso

Para el detalle de los modelos que se simularon, véase el anexo de este reporte. Como puede verse, se exploraron diferentes valores de hiperparámetros para lograr ver cómo esto afectaba la predicción. Muchos de los modelos no fueron sometidos, teniendo en cuenta que tenían resultados inferiores o similares a algunos que sí fueron cargados a Kaggle.

El mejor modelo para este enfoque fue el de redes neuronales, con la especificación 1. Sin embargo, este resultado es muy similar al obtenido con otras especificaciones y modelos. Así mismo, en Kaggle, no superó el valor de 0.56.

En redes neuronales se desarrollaron una cantidad mayor de variantes que en los otros modelos, teniendo en cuenta la cantidad de hiperparámetros que este tenía. Se identificó que el mejor modelo tenía dos capas intermedias, con tres su desempeño bajaba.

Otro aspecto importante es que, si bien el modelo de *Random Forest* tenía uno de los mejores resultados en F1-score, su valor en Kaggle fue de 0.44. Esto puede deberse a que su valor de *Recall* era muy inferior (0.55) a los de los modelos que tuvieron un valor en Kaggle más alto (entre 0.7 y 0.8).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID de modelo | Especificación | Modelo | Hiperparámetros | Accuracy | Precision | Recall | F1-score - Validation Data | F1-score - Test Data |
| 1 | Especificación 1 | Lasso | alfa = 1x10^-5 | 0.85 | 0.68 | 0.60 | 0.63 | 0.55 |
| 2 | Especificación 3 | Lasso | alfa = 1x10^-5 | 0.71 | 0.40 | 0.75 | 0.52 | 0.4 |
| 3 | Especificación 4 | Lasso | alfa = 1x10^-5 | 0.65 | 0.34 | 0.74 | 0.47 | No sometido |
| 4 | Especificación 1 | Ridge | alfa = 1x10^-5 | 0.86 | 0.69 | 0.59 | 0.63 | 0.55 |
| 5 | Especificación 1 | Ridge | alfa = 1x10^-10 | 0.86 | 0.69 | 0.59 | 0.63 | No sometido |
| 6 | Especificación 1 | Elastic Net | alfa = 1x10^-10; mixture = 0.5 | 0.85 | 0.68 | 0.60 | 0.63 | No sometido |
| 7 | Especificación 1 | Elastic Net | alfa = 1x10^-10; mixture = 0.75 | 0.85 | 0.68 | 0.60 | 0.63 | 0.55 |
| 8 | Especificación 1 | Random Forest | mtry = 7; min.node.size = 5 | 0.87 | 0.76 | 0.56 | 0.64 | 0.43 |
| 9 | Especificación 1 | Random Forest | mtry = 5; min.node.size = 5 | 0.86 | 0.77 | 0.51 | 0.61 | No sometido |
| 10 | Especificación 1 | Random Forest | mtry = 5; min.node.size = 5; Imputación de ceros (0) en EdadTrabajo = 0 | 0.86 | 0.77 | 0.51 | 0.62 | No sometido |
| 11 | Especificación 1 | Redes neuronales | Epocas: 500; Bache: 5112; Capa densa 1: 8; Activación 1: ReLu; Capa densa 2: 0; Activación 2: N/A; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.84 | 0.61 | 0.67 | 0.64 | No sometido |
| 12 | Especificación 1 | Redes neuronales | Epocas: 500; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 0; Activación 2: N/A; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.79 | 0.21 | 0.00 | 0.00 | No sometido |
| 13 | Especificación 1 | Redes neuronales | Epocas: 500; Bache: 5112; Capa densa 1: 8; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.83 | 0.59 | 0.70 | 0.64 | 0.56 |
| 14 | Especificación 1 | Redes neuronales | Epocas: 250; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.84 | 0.60 | 0.71 | 0.65 | 0.56 |
| 15 | Especificación 1 | Redes neuronales | Epocas: 500; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.84 | 0.61 | 0.68 | 0.64 | No sometido |
| 16 | Especificación 1 | Redes neuronales | Epocas: 250; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 16; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.83 | 0.59 | 0.67 | 0.63 | No sometido |
| 17 | Especificación 1 | Redes neuronales | Epocas: 200; Bache: 5112; Capa densa 1: 64; Activación 1: ReLu; Capa densa 2: 32; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.84 | 0.63 | 0.63 | 0.63 | No sometido |
| 18 | Especificación 1 | Redes neuronales | Epocas: 200; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 4; Activación 3: ReLu; Dropout: 0.5 | 0.81 | 0.53 | 0.80 | 0.64 | No sometido |
| 19 | Especificación 1 | Redes neuronales | Epocas: 500; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.83 | 0.58 | 0.72 | 0.64 | No sometido |
| 20 | Especificación 1 | Redes neuronales | Epocas: 500; Bache: 10224; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.84 | 0.60 | 0.72 | 0.65 | No sometido |
| 21 | Especificación 1 | Redes neuronales | Epocas: 750; Bache: 10224; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.84 | 0.60 | 0.69 | 0.64 | No sometido |
| 22 | Especificación 1 | Redes neuronales | Epocas: 500; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.25 | 0.84 | 0.63 | 0.63 | 0.63 | No sometido |
| 23 | Especificación 1 | Redes neuronales | Epocas: 250; Bache: 2556; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.25 | 0.84 | 0.62 | 0.64 | 0.63 | No sometido |
| 24 | Especificación 5 | Redes neuronales | Epocas: 500; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.32 | 0.24 | 1.00 | 0.38 | No sometido |
| 25 | Especificación 5 | Redes neuronales | Epocas: 100; Bache: 5112; Capa densa 1: 64; Activación 1: ReLu; Capa densa 2: 32; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.66 | 0.38 | 0.97 | 0.54 | No sometido |
| 26 | Especificación 5 | Redes neuronales | Epocas: 50; Bache: 5112; Capa densa 1: 64; Activación 1: ReLu; Capa densa 2: 32; Activación 2: ReLu; Capa densa 3: 16; Activación 3: ReLu; Dropout: 0.5 | 0.47 | 0.29 | 0.99 | 0.44 | No sometido |
| 27 | Especificación 4 | Redes neuronales | Epocas: 50; Bache: 5112; Capa densa 1: 64; Activación 1: ReLu; Capa densa 2: 32; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.66 | 0.38 | 0.96 | 0.55 | No sometido |
| 28 | Especificación 4 | Redes neuronales | Epocas: 50; Bache: 5112; Capa densa 1: 64; Activación 1: Sigmoid; Capa densa 2: 32; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.73 | 0.43 | 0.90 | 0.58 | No sometido |
| 29 | Especificación 4 | Redes neuronales | Epocas: 50; Bache: 5112; Capa densa 1: 64; Activación 1: Sigmoid; Capa densa 2: 32; Activación 2: Sigmoid; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.65 | 0.37 | 0.95 | 0.53 | No sometido |
| 30 | Especificación 4 | Redes neuronales | Epocas: 50; Bache: 5112; Capa densa 1: 64; Activación 1: ReLu; Capa densa 2: 32; Activación 2: Sigmoid; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.58 | 0.32 | 0.92 | 0.48 | No sometido |
| 31 | Especificación 4 | Redes neuronales | Epocas: 50; Bache: 5112; Capa densa 1: 64; Activación 1: tanh; Capa densa 2: 32; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.65 | 0.37 | 0.95 | 0.54 | No sometido |
| 32 | Especificación 2 | Redes neuronales | Epocas: 500; Bache: 5112; Capa densa 1: 16; Activación 1: ReLu; Capa densa 2: 8; Activación 2: ReLu; Capa densa 3: 0; Activación 3: N/A; Dropout: 0.5 | 0.84 | 0.61 | 0.69 | 0.65 | 0.56 |
| 33 | Especificación 2 | Ridge | alfa = 1x10^-10 | 0.86 | 0.69 | 0.59 | 0.63 | No sometido |