**Anexo 1. Análisis de Datos – Modelos de clasificación Base de datos training**

**Tabla 1**. Estadísticas descriptivas generales

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estadísticas por hogar | nbr.null | nbr.na | min | max | range | median | mean | std.dev |
| No. Personas | - | - | 1 | 28 | 27 | 3 | 3,30 | 1,78 |
| Li | - | - | 99544,84 | 131125,57 | 31580,72 | 121603,86 | 120415,70 | 7179,03 |
| Lp | - | - | 167222,48 | 303816,6902 | 136594,21 | 280028,7104 | 271604,40 | 33543,76 |
| No. cuartos por hogar | - | - | 1 | 18 | 17 | 3 | 3,39 | 1,21 |
| No. de personas por cuarto | - | - | 0,2 | 16 | 15,8 | 1,5 | 1,73 | 0,83 |
| Edad promedio | - | - | 5,67 | 102 | 96,33 | 33,5 | 37,45 | 16,88 |
| No. hijos | 55974 | - | - | 12 | 12 | 1 | 1,17 | 1,12 |
| No. de personas con trabajo formal | 93346 | - | - | 8 | 8 | - | 0,58 | 0,76 |
| Años educación promedio | 2605 | - | - | 27 | 27 | 10,5 | 10,30 | 4,10 |
| No. horas trabajadas promedio | 22138 | - | - | 130 | 130 | 44 | 39,18 | 19,65 |
| Ingreso total por hogar | 13220 | - | - | 85833333,33 | 85833333,33 | 1187717 | 1801440,60 | 2339945,38 |
| Participación en trabajo formal | 93346 | - | - | 1 | 1 | - | 0,34 | 0,43 |
| % personas con subsidio familiar | 80203 | - | - | 9 | 9 | 0,2 | 0,42 | 0,56 |
| Si alguna persona cuenta con segundo trabajo | 153588 | - | - | 1 | 1 | - | 0,07 | 0,25 |
| Si alguna persona cuenta con otros ingresos | 71127 | - | - | 1 | 1 | 1 | 0,57 | 0,50 |
| Si alguna persona cuenta con otros ingresos instituciones | 140035 | - | - | 1 | 1 | - | 0,15 | 0,36 |
| Tasa ocupación | 22138 | - | - | 1 | 1 | 0,5 | 0,58 | 0,34 |
| Tasa desempleo | 138670 | - | - | 1 | 1 | - | 0,09 | 0,23 |
| Tasa inactivas | 59582 | - | - | 1 | 1 | 0,33 | 0,35 | 0,33 |
| Tasa participación | 16498 | - | - | 1 | 1 | 0,67 | 0,65 | 0,33 |

Fuente: Elaboración propia

**Gráfico 1.** Ingresos promedio mes entre

hogares pobres y no pobres

Gráfico

Descripción generada automáticamente

Fuente: Elaboración propia

**Gráfico 2.** Box-Plot por número de personas

e ingresos en los hogares pobres y no pobres

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Fuente: Elaboración propia

**Gráfico 3.** Box-Plot por número de personas con trabajo formal y número de horas trabajadas en los hogares pobres y no pobres

**Gráfico 4.** Box-Plot por número de hijos e ingresos promedio en los hogares pobres y no pobres

Gráfico, Gráfico de dispersión

Descripción generada automáticamente Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Fuente: Elaboración propia Fuente: Elaboración propia

Las gráficas de caja muestran que el ingreso promedio por hogar depende del número de personas por hogar, se puede evidenciar que a mayor número de personas en el hogar, se presentan menores ingresos promedio hogar y aumenta el número de hogares pobres. Los hogares con menor número de personas reportan mayores ingresos, esta relación también se observa con el número de hijos promedio por hogar.

**Tabla 2**. Estadísticas descriptivas generales base de datos testing.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estadísticas por hogar | nbr.null | nbr.na | min | max | range | median | mean | std.dev |
| No. Personas | 0 | 0 | 1 | 22 | 21 | 3 | 3,312311087 | 1,79076566 |
| No. De personas por cuarto | 0 | 0 | 0,2 | 12 | 11,8 | 1,5 | 1,733690391 | 0,833106227 |
| Edad promedio | 0 | 0 | 7 | 101 | 94 | 33,4 | 37,43497851 | 16,87501555 |
| No. hijos | 22397 | 0 | 0 | 11 | 11 | 1 | 1,166651554 | 1,118659572 |
| No. de personas con trabajo formal | 38282 | 0 | 0 | 7 | 7 | 0 | 0,556945956 | 0,749621974 |
| Años educación promedio | 1075 | 0 | 0 | 27 | 27 | 10,4 | 10,20779081 | 4,097525066 |
| No. Horas trabajadas promedio | 8905 | 0 | 0 | 130 | 130 | 44 | 39,02967523 | 19,65453794 |
| % personas subsidio familiar | 32897 | 0 | 0 | 9 | 9 | 0,142857143 | 0,409732954 | 0,557371703 |
| Si alguna persona cuenta con segundo trabajo | 61678 | 0 | 0 | 1 | 1 | 0 | 0,067857575 | 0,251503241 |
| Si alguna persona cuenta con otros ingresos | 29200 | 0 | 0 | 1 | 1 | 1 | 0,558699069 | 0,496546217 |
| Si alguna persona cuenta con otros ingresos instituciones | 55770 | 0 | 0 | 1 | 1 | 0 | 0,157145448 | 0,363940597 |
| Tasa ocupación | 8905 | 0 | 0 | 1 | 1 | 0,5 | 0,579627289 | 0,338375389 |
| Tasa desempleo | 55445 | 0 | 0 | 1 | 1 | 0 | 0,089606273 | 0,227929196 |
| Tasa participación | 6651 | 0 | 0 | 1 | 1 | 0,666666667 | 0,644219633 | 0,326715732 |

Fuente: Elaboración propia

**Anexo 2.**

**Tabla 3**. Resultados modelos de clasificación

|  |  | **Validación cruzada - training** | | | | | | **Matriz de confusión fuera de muestra** | | | **Medidas fuera de muestra** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Modelo** | **Descripción** | **Cut-off** | **ROC** | **Sens** | **Espec** | **AUC** | **Kappa** |  | **No** | **Si** | **Sens** | **Espec** | **Accuracy** |
| **Logit** | Modelo 1 | 0,50 | 0,898 | 0,526 | 0,946 | 0,862 | 0,523 | Si | 3.154 | 3.450 | 0,5224 | 0,9475 | 0,8624 |
| No | 25.000 | 1.386 |
| Modelo 2 | 0,899 | 0,529 | 0,946 | 0,863 | 0,526 | Si | 3.126 | 3.478 | 0,5267 | 0,9461 | 0,8622 |
| No | 24.965 | 1.421 |
| **Logit rfThresh** | Modelo 1 | 0,21 |  | 0,825 | 0,803 | 0,897 |  | Si | 1.166 | 5.438 | 0,8234 | 0,8069 | 0,8102 |
| No | 21.290 | 5.096 |
| Modelo 2 | 0,21 |  | 0,825 | 0,805 | 0,897 |  | Si | 1.162 | 5.442 | 0,8240 | 0,8083 | 0,8114 |
| No | 21.327 | 5.059 |
| **Logit - Lasso** | Modelo 1 | 0,50 | 0,897 | 0,477 | 0,957 | 0,861 | 0,499 | Si | 3.479 | 3.125 | 0,4732 | 0,9562 | 0,8595 |
| No | 25.230 | 1.156 |
| Modelo 2 | 0,898 | 0,482 | 0,956 | 0,861 | 0,503 | Si | 3.435 | 3.169 | 0,4799 | 0,9556 | 0,8604 |
| No | 25.215 | 1.171 |
| **Logit - Lasso rfThresh** | Modelo 1 | 0,23 |  | 0,81 | 0,811 | 0,895 |  | Si | 1.244 | 5.360 | 0,8116 | 0,8153 | 0,8146 |
| No | 21.513 | 4.873 |
| Modelo 2 | 0,22 |  | 0,815 | 0,804 | 0,895 |  | Si | 1.204 | 5.400 | 0,8177 | 0,8110 | 0,8123 |
| No | 21.399 | 4.987 |
| **Logit - Ridge** | Modelo 1 | 0,50 | 0,898 | 0,525 | 0,947 | 0,863 | 0,524 | Si | 3.161 | 3.443 | 0,5214 | 0,9479 | 0,8625 |
| No | 25.010 | 1.376 |
| Modelo 2 | 0,899 | 0,528 | 0,947 | 0,863 | 0,526 | Si | 3.135 | 3.469 | 0,5253 | 0,9475 | 0,8630 |
| No | 25.000 | 1.386 |
| **Logit - Ridge rfThresh** | Modelo 1 | 0,21 |  | 0,827 | 0,801 | 0,896 |  | Si | 1.163 | 5.441 | 0,8239 | 0,8064 | 0,8099 |
| No | 21.278 | 5.108 |
| Modelo 2 | 0,21 |  | 0,826 | 0,803 | 0,897 |  | Si | 1.151 | 5.453 | 0,8257 | 0,8058 | 0,8098 |
| No | 21.263 | 5.123 |
| **Logit - ElasticNet** | Modelo 1 | 0,50 | 0,899 | 0,523 | 0,947 | 0,862 | 0,523 | Si | 3.168 | 3.436 | 0,5203 | 0,9478 | 0,8622 |
| No | 25.009 | 1.377 |
| Modelo 2 | 0,899 | 0,528 | 0,947 | 0,863 | 0,525 | Si | 3.146 | 3.458 | 0,5236 | 0,9469 | 0,8622 |
| No | 24.985 | 1.401 |
| **Logit - ElasticNet rfThresh** | Modelo 1 | 0,21 |  | 0,825 | 0,803 | 0,896 |  | Si | 1.175 | 5.429 | 0,8221 | 0,8077 | 0,8105 |
| No | 21.311 | 5.075 |
| Modelo 2 | 0,21 |  | 0,828 | 0,801 | 0,897 |  | Si | 1.141 | 5.463 | 0,8272 | 0,8041 | 0,8087 |
| No | 21.217 | 5.169 |
| **Logit - lasso  up sample** | Modelo 1 | 0,50 | 0,896 | 0,084 | 0,790 | 0,814 | 0,627 | Si | 1.092 | 5.512 | 0,8346 | 0,7928 | 0,8012 |
| No | 20.920 | 5.466 |
| Modelo 2 | 0,898 | 0,836 | 0,794 | 0,815 | 0,631 | Si | 1.099 | 5.505 | 0,8336 | 0,7969 | 0,8043 |
| No | 21.028 | 5.358 |
| **Logit - lasso  up sample rfThresh** | Modelo 1 | 0,52 |  | 0,818 | 0,801 | 0,894 |  | Si | 1.197 | 5.407 | 0,8187 | 0,8063 | 0,8088 |
| No | 21.275 | 5.111 |
| Modelo 2 | 0,51 |  | 0,826 | 0,795 | 0,895 |  | Si | 1.126 | 5.478 | 0,8295 | 0,8011 | 0,8068 |
| No | 21.137 | 5.249 |
| **logit - ridge  up sample** | Modelo 1 | 0,50 | 0,818 | 0,858 | 0,599 | 0,729 | 0,458 | Si | 921 | 5.683 | 0,8605 | 0,5954 | 0,6485 |
| No | 15.710 | 10.676 |
| Modelo 2 | 0,818 | 0,858 | 0,598 | 0,728 | 0,457 | Si | 921 | 5.683 | 0,8605 | 0,5954 | 0,6485 |
| No | 15.710 | 10.676 |
| **logit - ridge  up sample rfThresh** | Modelo 1 | 0,53 |  | 0,759 | 0,728 | 0,825 |  | Si | 1.683 | 4.921 | 0,7452 | 0,7226 | 0,7271 |
| No | 19.067 | 7.319 |
| Modelo 2 | 0,53 |  | 0,759 | 0,728 | 0,825 |  | Si | 1.683 | 4.921 | 0,7452 | 0,7226 | 0,7271 |
| No | 19.067 | 7.319 |
| **Logit - EN  up sample** | Modelo 1 | 0,50 | 0,897 | 0,839 | 0,789 | 0,814 | 0,627 | Si | 1.100 | 5.504 | 0,8334 | 0,7908 | 0,7993 |
| No | 20.865 | 5.521 |
| Modelo 2 | 0,889 | 0,839 | 0,774 | 0,806 | 0,612 | Si | 1.083 | 5.521 | 0,8360 | 0,7764 | 0,7884 |
| No | 20.487 | 5.899 |
| **Logit - EN  up sample rfThresh** | Modelo 1 | 0,53 |  | 0,813 | 0,808 | 0,895 |  | Si | 1.240 | 5.364 | 0,8122 | 0,8129 | 0,8128 |
| No | 21.450 | 4.936 |
| Modelo 2 | 0,52 |  | 0,818 | 0,793 | 0,889 |  | Si | 1.235 | 5.369 | 0,8130 | 0,7984 | 0,8013 |
| No | 21.066 | 5.320 |
| **Logit - lasso  down sample** | Modelo 1 | 0,50 | 0,894 | 0,836 | 0,787 | 0,811 | 0,623 | Si | 1.085 | 5.519 | 0,8357 | 0,7906 | 0,7996 |
| No | 20.861 | 5.525 |
| Modelo 2 | 0,895 | 0,835 | 0,789 | 0,812 | 0,625 | Si | 1.096 | 5.508 | 0,8340 | 0,7936 | 0,8017 |
| No | 20.940 | 5.446 |
| **Logit - lasso  down sample rfThresh** | Modelo 1 | 0,50 |  | 0,834 | 0,787 | 0,893 |  | Si | 1.092 | 5.512 | 0,8346 | 0,7922 | 0,8007 |
| No | 20.902 | 5.484 |
| Modelo 2 | 0,51 |  | 0,822 | 0,795 | 0,894 |  | Si | 1.141 | 5.463 | 0,8272 | 0,8004 | 0,8058 |
| No | 21.119 | 5.267 |
| **Logit - ridge  down sample** | Modelo 1 | 0,50 | 0,730 | 0,861 | 0,573 | 0,717 | 0,433 | Si | 885 | 5.719 | 0,8660 | 0,5644 | 0,6248 |
| No | 14.892 | 11.494 |
| Modelo 2 | 0,809 | 0,860 | 0,586 | 0,723 | 0,446 | Si | 909 | 5.695 | 0,8624 | 0,5813 | 0,6375 |
| No | 15.337 | 11.049 |
| **Logit - ridge  down sample rfThresh** | Modelo 1 | 0,50 |  | 0,849 | 0,606 | 0,739 |  | Si | 1.121 | 5.483 | 0,8303 | 0,5997 | 0,6459 |
| No | 15.824 | 10.562 |
| Modelo 2 | 0,53 |  | 0,737 | 0,738 | 0,818 |  | Si | 1.872 | 4.735 | 0,7167 | 0,7342 | 0,7307 |
| No | 19.372 | 7.014 |
| **Logit EN  down sample** | Modelo 1 | 0,50 | 0,887 | 0,838 | 0,772 | 0,805 | 0,609 | Si | 1.090 | 5.514 | 0,8349 | 0,7761 | 0,7879 |
| No | 20.479 | 5.907 |
| Modelo 2 | 0,888 | 0,837 | 0,773 | 0,805 | 0,610 | Si | 1.079 | 5.525 | 0,8366 | 0,7770 | 0,7889 |
| No | 20.502 | 5.884 |
| **Logit EN  down sample rfThresh** | Modelo 1 | 0,52 |  | 0,82 | 0,791 | 0,887 |  | Si | 1.231 | 5.373 | 0,8136 | 0,7967 | 0,8001 |
| No | 21.022 | 5.364 |
| Modelo 2 | 0,52 |  | 0,823 | 0,789 | 0,888 |  | Si | 1.210 | 5.394 | 0,8168 | 0,7952 | 0,7995 |
| No | 20.981 | 5.405 |
| **Logit -lasso smote oversampled** | Modelo 1 | 0,50 | 0,887 | 0,921 | 0,559 | 0,766 | 0,501 | Si | 3.710 | 2.894 | 0,4382 | 0,0789 | 0,1508 |
| No | 2.082 | 24.304 |
| Modelo 2 | 0,891 | 0,912 | 0,608 | 0,782 | 0,539 | Si | 3.866 | 2.738 | 0,4146 | 0,0880 | 0,1534 |
| No | 2.322 | 24.064 |
| **Logit -lasso smote oversampled rfThresh** | Modelo 1 | 0,56 |  | 0,819 | 0,772 | 0,879 |  | Si | 5.401 | 1.203 | 0,1822 | 0,2252 | 0,2166 |
| No | 5.943 | 20.443 |
| Modelo 2 | 0,57 |  | 0,811 | 0,769 | 0,875 |  | Si | 5.373 | 1.231 | 0,1864 | 0,2252 | 0,2175 |
| No | 5.943 | 20.443 |
| **Logit ridge SMOTE  Oversampled** | Modelo 1 | 0,50 | 0,5 | 1 | 0 | 0,571 | 0 | Si |  | 6604 | 1,0000 | - | 0,2002 |
| No |  | 26386 |
| Modelo 2 | 0,5 | 1 | 0 | 0,571 | 0 | Si |  | 6604 | 1,0000 | - | 0,2002 |
| No |  | 26386 |
| **Logit EN  SMOTE** | Modelo 1 | 0,50 | 0,904 | 0,843 | 0,785 | 0,818 | 0,629 | Si | 5062 | 1542 | 0,2335 | 0,1539 | 0,1698 |
| No | 4060 | 22326 |
| Modelo 2 | 0,907 | 0,849 | 0,795 | 0,826 | 0,644 | Si | 5009 | 1595 | 0,2415 | 0,1490 | 0,1675 |
| No | 3931 | 22455 |
| **Logit EN  SMOTE rfThresh** | Modelo 1 | 0,54 |  | 0,812 | 0,809 | 0,894 |  | Si | 5362 | 1242 | 0,1881 | 0,1867 | 0,1870 |
| No | 4927 | 21459 |
| Modelo 2 | 0,56 |  | 0,815 | 0,802 | 0,893 |  | Si | 5409 | 1195 | 0,1810 | 0,1935 | 0,1910 |
| No | 5105 | 21281 |
| **Tree** | Modelo 1 | N/A | 0,872 | 0,546 | 0,917 | 0,842 | 0,485 | Si | 1495 | 3414 | 0,6955 | 0,8864 | 0,8580 |
|  |  | No | 24891 | 3190 |
| Modelo 2 | 0,869 | 0,534 | 0,935 | 0,855 | 0,509 | Si | 1548 | 3433 | 0,6892 | 0,8868 | 0,8570 |
|  |  | No | 24838 | 3171 |
| **forest** | Modelo 1 | N/A | 0,906 | 0,575 | 0,942 | 0,869 | 0,559 | Si | 1510 | 3783 | 0,7147 | 0,8981 | 0,8687 |
|  |  |  |  |  | No | 24876 | 2821 |
| Modelo 2 | 0,898 | 0,568 | 0,939 | 0,864 | 0,545 | Si | 1585 | 3698 | 0,7000 | 0,8951 | 0,8639 |
|  |  |  |  |  | No | 24801 | 2906 |

Fuente: Elaboración propia

**Anexo 3.**

**Tabla 4**. Coeficientes modelo elastic-net up samplimg.

|  |  |
| --- | --- |
| **Predictores por hogar** | **Coeficiente** |
| % personas que reciben subsidio familiar | 0,8399 |
| Años educación promedio | 0,7573 |
| Valor arriendo mensual | 0,7474 |
| Tasa de ocupación | 0,6522 |
| Edad promedio | 0,5686 |
| % personas con trabajo formal | 0,5542 |
| Reciben otros ingresos | 0,5266 |
| Horas trabajadas promedio semana | 0,3854 |
| Interacción edad promedio y % mujeres | 0,1830 |
| Ciudad de residencia Pereira | 0,1378 |
| Interacción educación promedio y % mujeres | 0,1293 |
| Ciudad de residencia Bogotá | 0,1264 |

Fuente: Elaboración propia

La tabla anterior presenta las variables más representativas del modelo, en total se estimó el modelo con 76 variables, resultados reportados en el archivo excel denominado coeficientes que reposa en la carpeta 3.Stores.