

DAT102x: prediciendo la pobreza en todo el mundo alojada POR MICROSOFT

Descripción del problema

- Sobre los datos
- Variable objetivo
 - Formato de envío
 - Métrica de rendimiento
- Características
 - Fila de ejemplo
- Referencias

Sobre los datos

Su objetivo es predecir la probabilidad de que las personas en siete países diferentes vivan por debajo del umbral de pobreza en el umbral de \$ 2.50 / día, dados otros indicadores socioeconómicos. La probabilidad de estar en la pobreza se calculó utilizando el Índice de probabilidad de pobreza (PPI), que estima el estado de pobreza de un individuo utilizando 10 preguntas sobre las características de un hogar y la propiedad de los activos. Los datos restantes provienen de las encuestas de hogares de Financial Inclusion Insights realizadas por InterMedia.

INICIO

(/COMPETITIONS/15/PREDICTING-POVERTY/PAGE/46/)

DESCRIPCIÓN DEL PROBLEMA

(/COMPETITIONS/15/PREDICTING-POVERTY/PAGE/47/)

DE EXPORTACIÓN DE PUNTUACIÓN PARA EDX

(/COMPETITIONS/15/PREDICTING-POVERTY/PAGE/48/)

CLASIFICACIÓN RESULTADOS

(/COMPETITIONS/15/PREDICTING-POVERTY/LEADERBOARD/)

DE DATOS DESCARGAR

(/COMPETITIONS/15/PREDICTING-

Variable objetivo

Estamos tratando de predecir la variable `poverty_probability` para cada fila del conjunto de datos de prueba. `poverty_probability` es un número positivo de coma flotante (por ejemplo, 0.84) entre 0.0 y 1.0 inclusive.

Tu trabajo es:

1. Entrene un modelo utilizando las entradas en `train_values.csv` y las etiquetas `train_labels.csv`
2. Predecir flotantes para cada fila en `test_values.csv` para el que no conoce la verdadera probabilidad de pobreza.
3. Envíe sus predicciones en un formato que coincida **exactamente con la** `presentación_formato.csv` .
4. Sube tus predicciones a esta competencia para obtener un puntaje.
5. Exporte su token de calificación (haga clic en la pestaña "Exportar puntaje para EdX") y péguelo en el calificador de tareas en edX para obtener su calificación del curso.

Formato de envío

El formato para el archivo de envío es dos columnas con el `row_id` y el `poverty_probability` . El tipo de datos `poverty_probability` es flotante entre 0.0 y 1.0, **así que asegúrese de que haya un punto decimal en su envío** . Por ejemplo, `0.66` es un flotante válido que representa el 66%, pero *no* `66` es .

Si predijo que todas las personas tuvieron un cambio del 100% de estar en la pobreza, una probabilidad de 1.0, las primeras líneas del `.csv` archivo que envíe se verían así:

POVERTY/DATA/)

PRESENTACIONES

(/COMPETITIONS/15/PREDICTING-POVERTY/SUBMISSIONS/)

REGLAS OFICIALES

(/COMPETITIONS/15/PREDICTING-POVERTY/RULES/)

```
row_id,poverty_probability
0,1.0
1,1.0
2,1.0
3,1.0
4,1.0
⋮
```

Métrica de rendimiento

Estamos prediciendo una probabilidad (flotante continua), por lo que este es un problema de regresión. Para medir la regresión, usaremos una métrica llamada r-cuadrado, también conocida como coeficiente de determinación. Es una medida estadística de la bondad de ajuste de un modelo de regresión, por lo que un valor más alto es mejor (a diferencia de una métrica de error, donde un valor más bajo es mejor).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Donde SS_{res} es la suma de cuadrados de residuos y SS_{tot} es la suma total de cuadrados. El mejor puntaje posible es 1, pero el peor puntaje posible puede ser negativamente infinito.

Características

Hay 58 variables en este conjunto de datos. Cada fila del conjunto de datos representa a un individuo en un país determinado. Cada `country` está representado por una letra única. Hay siete países en total. Además del país, `religion` también está oscurecido, con cada religión correspondiente a una letra única.

Las variables son las siguientes:

DEMOGRAFÍA

- `country` - Identificador único para cada país
- `is_urban` - Área de residencia urbana versus rural
- `age` - Años
- `female` - Sexo (verdadero = femenino, falso = masculino)
- `married` - Estado civil
- `religion` - Identificador único para la religión.
- `relationship_to_hh_head` - Relación del demandado con el jefe del hogar.

EDUCACIÓN

- `education_level` - Nivel más alto de educación (0 = sin educación, 1 = educación primaria, 2 = educación secundaria, 3 = educación superior)
- `literacy` - Capacidad de leer y comprender
- `can_add` - Posibilidad de agregar
- `can_divide` - Capacidad de dividir
- `can_calc_percents` - Capacidad para calcular porcentajes
- `can_calc_compounding` - Capacidad para calcular el interés compuesto

EMPLEO

- `employed_last_year` - Si el encuestado estuvo empleado en el último año
- `employment_category_last_year` - Categoría de empleo el año pasado (por ejemplo, empleado, jubilado)
- `employment_type_last_year` - Tipo de empleo el año pasado (por ejemplo, asalariado, estacional)
- `share_hh_income_provided` - Parte de los ingresos del hogar proporcionados
- `income_ag_livestock_last_year` - Si el encuestado recibió ingresos de la agricultura o la ganadería en el último año
- `income_friends_family_last_year` - Si el encuestado recibió ingresos de amigos o familiares en el último año

- `income_government_last_year` - Si el encuestado recibió ingresos del gobierno en el último año
- `income_own_business_last_year` - Si el encuestado recibió ingresos de su propio negocio en el último año
- `income_private_sector_last_year` - Si el encuestado recibió ingresos del sector privado en el último año
- `income_public_sector_last_year` - Si el encuestado recibió ingresos del sector público en el último año

ECONÓMICO

- `num_times_borrowed_last_year` - Número de veces que el encuestado pidió prestado dinero en el último año
- `borrowing_recency` - Lo reciente de la última actividad de préstamo
- `formal_savings` - Tiene ahorros en una institución formal
- `informal_savings` - Tiene ahorros en una institución informal
- `cash_property_savings` - Tiene ahorros en efectivo o propiedad
- `has_insurance` - Tiene al menos una forma de seguro
- `has_investment` - Tiene al menos una forma de inversión
- `bank_interest_rate` - Tasa de interés pagada en un préstamo bancario
- `mm_interest_rate` - Tasa de interés pagada en un préstamo de dinero móvil
- `mfi_interest_rate` - Tasa de interés pagada en un préstamo de una institución microfinanciera (IMF)
- `other_fsp_interest_rate` - Tasa de interés pagada a otro proveedor de servicios financieros (FSP)
- `num_shocks_last_year` - Número de perturbaciones financieras experimentadas en el último año.
- `avg_shock_strength_last_year` - Resistencia media de los choques experimentados en el último año.
- `borrowed_for_emergency_last_year` - Pidió dinero prestado para una emergencia en el último año.
- `borrowed_for_daily_expenses_last_year` - Dinero prestado para gastos diarios en el último año

- `borrowed_for_home_or_biz_last_year` - Dinero prestado para gastos de hogar o negocio en el último año.

TELÉFONO

- `phone_technology` - Sofisticación del tipo de teléfono (0 = sin teléfono, 1 = teléfono básico, 2 = teléfono con función, 3 = teléfono inteligente)
- `can_call` - Posibilidad de hacer una llamada telefónica
- `can_text` - Capacidad de texto
- `can_use_internet` - Posibilidad de usar internet en el teléfono
- `can_make_transaction` - Posibilidad de realizar una transacción financiera en el teléfono.
- `phone_ownership` - Residencia telefónica (0 = sin teléfono, 1 = comparte teléfono, 2 = posee teléfono)
- `advanced_phone_use` - Capacidad para realizar tareas avanzadas en un teléfono

INCLUSIÓN FINANCIERA

- `reg_bank_acct` - Tiene una cuenta bancaria a su propio nombre
- `reg_mm_acct` - Tiene una cuenta de dinero móvil a su nombre
- `reg_formal_nbfi_account` - Tiene una cuenta en una institución financiera no bancaria (NBFI) a su propio nombre
- `financially_included` - Incluido financieramente, que se define como tener al menos uno de los siguientes: una cuenta bancaria registrada, una cuenta de dinero móvil registrada o una cuenta NBFI registrada
- `active_bank_user` - Ha utilizado su cuenta bancaria en los últimos 90 días.
- `active_mm_user` - Ha utilizado su cuenta de dinero móvil en los últimos 90 días.
- `active_formal_nbfi_user` - Ha usado su cuenta formal de NBFI en los últimos 90 días
- `active_informal_nbfi_user` - Ha realizado actividades financieras en un NBFI informal en los últimos 90 días.
- `nonreg_active_mm_user` - Ha utilizado una cuenta de dinero móvil a nombre de otra persona en los últimos 90 días (uso 'sin receta')
- `num_formal_institutions_last_year` - Número de instituciones financieras formales utilizadas en el último año.

- `num_informal_institutions_last_year` - Número de instituciones financieras informales utilizadas en el último año.
- `num_financial_activities_last_year` - Número de diferentes tipos de actividades financieras realizadas en el último año.

Fila de ejemplo

Aquí hay un ejemplo de una de las filas en el conjunto de datos para que pueda ver los tipos de valores que puede esperar en el conjunto de datos. Muchos son booleanos, algunos son numéricos, algunos son categóricos y pueden faltar valores.

	0 0
país	do
is_urban	Falso
años	18 años
hembra	Cierto
casado	Cierto
religión	PAG
relación a cabeza	Otro
Nivel de Educación	1
alfabetismo	Cierto
puede agregar	Cierto
puede_dividir	Cierto
can_calc_percents	Cierto
can_calc_compounding	Cierto

0 0

empleada_anterior_año	Falso
empleo_categoría_último_año	ama de casa o estudiante
Employment_type_last_year	no funciona
share_hh_income_provided	1
ingreso_ag_vivo_último_año	Falso
ingresos_amigos_familia_último_año	Falso
renta_gobierno_último_año	Falso
renta_propiedad_empresa_año_año	Falso
ingreso_privado_sector_último_año	Falso
renta_sector_de_pública_año_anterior	Falso
num_times_borrowed_last_year	0 0
préstamos	0 0
ahorros_formales	Falso
ahorros_informales	Falso
cash_property_savings	Falso
has_insurance	Falso
has_investment	Falso
bank_interest_rate	Yaya
mm_interest_rate	Yaya
mfi_interest_rate	Yaya

0 0

other_fsp_interest_rate	Yaya
num_shocks_last_year	0 0
avg_shock_strength_last_year	0 0
prestado_para_emergencia_último_año	Falso
prestado_para_paratos_experimentales_el_año	Falso
prestado_para_casa_o_biz_el_año_	Falso
phone_technology	0 0
puede llamar	Cierto
can_text	Cierto
can_use_internet	Falso
can_make_transaction	Falso
propiedad del teléfono	1
Advanced_phone_use	Falso
reg_bank_acct	Cierto
reg_mm_acct	Falso
reg_formal_nbfi_account	Falso
financieramente_incluido	Cierto
active_bank_user	Cierto
active_mm_user	Falso
active_formal_nbfi_user	Falso

0 0

active_informal_nbfi_user	Falso
nonreg_active_mm_user	Falso
num_formal_instituutions_last_year	1
num_informal_institutions_last_year	0 0
num_financial_activities_last_year	1

Referencias

- Innovaciones para la acción contra la pobreza (IPA). Sobre el PPI: una herramienta de medición de la pobreza. <https://www.povertyindex.org/about-ppi> (<https://www.povertyindex.org/about-ppi>)
- InterMedia Perspectivas de inclusión financiera. <http://finclusion.org/> (<http://finclusion.org/>)

Términos de uso (/termsofuse/) · Política de derechos de autor (/copyrightpolicy/) · Política de privacidad (/privacypolicy/)

Copyright © 2019 DrivenData, Inc. Microsoft® y el logotipo de Microsoft son marcas comerciales de Microsoft, Inc. EdX® y el logotipo de edX son marcas comerciales o marcas comerciales registradas de edX Inc. Todos los derechos reservados.

Trabaja con nuestra comunidad (<https://www.drivendata.org/>) · Trabaja con nuestro equipo (<http://www.drivendata.co/>)

Alojado por (<https://www.drivendata.org/>)