



DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
Inteligencia de negocios 202220 – Laboratorio 4
PROFESORA: Haydemar Nuñez

Nombres	Apellidos	Código	Login
María Sofía	Álvarez López	201729031	ms.alvarezl
Brenda Catalina	Barahona Pinilla	201812721	bc.barahona
Alvaro Daniel	Plata Márquez	201820098	ad.plata

Informe de laboratorio #4

Escenarios de prueba para el API y análisis:

Para probar los escenarios, puede encontrar cada uno de los Json en la carpeta "Test". En los escenarios se pondrán los pantallazos de la prueba y el resultado que dio. Además de puntualizar el nombre de la prueba (del archivo)

- **Predicción coherente:**

Nombre del archivo: R2Coherente1

POST /r2 GetR2

Parameters

No parameters

Request body required

```
{  "data": {    "data": [      {        "unnamed_0": "s",        "adult_mortality": 16.0,        "infant_deaths": 11.0,        "alcohol": 5.88,        "percentage_expenditure": 547.218141,        "hepatitis_B": 98.0,        "measles": 6871,        "tbi": 26.8,        "under_five_deaths": 12.0,        "polio": 99.0,        "total_expenditure": 4.11,        "diphtheria": 99.0,        "hiv_aids": 0.3,        "gdp": 4212.549288,        "population": 66881867.0,        "thinness 10-19 years": 8.2      }    ]  }
```

Code

Details

200

Response body

```
{  "r2": 0.9242782975164372}
```

Para este escenario calculamos en Jupyter el r^2 y este es el mismo que el r^2 que obtuvimos con la API.

Nombre del archivo: PrediccionCoherente1

POST /predict Make Predictions

Parameters

No parameters

Request body required

```
{
  "data": [
    {
      "unnamed_0": "s",
      "adult_mortality": 16,
      "infant_deaths": 11,
      "alcohol": 5.88,
      "percentage_expenditure": 547.218141,
      "hepatitis_B": 98.0,
      "measles": 6871,
      "bmi": 26.8,
      "under_five_deaths": 12,
      "polio": 99.0,
      "total_expenditure": 4.11,
      "diphtheria": 99.0,
      "hiv_aids": 0.3,
      "gdp": 4212.549288,
      "population": 66881867.0,
      "thinness_10_19_years": 8.3,
      "thinness_5_9_years": 8.5
    }
  ]
}
```

CodeDetails

200

Response body

```
{
  "predict": "[74.94561162835979, 61.731706271446886]"
}
```

En esta predicción se puede ver que se ponen a prueba dos casos, y podemos ver que el life expectancy que resulta es muy cercano al que es en realidad. Para el primer caso, el valor de la variable de interés es de 74.94, cuando en realidad debió ser de 73.7. Para la segunda, la predicción dio 61.73, siendo en realizada 59.2.

- **Predicción incoherente:**

Nombre del archivo: PrediccionIncoherente1

Parameters

No parameters

Request body required

```
{
  "data": [
    {
      "unnamed_0": "s",
      "adult_mortality": 160,
      "infant_deaths": 1100,
      "alcohol": 35.88,
      "percentage_expenditure": 947.210141,
      "hepatitis_B": 0,
      "measles": 0,
      "bmi": 98.8,
      "under_five_deaths": 12,
      "polio": 99.0,
      "total_expenditure": 4.11,
      "diphtheria": 99.0,
      "hiv_aids": 0.3,
      "gdp": 4212.549200,
      "population": 600.0,
      "thinness_10_19_years": 100,
      "thinness_5_9_years": 100
    }
  ]
}
```

CodeDetails

200

Response body

```
{
  "predict": "[77.90735364370474]"
}
```

Podemos ver que varios de estos datos no tienen sentido, entre estos la cantidad de población y la cantidad de muertes de niños no es coherente. Adicional a esto la delgadez en niños de 10 a 19 años y 5 a 9 años es alta. Sin embargo, se predice que la esperanza de vida es 77.9, lo cual no concuerda con las altas cifras de las variables mencionadas anteriormente.

Nombre del archivo: PrediccionIncoherenteMedia2

POST /predict Make Predictions

Parameters

No parameters

Request body required

```
{
  "data": [
    {
      "unnamed_0": 78,
      "adult_mortality": 1625,
      "infant_deaths": 31,
      "alcohol": 381,
      "percentage_expenditure": 2945742256,
      "hepatitis_B": 655,
      "measles": 2433,
      "bmi": 377,
      "under_five_deaths": 43,
      "polio": 820,
      "total_expenditure": 449,
      "diphtheria": 819,
      "hiv_aids": 16,
      "gdp": 1479315375,
      "population": 114609342,
      "thinness_10_19_years": 47,
      "thinness_5_9_years": 48
    }
  ]
}
```

Server response

CodeDetails

200

Response body

```
{
  "predict": "[6964.773366704773]"
}
```

Al calcular la media, vemos que los outliers la afectaron grandemente a la media, llegándola a sesgar debido a que esta es sensible a outliers. Con esto, se ve que la media termina representando datos incoherentes para un país. Esto también se ve reflejado en la variable de interés, ya que la esperanza de vida es de 6964.7773, lo cual no tiene sentido.

Nombre del archivo: R2Incoherencia1

POST /r2 GetR2

Parameters

No parameters

Request body required

```
{  "data": {    "data": [      {        "unnamed_0": 21,        "adult_mortality": 151.0,        "infant_deaths": 0.0,        "alcohol": 2,        "percentage_expenditure": 569.2953509,        "hepatitis_B": 11,        "measles": 0,        "bmi": 76.5,        "under_five_deaths": 0.0,        "polio": 75.6,        "total_expenditure": 5,        "diphtheria": 10,        "hiv_aids": 80,        "gdp": 3000,        "population": 200,        "thinness 10 19 years": 3.      }    ]  }
```

Code	Details
200	<div>Response body</div> <div><pre>{ "r^2": -4956438.933716847}</pre></div>

Podemos ver que, en este escenario, el r^2 toma un valor negativo que no tiene coherencia, esto se debe a que la predicción hecha y esperada es muy diferente. Este valor no tiene sentido este puede tener un valor máximo de 1.

- **Hay fallas**

Nombre del archivo: FallaUnnamed

POST

/predict

Make Predictions

Parameters

No parameters

Request body

required

```

{
  "data": [
    {
      "adult_mortality": 151.0,
      "infant_deaths": 0,
      "alcohol": 1.8,
      "percentage_expenditure": 423.2953589,
      "hepatitis_B": 9.0,
      "measles": 0,
      "tmi": 68.6,
      "under_five_deaths": 0.0,
      "polio": 91.0,
      "total_expenditure": 4.87,
      "diphtheria": 9.0,
      "hiv_aids": 0.1,
      "gdp": 2284.37858,
      "population": 146.0,
      "thinness_10_19_years": 0.1,
      "thinness_5_9_years": 0.1,
      "income_per_capita_of_residents": 693
    }
  ]
}

```

Code

Details

422

Error: Unprocessable Entity

Response body

```

{
  "detail": [
    {
      "loc": [
        "body",
        "data",
        0,
        "unnamed_0"
      ],
      "msg": "field required",
      "type": "value_error.missing"
    }
  ]
}

```

Podemos ver que ocurre un error, al realizar una prueba sin la columna "Unnamed". Esto ocurre porque en el modelo teníamos en los datos esta columna, que, en el pipeline eliminamos para hacer el modelo final de predicción. Con lo anterior generamos un problema ya que requerimos de esta columna para que el pipeline y por consiguiente el modelo funcione.

Estrategia por desarrollar sobre el software para mitigar incoherencias en el resultado y fallas en el sistema:

Una posible estrategia es mejorar el pipeline, en esta podríamos tener en cuenta la posibilidad de que no sea agregado la columna "Unnamed", por lo que, el modelo debería correr exitosamente con o sin esta columna.

Por otro lado, podríamos dar un mejor manejo a los valores nulos, nosotros los tratamos con la media, pero se podría probar alternativas como la eliminación de estos datos o la imputación con otro valor significativo que no sea la media.

Conclusiones:

Hay que tener en cuenta que Los datos solo funcionan para países vecinos a los Alpes y que sean de años cercanos, para que tengan las mismas estadísticas, ya que, debido a factores macro y micro, los índices o variables de estos países pudieron haber cambiado drásticamente, por lo que el modelo dejaría de predecir adecuadamente la expectativa de vida.

Bono:

1. Construir transformaciones personalizadas e incluirlas en el pipeline y garantizar que el proceso completo es correcto.

El pipeline que usamos para la realización de este laboratorio tiene transformaciones personalizadas, estas clases se pueden ver en el documento clases.py, en la carpeta Notebook.

2. Desplegar la API en un servidor gratuito como Heroku para que pueda prestar servicio a cualquiera haciendo uso de una URL.

Desplegamos la API en AWS, en este logramos que la IP sea estática. Puede acceder en : <http://3.228.160.169/docs>

3. Implementar la estrategia para mitigación de errores identificados en los escenarios y documentados por ustedes en el documento de entrega.