

Hormigones

Importante: la información que se te están entregando para resolver el caso a continuación es confidencial, por lo cual te comprometes a no compartirla bajo ningún motivo, no utilizarla para fines distintos a los de esta evaluación, y a eliminarla después de la presentación.

Instrucciones Generales

- El objetivo del caso es evaluar tus habilidades en modelamiento predictivo, para lo cual resolverás un problema de negocio utilizando la información que se te ha entregado.
- Tu modelo será evaluado comparando tu predicción con los valores del conjunto de prueba (test set). Respeta el formato de test enviado (no modifiques el archivo, cabeceras, formatos, encoding, orden de las filas, etc.). Envía tus resultados al correo francisco@brein.pe. La fecha límite para el envío de resultados es el martes 6 de Diciembre de 2022.
- El desempeño del modelo se evaluará mediante el indicador MAE (Mean Absolute Error), usando la definición estándar $MAE := n^{-1} \sum_{i=1}^n |y_i - \hat{y}_i|$.
- Adicionalmente, al archivo que envíes con tus predicciones deberás de preparar una presentación para mostrar tu proceso de resolución del caso. Te aconsejamos que en esta presentación consideres:
 - Explicación del problema y lo que estás intentado resolver.
 - Algunos insights que hayas podido encontrar en la información que se te ha enviado.
 - El proceso de desarrollo del modelo: la limpieza y transformaciones de datos, los modelos que has aplicado y tus resultados.
 - Como te imaginarías que se puede usar este modelo en la empresa. Que potenciales beneficios económicos se podrían generar de usar este modelo.

Problema

El problema de negocio que se está evaluando corresponde a Cementos Perú (CP), una importante empresa de materiales de construcción. CP cuenta con información histórica del proceso productivo del concreto y desea utilizar machine learning para encontrar mejoras en el diseño de este material.

Sobre el Concreto

El concreto es un material utilizado en construcción formado principalmente por cemento, áridos (los áridos pueden ser arena, grava o gravilla), agua y aditivos. CP cuenta con información de estos insumos y cuenta con un equipo de diseño que se encarga de definir la cantidad y proporción de cada uno de estos para cada tipo de concreto (existe más de un tipo de producto de concreto). Cabe mencionar que el diseño o receta se hace para un metro cúbico (m³) de concreto producido. Adicionalmente, la empresa cuenta con información otras características que podrían influir en el proceso

de elaboración del concreto: temperatura, distancia a obra, planta en la que fue producido, etc.

El concreto cuenta con varias propiedades que el cliente puede solicitar a la empresa. Sin embargo, la más importante es la **resistencia**. Esta resistencia es especificada por el cliente y debe ser cumplida por la empresa por normativa legal. Esto es muy importante ya que un incumplimiento puede tener consecuencias legales tales como multas y/o demolición de la obra. La del producto final se mide en un laboratorio externo mediante muestras tomadas en obra. En la obra se toman dos muestras aleatorias (R_{28_1} y R_{28_2}) y luego de 28 días de secado se mide la resistencia obtenida, la cual se calcula finalmente como el promedio de ambas muestras.

El objetivo de este problema es predecir, para los despachos producidos en la **región metropolitana**, y para los **productos normales**, la **resistencia promedio** en el día 28 del concreto. Estas consideraciones se explicarán más adelante.

Cuentas con los siguientes archivos para resolver este problema:

- CAA_AP_TABLA1.csv
- CAA_AP_TABLA2.csv
- CAA_AP_TABLA3.csv
- CAA_AP_TABLA4.csv
- CAA_AP_TABLA5.csv
- CAA_AP_TABLA6.csv
- CAA_AP_TABLA7.csv

Tabla 1

La tabla 1 (ver anexo) contiene todo el detalle de todo el **concreto producido entre el 2016 y 2018** para esta empresa. Cada fila es una “guía de despacho”, una producción única de concreto, es decir cada vez que un camión o mixer entra a una planta, recibe el concreto y lo despacha a alguna obra.

En esta tabla se encuentra el **código de sucursal**, y como para este problema debes considerar la **región metropolitana** solo debes tener en cuenta los siguientes códigos: 101, 111, 121, 122, 123, 131, 132, 133, 151.

Para este problema debes considerar los **productos normales**, estos se consideran como los productos que cumplan las siguientes características:

- El producto debe ser de tipo STD
- La descripción del producto debe ser exactamente 11 caracteres (el ejemplo de la tabla HB250902012 HIDRO tiene más de 11 caracteres, por lo que no lo debes considerar)
- Los primeros dos caracteres del producto deben ser GB, GR, HN o HB

Además de la descripción del producto, puedes sacar datos muy útiles, por ejemplo para el producto HB300902010, este se puede descomponer en:

- HB: Los primeros dos caracteres indican el tipo de concreto (4 tipos)
- 300: Los siguientes tres dígitos indican la **resistencia especificada por el cliente**, es decir la resistencia mínima que debe cumplir el concreto

- 90: Los siguientes dos dígitos son el nivel de confianza requerido por el cliente para el cumplimiento de la resistencia
- 20: Los siguientes dos dígitos son el tamaño del árido (20 ó 40)
- 10: Los últimos dos dígitos son el “cono de abrams” solicitado por el cliente, esto es una propiedad de la consistencia del concreto

Otras consideraciones que debes tomar en cuenta es que para este tipo de productos **la dosis mínima requerida de cemento es de 240 kilos por m³**.

Tabla 2

En la tabla 2 cuentas con todas las muestras tomadas de resistencia por los laboratorios externos. Ten en cuenta que no todas las guías de despacho tienen muestra de resistencia, por lo que hay muchas menos muestras que guías.

Para este problema solo debes considerar las muestras de **tipo de ensayo de compresión**, es decir **C**. Normalmente una guía tiene dos muestras y la resistencia final se calcula como el **promedio** de estas dos. En el caso que no haya segunda muestra, esta aparece como cero, debes tener cuidado de no promediarla, sino simplemente considerar el valor de la primera.

Tabla 3

En la tabla 3 cuentas con el detalle de los insumos **utilizados** para cada guía de despacho producida. Cada fila es un insumo específico utilizado para cada guía. Para cada insumo se sabe tanto la cantidad **requerida** (cantidad teórica que se debió utilizar) como la **consumida** (cantidad real que se utilizó) para el volumen total de concreto producido (en m³) en esa guía. Además, tienes una clasificación del tipo de insumo que puede ser **agua, cemento, arena, aditivo, grava, gravilla, fibra u otro**. No todos los productos de concreto tienen todos los **tipos de insumos**, y en algunas ocasiones pueden tener más de uno del mismo tipo.

Por ejemplo, la **guía 8062810 utilizó 6 insumos** (agua, cemento, arena, aditivo, grava y gravilla). Ten en cuenta que ambas cantidades mostradas en esta tabla son para el **total** de m³ producidos de esa guía. Debes tener en cuenta que para que tu modelo funcione correctamente, tienes que utilizar la lógica del diseño del concreto, conocida como **dosificación**, que es convertir todas las cantidades como insumos utilizados por 1 m³. Por ejemplo, para la guía 8062810 se produjeron 6.5 m³ (cantidad a facturar), por lo que deberás dividir todas las cantidades entre este número para hallar la dosis por 1m³.

Tabla 4

En la tabla 4 cuentas con las humedades tomadas para cada árido (arena, grava y/o gravilla) utilizado para cada guía de despacho.

Esta parte te ayudará a tener un modelo más fino, ya que no es lo mismo utilizar un árido más seco que uno más húmedo, ya que el agua total disponible para interactuar con el cemento influye en la resistencia del concreto.

Para esto, en la tabla 3 tú cuentas con la cantidad realmente ingresada de cada insumo. Por ejemplo en la guía 8062810 se ingresaron 6563 kilos de arena, 2741 kg de gravilla, y 4081 kg de grava. Estos datos son conocidos como **peso total del árido** ya que contiene el **peso seco del árido** más el **peso del agua por humedad** absorbida por el árido. Para seguir la lógica de diseño del hormigón debes calcular y utilizar el peso seco de cada árido siguiendo la siguiente fórmula:

$$\begin{aligned}\text{Peso_total} &= \text{Peso_seco} + \text{Peso_agua} \\ \text{Peso_total} &= \text{Peso_seco} + \text{Peso_seco} \times \text{Porcentaje_Humedad}\end{aligned}$$

Ten en cuenta que en la tabla 4 se encuentran las humedades para cada árido y debes notar que se encuentran como porcentaje de 0 a 100.

Por otro lado, siguiendo la misma lógica de diseño, en la misma guía se utilizó 689 kilos de agua, pero debes calcular el **agua total** que incluye también el agua que se encuentra absorbida en los áridos (arenas + grava + gravilla):

$$\text{Agua_total} = \text{Agua_utilizada} + \text{Peso_agua_aridos}$$

Acuérdate que, si todavía no has dividido las cantidades de los insumos entre la cantidad a facturar, luego de hallar los pesos secos y el agua total debes dividir estos valores entre la cantidad a facturar.

Tabla 5

En la tabla 5 cuentas con las **temperaturas** para la región metropolitana para todas las horas y fechas desde el 2016 al 2018. En este caso solo cuentas con una fecha y/o hora para poder relacionarlo con las guías de despacho.

Tabla 6

En la tabla 6 cuentas con muestras que se toman cada cierto tiempo de **algunas propiedades de los áridos** (arena, grava, gravilla), para cada código de material. Cuentas con los módulos de finura real y teóricos, estas son propiedades que indican que tan fino es el árido. En este caso solo cuentas con una fecha y/o hora para poder relacionarlo con las guías de despacho.

Tabla 7

En la tabla 7 tienes el **data test** con las guías de despacho para las cuales tienes que realizar la predicción de resistencia promedio en el día 28 y para las cuales debes subir tus resultados en el tablero.

Por favor respeta el formato de este archivo (nombres de las columnas, número de filas, orden de las filas, encoding, extensión del archivo) para que no tengas problemas a la hora de visualizar tus resultados. En esta data se encuentra tanto las guías que se calificarán en la tabla pública como privada.

Anexo

| CAMPO | DESCRIPCIÓN | EJEMPLO |
|----------------------|---|----------------------|
| CODIGO_GUIA_DESPACHO | Es el código que identifica cada producción única de concreto, esto se conoce como guía de despacho | 1510001 |
| FECHA_GUIA_DESPACHO | Es la fecha (YYYYMMDD) en la cual se despacho esa guía | 20180913 |
| CODIGO_PRODUCTO | Es el código del producto de concreto producido | 13372 |
| DESCRIPCION_PRODUCTO | Es la descripción del producto de concreto | HB250902012 HIDRO |
| PRODUCTO_VAP | Indica si un producto es STD (estándar) o VAP (valor agregado) | STD |
| CODIGO_SUCURSAL | Es el código de cada planta específica donde se produce el concreto (cada planta es una mezcladora específica) | 151 |
| ANILLO | El anillo clasifica zonas para identificar que tan lejos está la obra de la planta (a mayor número, está más lejos) | 402 |
| HORA_IMPRESO_GUIA | Es la hora a la que se imprimió (se generó) la guía de despacho | 08:31 |
| HORA_CARGA | Es la hora a la que se cargó efectivamente el concreto en el mixer (camión) | 08:43 |
| CANTIDAD_A_FACTURAR | La cantidad de m3 producidos de concreto | 6 |
| TAMANO | Una categorización del tamaño de m3 producidos de concreto | M |

Cuadro 1: Todos los despachos

| CAMPO | DESCRIPCIÓN | EJEMPLO |
|----------------------|---|-------------|
| NUMERO_GUIA | Es el código de la guía de despacho | 1510036 |
| LABORATORIO | El código del laboratorio que tomó la muestra | 2 |
| DESCRIPCION_PRODUCTO | Es la descripción del producto que fue muestreado | HB250902010 |
| TIPO_ENSAYO | Es el tipo de ensayo (C es de compresión) | C |
| R28_1 | Resultado de la resistencia de la primera muestra tomada al día 28 de producido el concreto | 272 |
| R28_2 | Resultado de la resistencia de la segunda muestra tomada al día 28 de producido el concreto | 274 |
| FLAG_MUESTRA_NULA | Flag indicando si la muestra es nula (1) o no (0) | 0 |

Cuadro 2: Muestras de resistencia

| CAMPO | DESCRIPCIÓN | EJEMPLO |
|---------------------------|---|-----------|
| CODIGO_GUIA_DESPACHO | Es el código de la guía de despacho | 8062810 |
| INSUMO_CANTIDAD_REQUERIDA | Es el total de insumo de cantidad requerido (para el total de m3 producidos de concreto) | 9.75 |
| INSUMO_CANTIDAD_CONSUMIDA | Es el total de insumo de cantidad finalmente consumido (para el total de m3 producidos de concreto) | 9.68 |
| INSUMO | Es el código del insumo utilizado | AD_HP-450 |
| TIPO_INSUMO | Es la clasificación del insumo | ADITIVO |

Cuadro 3: Insumos de concreto

| CODIGO_GUIA_DESPACHO | INSUMO_CANTIDAD_REQUERIDA | INSUMO_CANTIDAD_CONSUMIDA | INSUMO | TIPO_INSUMO |
|----------------------|---------------------------|---------------------------|------------|-------------|
| 8062810 | 9.75 | 9.68 | AD_HP-450 | ADITIVO |
| 8062810 | 6570 | 6563 | AR_STAROSA | ARENA |
| 8062810 | 1462 | 1464 | CE_MEXT | CEMENTO |
| 8062810 | 2742 | 2741 | GL_STROSA | GRAVILLA |
| 8062810 | 4074 | 4081 | GV_STROSA | GRAVA |
| 8062810 | 690 | 689 | WATER | WATER |

Cuadro 4: Ejemplo de filas de la tabla 3: insumos para una guía de despacho

| CAMPO | DESCRIPCIÓN | EJEMPLO |
|--------------------|---|------------|
| NUMERO_GUIA | Es el código de la guía de despacho | 6942085 |
| FACTOR_HUMEDAD | Es el porcentaje de humedad del árido señalado (está ya en %) | 2.1 |
| MATERIA_PRIMA | Es la materia prima o insumo a la cual se le tomó la humedad | GL_VICENTE |
| TIPO_MATERIA_PRIMA | Es la clasificación de la materia prima o insumo | GRAVILLA |

Cuadro 5: Humedad de los áridos

| CAMPO | DESCRIPCIÓN | EJEMPLO |
|-------|---|------------------|
| Fecha | Es la fecha y hora de la temperatura ambiente | 31/12/2015 19:00 |
| Temp | Es la temperatura ambiente tomada (en grados Celsius) | 19 |

Cuadro 6: Temperaturas

| CAMPO | DESCRIPCIÓN | EJEMPLO |
|------------------------|--|------------|
| CODIGO_MATERIAL | Código del material del árido en la muestra | AR_GARZAS |
| FECHA_MUESTRA | Fecha de toma de muestra del árido | 2016-01-04 |
| MODULO_FINURAS | Propiedad de que tan fino es el árido (medida real) | 3.4338 |
| MODULO_FINURAS_TEORICO | Propiedad de que tan fino es el árido (medida teórica) | 3.02 |

Cuadro 7: Áridos

| CAMPO | DESCRIPCIÓN | EJEMPLO |
|----------------------|---|---------|
| CODIGO_GUIA_DESPACHO | Código de guía de despacho a predecir | 1511844 |
| R28_PRED | Target a predecir (reemplazar con valor predicho) | 0 |

Cuadro 8: Test