

Manual de Uso — MDI (Material Design Intelligence)

Plataforma de Diseño Inteligente de Materiales Cementíceos

Febrero 2026

Contents

1 Introducción	3
2 Página 1: Data Explorer	4
2.1 Qué es	4
2.2 Cómo usarla	4
2.2.1 Seleccionar un dataset	4
2.2.2 Estadísticas descriptivas	4
2.2.3 Matriz de correlación	4
2.2.4 Histogramas de distribución	5
2.2.5 Tabla de datos	5
3 Página 2: Model Training	6
3.1 Qué es	6
3.2 Cómo entrenar un modelo	6
3.2.1 Paso 1: Configurar parámetros	6
3.2.2 Paso 2: Hacer click en “Train”	6
3.2.3 Paso 3: Revisar las métricas	6
3.2.4 Paso 4: Gráfico Actual vs Predicted	6
3.2.5 Paso 5: Predicción individual	7
3.3 Ejemplo práctico	7
4 Página 3: Explanations (Explicabilidad)	8
4.1 Qué es	8
4.2 Cómo usarla	8
4.2.1 Paso 1: Seleccionar modelo y computar SHAP	8
4.2.2 Paso 2: Feature Importance	8
4.2.3 Paso 3: SHAP Summary Plot	8
4.2.4 Paso 4: Dependence Plot	8
4.2.5 Paso 5: Explicación de predicción individual	9
5 Página 4: Exploration Lab	10
5.1 Qué es	10
5.2 Pestaña 1: Paramétrica	10
5.2.1 Para qué sirve	10

5.2.2 Cómo usarla	10
5.2.3 Cómo leer el resultado	10
5.3 Pestaña 2: Multivariable	10
5.3.1 Para qué sirve	10
5.3.2 Cómo usarla	10
5.3.3 Cómo leer el resultado	10
5.4 Pestaña 3: Comparación	11
5.4.1 Para qué sirve	11
5.4.2 Cómo usarla	11
5.4.3 Cómo leer el resultado	11
5.5 Pestaña 4: Microestructura	11
5.5.1 Para qué sirve	11
5.5.2 Cómo usarla	11
5.5.3 Cómo leer la imagen	12
5.5.4 Panel lateral de propiedades	12
6 Página 5: Configurations	13
6.1 Qué es	13
6.2 Cómo usarla	13
6.2.1 Ver configuraciones guardadas	13
6.2.2 Marcar como candidata	13
6.2.3 Eliminar una configuración	13
6.2.4 Exportar a CSV	13
7 Flujo de trabajo recomendado	14
7.1 Escenario: diseñar una mezcla con alta ceniza volante	14
7.1.1 1. Explorar datos (Data Explorer)	14
7.1.2 2. Entrenar modelo (Model Training)	14
7.1.3 3. Entender el modelo (Explanations)	14
7.1.4 4. Explorar diseños (Exploration Lab)	14
7.1.5 5. Exportar (Configurations)	15
8 Consejos útiles	16
9 Glosario	17

1 Introducción

MDI (Material Design Intelligence) es una plataforma web para el diseño de mezclas de concreto asistido por inteligencia artificial. Permite explorar datos, entrenar modelos predictivos, entender las predicciones mediante explicabilidad (SHAP), explorar el espacio de diseño de forma interactiva y gestionar configuraciones candidatas para validación experimental.

La aplicación tiene 5 páginas principales, accesibles desde el menú lateral izquierdo:

1. **Data Explorer** — exploración y visualización de los datos
2. **Model Training** — entrenamiento de modelos y predicción individual
3. **Explanations** — explicabilidad SHAP (por qué el modelo predice lo que predice)
4. **Exploration Lab** — exploración paramétrica, superficies de respuesta, comparación y microestructura
5. **Configurations** — gestión y exportación de mezclas candidatas

2 Página 1: Data Explorer

2.1 Qué es

Es la primera pantalla de la aplicación. Permite explorar los datos disponibles antes de entrenar modelos. Muestra estadísticas descriptivas, una matriz de correlación, histogramas de distribución y una tabla con los datos crudos.

2.2 Cómo usarla

2.2.1 Seleccionar un dataset

En la parte superior hay un **dropdown “Dataset”** con tres opciones:

	Dataset	Muestras	Descripción
concrete	1,030	Dataset real UCI de resistencia a compresión	
concrete_xai	500	Dataset extendido sintético	
geopolymer	400	Concreto geopolimérico sintético	

Al cambiar de dataset, toda la página se actualiza con los datos del dataset seleccionado.

2.2.2 Estadísticas descriptivas

Se muestran tarjetas para cada variable con:

- **Mean:** valor promedio
- **Std:** desviación estándar (cuánto varían los datos)
- **Range:** valor mínimo y máximo

Por ejemplo, si cement tiene Mean: 281.17 y Range: [102.0, 540.0], significa que en el dataset hay mezclas con tan poco como 102 kg/m³ de cemento hasta 540 kg/m³, con un promedio de 281 kg/m³.

2.2.3 Matriz de correlación

Deabajo de las estadísticas hay un **mapa de calor** que muestra la correlación entre todas las variables:

- **Rojo intenso (+1):** las dos variables aumentan juntas
- **Blanco (0):** no hay relación lineal
- **Azul intenso (-1):** cuando una sube, la otra baja

Observaciones típicas:

- cement y compressive_strength tienen correlación positiva (~0.5): más cemento tiende a dar más resistencia
- water y compressive_strength tienen correlación negativa (~-0.3): más agua tiende a reducir la resistencia
- superplasticizer y water pueden tener correlación negativa: el superplastificante permite reducir el agua

2.2.4 Histogramas de distribución

Gráficos de barras que muestran cómo se distribuyen los valores de cada variable. Permiten identificar:

- Si una variable está concentrada en un rango estrecho o dispersa
- Si hay valores atípicos
- Si la distribución es simétrica o sesgada

2.2.5 Tabla de datos

Al final de la página hay una tabla paginada con 100 filas del dataset. Permite ver los datos crudos fila por fila.

3 Página 2: Model Training

3.1 Qué es

Permite entrenar modelos de machine learning que predicen la resistencia a compresión del concreto a partir de su composición.

3.2 Cómo entrenar un modelo

3.2.1 Paso 1: Configurar parámetros

En la sección “Train New Model”:

- **Algorithm:** seleccionar Random Forest o Gradient Boosting
 - *Random Forest*: más robusto, buen punto de partida. Recomendado para empezar.
 - *Gradient Boosting*: puede ser más preciso pero es más sensible a los hiperparámetros.
- **n_estimators:** cantidad de árboles (por defecto 100). Más árboles = más preciso pero más lento. 100 es un buen valor.
- **max_depth:** profundidad máxima de cada árbol. Dejar vacío para sin límite. Si se quiere evitar sobreajuste, probar con valores como 10 o 15.
- **learning_rate** (solo Gradient Boosting): tasa de aprendizaje. Valores bajos (0.05-0.1) son más estables.

3.2.2 Paso 2: Hacer click en “Train”

El entrenamiento tarda unos segundos. Al terminar aparece un botón con el modelo creado mostrando su R².

3.2.3 Paso 3: Revisar las métricas

Tres tarjetas muestran:

- **R² Score** (verde): proporción de varianza explicada. Un R² de 0.90 significa que el modelo explica el 90% de la variabilidad en la resistencia. Valores aceptables: > 0.85.
- **MAE** (naranja): error absoluto medio en MPa. Si el MAE es 4.5, significa que en promedio el modelo se equivoca por ±4.5 MPa.
- **RMSE** (rojo): raíz del error cuadrático medio. Similar al MAE pero penaliza más los errores grandes.

3.2.4 Paso 4: Gráfico Actual vs Predicted

Un gráfico de dispersión donde:

- **Eje X:** resistencia real (medida en el laboratorio)
- **Eje Y:** resistencia predicha por el modelo
- **Línea diagonal punteada:** predicción perfecta

Cuanto más cerca estén los puntos de la diagonal, mejor es el modelo. Si los puntos forman una nube dispersa, el modelo no es confiable.

3.2.5 Paso 5: Predicción individual

En la sección “Individual Prediction”:

1. Completar los 8 campos con los valores de la mezcla deseada (en kg/m³ para componentes, días para edad)
2. Hacer click en “Predict”
3. El resultado aparece con formato: **34.5 MPa [29.1 – 39.9] (95% CI)**

¿Qué significa cada parte?

- **34.5 MPa**: predicción central del modelo
- **[29.1 – 39.9]**: intervalo de confianza al 95%. El modelo estima que la resistencia real caerá en este rango con 95% de probabilidad
- **(95% CI)**: nivel de confianza del intervalo

Regla práctica: si el intervalo es estrecho (ej: [33-37]), el modelo tiene alta confianza. Si es amplio (ej: [20-50]), hay mucha incertidumbre y conviene ser cauteloso.

3.3 Ejemplo práctico

Para probar una mezcla convencional:

Campo	Valor
cement	350
blast_furnace_slag	0
fly_ash	0
water	180
superplasticizer	6
coarse_aggregate	1000
fine_aggregate	700
age	28

Resultado esperado: ~35-45 MPa con intervalo de ±5 MPa.

4 Página 3: Explanations (Explicabilidad)

4.1 Qué es

Usa SHAP (SHapley Additive exPlanations) para explicar **por qué** el modelo predice lo que predice. Es la capa de “caja abierta” que permite entender el razonamiento del modelo.

4.2 Cómo usarla

4.2.1 Paso 1: Seleccionar modelo y computar SHAP

1. En el dropdown, seleccionar el modelo entrenado
2. Hacer click en “**Compute SHAP**”
3. Esperar unos segundos (SHAP analiza todas las muestras del conjunto de test)

4.2.2 Paso 2: Feature Importance

Gráfico de barras horizontales que muestra qué variables son más importantes **en general**, ordenadas de mayor a menor.

Lectura: si **cement** tiene importancia 4.5 y **water** tiene 3.2, significa que en promedio el cemento afecta la predicción en ±4.5 MPa y el agua en ±3.2 MPa.

Variables que típicamente aparecen arriba:

1. **cement**: componente principal del concreto
2. **age**: la resistencia crece con el tiempo de curado
3. **water**: relación inversa con resistencia (más agua = más poroso = menos resistente)
4. **superplasticizer**: mejora trabajabilidad y puede aumentar resistencia
5. **fly_ash**: efecto variable según cantidad y edad

4.2.3 Paso 3: SHAP Summary Plot

Gráfico de dispersión donde cada punto es una muestra del dataset:

- **Eje Y**: cada variable (ordenadas por importancia)
- **Eje X**: valor SHAP (contribución a la predicción)
 - Derecha (positivo) = aumenta la resistencia
 - Izquierda (negativo) = disminuye la resistencia
- **Color**: rojo = valor alto de la variable, azul = valor bajo

Cómo leerlo:

- Si para **cement** los puntos rojos (valor alto) están a la derecha (SHAP positivo): confirma que más cemento = más resistencia.
- Si para **water** los puntos rojos están a la izquierda: confirma que más agua = menos resistencia.

4.2.4 Paso 4: Dependence Plot

1. Seleccionar una variable en el dropdown (ej: **cement**)
2. Hacer click en “**Load Dependence Plot**”

El gráfico muestra:

- **Eje X:** valor de la variable seleccionada (ej: cemento de 100 a 540 kg/m³)
- **Eje Y:** contribución SHAP de esa variable
- **Color:** la variable con la que más interactúa (detectada automáticamente)

Permite ver relaciones no lineales: por ejemplo, la contribución del cemento puede ser casi plana hasta 200 kg/m³ y luego crecer fuertemente.

4.2.5 Paso 5: Explicación de predicción individual

1. Completar los 8 campos con una mezcla
2. Hacer click en “**Explain Prediction**”
3. Aparece:
 - Un texto en español explicando los 3 factores más influyentes
 - Un **gráfico waterfall** que muestra paso a paso cómo se llega a la predicción

El waterfall se lee de izquierda a derecha: arranca en el valor base (~35.8 MPa) y cada barra roja (positiva) o azul (negativa) muestra cuánto contribuye cada ingrediente hasta llegar a la predicción final.

5 Página 4: Exploration Lab

5.1 Qué es

El corazón del diseño de materiales. Permite explorar el espacio de composiciones de forma sistemática usando el modelo entrenado. Tiene 4 pestañas.

5.2 Pestaña 1: Paramétrica

5.2.1 Para qué sirve

Responde: “si fijo todos los ingredientes y varío uno solo, ¿cómo cambia la resistencia?”

5.2.2 Cómo usarla

1. **Seleccionar modelo** en el dropdown superior
2. En “Base Configuration”, poner los valores de una mezcla de referencia (viene precargados valores por defecto)
3. En “**Sweep Feature**”, seleccionar la variable a variar (ej: fly_ash)
4. Ajustar el **rango** con el slider (ej: 0 a 200)
5. Hacer click en “**Run Sweep**”

5.2.3 Cómo leer el resultado

- **Línea azul**: predicción del modelo para cada valor de la variable
- **Banda sombreada**: intervalo de confianza al 95%
- **Zona verde**: región óptima (valores que producen resistencia mayor o igual al 90% del máximo)

Ejemplo de lectura: “la ceniza volante entre 50 y 120 kg/m³ mantiene la resistencia por encima de 36 MPa. Por debajo de 50 no aporta, y por encima de 150 empieza a bajar la resistencia.”

5.3 Pestaña 2: Multivariable

5.3.1 Para qué sirve

Responde: “¿cómo interactúan dos ingredientes?” Genera un mapa de calor 2D.

5.3.2 Cómo usarla

1. En “Base Configuration”, poner los valores de referencia
2. Seleccionar **Variable 1** (ej: fly_ash) y ajustar su rango
3. Seleccionar **Variable 2** (ej: water) y ajustar su rango
4. Hacer click en “**Generate Surface**”

5.3.3 Cómo leer el resultado

Es un mapa de calor donde:

- **Eje X**: primera variable

- **Eje Y:** segunda variable
- **Color:** resistencia predicha (rojo = alta, amarillo = baja)

Permite identificar la **zona óptima** visualmente: la región más roja del mapa corresponde a la mejor combinación de las dos variables.

5.4 Pestaña 3: Comparación

5.4.1 Para qué sirve

Compara 2 mezclas lado a lado con predicción, incertidumbre y análisis SHAP.

5.4.2 Cómo usarla

1. En **Config A**, poner los valores de la primera mezcla (viene con valores por defecto)
2. En **Config B**, poner los valores de la segunda mezcla
3. Poner nombres descriptivos en “Label” (ej: “Convencional” y “Ecológica”)
4. Hacer click en “**Compare**”

5.4.3 Cómo leer el resultado

- **Gráfico de barras:** altura = predicción, barra de error = incertidumbre
- **Tarjetas detalladas:** cada configuración muestra:
 - Predicción con intervalo: “38.5 MPa [34.2 – 42.8]”
 - Top-3 factores SHAP: qué ingredientes influyen más y en qué dirección
- **Botón “Save as Candidate”:** guarda la mezcla para exportarla después

5.5 Pestaña 4: Microestructura

5.5.1 Para qué sirve

Genera una representación visual de la sección transversal del concreto según la composición. Permite “ver” cómo se vería el material por dentro. **No requiere un modelo entrenado.**

5.5.2 Cómo usarla

1. Elegir un **preset** haciendo click en los chips superiores:
 - *Concreto convencional*: mezcla estándar
 - *Alta ceniza volante*: alto reemplazo de cemento
 - *Alta resistencia*: bajo w/c, mucho cemento
 - *Bajo cemento / ecológico*: alta ceniza, poco cemento
 - *Joven (3 días)*: concreto recién hecho
 - *Alta relación a/c*: mucha agua
2. O modificar los valores manualmente en los campos
3. La imagen se regenera instantáneamente

5.5.3 Cómo leer la imagen

Elemento visual	Qué representa	Depende de
Polígonos grises grandes	Agregado grueso (grava)	coarse_aggregate
Círculos pequeños color arena	Agregado fino (arena)	fine_aggregate
Esfueras oscuras perfectas	Ceniza volante	fly_ash
Manchas negras	Poros / vacíos	Relación water/cement
Color del fondo	Pasta de cemento	cement y age
Puntos claros en el fondo	Granos de cemento sin hidratar	Visibles con age bajo

5.5.4 Panel lateral de propiedades

Al lado de la imagen se muestran propiedades calculadas:

- **Relación agua/cementante (w/c):** valor clave en tecnología del concreto. Menor w/c = mayor resistencia pero menor trabajabilidad.
- **Fracción de agregados:** porcentaje del volumen total que ocupan los agregados.
- **Reemplazo de ceniza volante:** qué porcentaje del cementante es ceniza volante en vez de cemento.

6 Página 5: Configurations

6.1 Qué es

Almacén de mezclas guardadas durante la exploración. Permite gestionar, marcar como candidatas y exportar a CSV para el laboratorio.

6.2 Cómo usarla

6.2.1 Ver configuraciones guardadas

La tabla muestra todas las mezclas guardadas con:

- **Label:** nombre descriptivo
- **Valores:** cada ingrediente
- **Prediction:** resistencia predicha
- **95% CI:** intervalo de confianza
- **Candidate:** checkbox de candidata a validación

6.2.2 Marcar como candidata

Hacer click en el **checkbox** de la columna “Candidate”. La fila se resalta en verde. Las candidatas son las mezclas que se llevarán al laboratorio para validación experimental.

6.2.3 Eliminar una configuración

Hacer click en el **ícono de papelera** de la fila.

6.2.4 Exportar a CSV

Dos botones disponibles:

- “**Export All CSV**”: descarga todas las configuraciones
- “**Export Candidates Only**”: descarga solo las marcadas como candidatas

El archivo CSV se descarga al computador y contiene todas las columnas necesarias: ingredientes, predicción, intervalo de confianza e identificador del modelo.

7 Flujo de trabajo recomendado

A continuación se describe el flujo completo para un caso de uso típico.

7.1 Escenario: diseñar una mezcla con alta ceniza volante

7.1.1 1. Explorar datos (Data Explorer)

1. Seleccionar dataset “concrete”
2. Observar que `fly_ash` tiene rango [0, 200.1] y promedio 54.19 kg/m³
3. En la matriz de correlación, verificar que `fly_ash` tiene correlación baja con `compressive_strength` — esto indica que su efecto no es lineal simple

7.1.2 2. Entrenar modelo (Model Training)

1. Seleccionar “Random Forest”, dejar `n_estimators` en 100
2. Click en “Train”
3. Verificar $R^2 > 0.88$ (buen modelo)

7.1.3 3. Entender el modelo (Explanations)

1. Click en “Compute SHAP”
2. En el Summary Plot, buscar `fly_ash`: los puntos rojos (valor alto) están distribuidos tanto a la derecha como a la izquierda — efecto no lineal
3. Cargar el Dependence Plot para `fly_ash`: se observa que la contribución es ligeramente positiva hasta ~100 kg/m³ y se vuelve negativa después

7.1.4 4. Explorar diseños (Exploration Lab)

1. Pestaña Paramétrica: configurar mezcla base y variar `fly_ash` de 0 a 250. Identificar que la región óptima está entre 50 y 120 kg/m³.
2. Pestaña Multivariable: variar `fly_ash` (0-200) y `water` (140-220). En el mapa de calor, identificar la zona roja: `fly_ash` ~80, `water` ~160.
3. Pestaña Comparación: crear dos mezclas:
 - Config A: “Convencional” — `cement`=350, `fly_ash`=0, `water`=180
 - Config B: “30% ceniza” — `cement`=245, `fly_ash`=105, `water`=165
 - Comparar: la Config B tiene menor resistencia (-3 MPa) pero mayor sustentabilidad. El SHAP muestra que la reducción de `water` en B compensa parcialmente la menor cantidad de cemento.
4. Guardar Config B como candidata: click en “Save as Candidate”
5. Pestaña Microestructura: comparar visualmente “Concreto convencional” vs “Alta ceniza volante” — se ven claramente las esferas de ceniza volante y menor porosidad por el superplastificante.

7.1.5 5. Exportar (Configurations)

1. Ir a Configurations
2. Marcar la mezcla como candidata (checkbox)
3. Click en “Export Candidates Only”
4. El CSV descargado contiene toda la información para reproducir la mezcla en el laboratorio

8 Consejos útiles

- **SHAP tarda unos segundos:** el cálculo de SHAP sobre el conjunto de test completo puede tardar 5-15 segundos la primera vez. Es normal. Una vez computado, queda en caché y no se recalcula.
- **R² bajo:** si el modelo tiene R² menor a 0.85, probar aumentando n_estimators a 200 o 300. Si se usa Gradient Boosting, reducir learning_rate a 0.05 y aumentar n_estimators a 300.
- **Intervalo de confianza amplio:** significa que el modelo tiene poca certeza para esa composición. Puede ser porque la mezcla está fuera del rango de los datos de entrenamiento. Conviene ser cauteloso al llevar esa mezcla al laboratorio.
- **Dataset recomendado:** el dataset “concrete” (UCI) es el más completo y confiable con 1,030 muestras reales. Los datasets “concrete_xai” y “geopolymer” son sintéticos y sirven para demostrar funcionalidades adicionales.
- **Orden recomendado:** Data Explorer -> Model Training -> Explanations -> Exploration Lab -> Configurations. Cada paso construye sobre el anterior.

9 Glosario

Término	Definición
Agregado fino	Arena, partículas menores a 4.75 mm
Agregado grueso	Grava, partículas mayores a 4.75 mm
Ceniza volante	Subproducto de centrales térmicas, usado como reemplazo parcial de cemento
Compressive strength	Resistencia a compresión, medida en MPa
Ensemble	Método de ML que combina múltiples modelos
Feature	Variable de entrada del modelo
Gradient Boosting	Algoritmo que construye árboles secuencialmente corrigiendo errores
Intervalo de confianza	Rango donde se espera que caiga el valor real con cierta probabilidad
ITZ	Interfacial Transition Zone, zona débil entre agregado y pasta
MAE	Mean Absolute Error, error promedio en unidades originales
MPa	Megapascal, unidad de resistencia (1 MPa aprox. 10.2 kg/cm ²)
R²	Coeficiente de determinación (0 = malo, 1 = perfecto)
Random Forest	Algoritmo que promedia predicciones de múltiples árboles
RMSE	Root Mean Squared Error, error cuadrático medio
SHAP	SHapley Additive exPlanations, método de explicabilidad
Superplastificante	Aditivo que mejora la fluidez sin agregar agua
w/c ratio	Relación agua/cimentante, parámetro clave de diseño de mezcla
Waterfall plot	Gráfico que descompone una predicción en contribuciones