

Guía para configurar y ejecutar el modelo predictivo.

Features

- Poder agregar los valores significativos para cada columna (se le pone 1 o 0)
- Verificar la forma en la que el proyecto podría trabajar con multiples documentos y las columnas que van a ayudar a hacer el merge
- Más modelos predictivos
- El

¿Cómo se usa?

Para usar esta herramienta no es necesario programar ni modificar el código. Solo debes ejecutar el siguiente comando en la consola:

```
./divtic.exe --config config.json
```

Nota: Este comando puede variar ligeramente dependiendo del sistema operativo que utilices. En Windows, asegúrate de que `divtic.exe` esté en la misma carpeta donde estás ejecutando el comando o que esté en tu PATH.

El parámetro `--config` indica el archivo de configuración que contiene todas las instrucciones necesarias para que el modelo funcione correctamente.

Campos principales.

Estos campos están en el archivo `config.json` y definen cómo se comportará la herramienta.

Campo	Descripción	Ejemplo
<code>file</code>	Ruta al archivo CSV de entrada.	<code>"CS_ADULTOS.csv"</code>
<code>target_column</code>	Columna objetivo que deseas predecir.	<code>"P4_1"</code>
<code>columns</code>	Lista de columnas específicas a usar. Si se deja vacía, se usa el	<code>["P2", "P3", "P5"]</code>

Campo	Descripción	Ejemplo
	análisis automático por correlación.	
default_save_path	Carpeta donde se guardarán los resultados.	"./results"
results_path	Ruta base para los archivos de salida.	"correlaciones_heatmap.png"
report_tittle	Título que se mostrará en el PDF generado.	"Modelo de predicción de hipertensión basado en ENSANUT"
save_fig_path	Carpeta donde se guardarán las gráficas.	"figures/"
corr_store_file	Nombre del archivo CSV con los resultados de correlación.	"features_correlacionadas"

Configuración Avanzada.

sample_config

Define cómo se genera la muestra de datos.

- `sample_size` : Número de casos positivos y negativos a incluir.
- `pos_condition` : Valor que representa a la clase positiva en la columna objetivo.
- `neg_condition` : Valor que representa a la clase negativa.
- `test_size` : Porcentaje de datos para prueba (entre 0 y 1).
- `percentage_per_corr` : Umbral de correlación para seleccionar variables.
- `random_state` : Semilla para mantener los resultados reproducibles.

graph_config

Personaliza la generación del gráfico de correlación.

- `graph_tittle` : Título del gráfico.
- `annot` : Mostrar o no los valores numéricos en el heatmap.
- `color_bar` : Mostrar o no la barra de colores.
- `color_map` : Paleta de colores del gráfico.
- `fmt` : Formato de los números en el gráfico.
- `hide_halfh_graph` : Ocultar mitad superior para facilitar lectura.
- `figname` : Nombre del archivo del gráfico.

- `img_format` : Formato del archivo gráfico (`svg` , `png` , `pdf`).
- `show_plot` : Mostrar o no el gráfico en pantalla.
- `size_w` y `size_h` : Dimensiones del gráfico en pulgadas.





`model_config`

Define parámetros específicos para los modelos de aprendizaje automático. Puedes configurar:

- **Regresión logística** (`logistic_regression_config`)
- **Random Forest** (`random_forest_config`)
- **SVM (Máquinas de Vectores de Soporte)** (`svc_config`)

Cada uno tiene parámetros detallados como `max_iter` , `C` , `kernel` , `criterion` , etc. Solo modifica estos si sabes qué estás haciendo o si te interesa afinar el rendimiento del modelo.

Notas Finales.

-  **¿Qué hago si hay un error al ejecutar?**
Asegúrate de que el archivo `.csv` existe, tiene encabezados válidos, y no contiene datos faltantes importantes. El modelo espera que todos los valores sean numéricos enteros.
-  **¿Dónde se guardan los resultados?**
Al ejecutar la herramienta se creará automáticamente una carpeta en `results/` con la fecha y hora del análisis. Dentro de ella encontrarás:
 - El PDF con los resultados y métricas.
 - Un archivo `.csv` con las correlaciones encontradas.
 - El gráfico de correlaciones.
 - Archivos `.pkl` con los modelos entrenados.
-  **¿Qué hago si no sé qué columnas usar?**
Deja el campo `"columns"` vacío. La herramienta detectará automáticamente las columnas más relevantes según su correlación con la variable objetivo.
-  **¿Necesito conexión a internet o instalar más cosas?**
No. Si usas el ejecutable (`.exe`) no necesitas instalar Python ni librerías. Solo asegúrate de tener los archivos necesarios en la misma carpeta.

Autoría y condiciones de uso.

Este software y la herramienta descrita fueron desarrollados por:

Josue Norberto Aguilar Islas

josue.aguilar@alumnos.udg.mx

16 de Mayo del 2025