**Áreas de Oportunidad (Ideas para el futuro)**

Estas no son críticas, sino sugerencias para hacer tu librería aún más potente, basadas en su sólida base actual.

1. **Comparación de Modelos:** El flujo actual parece ser "un reporte por modelo". Una funcionalidad increíblemente útil sería poder comparar varios modelos en un solo reporte. Imagina una API como esta:

Python

make\_comparison\_report(

y\_true,

predictions={

"Logistic Regression": y\_pred\_lr,

"Random Forest": y\_pred\_rf

},

probabilities={

"Logistic Regression": y\_proba\_lr,

"Random Forest": y\_proba\_rf

},

path="comparison.md"

)

Esto podría generar una tabla comparando las métricas de cada modelo y gráficos superpuestos (como varias curvas ROC en una misma imagen).

1. **Interactividad opcional (HTML):** Aunque la belleza de los PNG es su simplicidad, podrías ofrecer un formato de salida alternativo. Un output\_format="html" podría usar librerías como [Plotly](https://plotly.com/python/) o [Bokeh](https://bokeh.org/) para generar gráficos interactivos (con tooltips que muestren valores, etc.). Esto sería un complemento, no un reemplazo, del formato Markdown.
2. **Mayor personalización:** Permitir a los usuarios un mayor control sobre la salida podría ser útil. Por ejemplo:
   * **Selección de métricas:** Un parámetro metrics=['accuracy', 'f1\_macro', 'roc\_auc'] para incluir solo las deseadas.
   * **Estilo de los gráficos:** Un diccionario plot\_config para pasar argumentos a Matplotlib y personalizar colores, títulos, etc.
3. **Ampliación de Métricas y Gráficos:** Mencionas las limitaciones actuales, y ese es el camino a seguir. Algunas adiciones de alto impacto podrían ser:
   * **Gráfico de Calibración:** Muy importante en clasificación para saber si las probabilidades predichas son confiables.
   * **Lift & Gain Charts:** Útiles en contextos de negocio (ej. marketing).
   * **Métricas para Clasificación Multiclase:** Añadir roc\_auc\_ovr\_weighted, y quizás curvas ROC micro/macro promediadas.

**Métricas**

* AUPRC (binaria y multiclase micro/macro).
* ROC/PR **micro/macro** globales en multiclase (además de OvR).
* Balanced Accuracy, **MCC**, **Cohen’s Kappa**, **Log Loss**.
* Top-k accuracy (k=3/5).
* Calibración: **Brier score**, curvas de confiabilidad.
* Regresión: **MAPE**, **MdAPE**, **RMSLE**, **MedAE**.
* Forecast: **MASE estacional** robusto, **sMAPE 2**, pinball-loss (quantiles), cobertura PIs.

**Gráficos**

* Matriz de confusión **normalizada** y por clase.
* Curvas PR/ROC agregadas (micro/macro) y por clase en una sola figura.
* Histogramas de residuales y QQ-plot.
* Curvas de calibración y **gain/lift**.

**Umbrales y análisis**

* Búsqueda de umbral óptimo por métrica (F1, J-statistic).
* Curvas precisión-recobrado vs umbral.
* Reporte de **importancia por clase** (aplica a algunas métricas).

**Reportes**

* Salida **HTML** (autocontenida) además de Markdown.
* Temas/plantillas Jinja2 personalizables (branding).
* Índice/TOC y resumen ejecutivo.
* Exportar a **JSON**/YAML las métricas.

**API/CLI**

* **Multi-run**: comparar varios modelos en un mismo informe.
* Soporte **multi-label**.
* Argumentos para **filtrar muestras** (máscaras) y **estratos**.
* Lectura de **Parquet**/Feather/NPZ además de CSV.
* Config por archivo (.evalcards.toml).

**Integraciones**

* **MLflow**/Weights & Biases: log de métricas e imágenes.
* sklearn: classification\_report enriquecido.
* Pandas Styler para tablas bonitas cuando se usa Jupyter.

**Rendimiento/robustez**

* Vectorización y **paralelización** opcional.
* Manejo explícito de NaNs/outliers.
* Semillas reproducibles y warnings claros.

**Calidad/DevOps**

* Tipado typing completo + **mypy**.
* **ruff/black** + pre-commit.
* Cobertura con **pytest-cov** + badge.
* CI matrix por **SO** (Linux/Windows/macOS) y Python.
* **MkDocs** con Material + GitHub Pages.
* Ejemplos en notebooks (examples/).

**Internacionalización**

* i18n: ES/EN (mensajes, títulos, métricas).
* Localización de formatos numéricos.

**priorizado (7 items)** para las próximas versiones:

**v0.3 – Salida y métricas clave**

1. **Reporte HTML autocontenido + export JSON**
   * make\_report(..., format="md|html") y --export-json con todas las métricas y paths.
2. **Métricas ampliadas (alto impacto)**
   * Clasificación: **AUPRC**, **Balanced Accuracy**, **MCC**, **Log Loss**.
   * Regresión: **MAPE**, **MedAE**, **RMSLE**.

**v0.4 – Multiclase y umbrales**

1. **ROC/PR micro & macro (multiclase)**
   * Curvas agregadas + roc\_auc\_macro, average\_precision\_macro.
2. **Análisis de umbral**
   * Búsqueda de umbral óptimo (F1/J), tabla y curva precisión–recobrado–F1 vs umbral.

**v0.5 – Calidad de probabilidades y comparación**

1. **Calibración**
   * **Brier score**, curva de confiabilidad y (opcional) reporte de binning.
2. **Comparación multi-modelo en un informe**
   * Ingesta de múltiples pares (y\_pred/proba) y tabla comparativa + “mejor por métrica”.

**v0.6 – DX y extensibilidad**

1. **Mejoras de DX/CLI**
   * Soporte **Parquet/Feather/NPZ**, archivo de config (.evalcards.toml), plantillas de nombre de salida (timestamp/model/tag), y **MkDocs** con ejemplos ejecutables.

 **Soporte multi-label**  
Cada muestra puede tener **varias etiquetas verdaderas** a la vez.  
Requiere y\_true y y\_proba como matrices (n\_samples, n\_classes).  
Métricas típicas: **precision/recall/F1 micro/macro/samples**, subset accuracy, AUC/PR por clase y agregadas. Umbrales por clase.

 **Métricas de ranking (MAP/NDCG)**  
Evalúan **listas ordenadas por score** (recomendación, retrieval, detección de anomalías).  
**MAP**: promedio del *Average Precision* por consulta. **NDCG@k**: ganancia normalizada con relevancia binaria o graduada.  
Suelen necesitar un query\_id/grupo por muestra.

 **Curvas de calibración con bins configurables**  
Compara **probabilidades predichas vs frecuencia observada** por bins.  
Parámetros: número de bins, **equal-width** vs **equal-freq**.  
Métricas asociadas: **Brier score**, **ECE/MCE**. Útil para evaluar/calibrar y\_proba.

 **QQ-plot e histograma de residuales (regresión)**  
Diagnóstico de errores: distribución, colas, asimetría.  
**Histograma**: forma general de los residuales. **QQ-plot**: residuales vs cuantiles normales (normalidad).  
Opcional: estadísticos (p. ej., Jarque-Bera, Shapiro) y notas de heterocedasticidad.

 **i18n ES/EN (mensajes y etiquetas)**  
Diccionario de textos (ES/EN) y parámetro lang.  
Traduce títulos, nombres de métricas, ejes y leyendas; **formato local** de números/porcentajes.  
*Fallback* a inglés y opción de aportar traducciones personalizadas.