



UNIVERSIDAD
NACIONAL
DE COLOMBIA

PROYECTO CULTURAL, CIENTÍFICO Y COLECTIVO DE NACIÓN

Análisis sobre la justicia de los modelos más usuales en seguridad predictiva

Diego Alejandro Hernández Castañeda
Cristian Alejandro Pulido Quintero

Francisco Gómez PHD

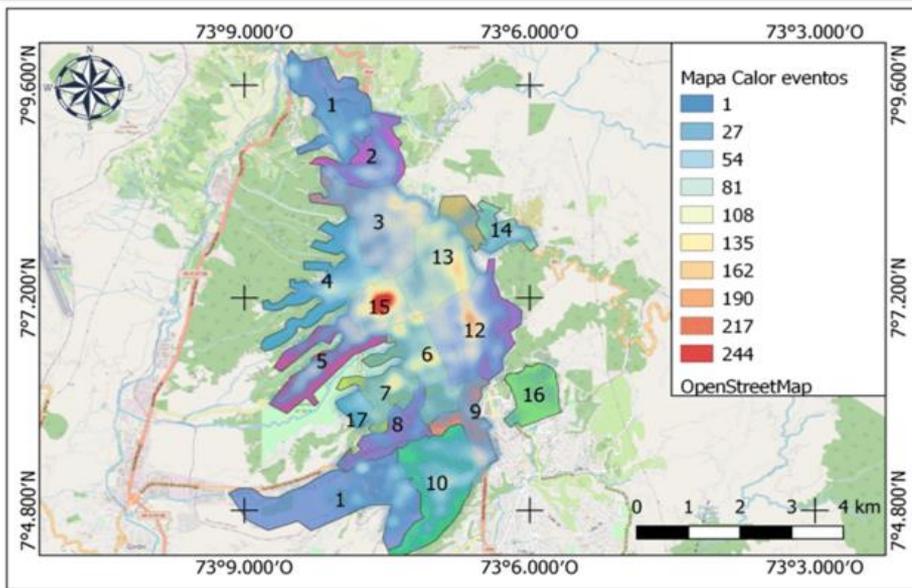
Facultad de Ingeniería, Departamento de Ingeniería de
Sistemas e Industrial Bogotá

Universidad Nacional de Colombia

PROYECTO CULTURAL, CIENTÍFICO Y COLECTIVO DE NACIÓN

Seguridad Predictiva

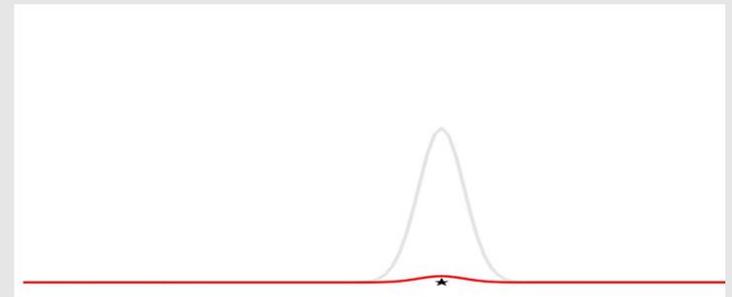
En la seguridad predictiva, se emplean datos y análisis para prever delitos potenciales, lo que ayuda a las autoridades a tomar medidas preventivas y asignar recursos de forma eficaz.



$$\lambda(s) = \frac{1}{n} \sum_{i=1}^n f(\|s - s_i\|)$$

$$\lambda(t, s) = \sum_{i:t_i < t} f(\|s - s_i\|)g(t - t_i)$$

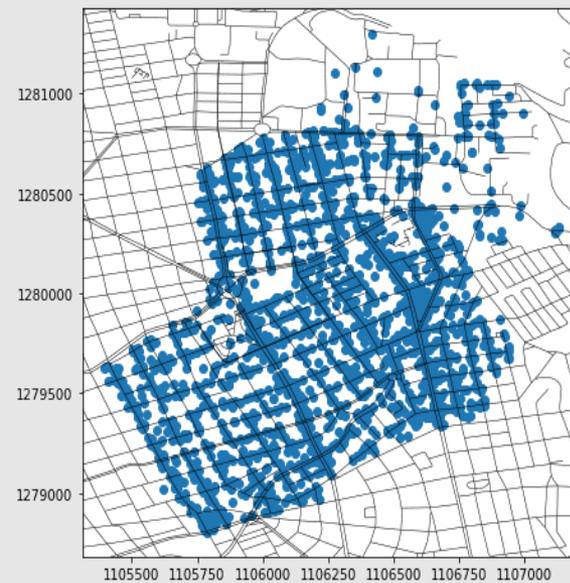
$$\lambda(t, x, y) = v(t)\mu(x, y) + \sum_{k:t_k < t} g(t - t_k, x - x_k, y - y_k)$$



Seguridad Predictiva

Factores que alteran el desempeño

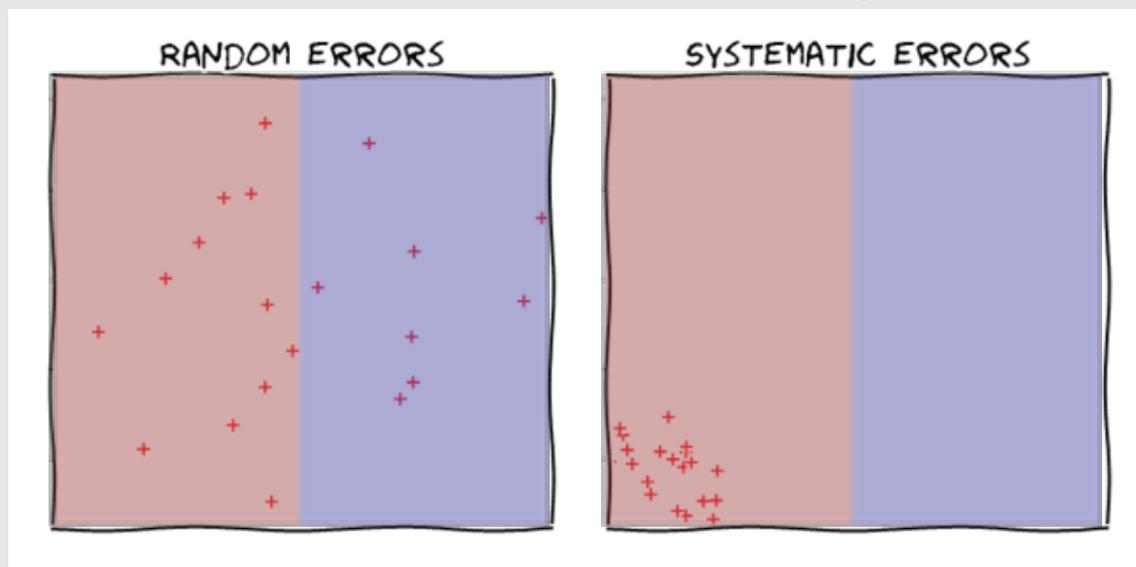
- Imposibilidad de registrar todos los eventos.
- La dinámica de reporte no es uniforme.
- Datos insuficientes para capturar la dinámica.
- Los supuestos del modelo implementado.



¡Estos sistemas pueden exponer sesgos en las predicciones, causando variaciones en el desempeño por regiones!

Predicción Crimen / Estructura usual

Dos Predicciones con 5% de error promedio

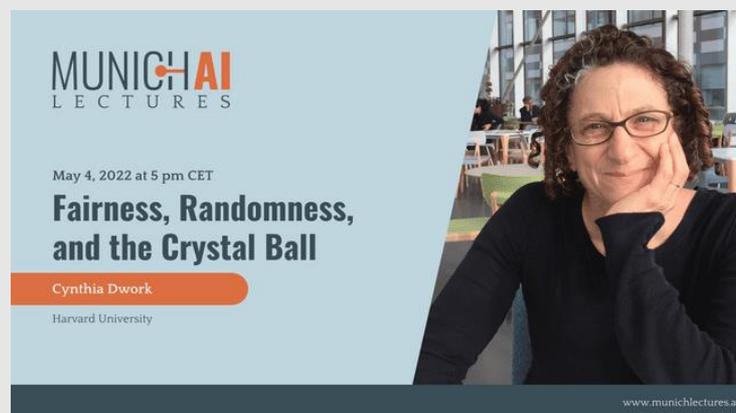


MORITZ HARDT, (2017)

¿Qué es justicia?

“Principio de oportunidad”, en este contexto, el desempeño de la predicción de un modelo debe ser independiente de las características de las zonas.

Cynthia Dwork, en su artículo "Fairness through awareness" (2011), fue precursora en abordar la justicia en modelos de machine learning.



Pequeño estado del arte

A technique to improve both fairness and accuracy in artificial intelligence
Methods that meet...

Achieving Fairness through Adversarial Learning: an Application to Recidivism Prediction

Christina Wadsworth
Stanford University
Stanford, CA
cwads@cs.stanford.edu

Francesca Vera
Stanford University
Stanford, CA
fvera@cs.stanford.edu

Chris Piech
Stanford University
Stanford, CA
piech@cs.stanford.edu

IBM Research AI
Yorktown Heights, NY, USA

A Surv

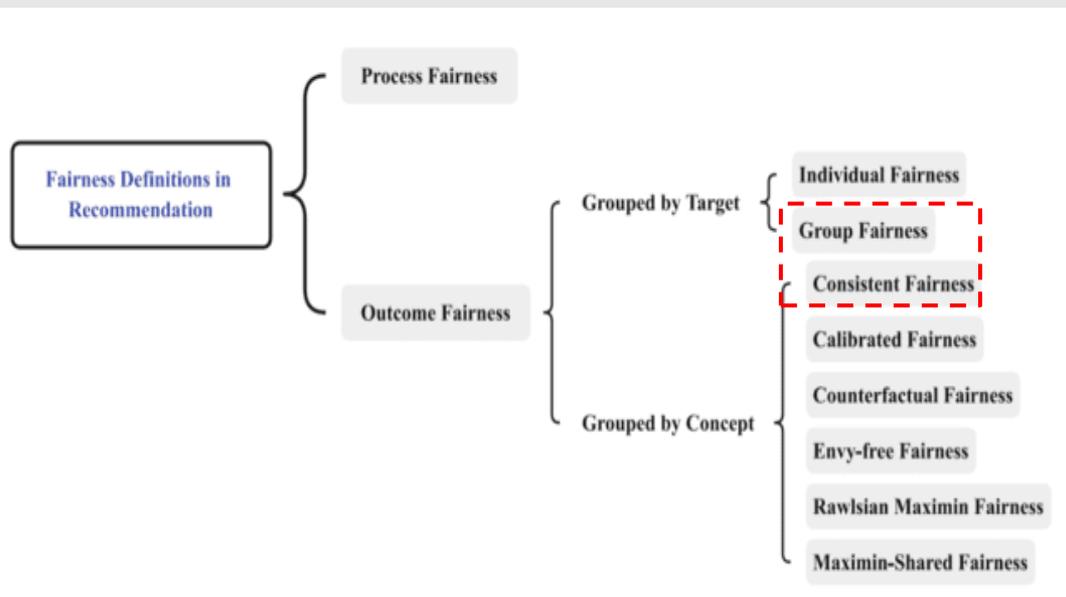
YIFAN WA

WEIZHI MA, Institute for AI Industry Research (AIR), Zhejiang University, China
MIN ZHANG, YIQUN LIU, and SHAOPING MA, Tsinghua University, China

2015 2016 2017

MIT (2022)

Métricas de justicia



WANG, Y. (2023)

- Actualmente no hay consenso sobre la definición.
- La mayoría de las definiciones se aplica sobre clasificadores.
- Variables protegidas
- Utilidad calcula un beneficio dado por el modelo en una región específica
- Lo cual permite comparar los beneficios otorgados entre regiones

Métricas de justicia

$$\text{Variance} = \frac{1}{|\mathcal{V}|^2} \sum_{v_x \neq v_y} (f(v_x) - f(v_y))^2$$

$$\text{MMD} = \max\{f(v), \forall v \in \mathcal{V}\} - \min\{f(v), \forall v \in \mathcal{V}\}$$

$$\text{Gini} = \frac{\sum_{v_x, v_y \in \mathcal{V}} |f(v_x) - f(v_y)|}{2|\mathcal{V}| \sum_v f(v)}$$

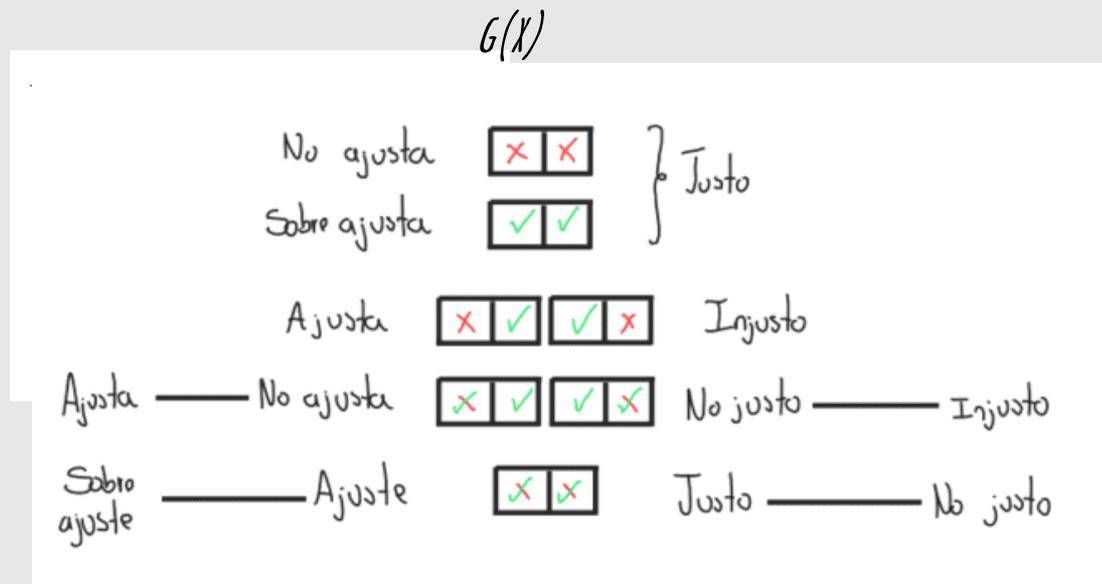
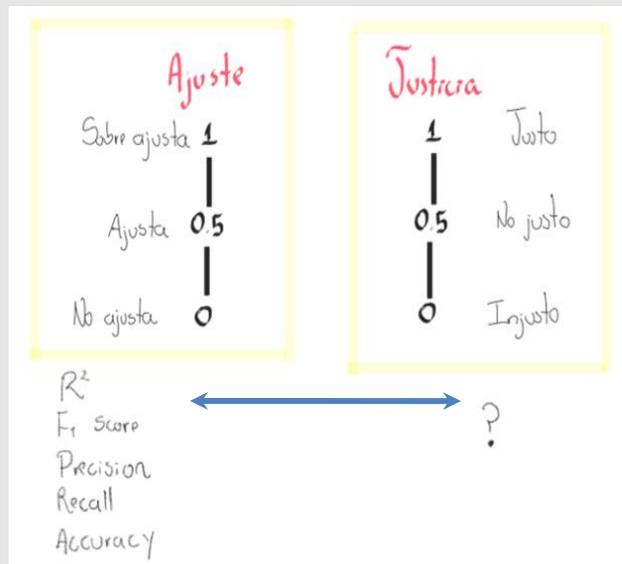
¿Cual es el promedio de diferencias entre la utilidad de los grupos?

¿Cual es la máxima diferencia de utilidad entre cualquier par de grupos?

¿Cómo se distribuye la utilidad entre los grupos?

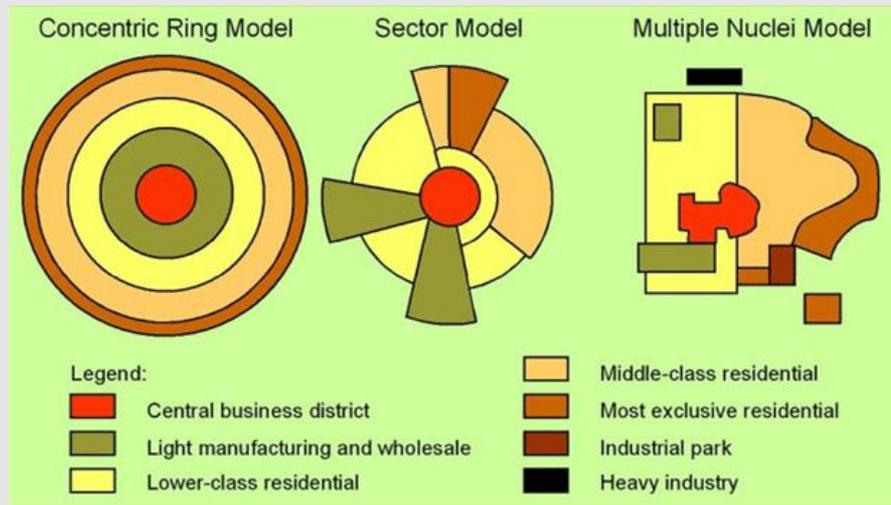
¿Qué tan justos son los modelos de predicción de crímenes más utilizados en el contexto local?

DADO UN MODELO $G(X)$:



Diseño experimental

1. Variable protegida



DAVIS, J. (2017)

- En el contexto latinoamericano existen muchas variables de segregación de la población.
- Uno de los mayor interés y estudio es el nivel de ingresos, donde se ha encontrado que el modelo sectorizado es el que mejor caracteriza esta dinámica en el contexto local.
- Variable protegida: *Población con ingresos bajos.*

Diseño experimental

2. Función de beneficio

Una zona se ve más beneficiada si su predicción se acerca a la realidad.

$f(\text{predicción, realidad, Grupo})$

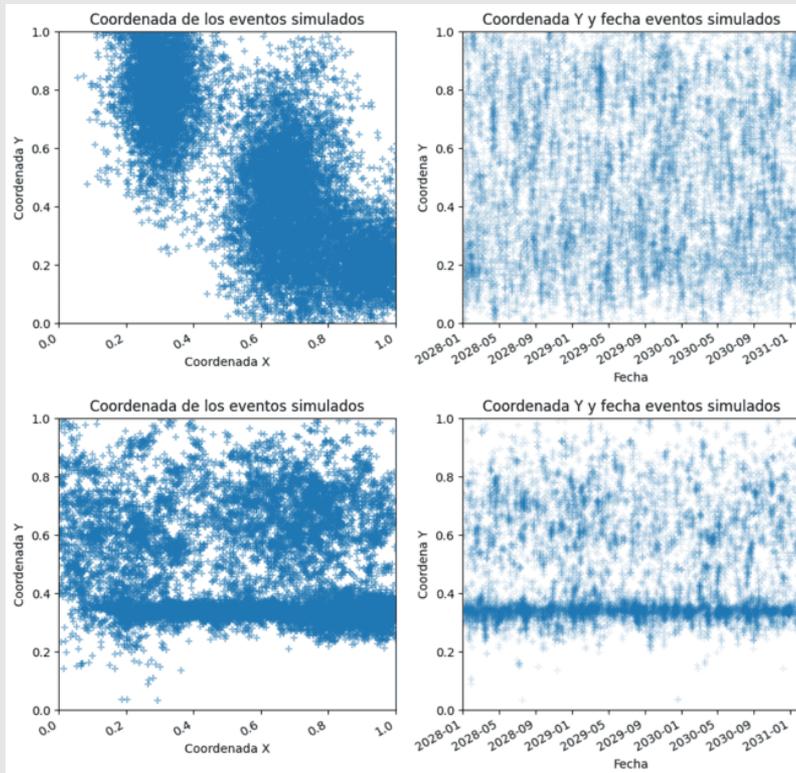
$$\frac{1}{|G|} \sum_{c \in G} |P(c) - Q(c)|$$

2. Justicia

- Variance
- Max Min Difference
- Gini

Diseño experimental

3. Datos y modelos



¿por qué?

- Capturan la dependencia temporal en los datos de eventos, lo que puede ser beneficioso para reflejar la dinámica real de crímenes, ya que los crímenes a menudo exhiben patrones de recurrencia.

¿por qué los modelos? (NAIVE, KDE, SEPP)

- Hay restricciones presupuestarias.
- No tan sofisticados con una efectividad aceptable.

Diseño experimental

4. Evaluación de desempeños

Earth mover's distance

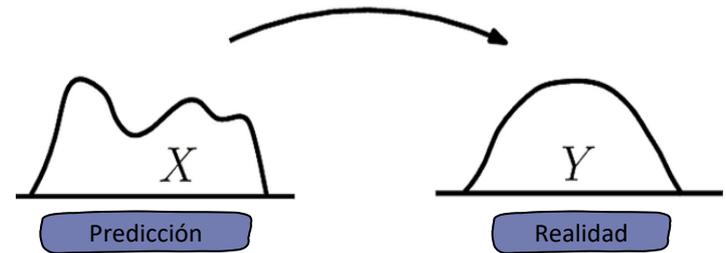
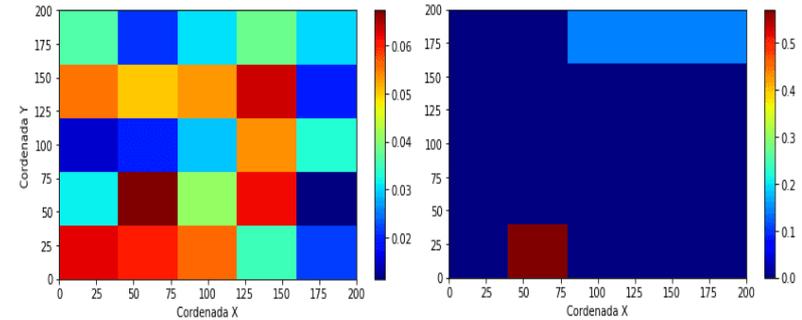
$$\min_{f_{i,j}} \sum_{i=1}^d \sum_{j=1}^d f_{i,j} d_{i,j}$$

$$\text{Sujeto a: } \sum_{i=1}^d f_{i,j} = w_{p_j}, \quad 1 \leq j \leq d$$

$$\sum_{j=1}^d f_{i,j} = w_{q_i}, \quad 1 \leq i \leq d$$

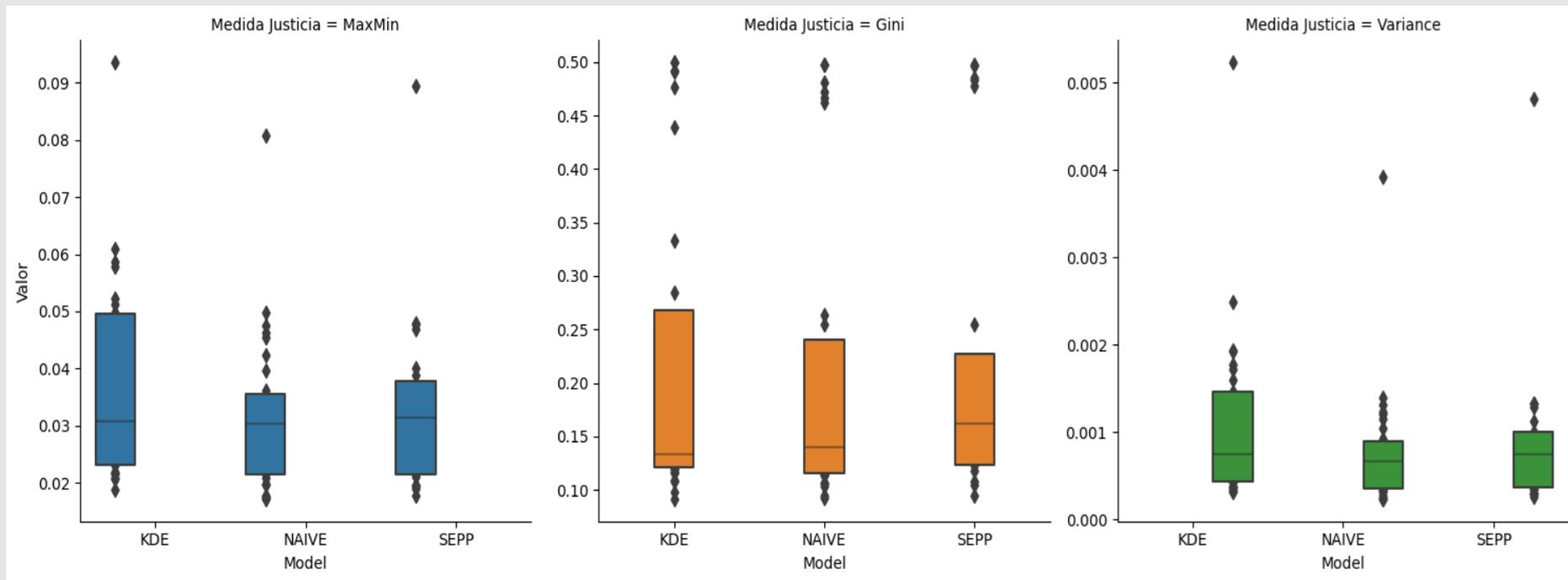
$$\sum_{i=1}^m \sum_{j=1}^n f_{i,j} = 1$$

$$EMD(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n f_{i,j} d_{i,j}}{\sum_{i=1}^m \sum_{j=1}^n f_{i,j}} = \sum_{i=1}^m \sum_{j=1}^n f_{i,j} d_{i,j}$$



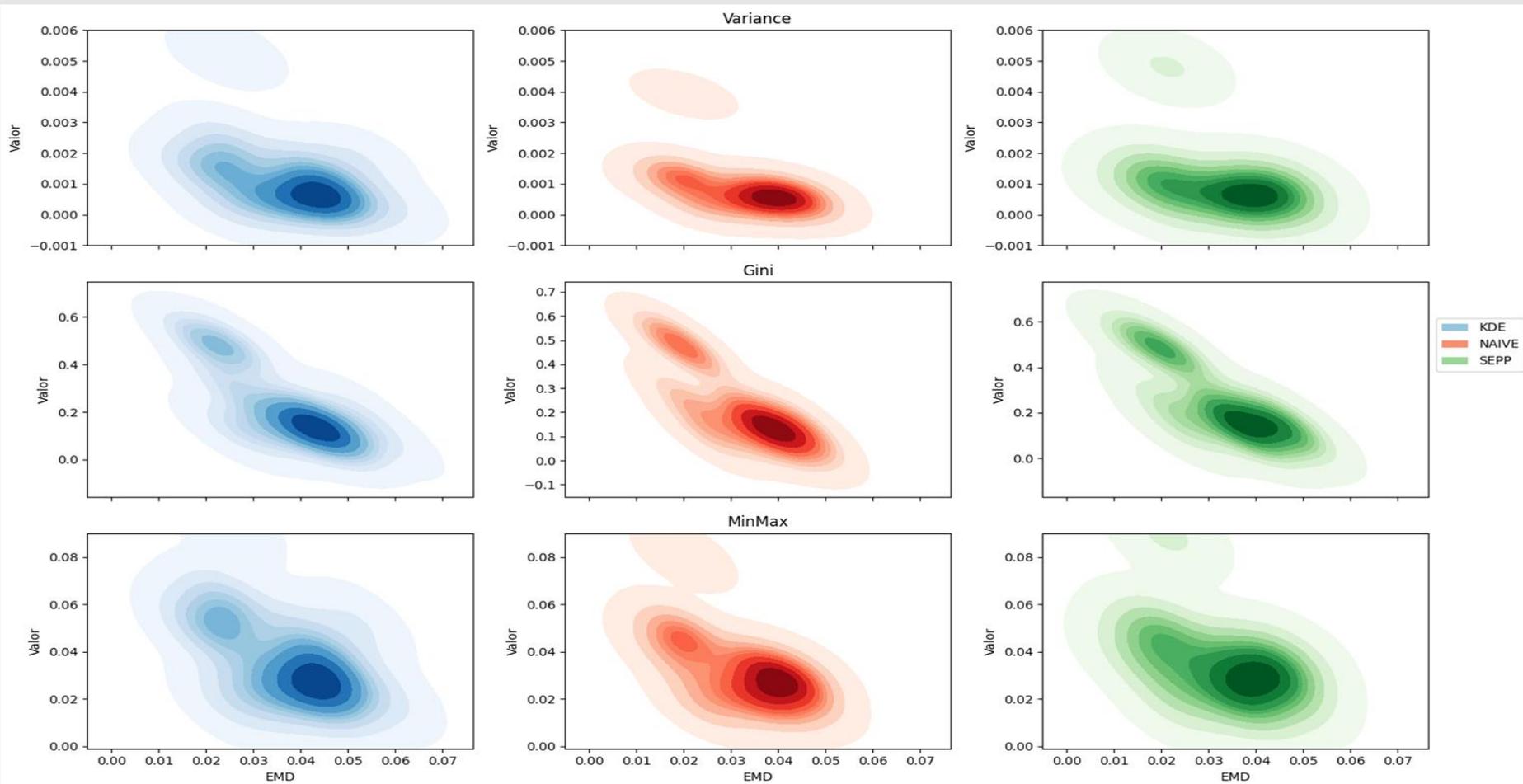
Resultados

Distribución de las medidas de justicia para los escenarios simulados



Resultados

Comparación entre las medidas de justicia y el desempeño de los modelos



Trabajo futuro

- Comportamiento de la justicia en datos reales.
- Funciones para establecer el beneficio.
- Nuevas configuraciones de variables protegidas.
- Modelos con más de una variable protegida.
- ¿Cómo el subreporte está afectado la justicia?

Referencias

- Crime mapping and public safety. (n.d.). <https://www.caliper.com/maptitude/crime/crime-mapping-software-uk.htm>
- U. M. Butt et al., "Spatio-Temporal Crime Predictions by Leveraging Artificial Intelligence for Citizens Security in Smart Cities," in IEEE Access, vol. 9, pp. 47516-47529, 2021, doi: 10.1109/ACCESS.2021.3068306.
- A technique to improve both fairness and accuracy in artificial intelligence. (2022, July 20). MIT News Massachusetts Institute of Technology. <https://news.mit.edu/2022/fairness-accuracy-ai-models-0720>
- Wang, Y., Ma, W., Zhang, M., Liu, Y., & Ma, S. (2023). A survey on the fairness of recommender systems. ACM Transactions on Information Systems, 41(3), 1–43. <https://doi.org/10.1145/3547333>
- Moritz Hardt, (2017). CS 294: Fairness in Machine Learning. (s. f.). <https://fairmlclass.github.io/1.html#/11>
- Davis, J. (2017). Spatial Models of Urban Land Use (Ch. 13). ppt video online download. <https://slideplayer.com/slide/4637186/>

Gracias

Universidad Nacional de Colombia

PROYECTO CULTURAL, CIENTÍFICO Y COLECTIVO DE NACIÓN