

Modelos caja negra

Modelos subrogados
(locales y globales)

Métricas (locales y
globales) del efecto de las
variables sobre la
predicción

Explicabilidad basada en
ejemplos

Modelos transparentes

Modelos intrínsecamente
interpretables, pero no
optimizados para
interpretabilidad

Modelos intrínsecamente
interpretables,
optimizados para
interpretabilidad