

Problem Set 3: Predicting Poverty

Curso: BIG DATA Y MACHINE LEARNING PARA ECONOMÍA APLICADA-UNIFICADO (MECA4107_1-2)

Criterios	Insuficiente	Regular	Suficiente	Sobresaliente	Puntuación del criterio
Introducción (5%)	1.25 puntos No plantea la introducción	2.5 puntos Presenta la introducción, pero no se describe el problema o se incluyen antecedentes. No contiene un resumen de los resultados y conclusiones principales.	3.75 puntos Se plantea la introducción, aunque existen algunas imprecisiones que no afectan la comprensión del problema. Contiene un resumen de los resultados y las principales conclusiones.	5 puntos La introducción enuncia con claridad y precisión el problema, además explica los antecedentes, datos y pertinencia para responder a dicho problema. Contiene un resumen de los resultados y las principales conclusiones.	/5

Criterios	Insuficiente	Regular	Suficiente	Sobresaliente	Puntuación del criterio
Descripción de los Datos (15%)	3.75 puntos No describe los datos	7.5 puntos Realizan una descripción de la base de datos, aunque no es muy claro el propósito de esta y la utilidad que puede representar para abordar el problema planteado en el taller. No es claro cómo se obtiene la muestra final de los datos. El análisis descriptivo no se encuentra articulado consistentemente con las tablas y/o gráficos presentados. No es claro el uso de conocimiento profesional que agregue valor a la sección.	11.25 puntos Existe una descripción básica de de la base de datos, aunque falta mayor precisión de su utilidad para abordar el problema planteado en el taller. Hay imprecisiones en la descripción de cómo se obtiene la muestra final. Aunque realiza el análisis descriptivo de los datos, algunas de las tablas o gráficos presentan inconsistencias o no permiten comprender plenamente la variación de los datos y la elección de las variables.	15 puntos Realizan una descripción de la base de datos y su utilidad en la solución del problema planteado en el taller. Se describe cómo se obtiene la muestra final y se realiza el análisis descriptivo de los datos, donde se presentan tablas y/o gráficos. Estos ayudan al lector a entender la variación de los datos y la elección de las variables. Se usa conocimiento profesional para agregar valor a la sección.	/15
Modelo basado en clasificación (10%)	2.5 puntos No se presenta la sección.	5 puntos La descripción del enfoque del modelo de clasificación no es precisa, no se incluyen los detalles suficientes de como fue abordado el problema.	7.5 puntos Se describe el enfoque del modelo de clasificación. Existen algunas impresiones que dificultan la comprensión del enfoque.	10 puntos Se describe el enfoque del modelo de clasificación. La descripción es clara y la estrategia es clara. Se presentan las ecuaciones correctas con sus debidas explicaciones	/10

Criterios	Insuficiente	Regular	Suficiente	Sobresaliente	Puntuación del criterio
Modelo basado en regresión (10%)	2.5 puntos No se presenta la sección	5 puntos La descripción del enfoque del modelo de clasificación no es precisa, no se incluyen los detalles suficientes de como fue abordado el problema.	7.5 puntos Se describe el enfoque del modelo de regresión. Existen algunas impresiones que dificultan la comprensión del enfoque.	10 puntos Se describe el enfoque del modelo de regresión. La descripción es clara y la estrategia es clara. Se presentan las ecuaciones correctas con sus debidas explicaciones	/10
Modelo Final (35%)	8.75 puntos No se presenta la sección	17.5 puntos La descripción del modelo final seleccionado para la competencia no es precisa, no se incluyen los detalles suficientes de como fue estimado, cuáles fueron las estrategias de sub-muestreo utilizadas, o en los hiper-parámetros relevantes. La descripción de las variables incluidas en el modelo final y en la discusión sobre la importancia relativa en la predicción no es la adecuada. No se hace la comparación con otros modelos.	26.25 puntos Se describe el modelo final seleccionado para la competencia en Kaggle, aunque existen imprecisiones sobre cómo fue estimado, cuáles fueron las estrategias de sub-muestreo utilizadas, o en los hiper-parámetros relevantes. Existen también imprecisiones en la descripción de las variables incluidas en el modelo final y en la discusión sobre la importancia relativa en la predicción. No se compara con al menos otras 2 especificaciones para cada enfoque.	35 puntos Se describe el modelo final seleccionado para la competencia en Kaggle Se incluyen detalles de la estimación, incluyendo cuáles fueron las estrategias de sub-muestreo utilizadas, o de los hiper-parámetros relevantes. Se describen las variables utilizadas en el modelo incluyendo la importancia relativa de las mismas en la predicción. Se compara con al menos otras 2 especificaciones para cada enfoque.	/35

Criterios	Insuficiente	Regular	Suficiente	Sobresaliente	Puntuación del criterio
Conclusión (5%)	1.25 puntos No presenta la conclusión.	2.5 puntos Presenta la conclusión, pero no se resumen los resultados adecuadamente, ni las recomendaciones que surgen sobre la utilidad de los diferentes modelos predictivos	3.75 puntos El resumen del trabajo no es claro, existen imprecisiones sobre la utilidad de los diferentes modelos predictivos.	5 puntos Se resumen claramente los resultados del trabajo y las recomendaciones que surgen sobre la utilidad de los diferentes modelos predictivos	/5
“Entrega” en Kaggle (10%)	2.5 puntos No se hizo entrega en Kaggle.	5 puntos Hay menos de 4 entregas en Kaggle de cada modelo y estas no están debidamente identificadas siguiendo los parámetros establecidos de la competencia.	7.5 puntos Hay menos de 6 entregas en Kaggle de cada modelo y estas no están debidamente identificadas siguiendo los parámetros establecidos de la competencia.	10 puntos Hay al menos 6 entregas en Kaggle 3 de clasificacion y 3 de regresion y estas están debidamente identificadas siguiendo los parámetros establecidos de la competencia.	/10

Criterios	Insuficiente	Regular	Suficiente	Sobresaliente	Puntuación del criterio
Documento y Repositorio de GitHub (10%)	2.5 puntos No contiene un repositorio	5 puntos El documento es autocontenido, pero excede las 10 páginas o las 8 tablas o figuras. Existe el repositorio y contiene un README. El repositorio no sigue el template, o el README no permite navegar el repositorio o reproducir el trabajo. No hay contribuciones significativas por parte de los miembros. El código no es fácilmente legible o reproducible.	7.5 puntos El documento no es completamente autocontenido, no excede las 10 páginas y las 8 tablas o figuras. Existe el repositorio siguiendo el template provisto. El README tiene algunas imprecisiones que no afectan la comprensión del lector para navegar y replicar el trabajo. La rama del repositorio principal muestra más de (2) y menos de cinco (5) contribuciones que aportan significativamente al taller, por parte de cada miembro del equipo. El código en el repositorio es reproducible y reproducible pero su legibilidad no es buena.	10 puntos El documento es autocontenido, no excede las 10 páginas y las 8 tablas o figuras. Existe el repositorio siguiendo el template provisto y contiene un README que ayuda al lector a navegar el repositorio e incluye instrucciones breves para replicar completamente el trabajo. La rama del repositorio principal muestra al menos cinco (5) contribuciones que aportan significativamente al taller, por parte de cada miembro del equipo. El código en el repositorio es reproducible, es legible y reproducible	/10

Total

/100

Puntuación general

Nivel 1

0 puntos mínimos

Nivel 2

5 puntos mínimos

Nivel 3

8 puntos mínimos

Nivel 4

11 puntos mínimos