

Actividades > Problem Set 2: Predicting Poverty

Problem Set 2: Predicting Poverty

Curso: BIG DATA Y MACHINE LEARNING PARA ECONOMÍA APLICADA-UNIFICADO (ECON4676_2-MECA4107_1)

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
-----------	--------------	---------	------------	-------	-----------	---------------	-------------------------

Categoría de grupo

Equipos Taller 1

Nombre de grupo

Equipo 4

Instrucciones

Este segundo taller tiene como objetivo aplicar los conceptos aprendidos sobre clasificación en datos reales que buscan predecir pobreza en los hogares colombianos. Se deberá entregar además del documento y el repositorio, las predicciones en la plataforma Kaggle.

Para unirse a la competición debe usar el siguiente link:

<https://www.kaggle.com/t/59ef509497064da4b855e6c0148af484>

Vence el 14 de abril de 2024 23:59

Disponible el 4 de marzo de 2024 00:01. Acceso restringido antes de que inicie la disponibilidad.

Disponible hasta 15 de abril de 2024 06:00. Acceso restringido después de que finalice la disponibilidad.

Datos adjuntos

 [Problem_Set_2.pdf](#) (110.35 KB)

Descargar todos los archivos

▼ [Ocultar rúbricas](#)

Nota: La evaluación de actividades se aplicará a todos los miembros de este grupo.

Nombre de rúbrica: Problem Set 2: Predicting Poverty

 Imprimir

Enviar actividad

Archivos para enviar *

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
Introducción (5%)	0 puntos El elemento no está presente	2 puntos Tres o más de los siguientes elementos están faltando y/o contiene imprecisiones que no permiten	3.5 puntos Dos de los siguientes elementos están faltando y/o contiene imprecisiones que afectan	4 puntos Uno de los siguientes elementos está faltando y/o contiene imprecisiones que afectan la	4.5 puntos Están presentes todos los siguientes elementos, pero contiene imprecisiones que no afectan a	5 puntos Están presentes todos los siguientes elementos y desarrollados de forma clara y precisa:	/5

Enviar

Cancelar

		1. Se enuncia con claridad y precisión el problema, 2. Se menciona n antecedentes relevantes y debidamente documentados 3. Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller. 4. Contiene un resumen de los resultados , menciona ndo los modelos “ganadores” enviados a Kaggle.	la sección: 1. Se enuncia con claridad y precisión el problema, 2. Se menciona n antecedentes relevantes y debidamente documentados 3. Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller. 4. Contiene un resumen de los resultados , menciona ndo los modelos “ganadores”	1. Se enuncia con claridad y precisión el problema, 2. Se menciona n antecedentes relevantes y debidamente documentados 3. Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller. 4. Contiene un resumen de los resultados , menciona ndo los modelos “ganadores” enviados a Kaggle.	1. Se enuncia con claridad y precisión el problema, 2. Se menciona n antecedentes relevantes y debidamente documentados 3. Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller. 4. Contiene un resumen de los resultados , menciona ndo los modelos “ganadores” enviados a Kaggle.	con claridad y precisión el problema, 2. Se menciona n antecedentes relevantes y debidamente documentados 3. Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller. 4. Contiene un resumen de los resultados , menciona ndo los modelos “ganadores” enviados a Kaggle. 5. Se presentan	
--	--	--	--	--	--	---	--

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
		5. Se presentan breves conclusiones del análisis.	5. Se enviados a Kaggle. Se presentan breves conclusiones del análisis.	5. Se presentan breves conclusiones del análisis.	5. Se presentan breves conclusiones del análisis.	breves conclusiones del análisis.	

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
Descripción de los Datos (15%)	0 puntos El elemento no está presente	7 puntos Tres o más de los siguientes elementos están faltando y/o contiene imprecisiones que no permiten la comprensión de la sección: 1. Se realiza una descripción completa de la fuente de datos, original y la provista para el taller. 2. Se discute la utilidad de la base para la solución del problema planteado en el taller. 3. Se describe como se obtiene la muestra final: incluyend o el unido de base, la transformación de variables y las variables incluidas en el análisis 4. Se realiza un análisis	10 puntos Dos de los siguientes elementos están faltando y/o contiene imprecisiones que afectan significativamente la comprensión de la sección: 1. Se realiza una descripción completa de la fuente de datos, original y la provista para el taller. 2. Se discute la utilidad de la base para la solución del problema planteado en el taller. 3. Se describe como se obtiene la muestra final: incluyend o el unido de base, la transformación de variables y las variables incluidas en el análisis	12 puntos Uno de los siguientes elementos está faltando y/o contiene imprecisiones que afectan la comprensión de la sección: 1. Se realiza una descripción completa de la fuente de datos, original y la provista para el taller. 2. Se discute la utilidad de la base para la solución del problema planteado en el taller. 3. Se describe como se obtiene la muestra final: incluyend o el unido de base, la transformación de variables y las variables incluidas en el análisis 4. Se realiza un análisis	14 puntos Están presentes todos los siguientes elementos, pero contiene imprecisiones que no afectan a la comprensión de la sección: 1. Se realiza una descripción completa de la fuente de datos, original y la provista para el taller. 2. Se discute la utilidad de la base para la solución del problema planteado en el taller. 3. Se describe como se obtiene la muestra final: incluyend o el unido de base, la transformación de variables y las variables incluidas en el análisis 4. Se realiza un análisis	15 puntos Están presentes todos los siguientes elementos y desarrollados de forma clara y precisa: 1. Se realiza una descripción completa de la fuente de datos, original y la provista para el taller. 2. Se discute la utilidad de la base para la solución del problema planteado en el taller. 3. Se describe como se obtiene la muestra final: incluyend o el unido de base, la transformación de variables y las variables incluidas en el análisis 4. Se realiza un análisis descriptivo de las	/15

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
		descriptiv o de las variables utilizadas en el análisis donde se presentan tablas y/o gráficos. 5. El análisis descriptiv o permite al lector entender la variación de los datos y la elección de las variables. 6. Se utiliza el conocimi ento profesion al y los antecedentes para agregar valor a la sección.	4. Se realiza un análisis descriptiv o de las variables utilizadas en el análisis donde se presentan tablas y/o gráficos. 5. El análisis descriptiv o permite al lector entender la variación de los datos y la elección de las variables. 6. Se utiliza el conocimi ento profesion al y los antecedentes para agregar valor a la sección.	descriptiv o de las variables utilizadas en el análisis donde se presentan tablas y/o gráficos. 5. El análisis descriptiv o permite al lector entender la variación de los datos y la elección de las variables. 6. Se utiliza el conocimi ento profesion al y los antecedentes para agregar valor a la sección.	descriptiv o de las variables utilizadas en el análisis donde se presentan tablas y/o gráficos. 5. El análisis descriptiv o permite al lector entender la variación de los datos y la elección de las variables. 6. Se utiliza el conocimi ento profesion al y los antecedentes para agregar valor a la sección.	variables utilizadas en el análisis donde se presentan tablas y/o gráficos. 5. El análisis descriptiv o permite al lector entender la variación de los datos y la elección de las variables. 6. Se utiliza el conocimi ento profesion al y los antecedentes para agregar valor a la sección.	

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
Modelo basado en clasificación (10%)	0 puntos El elemento no está presente	3 puntos Dos o más de los siguientes elementos están faltando y/o se utilizan solo un algoritmo y/o contiene imprecisiones que no permiten la comprensión de la sección: 1. Se explica claramente el intento de predecir directamente pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente el algoritmo utilizado para la entrega	5 puntos Dos de los siguientes elementos están faltando y/o se utilizan solo dos algoritmos distintos y/o contiene imprecisiones que afectan significativamente la comprensión de la sección: 1. Se explica claramente el intento de predecir directamente pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente los algoritmos utilizados para las entregas	7 puntos Uno de los siguientes elementos está faltando y/o contiene imprecisiones que afectan la comprensión de la sección: 1. Se explica claramente el intento de predecir directamente pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente los algoritmos utilizados para las entregas 4. Se utilizan al menos 3 algoritmos distintos	9 puntos Están presentes todos los siguientes elementos, pero contiene imprecisiones que no afectan a la comprensión de la sección: 1. Se explica claramente el intento de predecir directamente pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente los algoritmos utilizados para las entregas 4. Se utilizan al menos 3 algoritmos distintos	10 puntos Están presentes todos los siguientes elementos y desarrollados de forma clara y precisa: 1. Se explica claramente el intento de predecir directamente pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente los algoritmos utilizados para las entregas 4. Se utilizan al menos 3 algoritmos distintos	/10

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
Modelo basado en regresión (10%)	0 puntos El elemento no está presente	3 puntos Dos o más de los siguientes elementos están faltando y/o se utilizan solo un algoritmo y/o contiene imprecisiones que no permiten la comprensión de la sección: 1. Se explica clarament e el intento de predecir directamente y no pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente el algoritmo utilizado para la entrega	5 puntos Dos de los siguientes elementos están faltando y/o se utilizan solo dos algoritmos distintos y/o contiene imprecisiones que afectan significativamente la comprensión de la sección: 1. Se explica clarament e el intento de predecir directamente y no pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente los algoritmos utilizados para las entregas	7 puntos Uno de los siguientes elementos está faltando y/o contiene imprecisiones que afectan la comprensión de la sección: 1. Se explica clarament e el intento de predecir directamente y no pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente los algoritmos utilizados para las entregas 4. Se utilizan al menos 3 algoritmos distintos	9 puntos Están presentes todos los siguientes elementos, pero contiene imprecisiones que no afectan a la comprensión de la sección: 1. Se explica clarament e el intento de predecir directamente y no pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente los algoritmos utilizados para las entregas 4. Se utilizan al menos 3 algoritmos distintos	10 puntos Están presentes todos los siguientes elementos y desarrollados de forma clara y precisa: 1. Se explica clarament e el intento de predecir directamente y no pobres (1) y no pobres (0) 2. La explicación incluye las variables utilizadas en los modelos 3. Se describen brevemente los algoritmos utilizados para las entregas 4. Se utilizan al menos 3 algoritmos distintos	/10

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
Modelo Final (35%)	0 puntos El elemento no está presente	10 puntos Tres o más de los siguientes elementos están faltando y/o contiene imprecisiones que no permiten la comprensión de la sección: 1. Se describen detallada mente los dos mejores modelos o los elegidos para la evaluación en Kaggle. Un modelo es el del enfoque de clasificación y el otro modelo es el de clasificación 2. La descripción detalla el proceso de entrenamiento de los parámetros, la selección de los hiperparámetros y/o cualquier otra información	20 puntos Dos de los siguientes elementos están faltando y/o contiene imprecisiones que afectan significativamente la comprensión de la sección: 1. Se describen detallada mente los dos mejores modelos o los elegidos para la evaluación en Kaggle. Un modelo es el del enfoque de clasificación y el otro modelo es el de clasificación 2. La descripción detalla el proceso de entrenamiento de los parámetros, la selección de los hiperparámetros y/o cualquier otra	28 puntos Uno de los siguientes elementos está faltando y/o contiene imprecisiones que afectan la comprensión de la sección: 1. Se describen detallada mente los dos mejores modelos o los elegidos para la evaluación en Kaggle. Un modelo es el del enfoque de clasificación y el otro modelo es el de clasificación 2. La descripción detalla el proceso de entrenamiento de los parámetros, la selección de los hiperparámetros y/o cualquier otra información	32 puntos Están presentes todos los siguientes elementos, pero contiene imprecisiones que no afectan a la comprensión de la sección: 1. Se describen detallada mente los dos mejores modelos o los elegidos para la evaluación en Kaggle. Un modelo es el del enfoque de clasificación y el otro modelo es el de clasificación 2. La descripción detalla el proceso de entrenamiento de los parámetros, la selección de los hiperparámetros y/o cualquier otra información	35 puntos Están presentes todos los siguientes elementos y desarrollados de forma clara y precisa: 1. Se describen detallada mente los dos mejores modelos o los elegidos para la evaluación en Kaggle. Un modelo es el del enfoque de clasificación y el otro modelo es el de clasificación 2. La descripción detalla el proceso de entrenamiento de los parámetros, la selección de los hiperparámetros y/o cualquier otra información relevante para la	/35

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
		relevante para la comprensión de los modelos. 3. Se compara la performance predictiva de estos modelos a otros dos modelos por cada enfoque que se hayan enviado a Kaggle. 4. Se discute la importancia relativa de las variables utilizadas en cada uno de los dos mejores modelos. Se contrastan similitudes y diferencias. 5. Si se utiliza una estrategia de re-muestreo para el desbalance de clases. Esta esta explicada correctamente, con los detalles	información relevante para la comprensión de los modelos. 3. Se compara la performance predictiva de estos modelos a otros dos modelos por cada enfoque que se hayan enviado a Kaggle. 4. Se discute la importancia relativa de las variables utilizadas en cada uno de los dos mejores modelos. Se contrastan similitudes y diferencias. 5. Si se utiliza una estrategia de re-muestreo para el desbalance de clases. Esta esta explicada correctamente, con los detalles	relevante para la comprensión de los modelos. 3. Se compara la performance predictiva de estos modelos a otros dos modelos por cada enfoque que se hayan enviado a Kaggle. 4. Se discute la importancia relativa de las variables utilizadas en cada uno de los dos mejores modelos. Se contrastan similitudes y diferencias. 5. Si se utiliza una estrategia de re-muestreo para el desbalance de clases. Esta esta explicada correctamente, con los detalles	relevante para la comprensión de los modelos. 3. Se compara la performance predictiva de estos modelos a otros dos modelos por cada enfoque que se hayan enviado a Kaggle. 4. Se discute la importancia relativa de las variables utilizadas en cada uno de los dos mejores modelos. Se contrastan similitudes y diferencias. 5. Si se utiliza una estrategia de re-muestreo para el desbalance de clases. Esta esta explicada correctamente, con los detalles	comprensión de los modelos. 3. Se compara la performance predictiva de estos modelos a otros dos modelos por cada enfoque que se hayan enviado a Kaggle. 4. Se discute la importancia relativa de las variables utilizadas en cada uno de los dos mejores modelos. Se contrastan similitudes y diferencias. 5. Si se utiliza una estrategia de re-muestreo para el desbalance de clases. Esta esta explicada correctamente, con los detalles pertinentes.	

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
		pertinentes.	los detalles pertinentes.	pertinentes.	pertinentes. No se compara con al menos otras 2 especificaciones para cada enfoque.		

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
Conclusión (5%)	0 puntos El elemento no está presente	1 punto Tres o más de los siguientes elementos están faltando y/o contiene imprecisiones que no permiten la comprensión de la sección: 1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión 2. Se menciona cuales fueron los “mejores” modelos entrenado s. 3. Se distingue cual fue el “mejor” en términos predictivo s de todos los modelos. 4. Se reflexiona por qué este es el caso 5. Se reflexiona sobre las variables que incluye el “mejor” modelo	2 puntos Dos de los siguientes elementos están faltando y/o contiene imprecisiones que afectan significativamente la comprensión de la sección: 1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión 2. Se menciona cuales fueron los “mejores” modelos entrenado s. 3. Se distingue cual fue el “mejor” en términos predictivo s de todos los modelos. 4. Se reflexiona por qué este es el caso 5. Se reflexiona sobre las variables que incluye el	3 puntos Uno de los siguientes elementos está faltando y/o contiene imprecisiones que afectan la comprensión de la sección: 1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión 2. Se menciona cuales fueron los “mejores” modelos entrenado s. 3. Se distingue cual fue el “mejor” en términos predictivo s de todos los modelos. 4. Se reflexiona por qué este es el caso 5. Se reflexiona sobre las variables que incluye el “mejor” modelo	4.5 puntos Están presentes todos los siguientes elementos, pero contiene imprecisiones que no afectan la comprensión de la sección: 1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión 2. Se menciona cuales fueron los “mejores” modelos entrenado s. 3. Se distingue cual fue el “mejor” en términos predictivo s de todos los modelos. 4. Se reflexiona por qué este es el caso 5. Se reflexiona sobre las variables que incluye el “mejor” modelo	5 puntos Están presentes todos los siguientes elementos y desarrollados de forma clara y precisa: 1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión 2. Se menciona cuales fueron los “mejores” modelos entrenado s. 3. Se distingue cual fue el “mejor” en términos predictivo s de todos los modelos. 4. Se reflexiona por qué este es el caso 5. Se reflexiona sobre las variables que incluye el “mejor” modelo 6. Se presentan posibles	/5

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
		6. Se presentan posibles caminos para mejorar los resultados	“mejor” modelo 6. Se presentan posibles caminos para mejorar los resultados	6. Se presentan posibles caminos para mejorar los resultados	6. Se presentan posibles caminos para mejorar los resultados	caminos para mejorar los resultados	
“Entrega” en Kaggle (10%)	0 puntos No se hizo entrega en Kaggle.	2 puntos Hay 2 entregas o menos en Kaggle	5 puntos Hay entre 2 y 4 entregas en Kaggle	7 puntos Hay entre 6 y 8 entregas en Kaggle	9 puntos Hay entre 8 y 10 entregas en Kaggle	10 puntos Hay al menos 10 entregas en Kaggle: 5 para el enfoque de clasificación y 5 para el de regresión	/10

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
Documento y Repositorio de GitHub (10%)	0 puntos El elemento no está presente.	2 puntos Hay menos de (1) contribución de cada uno de los miembros del equipo y/o tres o más de los siguientes elementos está faltando y/o hay imprecisiones serias en el README y/o en el código que afectan significativamente la navegación del repositorio y/o la replicación del trabajo: 1. El document o es autocontenido, no excede las 10 páginas y las 8 tablas o figuras. 2. Existe el repositorio o siguiendo el template provisto 3. Contiene un README E que ayuda al lector a navegar el repositorio o e incluye instrucciones breves para replicar	5 puntos Hay menos de (3) pero más de (1) contribuciones de cada uno de los miembros del equipo y/o dos de los siguientes elementos está faltando y/o hay imprecisiones serias en el README y/o en el código que afectan significativamente la navegación del repositorio y/o la replicación del trabajo: 1. El document o es autocontenido, no excede las 10 páginas y las 8 tablas o figuras. 2. Existe el repositorio o siguiendo el template provisto 3. Contiene un README E que ayuda al lector a navegar el repositorio o e incluye instrucciones breves para replicar	7 puntos Hay menos de (5) pero más de (3) contribuciones de cada uno de los miembros del equipo y/o uno de los siguientes elementos está faltando y/o hay imprecisiones serias en el README y/o en el código que afectan significativamente la navegación del repositorio y/o la replicación del trabajo: 1. El document o es autocontenido, no excede las 10 páginas y las 8 tablas o figuras. 2. Existe el repositorio o siguiendo el template provisto 3. Contiene un README E que ayuda al lector a navegar el repositorio o e incluye instrucciones breves para replicar	9 puntos Están presentes todos los siguientes elementos, pero hay imprecisiones en el README y/o en el código que no afectan significativamente la navegación del repositorio y/o la replicación del trabajo: 1. El document o es autocontenido, no excede las 10 páginas y las 8 tablas o figuras. 2. Existe el repositorio o siguiendo el template provisto 3. Contiene un README E que ayuda al lector a navegar el repositorio o e incluye instrucciones breves para replicar completamente el trabajo. 4. La rama del	10 puntos Están presentes todos los siguientes elementos y desarrollados de forma clara y precisa: 1. El document o es autocontenido, no excede las 10 páginas y las 8 tablas o figuras. 2. Existe el repositorio o siguiendo el template provisto 3. contiene un README E que ayuda al lector a navegar el repositorio o e incluye instrucciones breves para replicar completamente el trabajo. 4. La rama del repositorio o principal muestra al menos cinco (5) contribuciones	/10

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy Bueno	Sobresaliente	Puntuación del criterio
		completamente el trabajo. 4. El código en el repositorio o es reproducible, es legible y reproducible	replicar completamente el trabajo. 4. El código en el repositorio o es reproducible, es legible y reproducible	replicar completamente el trabajo. 4. El código en el repositorio o es reproducible, es legible y reproducible	repositorio principal muestra al menos cinco (5) contribuciones que aportan significativamente al taller, por parte de cada miembro del equipo. 5. El código en el repositorio o es reproducible, es legible y reproducible	ones que aportan significativamente al taller, por parte de cada miembro del equipo. 5. El código en el repositorio o es reproducible, es legible y reproducible	

Total	/100
-------	------

Puntuación general

Nivel 1 0 puntos mínimos	Nivel 2 5 puntos mínimos	Nivel 3 8 puntos mínimos	Nivel 4 11 puntos mínimos
-----------------------------	-----------------------------	-----------------------------	------------------------------