

## Problem Set 3: Making Money with ML?

Curso: BIG DATA Y MACHINE LEARNING PARA ECONOMÍA APLICADA-UNIFICADO (ECON4676\_A-MECA4107\_A)

<div><div><div><div></div></div></div><div>os</div></div>		Insuficiente	Regular	Suficiente	Bueno	Muy bueno	Sobresaliente	Puntuación del criterio
Intrínseco (5%)	1 punto	2.5 puntos	3.5 puntos	4 puntos	4.5 puntos	5 puntos	/5	
	El elemento no está presente.	<p><b>Tres o más</b> de los siguientes elementos están faltando y/o contiene <b>imprecisiones que no permiten la comprensión</b> de la sección:</p> <ol style="list-style-type: none"><li>Se enuncia con claridad y precisión el problema,</li><li>Se mencionan antecedentes relevantes y debidamente documentados,</li><li>Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller,</li><li>Contiene un resumen de los resultados, mencionando el modelo “ganador” enviado a Kaggle,</li><li>Se presentan breves conclusiones del análisis.</li></ol>	<p><b>Dos</b> de los siguientes elementos están faltando y/o contiene <b>imprecisiones que afectan significativamente la comprensión</b> de la sección:</p> <ol style="list-style-type: none"><li>Se enuncia con claridad y precisión el problema,</li><li>Se mencionan antecedentes relevantes y debidamente documentados,</li><li>Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller,</li><li>Contiene un resumen de los resultados, mencionando el modelo “ganador” enviado a Kaggle,</li><li>Se presentan breves conclusiones del análisis.</li></ol>	<p><b>Uno</b> de los siguientes elementos está faltando y/o contiene <b>imprecisiones que afectan la comprensión</b> de la sección:</p> <ol style="list-style-type: none"><li>Se enuncia con claridad y precisión el problema,</li><li>Se mencionan antecedentes relevantes y debidamente documentados,</li><li>Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller,</li><li>Contiene un resumen de los resultados, mencionando el modelo “ganador” enviado a Kaggle,</li><li>Se presentan breves conclusiones del análisis.</li></ol>	<p>Están presentes <b>todos</b> los siguientes elementos, pero contiene <b>imprecisiones que no afectan a la comprensión</b> de la sección:</p> <ol style="list-style-type: none"><li>Se enuncia con claridad y precisión el problema,</li><li>Se mencionan antecedentes relevantes y debidamente documentados,</li><li>Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller,</li><li>Contiene un resumen de los resultados, mencionando el modelo “ganador” enviado a Kaggle,</li><li>Se presentan breves conclusiones del análisis.</li></ol>	<p>Están presentes <b>todos</b> los siguientes elementos y desarrollados de forma <b>clara y precisa</b>:</p> <ol style="list-style-type: none"><li>Se enuncia con claridad y precisión el problema,</li><li>Se mencionan antecedentes relevantes y debidamente documentados,</li><li>Se presentan brevemente los datos y la pertinencia para responder el problema planteado en el taller,</li><li>Contiene un resumen de los resultados, mencionando el modelo “ganador” enviado a Kaggle,</li><li>Se presentan breves conclusiones del análisis.</li></ol>		

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy bueno	Sobresaliente	Puntuación del criterio
Descriptores (30%)	6 puntos El elemento no está presente.	15 puntos <b>Tres o más</b> de los siguientes elementos están faltando y/o contiene <b>imprecisiones que no permiten la comprensión</b> de la sección:  1. Se realiza una descripción completa de la fuente de datos, tanto original como la utilizada en el taller, 2. Se discute la utilidad de la base para la solución del problema planteado en el taller, 3. Se describe cómo se obtiene la muestra final, incluyendo la unión entre las bases de datos, la transformación de variables, las variables incluidas en el análisis, y el tratamiento a observaciones faltantes y datos atípicos, 4. Se incluye al menos una variable proveniente de fuentes externas (como open street maps) y al menos una generada a partir del título o descripción de las propiedades, 5. Se realiza un análisis descriptivo de las variables utilizadas en el análisis mediante tablas y/o gráficos que son	21 puntos <b>Dos</b> de los siguientes elementos están faltando y/o contiene <b>imprecisiones que afectan significativamente la comprensión</b> de la sección:  1. Se realiza una descripción completa de la fuente de datos, tanto original como la utilizada en el taller, 2. Se discute la utilidad de la base para la solución del problema planteado en el taller, 3. Se describe cómo se obtiene la muestra final, incluyendo la unión entre las bases de datos, la transformación de variables, las variables incluidas en el análisis, y el tratamiento a observaciones faltantes y datos atípicos, 4. Se incluye al menos dos variables provenientes de fuentes externas (como open street maps) y al menos una generada a partir del título o descripción de las propiedades, 5. Se realiza un análisis descriptivo de las variables utilizadas en el análisis mediante tablas y/o gráficos que	24 puntos <b>Uno</b> de los siguientes elementos está faltando y/o contiene <b>imprecisiones que afectan la comprensión</b> de la sección:  1. Se realiza una descripción completa de la fuente de datos, tanto original como la utilizada en el taller, 2. Se discute la utilidad de la base para la solución del problema planteado en el taller, 3. Se describe cómo se obtiene la muestra final, incluyendo la unión entre las bases de datos, la transformación de variables, las variables incluidas en el análisis, y el tratamiento a observaciones faltantes y datos atípicos, 4. Se incluye al menos cuatro variables provenientes de fuentes externas (como open street maps) y al menos dos generadas a partir del título o descripción de las propiedades, 5. Se realiza un análisis descriptivo de las variables utilizadas en el análisis mediante tablas y/o gráficos que son autocontenidos	27 puntos Están presentes <b>todos</b> los siguientes elementos, pero contiene <b>imprecisiones que no afectan a la comprensión</b> de la sección:  1. Se realiza una descripción completa de la fuente de datos, tanto original como la utilizada en el taller, 2. Se discute la utilidad de la base para la solución del problema planteado en el taller, 3. Se describe cómo se obtiene la muestra final, incluyendo la unión entre las bases de datos, la transformación de variables, las variables incluidas en el análisis, y el tratamiento a observaciones faltantes y datos atípicos, 4. Se incluye al menos cuatro variables provenientes de fuentes externas (como open street maps) y al menos dos generadas a partir del título o descripción de las propiedades, 5. Se realiza un análisis descriptivo de las variables utilizadas en el análisis mediante tablas y/o gráficos que son autocontenidos y debidamente formateadas, con el formato sugerido. Las	30 puntos Están presentes <b>todos</b> los siguientes elementos y desarrollados de forma <b>clara y precisa</b> :  1. Se realiza una descripción completa de la fuente de datos, tanto original como la utilizada en el taller, 2. Se discute la utilidad de la base para la solución del problema planteado en el taller, 3. Se describe cómo se obtiene la muestra final, incluyendo la unión entre las bases de datos, la transformación de variables, las variables incluidas en el análisis, y el tratamiento a observaciones faltantes y datos atípicos, 4. Se incluye al menos cuatro variables provenientes de fuentes externas (como open street maps) y al menos dos generadas a partir del título o descripción de las propiedades, 5. Se realiza un análisis descriptivo de las variables utilizadas en el análisis mediante tablas y/o gráficos que son autocontenidos y debidamente formateadas, con el formato sugerido. Las tablas no son pantallazos de R/Python (o el	/30

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy bueno	Sobresaliente	Puntuación del criterio
		autocontenidos y debidamente formateadas, con el formato sugerido. Las tablas no son pantallazos de R/Python (o el software utilizado), y tienen título y notas que permitan entenderlas sin recurrir al documento. Las gráficas tienen título, ejes, leyenda y notas que permiten entenderlas sin recurrir al documento. Se incluye al menos un mapa junto con su interpretación. 6. El análisis descriptivo permite al lector entender la variación de los datos y la elección de las variables, 7. Se utiliza el conocimiento profesional y los antecedentes para agregar valor a la sección.	son autocontenidos y debidamente formateadas, con el formato sugerido. Las tablas no son pantallazos de R/Python (o el software utilizado), y tienen título y notas que permitan entenderlas sin recurrir al documento. Las gráficas tienen título, ejes, leyenda y notas que permiten entenderlas sin recurrir al documento. Se incluye al menos un mapa junto con su interpretación. 6. El análisis descriptivo permite al lector entender la variación de los datos y la elección de las variables, 7. Se utiliza el conocimiento profesional y los antecedentes para agregar valor a la sección.	y debidamente formateadas, con el formato sugerido. Las tablas no son pantallazos de R/Python (o el software utilizado), y tienen título y notas que permitan entenderlas sin recurrir al documento. Las gráficas tienen título, ejes, leyenda y notas que permiten entenderlas sin recurrir al documento. Se incluye al menos dos mapas junto con su interpretación. 6. El análisis descriptivo permite al lector entender la variación de los datos y la elección de las variables, 7. Se utiliza el conocimiento profesional y los antecedentes para agregar valor a la sección.	tablas no son pantallazos de R/Python (o el software utilizado), y tienen título y notas que permitan entenderlas sin recurrir al documento. Las gráficas tienen título, ejes, leyenda y notas que permiten entenderlas sin recurrir al documento. Se incluye al menos dos mapas junto con su interpretación. 6. El análisis descriptivo permite al lector entender la variación de los datos y la elección de las variables, 7. Se utiliza el conocimiento profesional y los antecedentes para agregar valor a la sección.	software utilizado), y tienen título y notas que permitan entenderlas sin recurrir al documento. Las gráficas tienen título, ejes, leyenda y notas que permiten entenderlas sin recurrir al documento. Se incluye al menos dos mapas junto con su interpretación. 6. El análisis descriptivo permite al lector entender la variación de los datos y la elección de las variables, 7. Se utiliza el conocimiento profesional y los antecedentes para agregar valor a la sección.	

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy bueno	Sobresaliente	Puntuación del criterio
Modo de uso (40%)	<p>8 puntos</p> <p>El elemento no está presente.</p>	<p>20 puntos</p> <p><b>Tres o más</b> de los siguientes elementos están faltando y/o contiene <b>imprecisiones que no permiten la comprensión</b> de la sección:</p> <ol style="list-style-type: none"> <li>Se describe, detalladamente, el algoritmo y especificación utilizado por el equipo que alcanzó el mejor puntaje en Kaggle.</li> <li>La descripción detalla las variables incluidas, el proceso de entrenamiento, la selección de los hiper-parámetro, la estrategia de validación, y/o cualquier otra información relevante para la comprensión de los modelos.</li> <li>Se compara los resultados predictivos al emplear una estrategia de validación típica vs. validación cruzada espacial, discutiendo el impacto de cada estrategia y mencionando los detalles pertinentes.</li> <li>Se compara el desempeño del envío que alcanzó el mejor puntaje en Kaggle contra otros nueve envíos del equipo en Kaggle, enfatizando en las diferencias</li> </ol>	<p>28 puntos</p> <p><b>Dos</b> de los siguientes elementos están faltando y/o contiene <b>imprecisiones que afectan significativamente la comprensión</b> de la sección:</p> <ol style="list-style-type: none"> <li>Se describe, detalladamente, el algoritmo y especificación utilizado por el equipo que alcanzó el mejor puntaje en Kaggle.</li> <li>La descripción detalla las variables incluidas, el proceso de entrenamiento, la selección de los hiper-parámetro, la estrategia de validación, y/o cualquier otra información relevante para la comprensión de los modelos.</li> <li>Se compara los resultados predictivos al emplear una estrategia de validación típica vs. validación cruzada espacial, discutiendo el impacto de cada estrategia y mencionando los detalles pertinentes.</li> <li>Se compara el desempeño del envío que alcanzó el mejor puntaje en Kaggle contra otros nueve envíos del equipo en Kaggle, enfatizando en</li> </ol>	<p>32 puntos</p> <p><b>Uno</b> de los siguientes elementos está faltando y/o contiene <b>imprecisiones que afectan la comprensión</b> de la sección:</p> <ol style="list-style-type: none"> <li>Se describe, detalladamente, el algoritmo y especificación utilizado por el equipo que alcanzó el mejor puntaje en Kaggle.</li> <li>La descripción detalla las variables incluidas, el proceso de entrenamiento, la selección de los hiper-parámetro, la estrategia de validación, y/o cualquier otra información relevante para la comprensión de los modelos.</li> <li>Se compara los resultados predictivos al emplear una estrategia de validación típica vs. validación cruzada espacial, discutiendo el impacto de cada estrategia y mencionando los detalles pertinentes.</li> <li>Se compara el desempeño del envío que alcanzó el mejor puntaje en Kaggle contra otros nueve envíos del equipo en Kaggle, enfatizando en las diferencias que mejoraron o</li> </ol>	<p>36 puntos</p> <p>Están presentes <b>todos</b> los siguientes elementos, pero contiene <b>imprecisiones que no afectan a la comprensión</b> de la sección:</p> <ol style="list-style-type: none"> <li>Se describe, detalladamente, el algoritmo y especificación utilizado por el equipo que alcanzó el mejor puntaje en Kaggle.</li> <li>La descripción detalla las variables incluidas, el proceso de entrenamiento, la selección de los hiper-parámetro, la estrategia de validación, y/o cualquier otra información relevante para la comprensión de los modelos.</li> <li>Se compara los resultados predictivos al emplear una estrategia de validación típica vs. validación cruzada espacial, discutiendo el impacto de cada estrategia y mencionando los detalles pertinentes.</li> <li>Se compara el desempeño del envío que alcanzó el mejor puntaje en Kaggle contra otros nueve envíos del equipo en Kaggle, enfatizando en las diferencias que mejoraron o empeoraron el desempeño</li> </ol>	<p>40 puntos</p> <p>Están presentes <b>todos</b> los siguientes elementos y desarrollados de forma clara y precisa:</p> <ol style="list-style-type: none"> <li>Se describe, detalladamente, el algoritmo y especificación utilizado por el equipo que alcanzó el mejor puntaje en Kaggle.</li> <li>La descripción detalla las variables incluidas, el proceso de entrenamiento, la selección de los hiper-parámetro, la estrategia de validación, y/o cualquier otra información relevante para la comprensión de los modelos.</li> <li>Se compara los resultados predictivos al emplear una estrategia de validación típica vs. validación cruzada espacial, discutiendo el impacto de cada estrategia y mencionando los detalles pertinentes.</li> <li>Se compara el desempeño del envío que alcanzó el mejor puntaje en Kaggle contra otros nueve envíos del equipo en Kaggle, enfatizando en las diferencias que mejoraron o empeoraron el desempeño predictivo. Por ejemplo, variables elegidas,</li> </ol>	/40

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy bueno	Sobresaliente	Puntuación del criterio
		<p>que mejoraron o empeoraron el desempeño predictivo. Por ejemplo, variables elegidas, especificación del modelo, hiperparámetros, estrategia de validación cruzada, y demás.</p> <p>5. Se discute la importancia relativa de las variables utilizadas en el mejor modelo, contrastando similitudes y diferencias.</p> <p>6. Se utiliza al menos 5 algoritmos distintos.</p>	<p>las diferencias que mejoraron o empeoraron el desempeño predictivo. Por ejemplo, variables elegidas, especificación del modelo, hiperparámetros, estrategia de validación cruzada, y demás.</p> <p>5. Se discute la importancia relativa de las variables utilizadas en el mejor modelo, contrastando similitudes y diferencias.</p> <p>6. Se utiliza al menos 6 algoritmos distintos.</p>	<p>empeoraron el desempeño predictivo. Por ejemplo, variables elegidas, especificación del modelo, hiperparámetros, estrategia de validación cruzada, y demás.</p> <p>5. Se discute la importancia relativa de las variables utilizadas en el mejor modelo, contrastando similitudes y diferencias.</p> <p>6. Se utiliza 7 algoritmos distintos.</p>	<p>predictivo. Por ejemplo, variables elegidas, especificación del modelo, hiperparámetros, estrategia de validación cruzada, y demás.</p> <p>5. Se discute la importancia relativa de las variables utilizadas en el mejor modelo, contrastando similitudes y diferencias.</p> <p>6. Se utiliza 7 algoritmos distintos.</p>	<p>especificación del modelo, hiperparámetros, estrategia de validación cruzada, y demás.</p> <p>5. Se discute la importancia relativa de las variables utilizadas en el mejor modelo, contrastando similitudes y diferencias.</p> <p>6. Se utiliza 7 algoritmos distintos.</p>	

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy bueno	Sobresaliente	Puntuación del criterio
Conclusión (8%)	1.6 puntos El elemento no está presente.	4 puntos <b>Tres o más</b> de los siguientes elementos están faltando y/o contiene <b>imprecisiones que no permiten la comprensión</b> de la sección:  1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión, 2. Se menciona cuáles fueron los “mejores” modelos entrenados, 3. Se distingue cuál fue el “mejor” en términos predictivos 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 reflexiona sobre la utilidad del modelo y los posibles caminos para mejorar los resultados.	5.6 puntos <b>Dos</b> de los siguientes elementos están faltando y/o contiene <b>imprecisiones que afectan significativamente la comprensión</b> de la sección:  1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión, 2. Se menciona cuáles fueron los “mejores” modelos entrenados, 3. Se distingue cuál fue el “mejor” en términos predictivos 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 reflexiona sobre la utilidad del modelo y los posibles caminos para mejorar los resultados.	6.4 puntos <b>Uno</b> de los siguientes elementos está faltando y/o contiene <b>imprecisiones que afectan la comprensión</b> de la sección:  1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión, 2. Se menciona cuáles fueron los “mejores” modelos entrenados, 3. Se distingue cuál fue el “mejor” en términos predictivos 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 reflexiona sobre la utilidad del modelo y los posibles caminos para mejorar los resultados.	7.2 puntos Están presentes <b>todos</b> los siguientes elementos, pero contiene <b>imprecisiones que no afectan la comprensión</b> de la sección:  1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión, 2. Se menciona cuáles fueron los “mejores” modelos entrenados, 3. Se distingue cuál fue el “mejor” en términos predictivos 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 reflexiona sobre la utilidad del modelo y los posibles caminos para mejorar los resultados.	8 puntos Están presentes <b>todos</b> los siguientes elementos y desarrollados de forma <b>clara y precisa</b> :  1. Se recuerda al lector la pregunta que se buscó responder con claridad y precisión, 2. Se menciona cuáles fueron los “mejores” modelos entrenados, 3. Se distingue cuál fue el “mejor” en términos predictivos 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 reflexiona sobre la utilidad del modelo y los posibles caminos para mejorar los resultados.	/8
Entrega en Kaggle (5%)	1 punto No se hizo entrega en Kaggle.	2.5 puntos Hay entre 1 y 2 entregas en Kaggle.	3.5 puntos Hay entre 3 y 4 entregas en Kaggle.	4 puntos Hay entre 5 y 6 entregas en Kaggle.	4.5 puntos Hay entre 7 y 9 entregas en Kaggle.	5 puntos Hay al menos 10 entregas en Kaggle.	/5

Criterios	Insuficiente	Regular	Suficiente	Bueno	Muy bueno	Sobresaliente	Puntuación del criterio
Documen- to y repo- sitorio de Git Hub (12%)	2.4 puntos  El elemento no está presente.	6 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:  <div>1. El documento es autocontenido, no excede las 10 páginas y las 10 tablas o figuras, 2. Existe el repositorio siguiendo el template provisto, 3. Contiene un README que ayuda al lector a navegar el repositorio e incluye instrucciones breves para replicar completamente el trabajo, 4. El código en el repositorio es legible, 5. El código es reproducible descargándose directamente del repositorio.</div>	8.4 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:  <div>1. El documento es autocontenido, no excede las 10 páginas y las 10 tablas o figuras, 2. Existe el repositorio siguiendo el template provisto, 3. Contiene un README que ayuda al lector a navegar el repositorio e incluye instrucciones breves para replicar completamente el trabajo, 4. El código en el repositorio es legible, 5. El código es reproducible descargándose directamente del repositorio.</div>	9.6 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:  <div>1. El documento es autocontenido, no excede las 10 páginas y las 10 tablas o figuras, 2. Existe el repositorio siguiendo el template provisto, 3. Contiene un README que ayuda al lector a navegar el repositorio e incluye instrucciones breves para replicar completamente el trabajo, 4. El código en el repositorio es legible, 5. El código es reproducible descargándose directamente del repositorio.</div>	10.8 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:  <div>1. El documento es autocontenido, no excede las 10 páginas y las 10 tablas o figuras, 2. Existe el repositorio siguiendo el template provisto, 3. Contiene un README que ayuda al lector a navegar el repositorio e incluye instrucciones breves para replicar completamente el trabajo, 4. La rama del 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 es legible, 6. El código es reproducible descargándose directamente del repositorio.</div>	12 puntos  Están presentes todos los siguientes elementos y desarrollados de forma clara y precisa:  <div>1. El documento es autocontenido, no excede las 10 páginas y las 10 tablas o figuras, 2. Existe el repositorio siguiendo el template provisto, 3. Contiene un README que ayuda al lector a navegar el repositorio e incluye instrucciones breves para replicar completamente el trabajo, 4. La rama del 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 es legible, 6. El código es reproducible descargándose directamente del repositorio.</div>	/12

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

Puntuación general

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