

ANEXO Evaluación Final Transversal

Instrucciones y Pauta de Evaluación

Sigla	Nombre Asignatura
BIY7121	MINERIA DE DATOS

CASO DE ESTUDIO

Australia es uno de los países con mayor extensión del mundo y el continente más seco y llano del planeta. Aproximadamente la mitad occidental de Australia es una extensa meseta árida. El clima predominante es el desértico y semiárido (se estima que un 40% del territorio lo forman dunas de arena), si bien el norte del país cuenta con un clima tropical. Las tierras más fértiles y el clima templado con cuatro estaciones se concentran en el sureste y suroeste. La principal cordillera del país, la Gran Cordillera Divisoria (The Great Dividing Range), se extiende a lo largo de más de 3.500 km entre los estados de Queensland y Victoria. El pico más alto es el Monte Kosciuszko con 2.228 metros. Los ríos más caudalosos nacen en la Gran Cordillera Divisoria y, en su gran mayoría, fluyen hacia el este y desembocan en el océano Pacífico. Por su longitud y caudal destacan dos ríos: el Murray y el Darling. Australia se divide en seis estados (New South Wales, Victoria, Queensland, South Australia, Western Australia y Tasmania) y dos territorios continentales (Northern Territory y Australian Capital Territory (ACT)). La ciudad de Sídney, con sus 5,73 millones de habitantes, es la más poblada de Australia y es la capital del estado de Nueva Gales del Sur. Es a su vez un importante centro industrial y puerto comercial de primera magnitud. La ciudad es muy extensa, con un radio aproximado de 40 km, incluyendo los suburbios. Otras ciudades importantes son Melbourne (5,19 millones), Brisbane (2,56 millones), Perth (2,38 millones) y Adelaide (1,41 millones). Casi el 90% de la población se concentra en las zonas urbanas. De sus más de 25 millones de habitantes, un 65,5% se encuentra en edad de trabajar (de 15 a 64 años), mientras que aquellos con más de 65 años y con menos de 15 años suponen el 15,7% y el 18,8% de la población respectivamente. La superficie del país es de aproximadamente 7.741.220 km², quince veces superior a España, representando el 5% del total de la superficie terrestre emergida. Es el sexto país del mundo en extensión. La distancia máxima del continente australiano de este a oeste es de 4.100 km y de norte a sur, de 3.200 km. La longitud costera es de 36.735 km. Su vasta extensión inevitablemente condiciona el tráfico de personas, mercancías y servicios. Australia es, además, el más llano de los continentes, con una altitud media de menos de 300 metros, siendo The Great Dividing Range la única gran cordillera de importancia. Casi dos tercios del territorio

carecen de corrientes de agua hacia el mar. La principal cuenca fluvial es la del Murray-Darling con 1.061.469 km², que drena las vertientes sur occidentales de la Gran Cordillera. El país cuenta con una gran diversidad climática que abarca desde el clima tropical en el norte, que representa el 39% del territorio, al clima templado-continental en el sudeste y Tasmania. En las zonas del centro predomina el clima desértico. La zona más fértil es la franja costera entre Sídney y Adelaida, con lluvias moderadas todo el año. Australia está menos sujeta a climas extremos que otros países en su misma latitud, debido a los efectos moderadores de los mares y océanos circundantes y a la ausencia de grandes montañas. Las estaciones varían con la latitud, pero aproximadamente son: primavera (septiembre y octubre), verano (noviembre-marzo), otoño (abril y mayo) e invierno (junio-agosto). Las temperaturas medias oscilan entre los 27°C en la zona norte y los 13°C en las zonas más al sur, alcanzando las máximas en la zona centro (38°C). Por tanto, las principales características climáticas australianas son inviernos suaves y veranos cálidos, así como abundante sol y poca humedad. Las precipitaciones son muy escasas en el interior y aumentan en las zonas costeras, de modo que las zonas mejor regadas son los litorales norte, este, sudeste y sudoeste. En el norte del país hay dos estaciones, seca en invierno y húmeda en verano con la irrupción de lluvias monzónicas. Australia es, después de la Antártida, el continente más seco. Aun así, y debido a su diversidad climática, se dan todo tipo de fenómenos naturales extremos como sequías, inundaciones, ciclones tropicales, vendavales, incendios forestales (en lo que se conoce como Bush o monte en Australia) y, ocasionalmente, tornados. En cuanto a recursos naturales y minería, Australia es uno de los principales productores y 3 exportadores de minerales y productos energéticos a nivel mundial. El sector minero representa alrededor del 10% del PIB con un valor de 148.000 millones de AUD. En 2018-19 presentó una tasa de crecimiento del 6%. El valor de las exportaciones totales de productos mineros y energéticos se espera que alcance los 299.000 millones AUD en el periodo 2019-20. Las exportaciones de mineral de hierro alcanzaron durante 2018-19 los 100.000 millones de AUD. Australia está entre los cinco principales países exportadores del mundo de bauxita, alúmina, mineral de hierro, zinc, carbón y de gas natural licuado (GNL).

Texto extraído del sitio <https://www.icex.es/> ministerio de industria comercio y turismo de España.

Se dispone de un set de datos de observaciones meteorológicas diarias de múltiples ubicaciones en Australia, obtenidas de la Oficina de Meteorología de la Commonwealth de Australia y procesadas para crear este conjunto de datos de muestra, los datos se han procesado para proporcionar una variable objetivo RainTomorrow (si hay lluvia al día siguiente - No / Sí) y una variable de riesgo RISK_MM (cuánta lluvia registrada en milímetros), esta información se ha dejado disponible para que usted la explore y busque información que sea relevante para demostrar su aprendizaje de Minería de datos.

EL DATASET POSEE LOS SIGUIENTES CAMPOS:

Columna	Descripción
Fecha	Fecha de la observación
Ubicación	Ubicación de la estación meteorológica
MinTemp	Temperatura mínima en grados Celsius
MaxTemp	Temperatura máxima en grados Celsius
Lluvia	Cantidad de lluvia registrada ese día en mm.
Evaporación	Evaporación (mm) en 24 horas
Sol	Número de horas de sol brillante en el día
DirRafaga	Dirección de la ráfaga de viento más fuerte en 24 horas.
VelRafaga	Velocidad (km/hr) de la ráfaga de viento más fuerte en 24 horas.
Dir9am	Dirección del viento a las 9am
Dir3pm	Dirección del viento a las 3pm
Vel9am	Velocidad (km/hr) del viento a las 9am
Vel3pm	Velocidad (km/hr) del viento a las 3pm
Hum9am	Porcentaje de humedad a las 9am
Hum3pm	Porcentaje de humedad a las 3pm
Pres9am	Presión atmosférica (hpa) a nivel del mar a las 9am
Pre3pm	Presión atmosférica (hpa) a nivel del mar a las 3pm
Nub9am	Fracción del cielo cubierto por nubes a las 9am. Se mide en "octavos", de manera que un valor 0 indica cielo totalmente despejado y 8, cielo totalmente cubierto.
Nub3pm	Fracción del cielo cubierto por nubes a las 3pm. Se mide en "octavos", de manera que un valor 0 indica cielo totalmente despejado y 8, cielo totalmente cubierto.
Temp9am	Temperatura en grados Celsius a las 9am
Temp3pm	Temperatura en grados Celsius a las 3pm
Lluvia Hoy	Variable indicadora que toma el valor 1 si la precipitación es en mm. en las últimas 24 hrs. excede 1 mm. y 0 si no.
RISK_MM	La cantidad de lluvia. Una especie de medida del "riesgo".
Lluvia Man	Variable indicadora que toma el valor 1 si al día siguiente llovió y 0 si no.

3. Pauta de Evaluación

Categoría	% logro	Descripción niveles de logro
Muy buen desempeño	100%	Demuestra un desempeño destacado, evidenciando el logro de todos los aspectos evaluados en el indicador.
Buen desempeño	80%	Demuestra un alto desempeño del indicador, presentando pequeñas omisiones, dificultades y/o errores.
Desempeño aceptable	60%	Demuestra un desempeño competente, evidenciando el logro de los elementos básicos del indicador, pero con omisiones, dificultades o errores.
Desempeño incipiente	30%	Presenta importantes omisiones, dificultades o errores en el desempeño, que no permiten evidenciar los elementos básicos del logro del indicador, por lo que no puede ser considerado competente.
Desempeño no logrado	0%	Presenta ausencia o incorrecto desempeño.

Indicador de Evaluación	Categorías de Respuesta					Ponderación Indicador de Evaluación
	Muy buen desempeño 100%	Buen desempeño 80%	Desempeño aceptable 60%	Desempeño incipiente 30%	Desempeño no logrado 0%	
1. Identifica qué es el proceso de extracción del conocimiento, usando diversas fuentes de datos de la organización a través de reconocimiento de patrones.	Reconoce todas las fases de la metodología CRISP-DM abarcando a lo largo del proyecto hasta la fase de Evaluación.	Reconoce la mayoría de las fases de la metodología CRISP-DM abordándolas a lo largo del proyecto.	Reconoce la mitad de las fases de la metodología CRISP-DM abordándolas a lo largo del proyecto.	Identifica menos de la mitad de las fases de la metodología CRISP-DM para la implementación del proyecto.	No identifica las fases de la Metodología CRISP-DM para proyectos de Minería de Datos.	15%
2. Realiza la exploración y selección de los datos utilizando herramientas de estadística descriptiva para identificar información relevante del negocio.	Explora los datos calculando más de 5 estadísticos que resumen campos clave en los datos además de missing values y outliers.	Explora los datos calculando entre 4 y 5 estadísticos que resumen campos clave en los datos además de missing values y outliers.	Explora los datos calculando 3 estadísticos que resumen campos clave en los datos, pero no identifica outliers o solo missing values.	Explora los datos calculando menos de 3 estadísticos que resumen campos clave en los datos, pero no identifica outliers ni missing values.	No calcula estadísticos que resuman los campos clave, ni tampoco identifica outliers ni missing values.	15%
3. Realiza limpieza y transformación de los datos, utilizando lenguaje de programación Python, procesándolos de acuerdo con sus características.	Realiza la transformación y exploración del total de los datos, discriminando su relevancia en el análisis de acuerdo con las características que poseen los datos.	Realiza la transformación y exploración la mayoría de los datos, discriminando su relevancia en el análisis de acuerdo con las características que poseen los datos.	Realiza la transformación y exploración de la mitad de los datos, discriminando su relevancia en el análisis de acuerdo con las características que poseen los datos.	Realiza la transformación y exploración de menos de la mitad de los datos, discriminando su relevancia en el análisis de acuerdo con las características de los datos.	No realiza la transformación y exploración de los datos.	15%
4. Aplica modelos matemáticos sobre los datos, utilizando bibliotecas especializadas de Python, para obtener información significativa del negocio.	Construye correctamente más de 3 modelos supervisados para obtener información relevante sobre los datos en apoyo al negocio.	Construye correctamente entre 2 y 3 modelos supervisados para obtener información relevante sobre los datos en apoyo al negocio.	Construye correctamente solo 2 modelos supervisados para obtener información relevante sobre los datos en apoyo al negocio.	Construye correctamente solo un modelo supervisado para obtener información relevante sobre los datos en apoyo al negocio.	No construye ningún modelo matemático supervisado aplicados a los datos.	15%

5. Identifica insights a partir del trabajo con los datos, de acuerdo con las características de los datos y del negocio para proporcionar valor competitivo.	Construye correctamente 3 modelos no supervisados obteniendo información relevante sobre los datos.	Construye 3 modelos no supervisados con algunos errores, obteniendo información relevante sobre los datos.	Construye correctamente solo 2 modelos no supervisados obteniendo información relevante sobre los datos.	Construye correctamente solo un modelo no supervisado obteniendo información relevante sobre los datos.	No construye ningún modelo matemático no supervisado aplicados a los datos.	15%
6. Aplica las funciones en el lenguaje definido, que le permiten la obtención de coeficientes para evaluar los modelos estadísticos según las técnicas aplicadas.	Realiza una correcta evaluación de los modelos interpretando los resultados y utilizando las métricas correspondientes al tipo de aprendizaje aplicado. Y selecciona el modelo más adecuado al contexto.	Realiza una correcta evaluación de los modelos utilizando las métricas correspondientes al tipo de aprendizaje aplicado, pero interpreta con algunos errores los resultados obtenidos. Y selecciona el modelo más adecuado al contexto.	Realiza la evaluación de los modelos utilizando las métricas que corresponden al tipo de aprendizaje, pero comete un mínimo error en su aplicación, Y selecciona el modelo más adecuado al contexto.	Realiza la evaluación de los modelos utilizando las métricas que corresponden al tipo de aprendizaje, pero comete un mínimo error en su aplicación, Y selecciona el modelo más adecuado al contexto.	No realiza la evaluación de los modelos implementados y por ende no logra seleccionar el mejor.	15%
7. Presenta los resultados obtenidos del modelo analizado al cliente, indicando posibles automatizaciones, mejoras y/o insights.	Presenta la documentación y contextualización del total de los resultados obtenidos de los modelos estadísticos en el entorno organizacional, indicando posibles automatizaciones, mejoras y/o insights.	Presenta la documentación y contextualización de la mayoría de los resultados obtenidos de los modelos estadísticos en el entorno organizacional, indicando posibles automatizaciones, mejoras y/o insights.	Presenta la documentación y contextualización de la mitad de los resultados obtenidos de los modelos estadísticos en el entorno organizacional, indicando posibles automatizaciones, mejoras y/o insights.	Presenta la documentación y contextualización de menos de la mitad de los resultados obtenidos de los modelos estadísticos en el entorno organizacional, sin indicar posibles automatizaciones, mejoras y/o insights.	No presenta la documentación ni contextualizar los resultados obtenidos de los modelos estadísticos en el entorno organizacional.	10%
Total						100%