



RETOS ASOCIADOS CON EL BI

Actividad de transferencia

INTEGRACIÓN DEL BUSINESS INTELLIGENCE, BIG DATA Y MACHINE LEARNING

1. Buscar en Google herramientas de Machine Learning Low Code, evaluar sus ventajas y desventajas.

Herramientas Low Code de Machine Learning:

1. **Google Teachable Machine:** Teachable Machine es una herramienta web de Google que permite crear modelos de aprendizaje automático con interfaces visuales. Es ideal para tareas de clasificación de imágenes, sonidos y posturas, y no requiere conocimientos avanzados de ML.

Ventajas:

- **Interfaz simple:** Extremadamente fácil de usar, incluso para principiantes.
- **Compatibilidad:** Extremadamente fácil de usar, incluso para principiantes.
- **Rápido:** Los modelos se pueden entrenar en minutos.
- **Sin instalación:** Funciona completamente en el navegador.

Desventajas:

- **Limitaciones en personalización:** No es adecuada para modelos complejos o entrenamientos avanzados.
- **Restricción en datasets:** No permite trabajar directamente con grandes volúmenes de datos ni con estructuras complejas.
- **Uso específico:** Más útil para casos sencillos como clasificación, menos adecuado para tareas como regresión o series temporales.
- **Google Teachable Machine:** Teachable Machine es una herramienta web de Google que permite crear modelos de aprendizaje automático con interfaces visuales. Es ideal para tareas de clasificación de imágenes, sonidos y posturas, y no requiere conocimientos avanzados de ML.



2. **Microsoft Lobe:** Lobe es una herramienta gratuita de Microsoft para crear modelos ML con un enfoque visual. Es particularmente adecuada para tareas de clasificación de imágenes.

Ventajas:

- **Fácil de usar:** Su diseño se enfoca en arrastrar y soltar, sin necesidad de escribir código.
- **Offline:** Funciona localmente, lo que reduce preocupaciones sobre privacidad.
- **Integración:** Exporta modelos para plataformas como TensorFlow, PyTorch y ONNX.
- **Educativa:** Buen recurso para aprender sobre el entrenamiento de modelos.

Desventajas:

- **Soporte limitado:** Actualmente se centra solo en la clasificación de imágenes.
- **Requiere hardware:** Entrenar modelos localmente puede ser lento en computadoras sin GPUs potentes.
- **Menos versatilidad:** Carece de opciones para trabajar con datos no relacionados con imágenes.

3. **KNIME Analytics Platform:** KNIME es una plataforma de análisis de datos visual que permite construir flujos de trabajo ML con componentes predefinidos. Es más versátil que las otras opciones.

Ventajas:

- **Multidisciplinaria:** Compatible con múltiples tipos de datos (tabulares, imágenes, texto, etc.).
- **Extensiones:** Ofrece integración con Python, R y herramientas avanzadas.
- **Visual y poderoso:** Proporciona nodos preconstruidos para tareas como regresión, clasificación y clustering.
- **Escalable:** Útil para proyectos pequeños y avanzados.

Desventajas:

- **Curva de Aprendizaje:** Su interfaz, aunque visual, puede ser intimidante para principiantes.
- **Requiere instalación:** A diferencia de Teachable Machine, no está completamente basada en navegador.
- **Consumo de recursos:** Puede volverse lenta con datasets muy grandes.



2. Usar <https://www.obviously.ai/> para una simulación de predicción de datos. Piense en su propia empresa, en datos que tengan en tablas en base de datos o hojas de cálculo (.xls o .csv). Seleccione en función a su data un dato que desee predecir y en base a los resultados presente un análisis con algunas hipótesis y conclusiones.

Este análisis no se pudo realizar en obviously.ai por razones desconocidas. Sin embargo, se hizo una pequeña implementación en python utilizando el algoritmo de ML *Random Forest* (el cual toma los famosos y mencionados en clase: *Decision Tree* muchas veces y calcula un promedio de los resultados para dar una predicción aún mejor). Esta implementación se realizó en precios de casas en ciertas ciudades (dado que por las políticas de confidencialidad de mi lugar de laburo, no podría utilizar estos datos). El dataset puede ser encontrado en el siguiente link:

https://github.com/DSarceno/ADEN/blob/main/Maestria_BigData_and_BusinessAnalytics/2%20-%20BI%20-%20Inteligencia%20de%20Negocios/S2_2/Cleaned_data_for_model.csv dando al botón de descargar. Y el código de la implementación así como las conclusiones pueden ser encontradas aquí:

https://github.com/DSarceno/ADEN/blob/main/Maestria_BigData_and_BusinessAnalytics/2%20-%20BI%20-%20Inteligencia%20de%20Negocios/S2_2/house_prices.ipynb.