

Machine Learning Aplicado

Desarrollo de productos ML

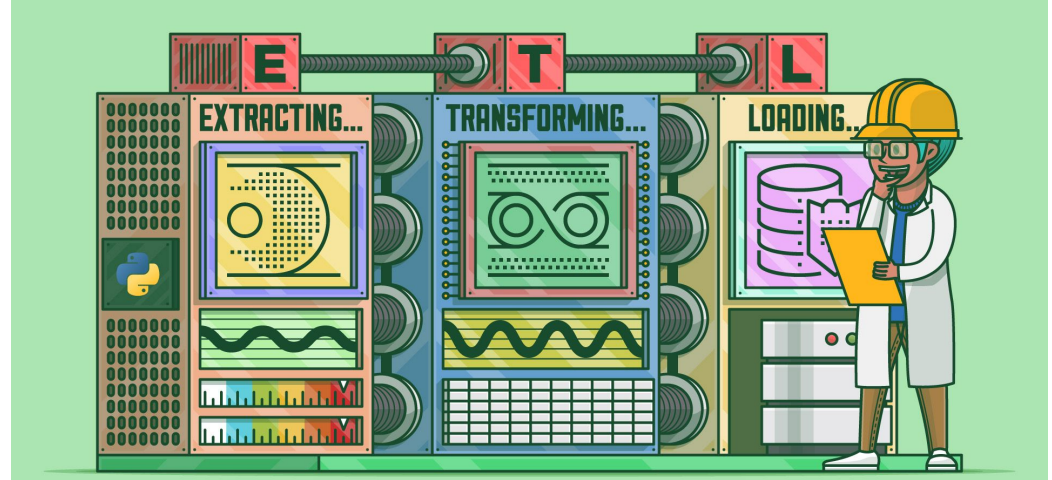
Ingeniería Civil Informática
Escuela de Ingeniería Informática
Universidad de Valparaíso

Profesor: Aaron Ponce Sandoval
correo: aaron.ponce@uv.cl

Perfiles de Desarrolladores:

- ❑ **Desarrolladores de modelos:** Científicos de datos altamente capacitados en estadística , álgebra lineal y cálculo que entrenan, evalúan y seleccionan modelos de alto rendimiento.
- ❑ **Desarrolladores de Implementación de modelos:** Desarrolladores altamente capacitados en diseño e ingeniería de software que construyen sistemas de software. Implementan/Integran los modelos en la nube y lo escalan para atender una gran cantidad de solicitudes de inferencias de manera concurrente.

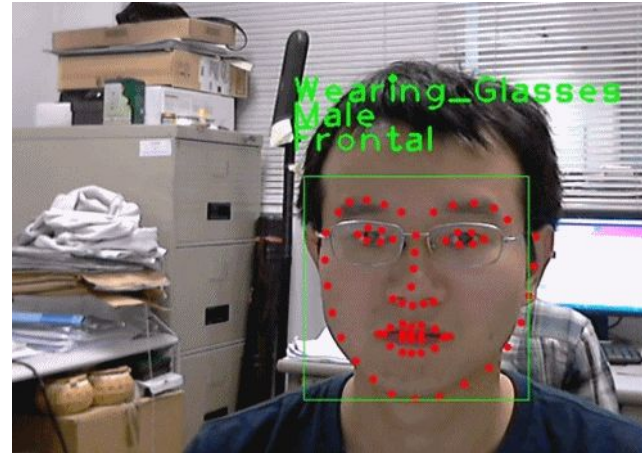
- ❑ **Ingeniería de datos:**
 - ❑ Creación de canalizaciones de datos para recopilar datos de diferentes fuentes,
 - ❑ Curación de datos
 - ❑ Transformación en datos limpios y homogéneos ***que se puedan usar de manera segura para entrenar modelos.***



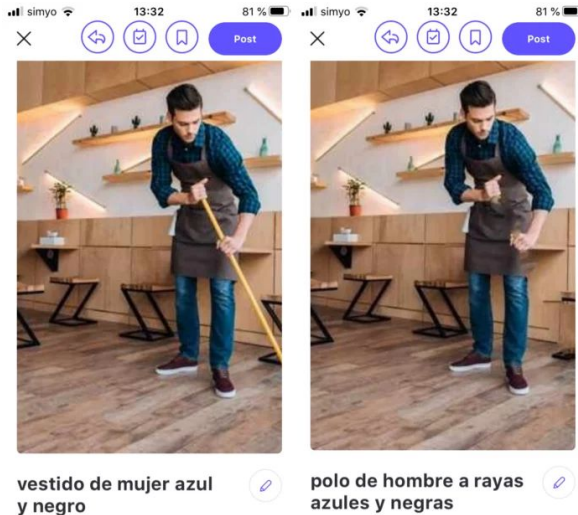
- ❑ **Diseño de productos:**
 - ❑ Comprender las necesidades comerciales
 - ❑ Identificar objetivos impactantes y matrices comerciales relevantes
 - ❑ Definir características de productos o historias de usuarios para esos objetivos
 - ❑ Reconocer los problemas subyacentes para definir que ML es más adecuado
 - ❑ Diseñar la experiencia del usuario.

❑ **Análisis de seguridad:**

- ❑ Asegurar de que el sistema de software, los datos y el modelo estén seguros y que no se revele información de identificación personal (PII) o datos sensibles.

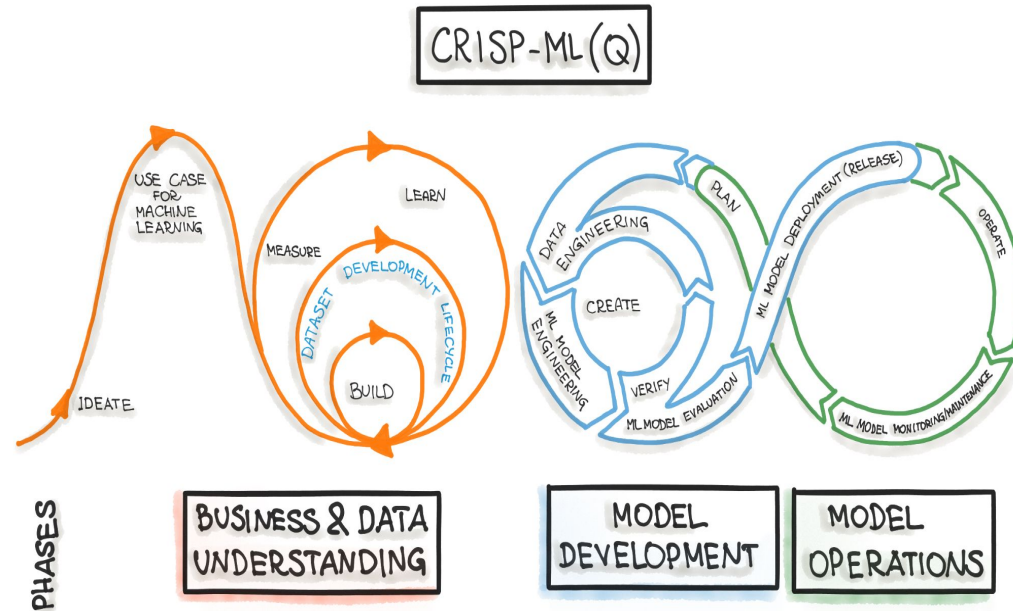


- ❑ **Ética de IA:** garantizar el cumplimiento de todas las leyes aplicables y agregar medidas para proteger contra cualquier tipo de sesgo (por ejemplo, limitar el alcance del modelo, agregar supervisión humana, etc.)



Ciclo de vida de desarrollo de modelos

- ❑ Business and Data Understanding (Planning)
- ❑ Data Engineering (Data Preparation)
- ❑ Machine Learning Model Engineering
- ❑ Evaluating Machine Learning Models
- ❑ Model Deployment
- ❑ Monitoring and Maintenance



Business and Data Understanding (Planning)

- ❑ Definir objetivos de negocio. -
- ❑ Traducir los objetivos comerciales en objetivos de ML -
- ❑ Recopilar y verificar datos. -
- ❑ Evaluar la viabilidad del proyecto. -

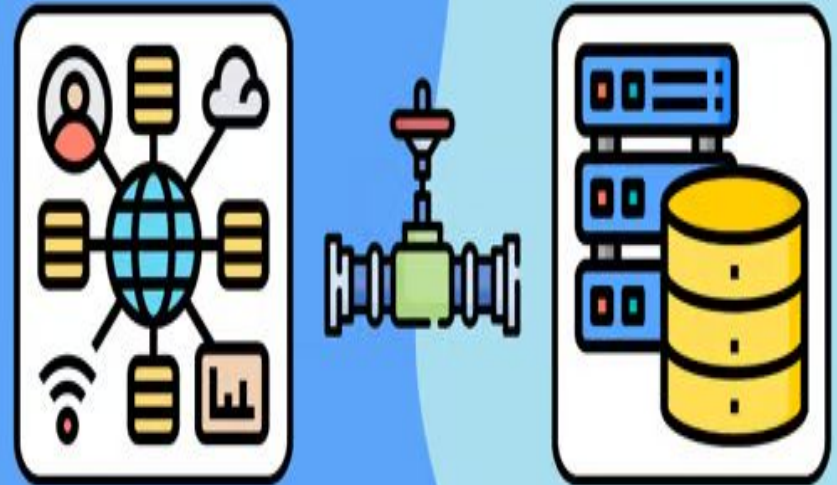


Debemos analizar

- ☐ Disponibilidad de los datos: ¿tenemos suficientes datos disponibles para entrenar el modelo?,
- ☐ Aplicabilidad: ¿esta solución resolverá el problema o mejorará el proceso actual?, ¿Podemos incluso usar el aprendizaje automático para resolver este problema?
- ☐ Restricciones legales: ¿Estamos siguiendo una forma ética de recopilar los datos? ¿Cuál será el impacto de esta aplicación en la sociedad?
- ☐ Robustez y escalabilidad: ¿Es esta aplicación lo suficientemente robusta? ¿Es escalable?
- ☐ Disponibilidad de recursos: ¿tenemos suficientes recursos informáticos, de almacenamiento, de red y humanos? ¿Tenemos profesionales cualificados?

Data Engineering (Data Preparation)

- ❑ Data collection and labeling
- ❑ Data Cleaning
- ❑ Data processing
- ❑ Data management



Data collection and labeling

Recolección de datos:

- ❑ ¿Cómo recopilaremos los datos ?
- ❑ ¿Usaremos datos internos, de código abierto, comprándolos o usaremos datos sintéticos ?

Etiquetado de datos:

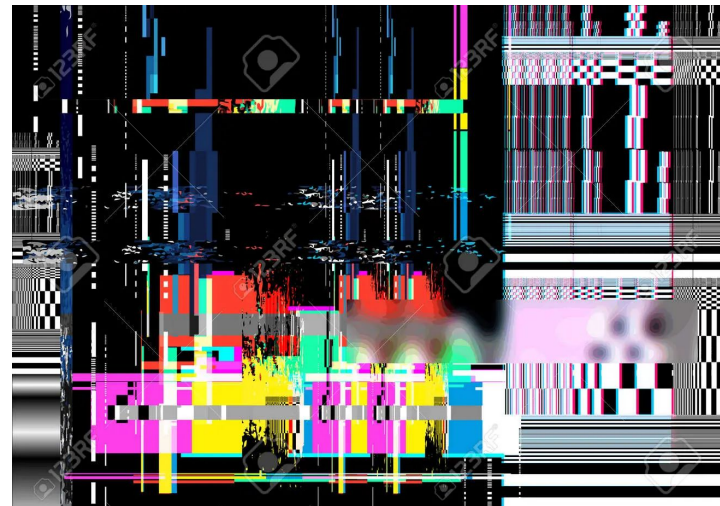
- ❑ La recopilación y el etiquetado de datos requieren la mayor parte de los recursos de la empresa: dinero, tiempo, profesionales, expertos en la materia y acuerdos legales.



- ❑ Limpieza de los datos imputando valores faltantes,
- ❑ Análisis de datos mal etiquetados,
- ❑ Eliminación de valores atípicos y reducción de ruido.

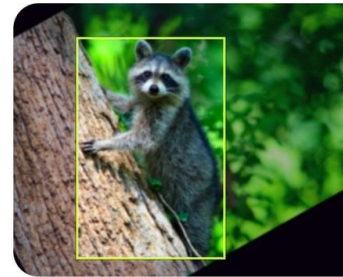
Creación de una canalización (pipeline) de datos para automatizar el proceso

Realizar la verificación de la calidad de los datos.



Data processing

- ❑ Implica la selección de características,
- ❑ Tratamiento de clases desequilibradas,
- ❑ Ingeniería de características,
- ❑ Aumento de datos
- ❑ Normalización y escalado de los datos.



Flip



Crop



Grayscale



Hue



Exposure



Blur



90 Rotate



Shear



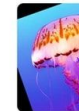
Noise



Brightness



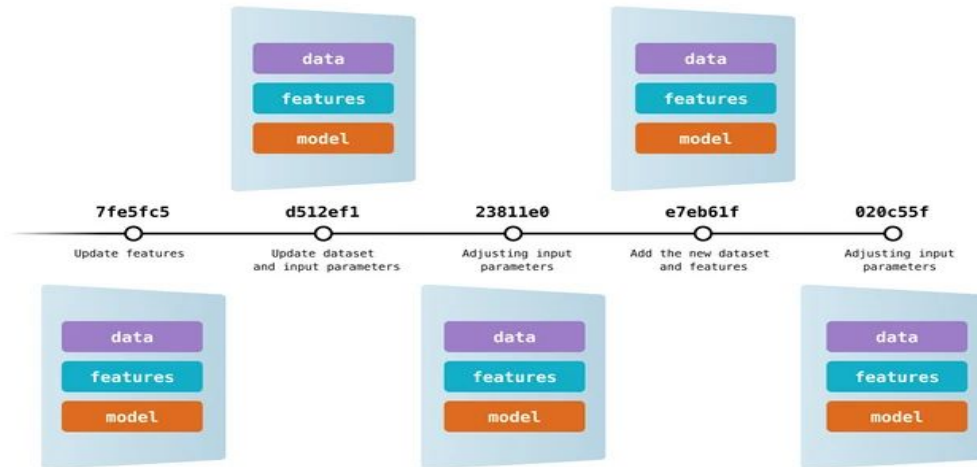
Cutout



Rotatic

Data management

- ❑ Solución de almacenamiento de datos
- ❑ Control de versiones de datos para reproducibilidad
- ❑ almacenamiento de metadatos y creación de canalizaciones ETL



- ❑ Construcción de arquitecturas de modelo mediante una amplia investigación.
- ❑ Definición de métricas del modelo
- ❑ Entrenamiento y validación del modelo en el conjunto de datos de entrenamiento y validación.
- ❑ Seguimiento de experimentos, cambios de código y canalizaciones de aprendizaje automático.
- ❑ Interpretar los resultados mediante la incorporación de expertos en conocimiento del dominio

Versionado y reproducibilidad:

- ❖ Los hiperparámetros,
- ❖ los experimentos de ML,
- ❖ la arquitectura del modelo,
- ❖ el entorno de desarrollo y los metadatos se almacenan

Model Evaluation

- ❑ Prueba de Métricas
- ❑ Prueba del modelo en un conjunto de datos de prueba
- ❑ Pruebas con datos aleatorios y mundo real
- ❑ Pruebas de rendimiento (Hardware)
- ❑ Toma de decisión de implementación

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN



- ❑ Integración del modelo ML en el sistema de software existente
- ❑ Modo de mantención y predicción, (por lotes o en línea)
- ❑ Exponer su funcionalidad predictiva como tableros interactivos, predicciones precalculadas
- ❑ Envolver el modelo ML como un componente en una arquitectura de software(web service, Microservicios)

- ❑ Supervisión del rendimiento del modelo y mantenimiento
- ❑ Mitigar el efecto de degradación del modelo o "obsolescencia del modelo"
- ❑ Supervisión de los recursos computacionales

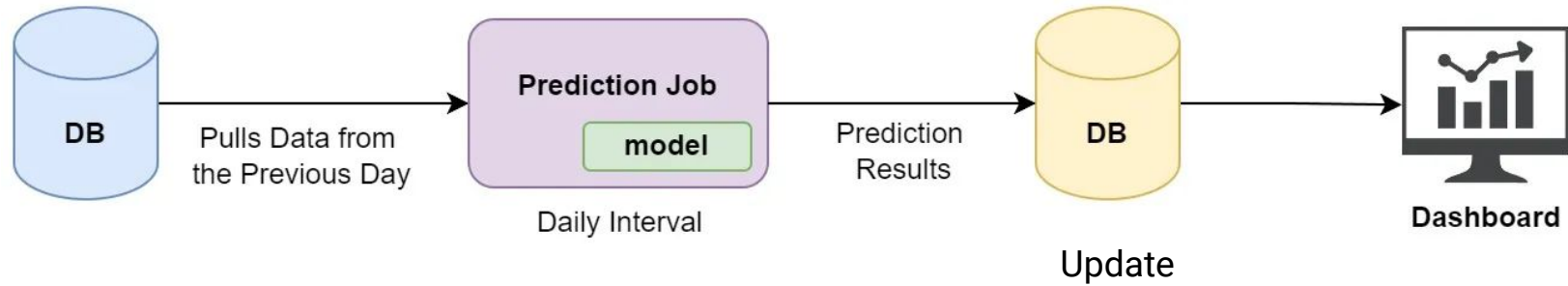
La decisión de las tareas de monitoreo conduce a la segunda tarea:

- Actualizar el modelo ML
- Reflexionar sobre el caso de uso comercial y Tarea

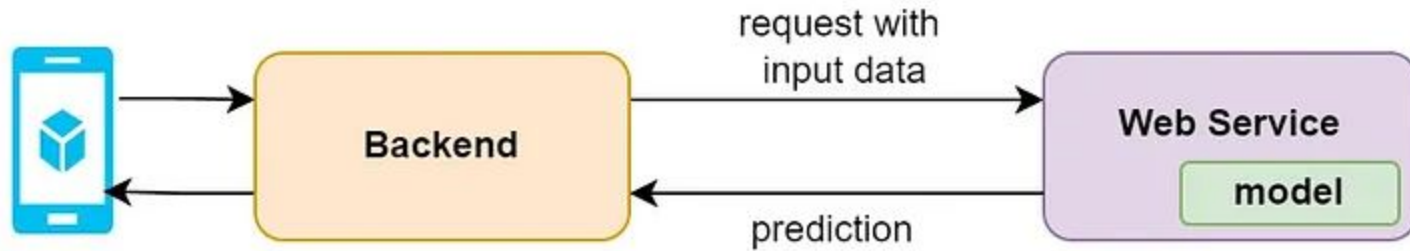
Primero, debemos preguntarnos si queremos que las predicciones se hagan de inmediato o si pueden esperar una hora, un día.

- ❑ Por lotes o fuera de línea : En caso de que podamos esperar un poco
- ❑ Implementación en línea: En caso que necesitemos las predicciones del modelo lo antes posible, el modelo estará funcionando todo el tiempo.
- ❑ Servicio web: En este caso implementamos nuestro modelo como un servicio web y podemos enviar solicitudes HTTPS y obtener la predicción de eso
- ❑ Streaming: En el caso que haya un flujo de eventos, el modelo escucha los eventos y reacciona ante ellos.

Por lotes o fuera de línea



Web Service



Streaming

