

Tópicos avanzados en IA

Vistazo general del módulo

Jose Laruta

Tópicos selectos en IA - UPB - Septiembre 2024

Agenda

1. Presentación
2. Logística
3. Generalidades

Acerca de mi

Machine Learning Engineer @
Jalasoft

- Deep learning
- MLOps
- Backend Dev
- Cloud
- Python, SQL, Scala



Objetivos

Introducir conceptos de preparación, evaluación y despliegue de modelos de ML/AI para aplicaciones de la vida real.

Explorar tecnologías y herramientas para facilitar la gestión del ciclo de vida de un modelo de ML/AI.

Requisitos

- Programación en Python
- Conceptos de AI/ML
- Desarrollo web (Backend)
- Cloud Providers (Google Cloud Platform)
- Ideas frescas

Contenido general:

1. Intro y Herramientas
2. Despliegue de modelos de visión.
3. Despliegue de modelos de lenguaje.
4. IA embebida.
5. Aplicaciones y proyecto final.

Logística

Sesiones presenciales en vivo:

- De lunes a viernes: **7:15 - 9:15** (Aula Topater, Campus Posgrado)
- Tolerancia: 10 minutos como máximo.
- Comunicación: Grupo de Telegram, sesiones asíncronas a coordinar.

Logística

Herramientas y entorno

El entorno de desarrollo se basa en herramientas gratuitas y open source, es necesario contar con un entorno uniforme para poder seguir los tutoriales y realizar los ejercicios de programación.

- **Hardware**: 4GB RAM, quad-core, 64bits (*mínimo, recomendado 8+GB*)
- **Sistema operativo**: indistinto (*Recomendado: Ubuntu 22.04*)
- **Lenguaje de Programación**: Python 3 (v3.10+)
- **Entorno de desarrollo integrado**: Visual Studio Code, Pycharm

Logística

Herramientas y entorno

Es importante contar con las herramientas adecuadas (laptop, acceso a plataformas) para poder aprovechar al máximo los contenidos impartidos, los ejercicios y las tareas.

⚠ Si se tiene una dificultad con alguna herramienta/plataforma, por favor informe inmediatamente al profesor!

Logística

Fechas Importantes

Primer Parcial:

Viernes 4 de Octubre de 2024

Segundo Parcial:

Viernes 18 de Octubre de 2024

Presentación Proyecto final:

Lunes 4 de Noviembre de 2024

Logística

Evaluación

1. Primer y segundo parcial (30%):

- a. Laboratorio - 70%
- b. Tarea - 30%
- c. Participación - 5%

2. Proyecto Final (40%):

- a. Código fuente - 50%
- b. Informe - 30%
- c. Presentación - 20%

Logística

Desempeño en la materia

1. Sobresaliente (90 - 100): Va más allá de lo esperado.
2. Bueno (70 - 90): Cumple con todo lo exigido.
3. Meh (< 70): Cumple lo imprescindible.

Logística

Actividades evaluadas

1. **Laboratorio**: Mini proyectos y aplicaciones relacionados a temas vistos en clase.
2. **Tareas**: Retos de código en clase o para la casa. Trabajos de investigación y análisis.
3. **Participación**: Participación general en clase, preguntas realizadas y respondidas.

Logística

Actividades evaluadas

4. **Proyecto final**: Aplicación completa que use los algoritmos y técnicas aprendidas en la materia.

Contenido

Introducción y herramientas

1. Introducción al despliegue de IA
2. Vistazo general de modelos de Computer Vision
3. Vistazo general de modelos de lenguaje
4. Herramientas, plataformas y frameworks

Contenido

Despliegue de modelos de visión

1. Introducción a modelos de visión artificial
2. Preparación de datos
3. Entrenamiento
4. Evaluación y métricas
5. Conversión a formatos desplegados
6. Despliegue de servicios

Contenido

Despliegue de modelos de lenguaje

1. Introducción a NLP y modelos de lenguaje
2. Preprocesamiento de texto
3. Entrenamiento
4. Fine tuning
5. Despliegue de servicios

Contenido

Tópicos avanzados en despliegue

1. Escalabilidad y balanceo de carga
2. Optimización y plataformas de cómputo
3. Despliegue en servicios cloud
4. Consideraciones generales

Contenido

IA embebida

1. Sistemas embebidos y edge computing
2. Optimización y cuantización
3. Despliegue de modelos de visión
4. Despliegue de modelos NLP - LM

Contenido

Proyecto y aplicaciones en el mundo real

1. Definición y elaboración de proyectos
2. Presentación y evaluación
3. Tendencias y tecnologías emergentes

Despliegue de modelos de ML/AI

