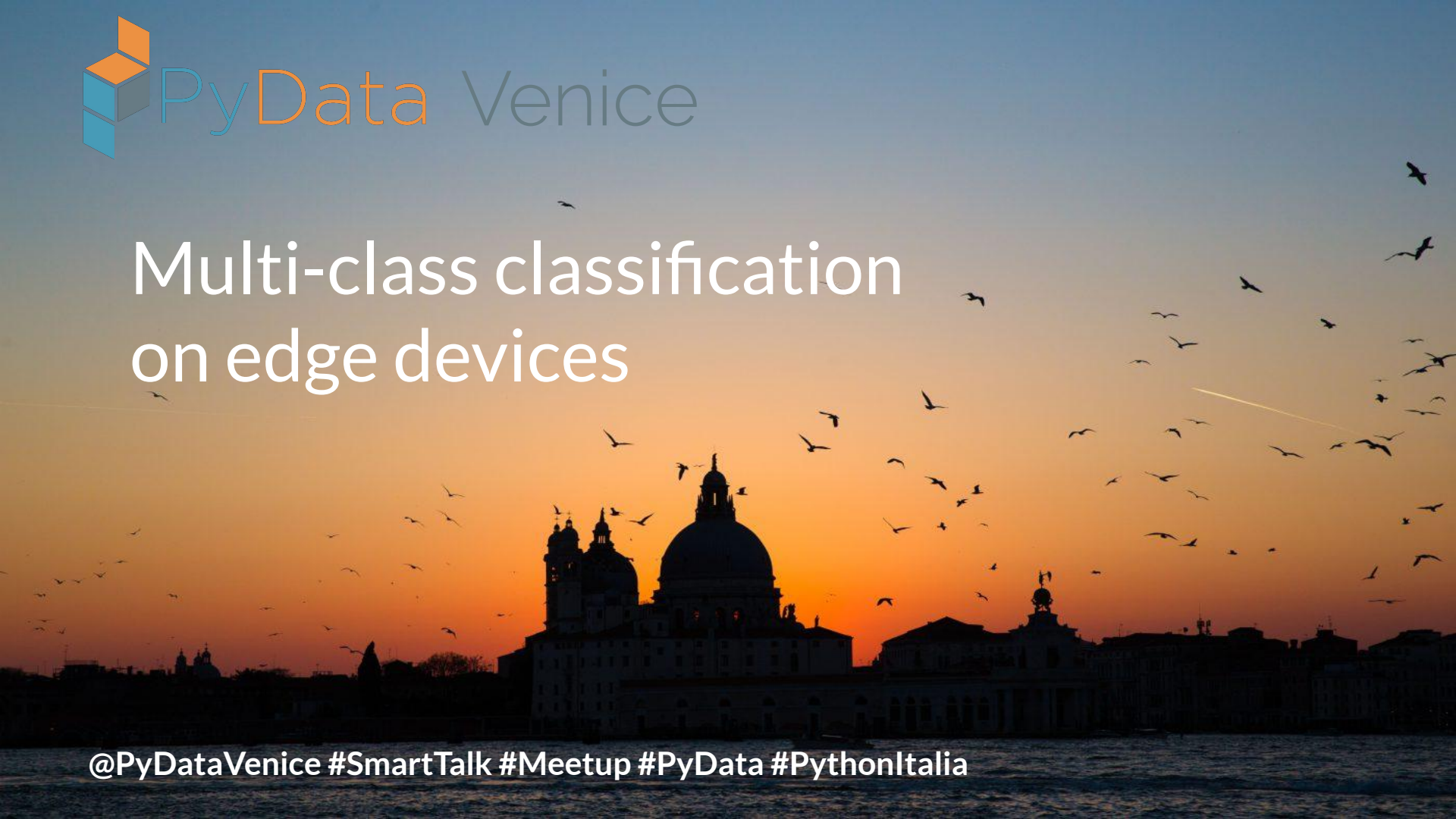
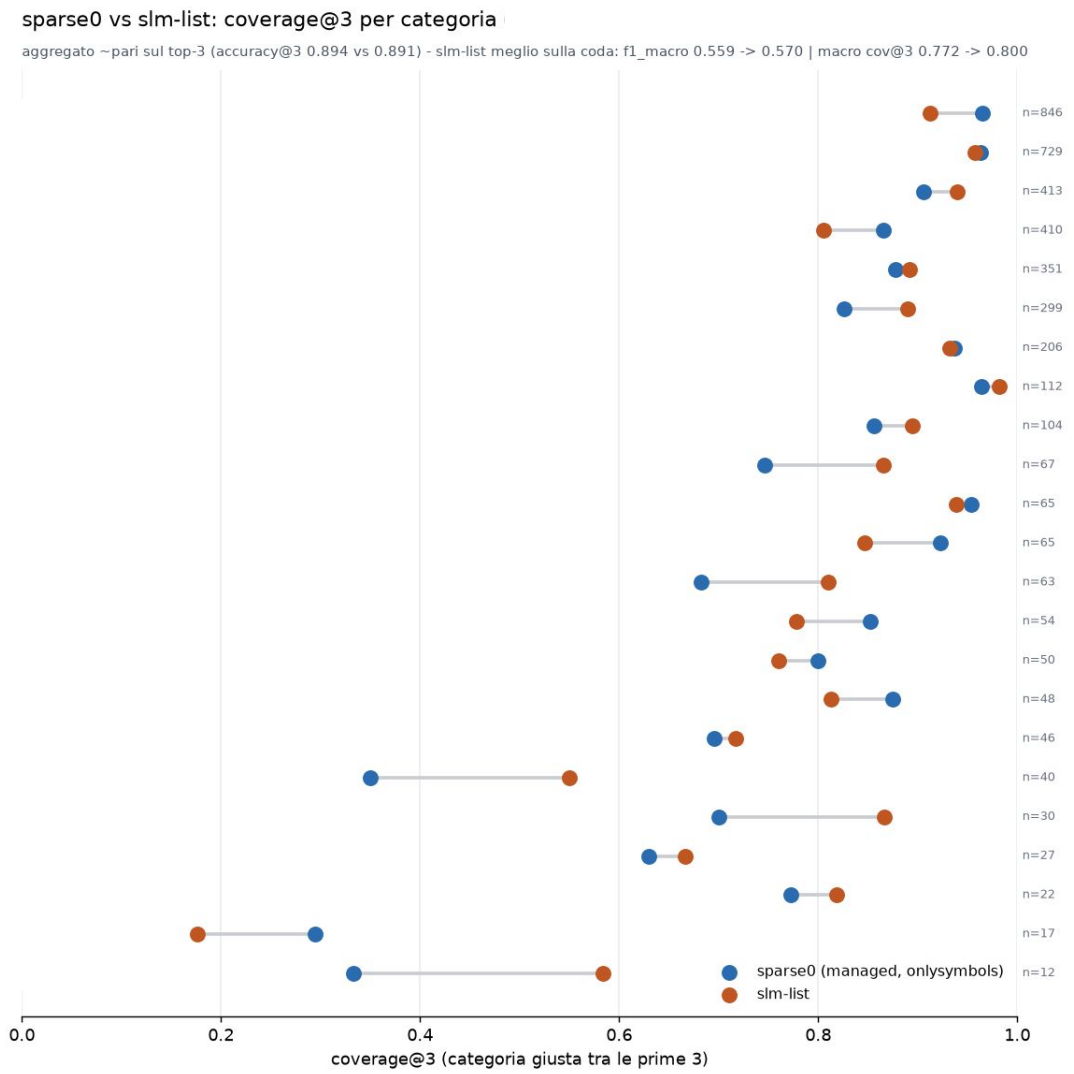

PyData Venice
Multi-class classification
on edge devices
@PyDataVenice #SmartTalk #Meetup #PyData #PythonItalia

PyData

Classification

- Classic algorithm
- SLM



Models







Inference

Modello / Caso d'uso	Risorse indicative (GPU / RAM)	Esempio pratico
Algoritmi classici (es. ML tradizionale, regressioni, clustering)	**Raspberry Pi** (2–8 GB RAM, pochi GB di storage)	 Raspberry Pi 4/5 (4–8 GB)
SLM 0.3B (es. LLM 300M–400M)	**Mobile / smartphone** (4–8 GB RAM, 1–3 GB storage per modello)	 Smartphone moderno (8 GB RAM o più)
1.5B (es. LLM 1.5 miliardi)	**~ 8–16 GB di RAM** (meglio 16 GB+ per più fluidità e contesto)	 Laptop / mini PC (16–32 GB RAM)
Opus 5.5 (in locale) (modello molto grande)	**~ 64–128 GB di RAM** (dipende dalla quantizzazione e dal contesto), più storage SSD	 Desktop / workstation (64–128 GB RAM)

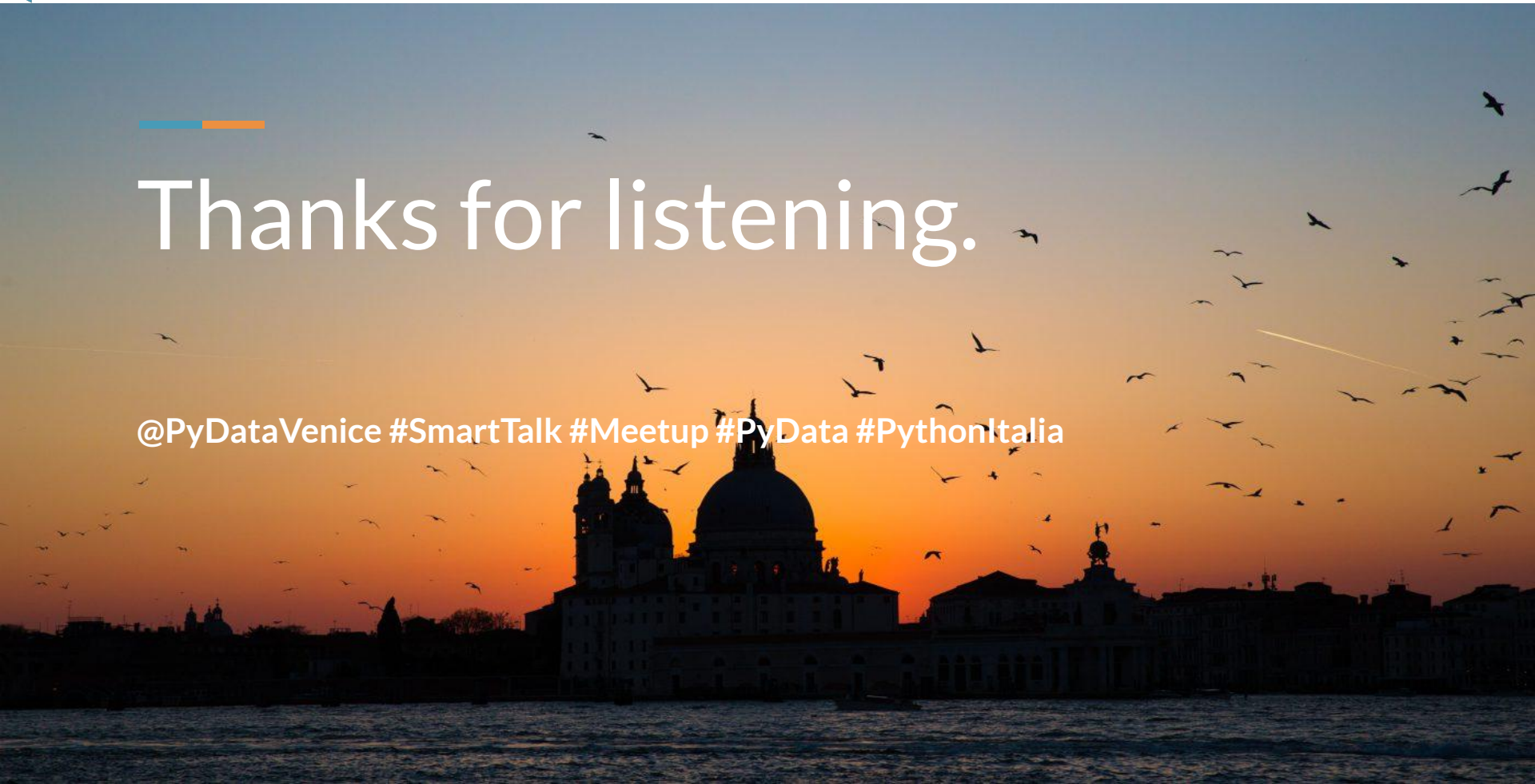
Training

Modello / Caso d'uso	Risorse indicative (GPU / RAM)	Esempio pratico
Algoritmi classici (non serve training pesante)	**CPU o GPU leggera** (anche solo 8–16 GB RAM)	 Laptop / desktop qualsiasi
SLM 0.3B (fine-tuning leggero)	**1× GPU** (8–16 GB) + 16–32 GB RAM	 GPU consumer (es. RTX 3060/4060)
1.5B (fine-tuning medio)	**1–2× GPU** (24–48 GB) + 32–64 GB RAM	 2× GPU (es. RTX 4090) o cloud equivalente
Opus 5.5 (training completo)	**8–64× GPU** (a seconda della disponibilità e del tempo) + **256 GB – 1 TB RAM** + storage veloce (NVMe)	 Cluster / cloud (es. AWS, GCP, Azure) o data center dedicato

venice.pydata.org



PyData



Thanks for listening.

@PyDataVenice #SmartTalk #Meetup #PyData #PythonItalia