

شرح تمرین

- **Triton Inference Server** یک پلتفرم ارائه مدل یادگیری عمیق open-source است که توسط NVIDIA برای سادهسازی استقرار یک مدل، توسعه یافته است. **triton** یک راه حل قوی و مقیاس پذیر برای میزبانی و مدیریت مدل های یادگیری عمیق ارائه می دهد.
- در این تمرین، هدف استقرار یک مدل یادگیری عمیق با استفاده از **triton** می باشد.
- مدل استفاده شده در تمرین اول برای طبقه بندی تصاویر گربه و سگ را به عنوان مدل یادگیری عمیق در این تمرین استفاده کنید.