

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

# **(تمارین درس مباحث ویژه)**

## **بخش ششم**

**مدرس: مهندس احمدزاده**

**دانشجویان:**

**عباس کارگر جهرمی / مهدی فرازنده شهرکی**

**دانشکده ملی مهارت میناب**

**بهمن ۱۴۰۳**

## A. چرا Distributed Ledger Technology (DLT) در مدیریت داده‌ها کاربرد دارد؟

DLT یا فناوری دفتر کل توزیع‌شده ساختاری است که داده‌ها به صورت مشترک، غیرمتمرکز و در گره‌های مختلف ذخیره می‌شوند.

دلایل کاربرد در مدیریت داده:

- عدم تمرکز: نیاز به سرور مرکزی را از بین می‌برد.
- شفافیت و ردیابی بالا: هر تراکنش قابل پیگیری و تغییرناپذیر است.
- امنیت بالا: چون داده‌ها در چندین نسخه و گره ذخیره می‌شوند، تغییر یا دستکاری سخت است.
- قابلیت اطمینان: حتی اگر یکی از گره‌ها از کار بیفتد، داده از بین نمی‌رود.
- کاربرد: در سیستم‌های مالی، زنجیره تأمین، هویت دیجیتال و سلامت.

## B. چرا Blockchain برای ذخیره‌سازی داده‌های امن استفاده می‌شود؟

Blockchain نوعی DLT است که داده‌ها را در بلوک‌هایی زنجیروار ذخیره می‌کند. ویژگی‌های امنیتی آن:

- غیرقابل تغییر بودن: (Immutability) پس از ثبت، تغییر یک بلوک باعث بی‌اعتباری کل زنجیره می‌شود.
- استفاده از رمزنگاری: هر بلوک شامل هش رمزنگاری شده‌ای است.
- اجماع: (Consensus) قبل از ثبت هر بلوک، شبکه باید آن را تأیید کند.
- مناسب برای داده‌های حساس: مثل سوابق پزشکی، قراردادهای دیجیتال و دارایی‌های دیجیتال.

## C. چرا GANs (Generative Adversarial Networks) در علم داده پیشرفته کاربرد دارند؟

GANs از دو شبکه عصبی تشکیل شده‌اند:

**Generator** (تولیدکننده) و **Discriminator** (تشخیص‌دهنده) که باهم رقابت می‌کنند. کاربردهای پیشرفته:

- تولید تصاویر، ویدیو، صدا و متن مصنوعی واقع‌گرایانه
- توسعه داده‌های مصنوعی (Synthetic Data) برای آموزش مدل‌ها
- ترمیم تصاویر آسیب‌دیده
- ترجمه سبک تصاویر یا صدا (مانند تبدیل عکس روز به شب)

- در پزشکی، طراحی دارو، بازی سازی و هنر دیجیتال هم بسیار کاربرد دارند.

## D. t-SNE و PCA چه تفاوتی دارند؟

هر دو ابزار کاهش ابعاد هستند، اما تفاوت های کلیدی دارند:

t-SNE	PCA	ویژگی
غیر خطی	خطی	نوع
حفظ فاصله های محلی در ابعاد پایین	تحلیل واریانس ویژگی ها	روش کار
کندتر و سنگین تر	سریع و سبک تر	سرعت
پایین (ابعاد جدید قابل تفسیر نیستند) بالا (محورهای جدید قابل توضیح اند)	تفسیر پذیری	
بیشتر برای Visualization	قبل از مدل سازی	کاربرد

## E. چرا UMAP برای Dimensionality Reduction استفاده می شود؟

UMAP (Uniform Manifold Approximation and Projection) یکی از پیشرفته ترین روش ها برای کاهش ابعاد است.  
مزایای آن:

- حفظ بهتر ساختار محلی و جهانی داده ها نسبت به t-SNE
- سرعت بیشتر از t-SNE
- مناسب برای Visualizing دیتاست های بزرگ با ابعاد بالا
- کاربرد: در بیوانفورماتیک، بینایی ماشین، یادگیری بدون نظارت، و NLP

## F. NetworkX چرا برای تحلیل گراف ها کاربرد دارد؟

NetworkX یک کتابخانه پایتونی برای تحلیل شبکه ها و گراف ها است.  
قابلیت ها:

- ساخت و تحلیل گراف های جهت دار و بدون جهت
- محاسبه شاخص هایی مثل Degree ، Centrality ، Clustering

- تجسم گرافها (Visualization)
- جستجو در گرافها با الگوریتمهایی مانند DFS ، BFS و Dijkstra
- مناسب برای کاربردهایی مثل تحلیل شبکه اجتماعی، مسیریابی، و بیوانفورماتیک

---

**G. NLTK** چرا برای پردازش زبان طبیعی (NLP) استفاده می‌شود؟

NLTK (Natural Language Toolkit) یکی از کتابخانه‌های پایه‌ای پایتون برای NLP است. قابلیت‌ها:

- Tokenization ( شکستن متن به کلمات یا جملات )
- حذف stop words
- stemming ( و lemmatization ریشه‌یابی کلمات )
- POS tagging ( برچسب‌گذاری نقش دستوری )
- تحلیل گرامری و درخت‌ساز
- مناسب برای یادگیری مفاهیم NLP ، آموزش و تحقیق

---

**H. Scrapy** برای Web Scraping کاربرد دارد؟

Scrapy یک فریم‌ورک قدرتمند پایتونی برای خزیدن (crawl) و استخراج داده از وب است. ویژگی‌ها:

- سریع و مقیاس‌پذیر
- پشتیبانی از چندین درخواست همزمان
- تعریف ساده spider ها برای استخراج داده
- قابلیت ذخیره خروجی به JSON ، CSV ، یا دیتابیس
- مناسب برای استخراج داده‌های ساخت‌یافته از سایت‌های پیچیده و بزرگ.

**BeautifulSoup**. اجرا برای کاری استفاده می‌شود؟

BeautifulSoup یک کتابخانه سبک برای تجزیه HTML (parse) و XML در پایتون است.  
کاربرد:

- استخراج اطلاعات از صفحات وب با استفاده از تگ‌ها، کلاس‌ها یا ID ها
- مناسب برای پروژه‌های سبک‌تر یا زمانی که داده‌ها به صورت استاتیک هستند
- ترکیب آن با کتابخانه‌هایی مثل requests ساده و قدرتمند است
- مثال کاربرد: استخراج عناوین مقالات از سایت خبری یا قیمت‌ها از یک فروشگاه آنلاین