



نام و نام خانوادگی : مرضیه نجفی \_ هدا مهرانی پور

نام استاد : محمد احمدزاده

موضوع : بخش ۶ Advanced Topics

تاریخ : ۱۴۰۴/۰۱/۲۰

## چرا (distributed ledger technology (dlt در مدیریت داده ها کاربرد دارد؟

(Distributed Ledger Technology (DLT) یا فناوری دفتر کل توزیع شده، یک سیستم امن و شفاف برای ثبت و مدیریت داده‌ها است که به وسیله شبکه‌ای از نودها (nodes) عمل می‌کند. در این نوع فناوری، داده‌ها به صورت توزیع شده در بین چندین قابل اعتماد ثبت می‌شوند، که این ویژگی‌ها باعث می‌شود DLT در مدیریت داده‌ها بسیار کارآمد باشد.

## blockchain چرا برای ذخیره سازی داده های امن استفاده می شود؟

بلاکچین برای امنیت داده‌ها خوبه، چون:

- پراکنده است : اطلاعات همه جا پخش شده، هک کردن سخت‌تره.
- رمزنگاری شده : داده‌ها قفل شدن، کسی نمی‌تونه دست بزنه.
- تغییرناپذیره : چیزی که ثبت شد، دیگه پاک نمی‌شه.
- شفافه : همه می‌تونن ببینن چی شده.
- امضای دیجیتال : هر کسی مسئول کار خودشه.
- اجماع : همه باید توافق کنن تا چیزی ثبت بشه.

## چرا GANs (Generative Adversarial Networks) در علم

### داده پیشرفته کاربرد دارند؟

GANs به دلیل ویژگی‌های خاص خود در علم داده پیشرفته به کار می‌روند.

۱. تولید داده‌های مصنوعی: شبیه‌سازی داده‌های واقعی برای افزایش حجم و تنوع داده‌ها.

۲. تعادل داده‌ها: تولید داده برای کلاس‌های کم‌نمایش مانند در مسائل پزشکی.

۳. بهبود کیفیت: تبدیل داده‌های با کیفیت پایین به داده‌های با کیفیت بالا.

۴. افزایش دقت مدل‌ها: آموزش مدل‌های یادگیری ماشین با داده‌های تولیدی برای افزایش دقت.

۵. کاربرد در هنر: خلق هنر دیجیتال و طراحی خودکار.

۶. درک ویژگی‌ها: استخراج الگوها و ویژگی‌های پیچیده داده‌ها.

۷. یادگیری تقویتی: شبیه‌سازی داده‌ها برای آموزش عاملی‌ها.

۸. مدل‌سازی توزیع‌ها: مدل‌سازی توزیع‌های پیچیده و غیر معمول.

نتیجه: GANها به عنوان ابزاری کارآمد در تولید و بهبود داده‌ها، به حل مسائل مختلف در علم داده کمک می‌کنند.

### t\_SNE و PCA چه تفاوتی دارند؟

PCA (تحلیل مؤلفه‌های اصلی):

- روش: خطی، حفظ بیشترین واریانس داده‌ها.

- هدف : کاهش ابعاد برای ساده‌سازی داده‌ها و پیش‌پردازش.

- خروجی : امکان تعیین تعداد ابعاد.

- سرعت : سریع.

(t-SNE (t-distributed Stochastic Neighbor Embedding :

- روش : غیر خطی، حفظ ساختار محلی و خوشه‌بندی داده‌ها.

- هدف : تجسم داده‌های با ابعاد بالا و کشف الگوها.

- خروجی : معمولاً ۲ یا ۳ بعد.

- سرعت : کند.

PCA برای کاهش ابعاد و ساده‌سازی داده‌ها، t-SNE برای تجسم و کشف الگوها در

داده‌های پیچیده.

## چرا UMAP برای Dimensionality Reduction استفاده می

### شود؟

۱. حفظ ساختار محلی : UMAP نقاط نزدیک در فضای اصلی را در فضای کاهش یافته

نیز نزدیک نگه می‌دارد.

۲. سرعت و کارایی : UMAP سریع‌تر و مقیاس‌پذیرتر از روش‌های مشابه مثل t-SNE

است.

۳. انعطاف‌پذیری : این الگوریتم دارای پارامترهای قابل تنظیم است که به کاربران

امکان بهینه‌سازی می‌دهد.

۴. تئوری قوی : UMAP بر اساس نظریه منیفلد طراحی شده که دقت بالایی در حفظ

ساختار داده‌ها دارد.

۵. عملکرد بهتر در تجسم : UMAP قابلیت کشف خوشه‌ها و نمایش بهتر داده‌ها را دارد.

۶. قابلیت تعمیم‌پذیری : می‌توان آن را بر روی داده‌های جدید اعمال کرد که مناسب استفاده مداوم است.

UMAP به دلیل این ویژگی‌ها به عنوان ابزاری مؤثر برای کاهش ابعاد در علم داده و یادگیری ماشین شناخته می‌شود.

## **NetworkX چرا برای تحلیل گراف‌ها کاربرد دارد؟**

NetworkX برای تحلیل گراف‌ها به این دلیل کاربرد دارد:

- انعطاف‌پذیری : انواع گراف‌ها را پشتیبانی می‌کند.
  - الگوریتم‌های زیاد : الگوریتم‌های مختلف تحلیل گراف را دارد.
  - سادگی : کار با آن آسان است.
  - تجسم : گراف‌ها را می‌توان دید.
  - قابل گسترش : می‌توان الگوریتم‌های جدید به آن اضافه کرد.
  - پشتیبانی : جامعه کاربری قوی دارد.
- NetworkX یک ابزار قدرتمند برای تحلیل گراف‌ها در زمینه‌های مختلف است.

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

NLTK برای پردازش زبان طبیعی (NLP) استفاده می‌شود چون:

- ابزارهای متنوع : برای کارهایی مثل توکن‌بندی، ریشه‌یابی، و برچسب‌گذاری دارد.

- منابع زبانی : واژه‌نامه‌ها و پیکره‌های متنی از پیش آماده دارد.
  - ساده : یادگیری و استفاده از آن آسان است.
  - قابل توسعه : می‌توانید ابزارهای جدید به آن اضافه کنید.
  - سازگار : با سایر کتابخانه‌های پایتون کار می‌کند.
- در کل، NLTK یک ابزار قدرتمند و همه‌کاره برای انجام کارهای مختلف NLP است.

## چرا Scrapy برای Web Scraping کاربرد دارد ؟

Scrapy برای Web Scraping خوبه، چون:

- قدرتمند : به چارچوب کامل برای جمع‌آوری منظم اطلاعاته.
  - سریع و بزرگ : خیلی سریع می‌تونه اطلاعات زیادی رو جمع کنه.
  - انعطاف‌پذیر : می‌تونه داده‌ها رو از صفحات پیچیده استخراج کنه.
  - مدیریت خودکار : درخواست‌ها و پاسخ‌ها رو خودش مدیریت می‌کنه.
  - پردازش داده : داده‌ها رو تمیز و مرتب می‌کنه و ذخیره می‌کنه.
  - تنظیم‌پذیر : می‌شه رفتارش رو تغییر داد و خطاها رو مدیریت کرد.
  - جامعه قوی : پشتیبانی خوبی داره و خیلی‌ها ازش استفاده می‌کنن.
- Scrapy به ابزار سریع، قدرتمند و قابل تنظیم برای جمع‌آوری اطلاعات از وب سایت هست.

## BeautifulSoup چرا برای کاری استفاده می‌شود؟

Beautiful Soup بیشتر برای کارهای زیر استفاده می‌شود:

تجزیه HTML و BeautifulSoup

پیمایش در درخت DOM

جستجو و فیلتر کردن المان‌ها

استخراج داده

Beautiful Soup ابزاری برای تجزیه، پیمایش و استخراج داده از صفحات HTML و XML است. معمولاً به عنوان بخشی از یک فرآیند web scraping استفاده می‌شود.

Beautiful Soup برای:

خوندن و فهمیدن HTML و XML : مثل این‌ها که به نقشه برای صفحات وب داشته باشی.  
گشتن تو صفحات وب : راحت می‌تونی بین قسمت‌های مختلف صفحه بگردی.  
پیدا کردن چیزهای خاص : سریع می‌تونی تگ‌ها و اطلاعات مورد نظرت رو پیدا کنی.  
برداشتن اطلاعات : بعد از پیدا کردن، می‌تونی اطلاعات رو از صفحه برداری.  
در کل، Beautiful Soup ابزاری ساده برای کار کردن با محتوای صفحات وب و استخراج اطلاعات از اون‌هاست.