

بسم نام خدا

آفریننده ملی مهارت میناب

تدریسات بخش ۴۰

نام و نام خانوادگی: فردین صداقت

واحد درسی: مباحث ویژه

رشته: مهندسی کامپیوتر

مدرس: مهندس محمد احمدزاده

فردین: ۱۴۰۴

A - چرا Distributed Ledger Technology (DLT) در مدیریت

دادها کاربرد دارد؟

دفترکل توزیع شده (Distributed ledger) یا دفترکل اشتراکی، یک پایگاه داده

مخزن اطلاعات است که بین چندین فرد (Node) یا دستگاه های محاسباتی پخش می شود. هر بار که نسخه جدیدی از اطلاعات ثبت می شود، کپی ها در دستگاه های دیگر به روز می شوند.

در هر روز زمانی می کنند. همه نسخه ها در دفترکل توزیع شده با هم برابر هستند و سرور مرکزی وجود ندارد.

دفترکل توزیع شده چیزی جز یک دفترچه اشتراکی نیست (دفترچه عمومی) و یک سیستم

برای مدیریت معاملات مانند خرید و فروش ارز دیجیتال است اما تفاوت در ماهیت

چگونه نگه داری اطلاعات و مدیریت دفترچه نهفته است. به عنوان مثال: در مورد کالک های

دیجیتالی شده مانند بیت کوین (Bitcoin)، کپی ها در شبکه (هرگز یک نسخه در سیستم) نسخه

اصلی دفترچه را دارد و می تواند سرابع تراکنش های خود را ببیند.

B - Blockchain چرا برای ذخیره سازی داده های امن استفاده می شود؟

ذخیره سازی بلاک چین یک جایگزین بالقوه از آن تراکنش های غیرقابل اعتماد تراست برای ذخیره

سازی آبروی متراست است.

ذخیره داده ها در دما گره ها، مزایای امنیتی، ذخیره سازی بلاک چین، مانع از سرافراز می شود.

فایل ها و توزیع آنها در شبکه دفترکل دسترسی هر چه به داده ها ساخت می کنند. هیچ نهاد مرکزی

برای کنترل دسترسی به فایل ها، داشتن کلیدهای مورد نیاز برای رمزگشایی فایل ها وجود ندارد. کلید

های خصوصی به طور کامل توسط کاربر کنترل می شود و از نظر شوری دسترسی شخص ثالث

به قابل یا غیر ممکن است. بنابراین امنیت و حریم خصوصی را ارتقا می دهد. ذخیره سازی بهای چین و چین می تواند به سیستم های ذخیره سازی سر و سر و قابل تنظیم تر اجازه دهد زیرا که این ترانزیت و تطبیق مانند سرعت باز و بسته داف و رنگی را دستکاری می کنند.

C- چرا Gans (Generative Adversarial Networks)

در علم داده پیشرفت کاربرد دارند!

شبکه های متخاصم مولد (Gans) یکی از قدرتمندترین مدل های مولد محسوب می شوند. در معماری شبکه Gan دو مولف وجود دارد. قطبی باز و بسته D و G که باید در تعادل قرار می گیرند. این دو مولف را می توان با دو شبکه عصبی مجزا می توان ساخت. یکی تولید کننده D و دیگری مولد G است. آنها با هم رقابت می کنند تا یکدیگر را حذف یا از بین ببرند و در نهایت به تعادل می رسند.

D - PCA و t-SNE چه تفاوتی دارند؟

t-Distributed Stochastic Neighbor t-SNE

Embedding یک تکنیک کاهش ابعاد غیر خطی است که برای حفظ ساختار محلی داده ها طراحی شده است. به این معنی که t-SNE تلاش می کند تا همسایگی های محلی داده ها را در فضای با ابعاد کمتری حفظ کند. PCA برعکس، یک تکنیک کاهش ابعاد خطی است که برای یافتن جهت های بیشترین واریانس در داده ها طراحی شده است. t-SNE معمولاً برای تجسم داده ها استفاده می شود زیرا می تواند ساختار محلی داده ها را بهتر از PCA حفظ کند. PCA معمولاً برای کاهش ابعاد داده ها از قبل از استفاده از الگوریتم های ماشین (Preprocessing) استفاده می شود زیرا می تواند پیچیدگی محاسباتی را کاهش دهد.

E- چرا UMAP برای Dimensional Reduction استفاده می شود؟

این روش جدید کاهش ابعاد باغتران برآورد ضعیفد یکپارخت و افکندش (UMAP) ارائه شده است. این تکنیک غیرخطی است و از ایده t-SNE برای ایجاد فضای بند

پایین برای نمایش داده ها پیروی می کند. با این حال، تفاوت های مهمی دارد.

مدل ریاضی آن براساس هندسه و توپولوژی ریانی است. UMAP از نرمال سازی در خودی احتمال استفاده می کند. در مقابل، یک نسخه معوار از کلا- نزدیکترین همسایه استفاده می شود.

UMAP از نظر محکد نسبت به t-SNE چندین مزیت دارد.

F- چرا برای تحلیل گراف کاربرد دارد؟

Network X یک کتابخانه پایتون برای ایجاد، دست و زدن، مطالعه ساختار و تجزیه و تحلیل شبکه های پیچیده بزرگ است. Network X علاوه بر معور سازی های دوبندی و سه بندی بسیار خوب، بسیار از معیار ها و الگوریتم های استاندارد گراف دارد. اختیار کاربر قرار می دهد که ما فزدر کوتاه ترین مسیر، مرکزیت، رده بندی صفحه و غیره.

G- NLTK چرا برای پردازش زبان طبیعی NLP استفاده می شود؟

NLTK کتابخانه سرنام مجموعه ای است که شامل کتابخانه ها و سرنامه های برای پردازش زبان های آماری است. یکی از قدرتمندترین گزینه ها در این زمینه است که شامل بسته های است تا ماشین ها زبان انسان را بفهمند و با بیانی شیرا به آن پاسخ بدهند.

پردازش زبان طبیعی (NLP) رشته ای است که بر روی قابل استفاده کردن زبان طبیعی انسان را توسط برنامه ها میسر می کند. در این میان NLTK یک بسته پایتون است که می توانید در ارتباط با پردازش زبان طبیعی از آن استفاده کنید.

H-چرا Scrapy برای Web Scraping کاربرد دارد ؟

Scrapy یکی از محبوب ترین کتابخانه های پایتون برای استخراج داده ها از وبسایت ها است. این کتابخانه دریافت داده ها از وبسایت به شیوه ای کارآمد کمک می کند ^{Scrapy} که می تواند تا داده های ماسخدار یافته ای را از وب دریافت کرد که بدوای توان از آنها برای مدل یادگیری ماشین خود استفاده کرده.

1- Beautiful Soup - چرا برای 'لایه' (انتظامی) می شود؟

Beautiful soup یکی از بهترین و محبوب ترین کتابخانه های خذنده وب است که می تواند برای استخراج داده ها از فایل های XML و HTML استفاده شود. این ابزار کمک می کند اسنادی را که از وب (استخراج شده) تصنیف کرده و تجزیه کنید. این امر، سبب صرفه جویی در ساعت ها روزها یا ماه ها در برنامهنویسی و تحلیل گران می شود.