



Namatek
True Education



CENTROID

SSE



www.namatek.com

Sci-kit learn library

کتابخانه sci-kit learn

فهرست مطالب

۱. کتابخانه sci-kit learn چیست؟
۲. کاربرد کتابخانه sci-kit learn
۳. علل اهمیت کتابخانه sci-kit learn
۴. نحوه نصب کتابخانه sci-kit learn
۵. مزایا و معایب استفاده از کتابخانه sci-kit learn
۶. چشم انداز آینده برای کتابخانه sci-kit learn

اگر یک دولوپر یا توسعه دهنده پایتون باشید که به دنبال ارتقای جعبه ابزار هوش مصنوعی خود هستید، پس یک بلیط طلایی برای توسعه راه حل‌های یادگیری ماشینی به صورتی قدرتمند دارید. استفاده از کتابخانه sci-kit learn بلیط برنده شما است. این کتابخانه، فرآیندهایی مانند طبقه بندی، رگرسیون و خوشه بندی را بسیار ساده تر خواهد کرد. زمانی که برای اولین بار از این کتابخانه استفاده می کنید، برای زمانی که قبلاً برای آماده سازی داده های خود برای یادگیری ماشین صرف می کردید، حتماً غبطه خواهید خورد. در این مقاله به بررسی کتابخانه sci-kit learn، مفاهیم اولیه آن، قابلیت های آن، کاربردهای کتابخانه sci-kit learn، علل اهمیت آن، نحوه نصب، مزایا و معایب استفاده از این کتابخانه و چشم انداز آینده sci-kit learn خواهیم پرداخت.

کتابخانه sci-kit learn چیست؟



کتابخانه sci-kit learn، یک کتابخانه منبع باز است که با ارائه طیف گسترده ای از ابزار به منظور یادگیری تحت نظارت و بدون نظارت، یادگیری ماشین را آسان تر و در دسترس تر می کند. با استفاده از این کتابخانه، انجام فرآیندهایی همچون تجزیه و تحلیل داده ها و مدل سازی های پیش بینی به

مراتب آسان‌تر خواهد شد. یکی از بهترین ویژگی‌های این کتابخانه، چگونگی ادغام یکپارچه آن با سایر استک‌های (Stacks) علم داده در پایتون است. این موضوع، هم افزایی عملکرد و سهولت استفاده از آن را افزایش می‌دهد. در نتیجه به علت سادگی، کارایی و ادغام دقیق با موارد زیر به یک ابزار اساسی برای استفاده از یادگیری ماشین در پایتون تبدیل شده است:

• SciPy

• NumPy

• Matplotlib

رابط بصری و مستندات عالی‌ای که در این کتابخانه وجود دارد، آن را به یک کتابخانه برای آموزش یادگیری ماشین و ابزاری قابل اعتماد به منظور استفاده از یادگیری ماشین برای حل مشکلات دنیای واقعی در صنایع و دانشگاه‌ها مبدل ساخته است. استحکام و انعطاف پذیری ای که این کتابخانه دارد، آن را در صنایع مختلف از امور مالی گرفته تا مراقبت‌های بهداشتی ضروری کرده است.

تقاضای فزاینده‌ای که برای هوش مصنوعی پیش بینی کننده و تصمیم گیری‌های مبتنی بر داده وجود دارد، نقش حیاتی sci-kit learn را در تحول دیجیتال و همسویی استراتژیک کسب و کار برجسته کرده است. بنابراین چه مبتدی باشید چه سینیور، این کتابخانه برای ارائه، موارد زیادی را در خود جای داده است. مستندات جامع باعث شده تا با تسلط بر قابلیت‌های آن، سریعاً به سمت پیشرفت حرکت کنید. پس آن را امتحان کنید و ببینید که چگونه می‌تواند پروژه‌های یادگیری ماشین را ارتقا دهد.

مفاهیم اولیه sci-kit learn

برای استفاده از sci-kit learn، ابتدا بهتر است با برخی از اصطلاحات که معمولاً در پروژه های یادگیری ماشین به کار می رود، آشنا شوید:

- **دقت:** نسبت کسری از پیش بینی هایی که یک مدل طبقه بندی به درستی انجام داده است و عبارت اند از:

- در طبقه بندی چند کلاسه، دقت برابر است با تقسیم تعداد کل نمونه ها بر پیش بینی های صحیح

- در طبقه بندی باینری، دقت برابر است با تقسیم تعداد کل مثال ها بر مجموع مثبت های واقعی به علاوه منفی های واقعی

- **داده های نمونه:** نمونه یا ویژگی خاص هر داده، که به صورت x تعریف می شوند. دو دسته از نمونه داده وجود دارد که عبارت اند از:

- داده های برچسب گذاری شده: شامل دو ویژگی و برچسب است که به صورت (x, y) : {Feature & Label} تعریف می شوند.

- داده های بدون برچسب: شامل ویژگی هستند؛ اما برچسب ندارند و به صورت $(x, ?)$: {Feature & ?} تعریف می شوند.

- **ویژگی:** یک متغیر ورودی است که یک ویژگی آن قابل اندازه گیری است. در هر پروژه ماشین لرنینگ یک یا چند ویژگی وجود دارد.

- **خوشه بندی:** تکنیکی است که نقاط داده را بر اساس شباهت های موجود بین آنها، گروه بندی می کند. هر گروه یک Cluster نامیده می شود.

- **K-Mean Clustering**: یک تکنیک یادگیری بدون نظارت است که به دنبال تعداد ثابت K در میانگین نقاط داده است و آنها را به نزدیک ترین خوشه اختصاص خواهد داد.
- **مدل**: رابطه بین ویژگی و برجسب را تعریف می‌کند.
- **رگرسیون در مقابل طبقه بندی**: هر دو مدل‌هایی هستند که به کاربر این امکان را می‌دهند که پیش بینی‌هایی را انجام دهد که به سؤالات مورد نظر وی پاسخ دهد. به عنوان مثال، کدام تیم در یک رویداد ورزشی خاص برنده خواهد شد:
- مدل‌های رگرسیون: یک مقدار عددی یا پیوسته را ارائه می‌دهند.
- مدل‌های طبقه بندی: یک مقدار مقوله ای یا گسسته را ارائه می‌دهند.

قابلیت های کتابخانه sci-kit learn

قابلیت‌هایی که کتابخانه sci-kit learn ارائه می‌دهد، عبارت اند از:

- **رگرسیون**: از جمله رگرسیون خطی و لجستیک
- **طبقه بندی**
- **خوشه بندی**: از جمله $K - Means$
- **انتخاب و ارزیابی مدل**
- **تکنیک کاهش ابعاد**
- **پیش پردازش**: از جمله نرمال سازی حداقل و حداکثر

کاربرد کتابخانه sci-kit learn



کتابخانه sci-kit learn، فقط یک ابزار ضروری برای دانشمندان داده نیست؛ بلکه یک کتابخانه نیروگاهی است که سبب نوآوری در صنایع مختلف می‌شود. از مراقبت‌های مالی و بهداشتی گرفته تا بازاریابی و تحقیق، سازمان‌ها به دنبال بهره‌برداری از قدرت sci-kit learn برای ساخت راه‌حل‌های یادگیری ماشینی پیشرفته هستند. برخی از کاربردهای کتابخانه sci-kit learn در دنیای واقعی به صورت زیر هستند:

تحول در مراقبت‌های بهداشتی با کشف سریع تر دارو

استفاده از الگوریتم‌های sci-kit learn در صنعت مراقبت‌های بهداشتی، سبب انقلابی بزرگ در کشف دارو شده است. یادگیری ماشینی به منظور پیش بینی نحوه تعامل ترکیبات شیمیایی با پروتئین‌های هدف، سبب شده تا سازمان‌هایی همانند Atomwise بیشتر از همیشه به شناسایی سریع‌تر تمامی گزینه‌های دارویی، امیدوار شوند. این سرعت در روند شناسایی

داروها سبب رسیدگی سریع تر به درمان بیماران و کاهش هزینه‌های تحقیق و توسعه دارویی شده است.

افزایش سرعت در کشف تقلب در امور مالی

کتابخانه sci-kit learn، به یکی از ابزارهای مهم و حیاتی در امر مبارزه با تقلب، تبدیل شده است.

سازمان‌هایی همانند JPMorgan از sci-kit learn به منظور تجزیه و تحلیل حجم وسیعی از داده‌های تراکنش و شناسایی الگوریتم‌های قدرتمند در الگوهای غیرعادی استفاده می‌کنند که می‌تواند، نشان دهنده فعالیت متقلبانه آنها باشد. این نوع مؤسسات با کشف هر چه سریع تر کلاهبرداری‌های صورت گرفته می‌توانند، میزان ضرر و زیان را تا حد زیادی کاهش دهند و اعتماد مشتری را حفظ کنند.

تقویت بازاریابی شخصی

کتابخانه sci-kit learn، همچنین با فعال کردن تجربیات شخصی مشتریان، دنیای بازاریابی را متحول کرده است. سازمان‌هایی همانند Spotify و Booking.com از کتابخانه sci-kit learn به منظور ساخت موتورهای نظریه‌ای پیچیده‌ای استفاده می‌کنند که متناسب با ترجیحات هر کاربر، محصولات یا محتوایی متناسب با آنها را ارائه می‌دهد. بازاریابان نیز از کتابخانه sci-kit learn به منظور تقسیم بندی مشتریان و مدل سازی پیش بینی کننده استفاده می‌کنند تا بتوانند پیام مناسب را در زمان مناسب به مخاطبان ارائه دهند.

تسريع تحقیقات علمی

کتابخانه sci-kit learn در قلمروی تحقیقات علمی به شکستن مرزها و انجام اکتشافات پیشگامانه و جدید کمک می‌کند. sci-kit learn به منظور تجزیه و تحلیل مجموعه‌ای از داده‌های پیچیده در فیزیک و نجوم گرفته تا پیش بینی در زمینه‌هایی همچون ژنومیک و علوم اعصاب کمک می‌کند و محققان به منظور استخراج بینش و هدایت نوآوری به ابزارهای همه کاره sci-kit learn تکیه می‌کنند. با ساده سازی جریان‌های کاری تجزیه و تحلیل داده‌ها، کتابخانه sci-kit learn به دانشمندان کمک می‌کند تا بر آنچه که به بهترین شکل ممکن انجام می‌دهند، تمرکز کنند که عبارت اند از پرسیدن سوالات مهم و کشف دانش جدید. مواردی که گفته شد تنها چند نمونه از کاربردهای sci-kit learn در دنیای واقعی و در صنایع مختلف بود. از آنجایی که سازمان‌های بیشتری در حال یادگیری ماشینی هستند و به قدرت آن پی برده‌اند، بدون شک یادگیری sci-kit learn نقش مهمی در شکل دهی آینده نوآوری مبتنی بر داده‌ها خواهد داشت.

علل اهمیت کتابخانه sci-kit learn

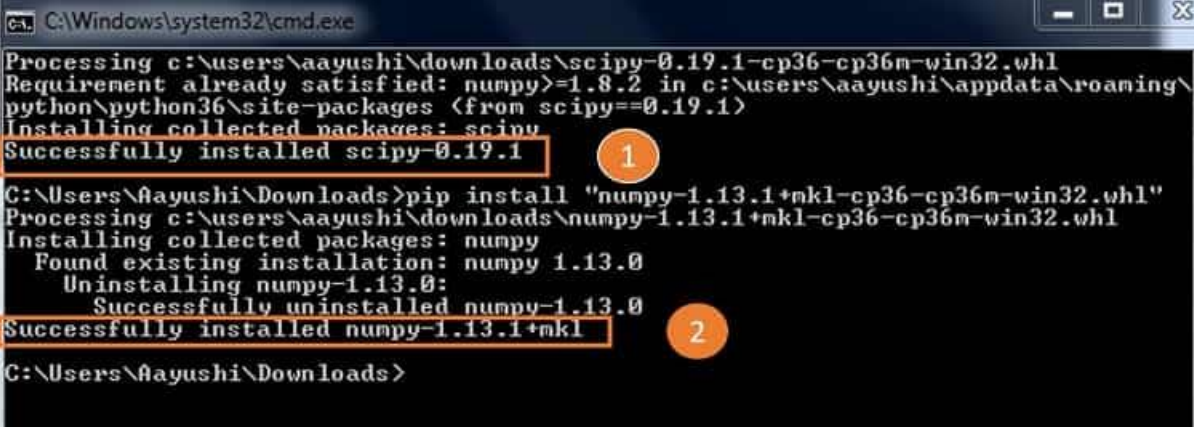


از جمله علل اهمیت کتابخانه sci-kit learn می‌توان به طیف گسترده ای از الگوریتم‌ها برای برنامه‌های مختلف یادگیری ماشین اشاره کرد و همانگونه که در بخش‌های قبلی گفته شد عبارت اند از:

- طبقه بندی
- خوشه بندی
- رگرسیون
- کاهش ابعاد
- انتخاب مدل

کتابخانه sci-kit learn، کاربر را به ابزارهایی به منظور هدایت نوآوری در صنایع مجهز می‌کند؛ اما شاید مهمتر از همه این موارد، تسلط کاربر بر این کتابخانه‌ها است که می‌تواند مدل‌های یادگیری ماشینی قدرتمندی را به منظور حل مشکلات در دنیای واقعی و پیشرفت‌های حرفه‌ای در علم داده بسازد.

نحوه نصب کتابخانه sci-kit learn



```
C:\Windows\system32\cmd.exe
Processing c:\users\ayayushi\downloads\scipy-0.19.1-cp36-cp36m-win32.whl
Requirement already satisfied: numpy>=1.8.2 in c:\users\ayayushi\appdata\roaming\python\python36\site-packages (from scipy==0.19.1)
Installing collected packages: scipy
Successfully installed scipy-0.19.1

C:\Users\Aayushi\Downloads>pip install "numpy-1.13.1+mk1-cp36-cp36m-win32.whl"
Processing c:\users\ayayushi\downloads\numpy-1.13.1+mk1-cp36-cp36m-win32.whl
Installing collected packages: numpy
Found existing installation: numpy 1.13.0
Uninstalling numpy-1.13.0:
Successfully uninstalled numpy-1.13.0
Successfully installed numpy-1.13.1+mk1

C:\Users\Aayushi\Downloads>
```

مراحل نصب کتابخانه sci-kit learn در سیستم عامل ویندوز به صورت زیر است:

۱. پایتون را نصب کنید. سپس، ترمینال را با جستجوی cmd باز کنید. در خط فرمان نسخه پایتون را وارد کنید.
۲. در این مرحله NumPy را نصب کرده و سپس آن را اجرا کنید.
۳. سپس نصب کننده یا SciPy Installer را نصب کنید.
۴. در این مرحله Pip را با تایپ عبارت `python get-pip.py` در خط فرمان نصب کنید.
۵. در آخرین مرحله با تایپ عبارت `pip install scikit learn` در خط فرمان `sci-kit learn` را نصب کنید.

مزایا و معایب استفاده از کتابخانه sci-kit learn



مزایا و معایب استفاده از کتابخانه sci-kit learn به صورت زیر هستند.

مزایا

کتابخانه sci-kit learn در هسته خود، مجموعه ای از ابزار قدرتمند و بصری را ارائه می‌دهد که همه چیز را از آماده سازی داده‌ها گرفته تا ارزیابی و آموزش مدل‌ها مدیریت خواهد کرد. در این بخش نگاهی دقیق به مزایای استفاده از این کتابخانه خواهیم کرد.

1) جعبه ای مدولار و انعطاف پذیر

یکی از مزایای مهم در کتابخانه sci-kit learn انعطاف پذیر بودن آن و همچنین نحوه طراحی آن به صورت ماژولار است؛ به این معنا که از بلوک‌های ساختمانی مختلفی تشکیل شده که می‌توان به صورت مستقل از آنها استفاده کرد یا آنها را همانند لگو به یکدیگر چسباند. این بلوک‌های ساختمانی شامل ابزارهایی هستند که در موارد زیر به کار گرفته می‌شوند:

- پیش پردازش داده‌ها
 - طیف گسترده ای از برآوردگرها به منظور آموزش مدل‌های یادگیری ماشینی
 - ارزیابی عملکرد مدل‌ها
- این نوع طراحی مدولار به توسعه دهندگان این امکان را خواهد داد که راه حل‌های یادگیری ماشین را به صورت سفارشی درآورند.

2) گردش کار آزمایش شده و دقیق

یکی دیگر از مزایای کتابخانه sci-kit learn، این است که گردش کار استاندارد را فراهم می‌کند؛ چه در حال ساخت یک فیلتر یا یک طبقه بندی کننده تصویر باشید، مراحل اولیه یکسانی را دنبال خواهید کرد که عبارت اند از:

۱. داده‌های خود را بارگیری کنید.
۲. آن را از قبل پردازش کنید تا برای تجزیه و تحلیل آماده شوند.
۳. مدل یادگیری ماشینی خود را آموزش دهید.
۴. نتایج را ارزیابی کنید تا ببینید مدل شما تا چه حد خوب عمل خواهد کرد.

وجود یک فرآیند منسجم همانند آنچه در کتابخانه sci-kit learn است، بسیار مفید است، به خصوص زمانی که شما تازه کار هستید و به تازگی شروع به یادگیری ماشین کرده‌اید.

(3) ادغام با سایر کتابخانه‌های کلیدی پایتون

sci-kit learn برای ادغام یکپارچه با سایر کتابخانه‌های کلیدی پایتون همانند Pandas و NumPy طراحی شده است. این یک مزیت بزرگ برای این کتابخانه است؛ زیرا به این معنا است که می‌توانید از نقاط قوت آن کتابخانه‌ها برای تجزیه و تحلیل و ساماندهی داده‌های خود به صورتی کارآمد استفاده کنید. این موضوع درست همانند این است که یک تیم از ابر قهرمانان علم داده با یکدیگر کار کنند.

(4) قابل استفاده در دنیای واقعی

کتابخانه sci-kit learn یک اسباب بازی برای سر هم کردن موارد نیست؛ بلکه یک ابزار جدی است که برای حل مشکلات در دنیای واقعی طراحی شده است. به عنوان مثال، می‌توان از این کتابخانه به منظور ایجاد سیستمی استفاده کرد که به صورت اتوماتیک، ایمیل‌های اسپم (Spam) را فیلتر می‌کند. در برخی از موارد نیز می‌توان با استفاده از این کتابخانه، برنامه‌ای ایجاد کرد که قادر است به یک تصویر نگاه کند و اشیا موجود در آن تصویر را شناسایی کند؛ همانند ماشین‌ها، درختان یا ساختمان‌ها. با استفاده از این کتابخانه، امکانات زیادی در اختیار کاربران قرار خواهد گرفت.

(5) قابل دسترس برای همگان

یکی از بهترین مزایای کتابخانه sci-kit learn، این است که تیم سازنده آن به سختی کار می‌کند تا بتواند آن را در اختیار همگان قرار دهد. چه یک فرد

مبتدی و تازه کار برای یادگیری ماشین باشید، چه حرفه ای و سینیور، وجود API های مستند و مجموعه ای گسترده از آموزش ها شما را هیجان زده خواهد کرد. با وجود طیف گسترده ای از الگوریتم ها که همه چیز را از طبقه بندی گرفته تا کاهش ابعاد، پوشش می دهد، این کتابخانه فوق العاده توانسته، طیف وسیعی از ابزارها به منظور کار کردن و یادگیری ماشین را در اختیار کاربران قرار دهد. این کتابخانه با توجه به امکاناتی که دارد، اولین ایستگاه برای یادگیری ماشین در پایتون خواهد بود.

چالش ها

قبل استفاده از کتابخانه sci-kit learn، بهتر است با چالش ها، محدودیت ها و ملاحظات اخلاقی مربوط به استفاده از آن آشنا شوید. در حالی که این ابزار به صورت غیرقابل انکاری قدرتمند است؛ اما دارای معایب قابل توجهی است که هر دانشمند داده باید از آن آگاه باشد. از نظر فنی sci-kit learn به دلایل زیر با چالش هایی مواجه است:

- پشتیبانی محدودی برای وظایف یادگیری عمیق
- مشکل در مدیریت مؤثر داده ها با ابعاد بالا
- مشکلات مقیاس پذیری هنگام کار با مجموعه ای از داده های بسیار

بزرگ

- ادغام دست و پا گیر با برخی از ابزارهای صنعتی
- این محدودیت ها می توانند بر عملکرد کارهای پیچیده ای همانند تشخیص تصویر یا پردازش زبان طبیعی تأثیر بگذارند و اغلب به تکنیک های پیشرفته تری نیز نیاز خواهند داشت. همچنین ممکن است، مدل ها به منظور مقیاس بندی کارآمد در محیط های کلان داده با مشکل مواجه شوند.

با این حال، پیامدهای اخلاقی استفاده از sci-kit learn شاید حتی مهم تر از این موارد باشد. یکی از نگرانی‌های اصلی پتانسیل سوگیری الگوریتمی در مدل‌های ایجاد شده با این کتابخانه است.

اگر این الگوریتم‌ها به دقت مدیریت نشود، این سوگیری‌ها می‌توانند منجر به ایجاد نتایج ناعادلانه یا تبعیض آمیز، به ویژه در حوزه‌های حساسی همانند مراقبت‌های بهداشتی شوند. بیمارستانی را تصور کنید که از مدلی برای پیش‌بینی سطوح خطر در بیمار استفاده می‌کند. اگر آن مدل روی داده‌هایی آموزش داده شود که اطلاعات جمعیتی خاصی را کمتر از میزان واقعی نشان دهد، ممکن است به صورت سیستماتیک، میزان خطر را برای آن گروه دست کم بگیرد و سبب شود تا مراقبت‌های کافی انجام نشود. جامعه دانش‌آموزان نیز با مسئله تعصب و انصاف دست و پنجه نرم می‌کنند؛ اما هیچ پاسخ آسانی برای آن وجود ندارد. تلاش‌ها برای تفسیرپذیرتر و شفاف تر کردن مدل‌ها به منظور کاهش این خطرات و اطمینان از پاسخ گویی آنها بسیار مهم خواهد بود.

چشم انداز آینده برای کتابخانه sci-kit learn



چشم انداز آینده برای کتابخانه sci-kit learn به صورت زیر خواهد بود:

افزایش مقیاس و ادغام یکپارچه

با رشد حجم داده‌ها، انتظار می‌رود نسخه‌های آینده‌ی sci-kit learn، مقیاس پذیری بیشتری را ارائه دهند. این امر به سازمان‌ها کمک خواهد کرد تا به صورت مؤثرتری مجموعه‌ای از داده‌های بزرگ را مدیریت کنند. علاوه بر این، ادغام یکپارچه‌تر با کتابخانه‌های یادگیری عمیق، در همین راستا خواهد بود. از جمله مزیت‌های ادغام پیشرفته می‌توان به ایجاد ساده‌تر مدل‌های هوش مصنوعی پیشرفته اشاره کرد که از تکنیک‌های کلاسیک یادگیری ماشین و رویکرد یادگیری عمیق پیشرفته استفاده می‌کنند.

بهبود قابلیت تفسیر مدل

با استفاده هر چه بیشتر از مدل‌های یادگیری ماشین در کسب و کارها، تقاضا برای شفاف سازی افزایش می‌یابد و احتمالاً بر توسعه ابزارهای مفیدتر به منظور تفسیر پذیری مدل، تمرکز خواهد کرد. این امر به دانشمندان کمک خواهد کرد تا مدل‌های خود را اصلاح کنند و صنایع را قادر خواهد ساخت تا از مقرراتی پیروی کنند که به هوش مصنوعی قابل توضیح نیاز دارند.

تولید داده‌های مصنوعی

گسترش sci-kit learn به تکنیک‌های هوش مصنوعی مولد می‌تواند، نحوه رویکرد کسب و کارها به سنتز و تقویت داده‌ها را متحول کند. با تولید داده‌های مصنوعی، سازمان‌ها می‌توانند آموزش مدل را بدون نیاز به خطر انداختن حریم خصوصی یا امنیت ارتقا دهند. این امر نیاز به جمع آوری

گسترده داده در دنیای واقعی را کاهش می‌دهد که می‌تواند سبب کاهش هزینه‌ها و سرعت بخشیدن به توسعه راه‌حل‌های هوش مصنوعی قوی شود.