ابزارهای برتر برای دانشمندان داده: ابزارهای تجزیه‌وتحلیل، ابزار مصورسازی داده، ابزار پایگاه داده و موارد دیگر

مقدمه مترجم:
دانشمند داده فردی است که می‌تواند با بررسی سؤالات نامحدود و دگرگون‌پذیر و همچنین استفاده از دانشی که دارد در مورد آمار و الگوریتم‌های پیشرفته، ارزش زیادی را خلق کند؛ اما یک دانشمند داده برخلاف تحلیلگر داده است چرا که تحلیلگر داده بر درک و فهمیدن داده‌ها از دیدگاه و چشم‌انداز گذشته و حال متمرکز می‌شود، در حالی که یک دانشمند داده بر تولید پیش‌بینی‌های معتبر و قابل اتکا برای آینده تمرکز می‌کند؛ اما از طرفی یک دانشمند داده باید درست مانند یک تحلیلگر داده قادر به پاکسازی، تجزیه‌وتحلیل و مصورسازی داده‌ها باشد. با این حال، یک دانشمند داده، عمق و تخصص بیشتری در این مهارت‌ها دارد و همچنین قادر به آموزش و بهینه‌سازی مدل‌های یادگیری ماشین است.

اما اگر می‌خواهید دانشمند داده باشید، باید فنونی که یک مهندس داده و یک تحلیلگر داده به آن مسلط است را بدانید و در کنار آن ریاضیات، آمار، کامپیوتر و برنامه‌نویسی برای شما ضرورت دارد؛ اما تمام دانشمندان داده با برنامه‌نویسی آشنایی ندارند. با این حال، دانستن ابزارهایی مانند ابزارهای تجزیه‌وتحلیل، مصورسازی یا تجسم داده‌ها و موارد دیگر که نیازی به دانستن برنامه‌نویسی ندارند بسیار ضرورت دارد. این ابزارهای برتر و مهم در اینجا آورده شده است.

مقدمه

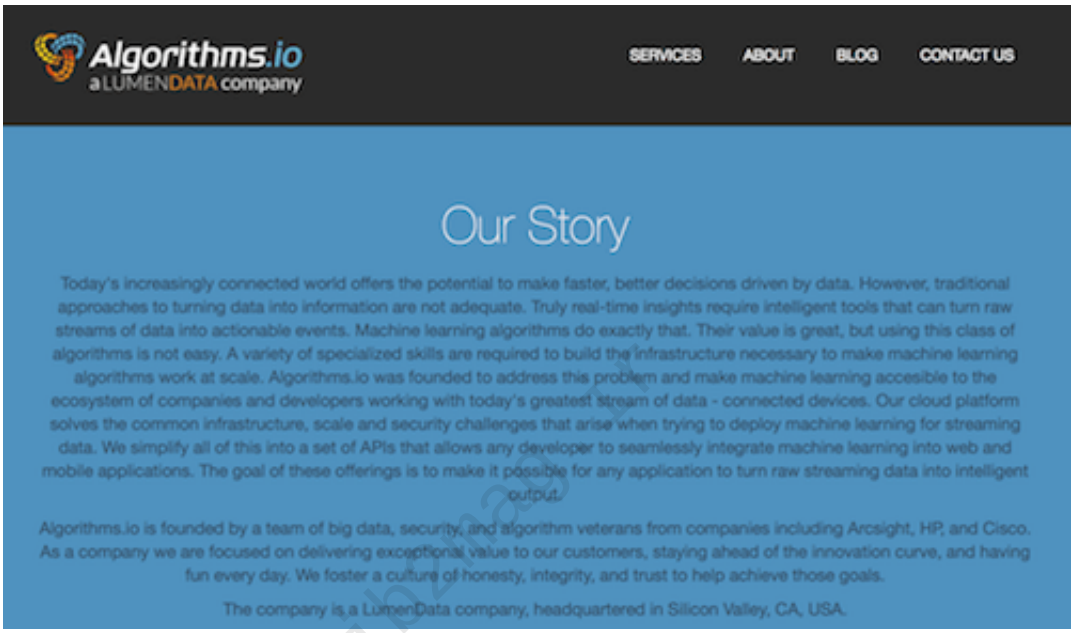
دانشمندان داده کنجکاو هستند و اغلب به دنبال ابزارهای جدیدی هستند که به آن‌ها در یافتن پاسخ کمک می‌کند. آن‌ها همچنین باید در استفاده از ابزارهای تجارت مهارت داشته باشند، حتی اگر ده‌ها مورد از آن‌ها وجود داشته باشد. به طور کلی، دانشمندان داده باید دانش کاملی از زبان‌های برنامه‌نویسی آماری برای ساخت سیستم‌های پردازش داده، پایگاه داده و ابزار مصورسازی داشته باشند. بسیاری از افراد در این زمینه نیزدانش برنامه‌نویسی را بخشی جدایی‌ناپذیر از علم داده می‌دانند. با این حال، همه دانشجویان دانشمند داده، برنامه‌نویسی را مطالعه نمی‌کنند، بنابراین آگاهی از ابزارهایی که برنامه‌نویسی را دور می‌زنند و شامل یک رابط گرافیکی کاربر پسند هستند، مفید است، به طوری که دانش دانشمندان در مورد الگوریتم‌های داده برای کمک به آن‌ها در ساخت مدل‌های پیش‌بینی کافی است.

با وجود همه چیزهایی که روی صفحه یک دانشمند داده است، شما وقت ندارید به جستجوی ابزارهای تجاری که می‌تواند به شما در انجام کار کمک کند، بپردازید. به همین دلیل ما ابزارهایی را جمع‌آوری کرده‌ایم که به مصورسازی داده‌ها، الگوریتم‌ها، زبان‌های برنامه‌نویسی آماری و پایگاه‌های داده کمک می‌کنند. ما ابزارها را بر اساس سهولت استفاده، محبوبیت، شهرت و ویژگی‌های آن‌ها انتخاب کرده‌ایم. ما ابزارهای برتر خود را برای دانشمندان داده به ترتیب حروف الفبا لیست کرده‌ایم تا جستجوی شما ساده شود؛ بنابراین، آن‌ها با هیچ رتبه‌بندی یا درجه‌بندی لیست نشده‌اند.

1. Algorithms.io

@algorithms_io

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.



الگوریتمزآی‌او[۱] ، یک

شرکت لومن‌دیتا[۲] [۳] است که یادگیری ماشین را به عنوان خدماتی برای جریان داده از دستگاه‌های متصل ارائه می‌دهد. این ابزار داده‌های خام را به چشم‌انداز واقعی و رویدادهای عملی تبدیل می‌کند تا شرکت‌ها در موقعیت بهتری برای استقرار یادگیری ماشین برای جریان داده‌ها باشند.

هزینه:دارد

2. Apache Giraph

آپاچی‌گیراف[۴] که یک

سیستم پردازش نمودار تکراری برای مقیاس‌پذیری بالا طراحی شده است، به عنوان همتای منبع‌باز پریگل[۵] آغاز شد؛ اما ویژگی‌های متعددی فراتر از مدل پایه پریگل اضافه می‌کند. دانشمندان داده از گیراف استفاده می‌کنند تا «پتانسیل مجموعه داده‌های ساختاریافته را در مقیاس گسترده» به کار گیرند.

هزینه:رایگان

3. Apache Hadoop

@hadoop

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

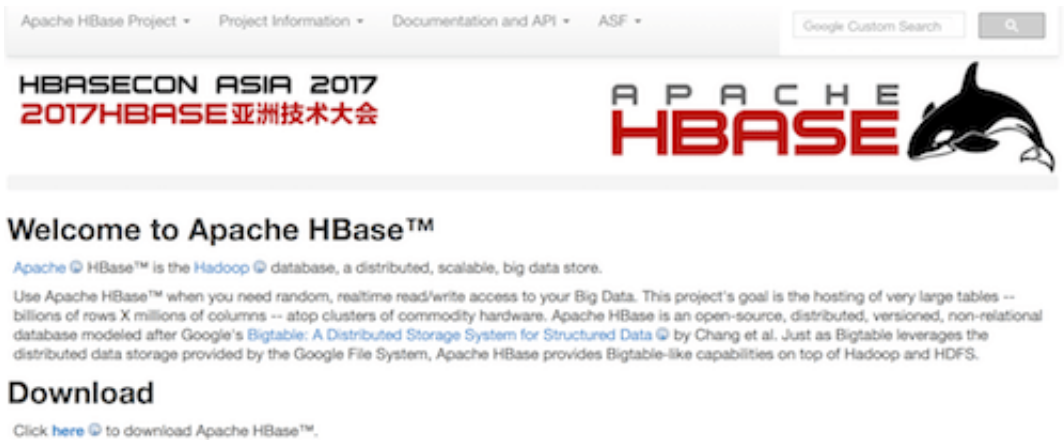
۴. یک آپاچی‌هادوپ[۶]

نرم‌افزار منبع‌باز برای محاسبات قابل اعتماد، توزیع شده و مقیاس‌پذیر است. چارچوبی که امکان پردازش توزیع شده مجموعه داده‌های بزرگ در خوشه‌های رایانه را فراهم می‌کند، کتابخانه نرم‌افزار نیز از مدل‌های ساده برنامه‌نویسی استفاده می‌کند. هادوپ برای تحقیق و تولید مناسب است.

هزینه:رایگان

۴. Apache Hbase

@ApacheHBase



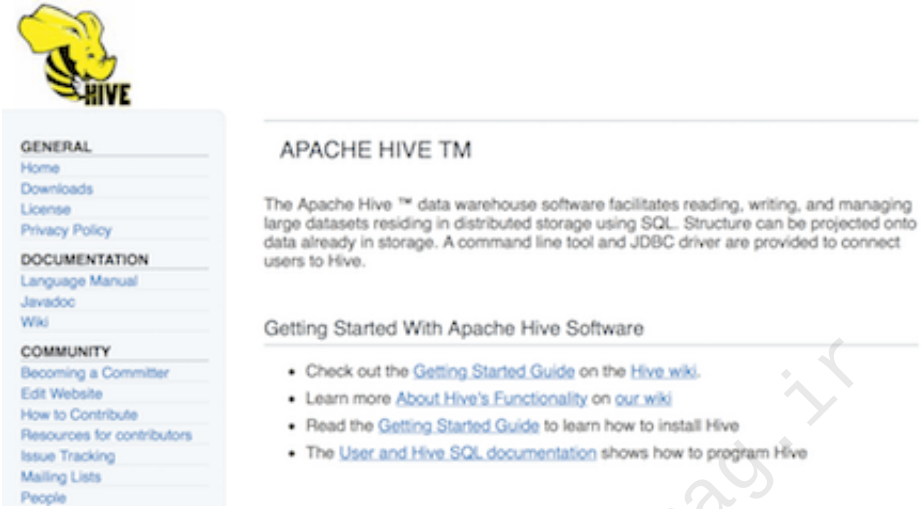
۵. آپاچی‌اچ‌بیس[۷] پایگاه‌داده

هادوپ، یک فروشگاه داده بزرگ، توزیع شده و مقیاس‌پذیر است. دانشمندان داده در صورت نیاز به دسترسی خواندن/نوشتن به صورت تصادفی، بلادرنگ به ابرداده، از این ابزار منبع باز استفاده می‌کنند. آپاچی‌اچ‌بیس قابلیت‌های مشابه بیگ‌تیبل[۸] را در بالای هادوپ و اچ‌دی‌اف‌اس[۹] نیز فراهم می‌کند.

هزینه:رایگان

۵. Apache Hive

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.



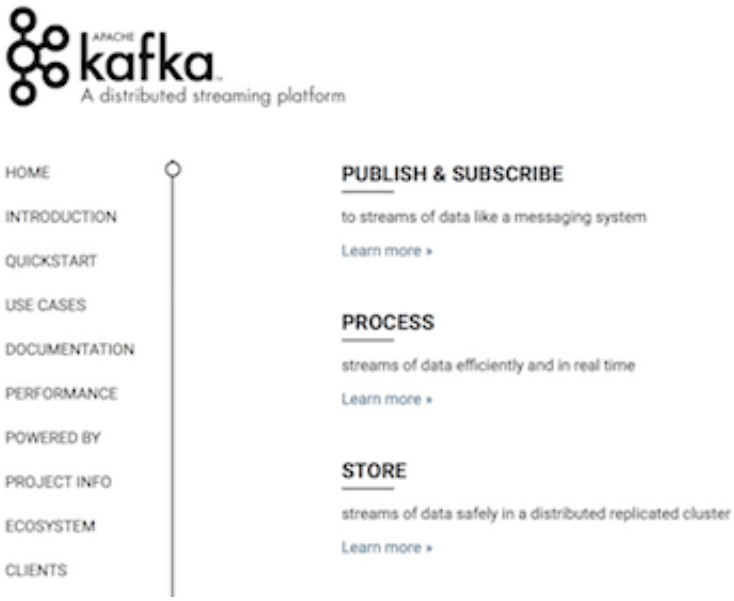
@ApacheHive

آپاچی‌هایو[10] به عنوان یک پروژه بنیادی نرم‌افزار آپاچی، به عنوان زیر پروژه آپاچی‌هادوپ آغاز به کار کرد و اکنون خود یک پروژه سطح بالا است. این ابزار یک نرم‌افزار انبارداده است که به خواندن، نوشتن و مدیریت مجموعه داده‌های بزرگی که در فضای ذخیره‌سازی توزیع شده با استفاده از زبان برنامه‌نویسی اس‌کیوال[11]ساکن هستند، کمک می‌کند.

هزینه:رایگان

Apache Kafka .6

@apachekafka



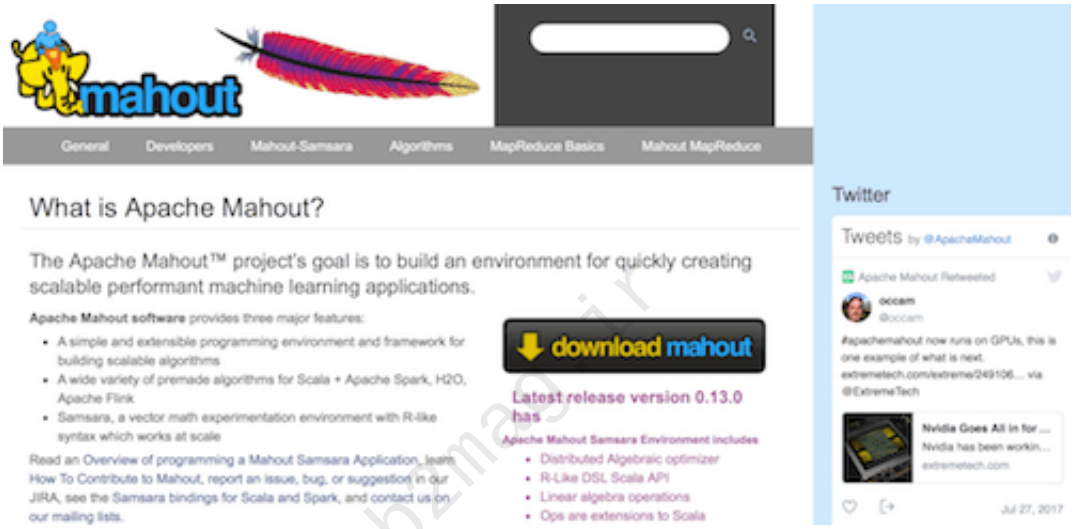
آپاچی‌کافکا[12] که یک پلتفرم جریانی توزیع شده است، جریان داده‌ها را در زمان واقعی پردازش می‌کند. دانشمندان داده از این ابزار برای ساخت خطوط انتقال داده و برنامه‌های جریانی واقعی استفاده می‌کنند، زیرا به شما امکان می‌دهد جریان‌های رکوردها را منتشر کرده و در آن‌ها مشترک شوید، جریان‌های رکوردها را به روشی تحمل‌پذیر در برابر خطا ذخیره کرده و جریان‌های رکوردها را در صورت بروز پردازش کنید.

هزینه:رایگان

Apache Mahout .7

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

@ApacheMahout

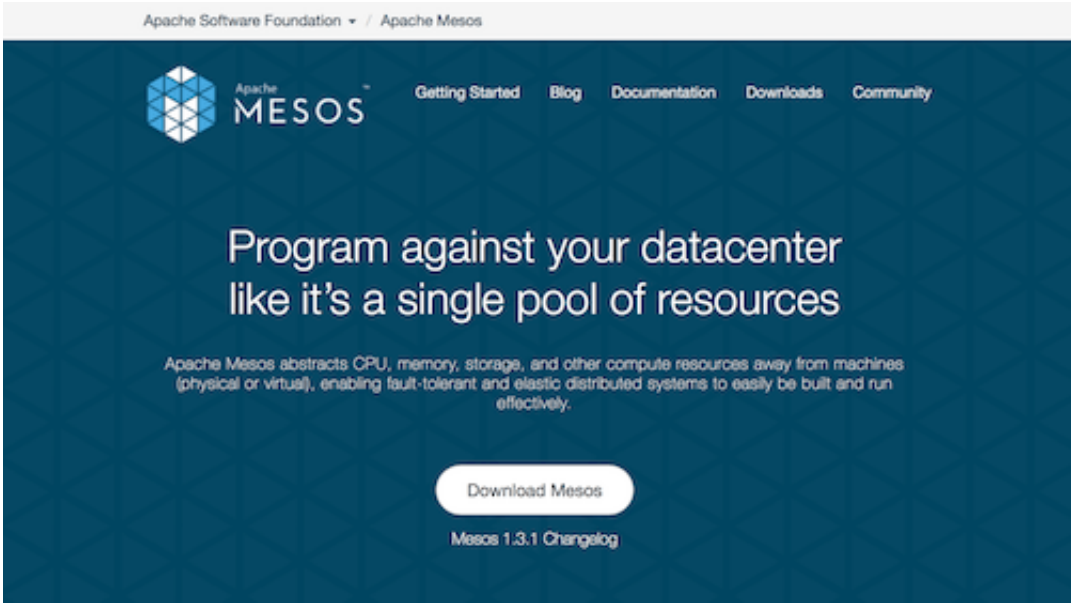


آپاچی‌ماهوت[۱۳] یک پروژه منبع باز آپاچی برای یادگیری ماشین است و هدف آن این است که یادگیری ماشین را مقیاس‌پذیر و داده‌کاوی را امکان‌پذیر کند. به طور خاص، هدف این پروژه «ایجاد محیطی برای ایجاد سریع برنامه‌های یادگیری ماشین با عملکرد مقیاس‌پذیر» است.

هزینه:رایگان

8. Apache Mesos

@ApacheMesos

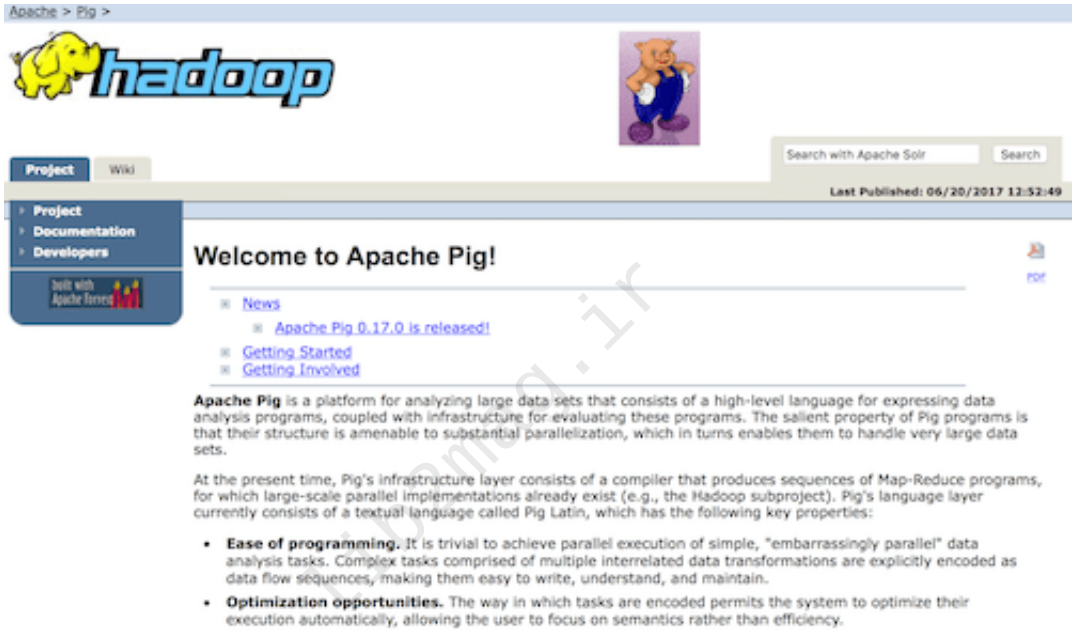


آپاچی‌مسس[۱۴] به‌عنوان مدیر خوشه، جداسازی و به اشتراک‌گذاری کارآمد منابع را در میان برنامه‌ها یا چارچوب‌های توزیع شده فراهم می‌کند. پردازنده مسس، حافظه، ذخیره‌سازی و سایر منابع را به دور از ماشین‌های فیزیکی یا مجازی جمع می‌کند تا سیستم‌های توزیع شده الاستیک متحمل در برابر خطا را به راحتی ساخته و به طور مؤثر اداره کند.

هزینه:رایگان

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

9. Apache Pig

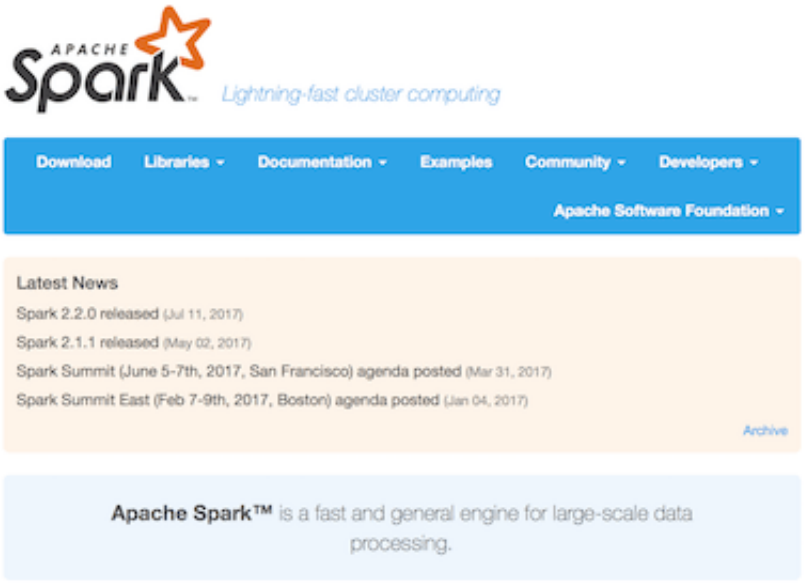


آپاچی‌پیگ[۱۵] که برای تجزیه‌وتحلیل مجموعه داده‌های بزرگ طراحی شده است، از زبان سطح بالایی برای بیان برنامه‌های تجزیه‌وتحلیل داده تشکیل‌شده است؛ که همراه با زیرساخت برای ارزیابی چنین برنامه‌هایی است. از آنجا که ساختار برنامه‌های پیگ می‌توانند از موازی‌سازی قابل توجهی استفاده کنند، بنابراین می‌توانند مجموعه‌های داده بزرگ را کنترل کنند.

هزینه:رایگان

10. Apache Spark

@ApacheSpark



آپاچی‌اسپارک[۱۶] «محاسبات خوشه‌ای سریع» را ارائه می‌دهد. طیف وسیعی از سازمان‌ها از اسپارک برای پردازش مجموعه داده‌های بزرگ استفاده می‌کنند و دانشمند داده با این ابزار می‌تواند به منابع مختلف داده مانند اچ‌دی‌اف‌اس، کاساندرا[۱۷]، اچ‌بیس و اس۳[۱۸] دسترسی پیدا کند.

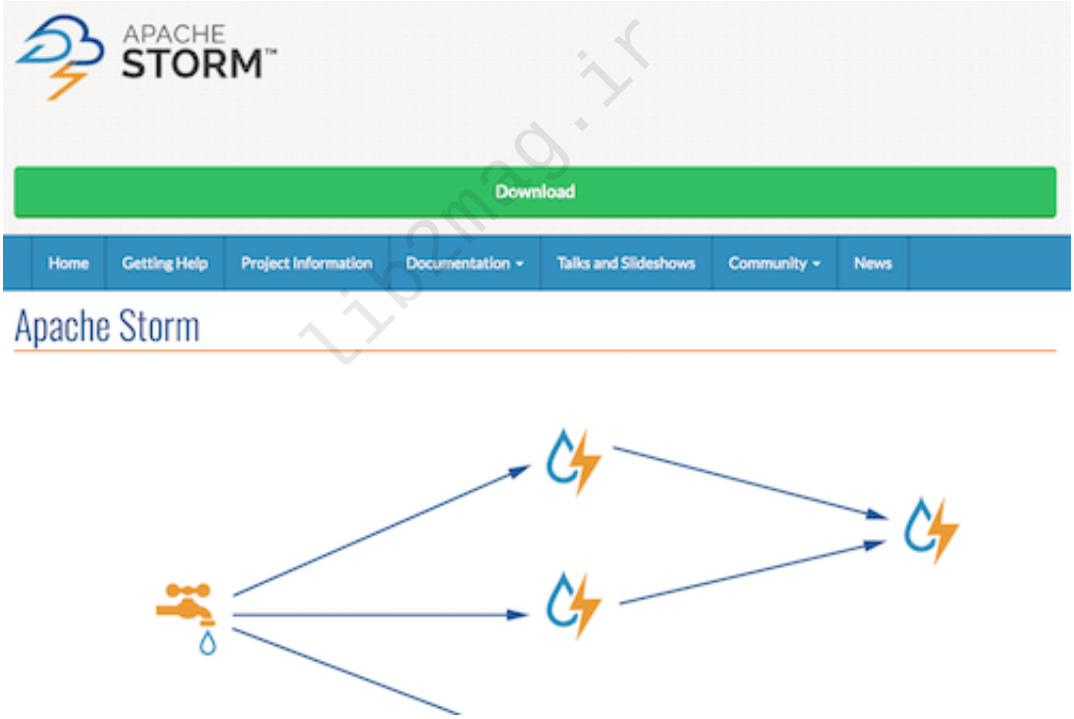
تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

هزینه:رایگان

11. Apache Storm

@ApacheStorm

@stormprocessor



آپاچی‌استورم[۱۹] ابزاری برای دانشمندان داده است که برای محاسبه زمان واقعی خطا و انتشار به کار می‌رود. همچنین پردازش جریان، محاسبه مداوم، آرپی‌سی[۲۰]توزیع شده و موارد دیگر را کنترل می‌کند.

هزینه:رایگان

12. BigML

@bigmlcom

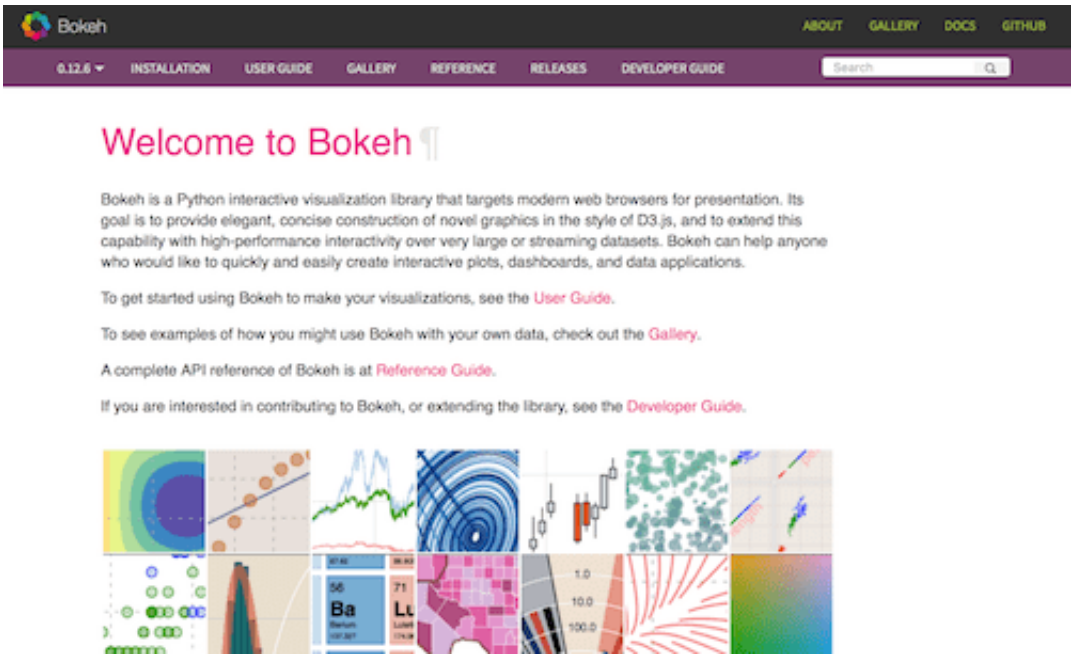
تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

بیگ‌ام‌ال[۲۱] یادگیری ماشینی را ساده می‌کند. این پلتفرم در کل شرکت برای اجرای عملی یادگیری ماشین در سازمان‌ها در فضای ابری یا محلی کار می‌کند. بیگ‌ام‌ال حل و طبقه‌بندی خودکار، رگرسیون، تجزیه‌وتحلیل خوشه‌ای، تشخیص ناهنجاری، کشف ارتباط و وظایف مدل‌سازی موضوع را ساده می‌کند.

هزینه:دارد

Bokeh .13

@BokehPlots



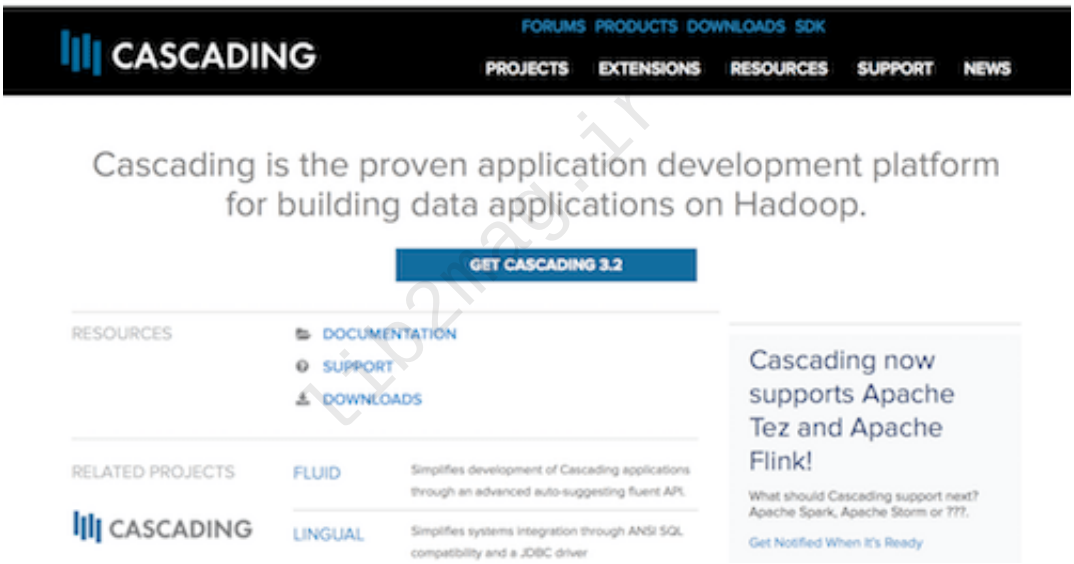
بوکه[۲۲] ، یک کتابخانه تجسم تعاملی بر پایه پایتن[۲۳] ، مرورگرهای وب مدرن را برای اهداف خود به کار می‌گیرد و به کاربران کمک می‌کند تا طرح‌های تعاملی، داشبورد و برنامه‌های داده را به راحتی ایجاد کنند.

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

هزینه:رایگان

14. Cascading

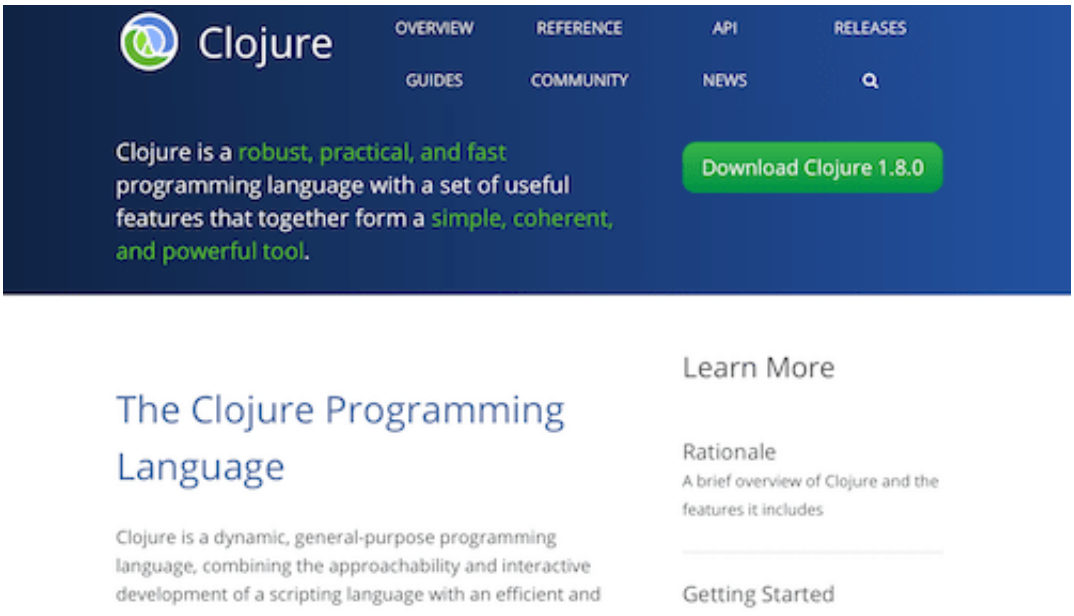
@cascading



کس‌کیدینگ[۲۴] یکپلتفرم
توسعه برنامه برای دانشمندان داده است که در آپاچی‌هادوپ، برنامه‌های ابرداده‌ای را ایجاد می‌کنند. کاربران می‌توانند مشکلات ساده و پیچیده داده‌ها را با کس‌کیدینگ حل کنند زیرا دارای موتور محاسبه، چارچوب یکپارچه‌سازی سیستم‌ها، پردازش داده‌ها و قابلیت‌های برنامه‌ریزی است.

هزینه:رایگان

15. Clojure



کلوژور[۲۵] که یک زبان برنامه‌نویسی قوی و سریع است، ابزاری کاربردی است که با توسعه تعاملی یک زبان برنامه‌نویسی با زیرساخت کارآمد برای برنامه‌نویسی چند رشته‌ای زوج می‌شود. کلوژور از این نظر بی‌نظیر است که یک زبان تألیفی است اما با هر ویژگی پشتیبانی شده در زمان اجرا پویا می‌ماند.

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

هزینه:رایگان

16. D3.js

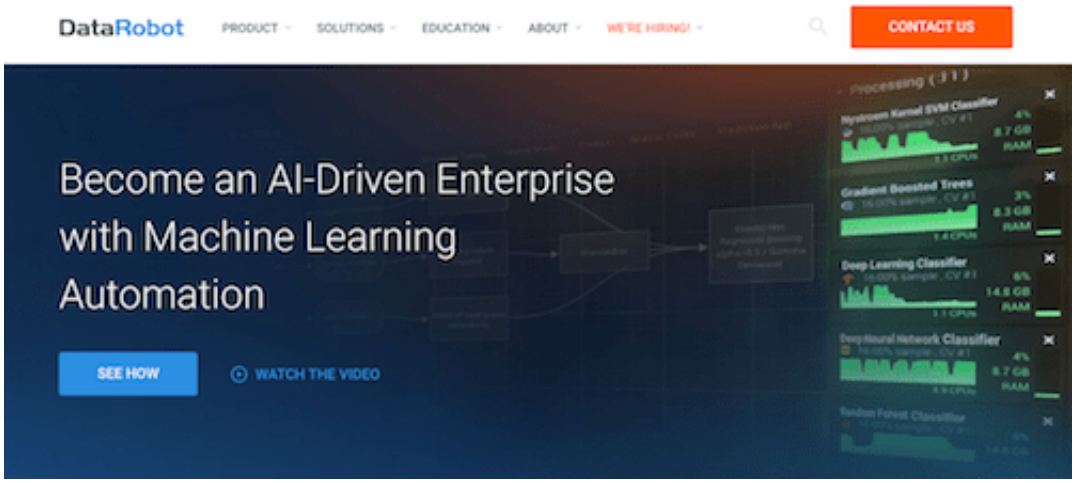
@mbostock

مایکبستوک[۲۶] متعهدبه
«کد و داده برای انسان»، دی‌تری.جی‌اس را ایجاد کرد. دانشمندان داده از این ابزار، یک کتابخانه جاوا اسکریپت برای دست‌کاری اسناد بر اساس داده‌ها، برای افزودن زندگی به داده‌های خود با اس‌وی‌جی[۲۷]، کانواس[۲۸]و اچ‌تی‌ام‌ال [۲۹] استفاده می‌کنند.

هزینه:رایگان

17. DataRobot

@DataRobot



دیتاروبات[۳۰] که یک
پلتفرم خودکار یادگیری ماشین پیشرفته است، به دانشمندان داده کمک می‌کند تا مدل‌های پیش‌بینی بهتر را سریع‌تر بسازند. هنگام استفاده از دیتاروبات می‌توانید به راحتی با اکوسیستم رو به گسترش الگوریتم‌های یادگیری ماشین

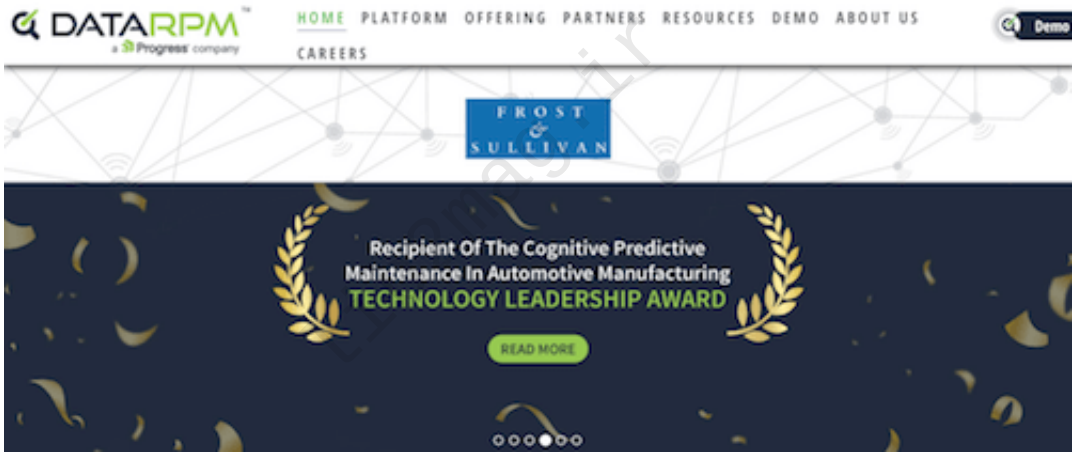
تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

همراه باشید.

هزینه:دارد

18. DataRPM

@DataRPM

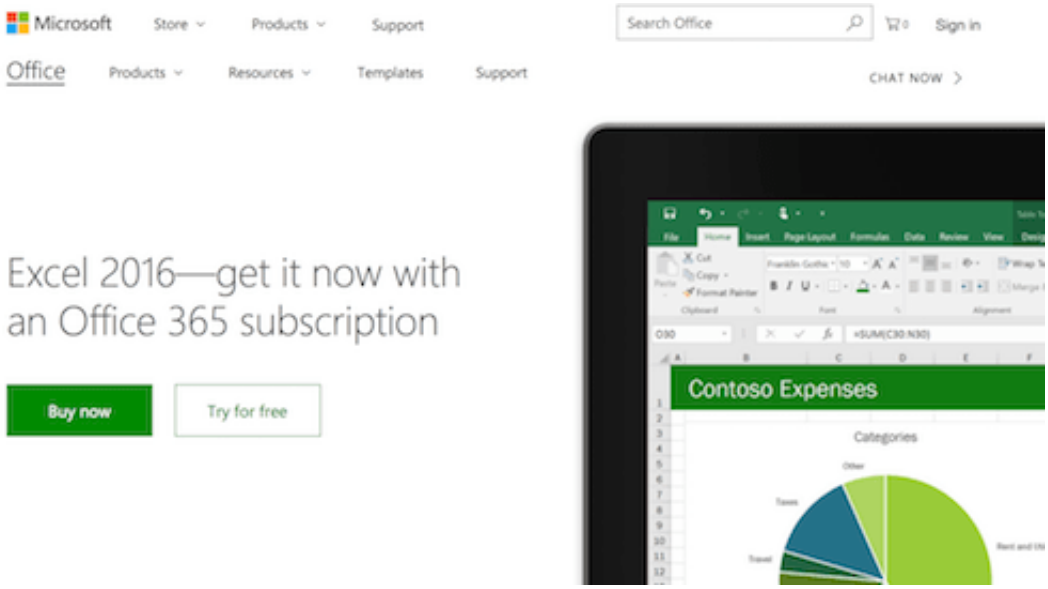


دیتاآرپی‌ام[۳۱] اولین‌وتنها
بستر نگهداری پیش‌بینی شناختی صنعت برای اینترنت اشیا صنعتی است. دیتاآرپی‌ام همچنین دریافت کننده جایزه رهبری فناوری ۲۰۱۷ برای پیش‌بینی شناختی نگهداری در ساخت خودرو از فراست و سالیوان[۳۲] است.

هزینه:دارد

19. Excel

@Office

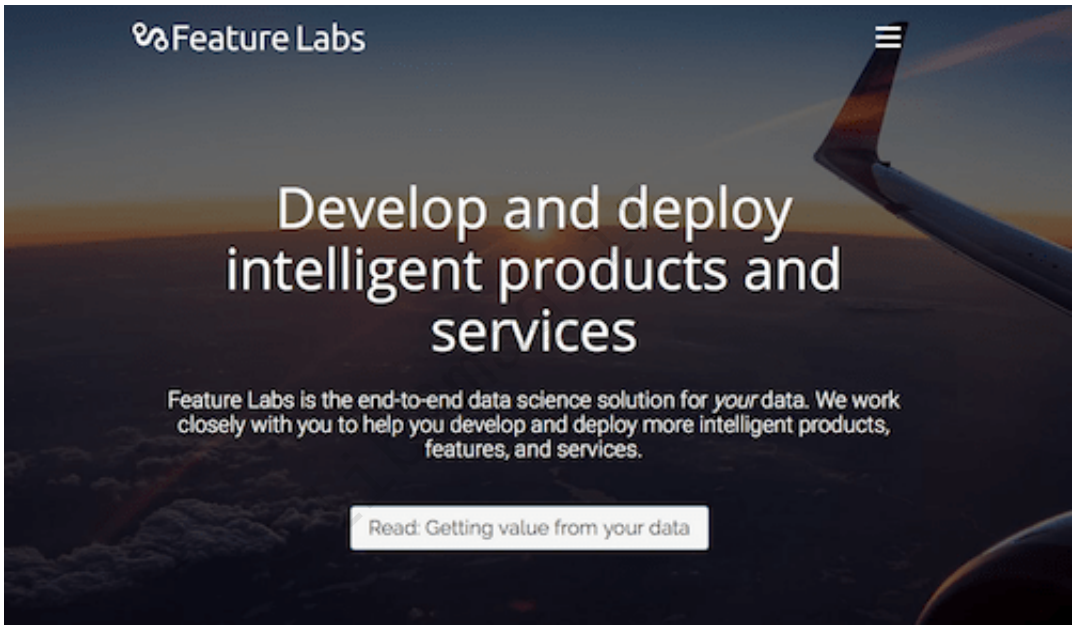


بسیاری از دانشمندان
اطلاعات، اکسل[۳۳] را به عنوان یک سلاح مخفی می‌دانند. این یک ابزار آشنا است که دانشمندان می‌توانند برای مرتب‌سازی سریع، فیلتر کردن و کار با داده‌های خود به آن اعتماد کنند. این برنامه در تمام رایانه‌ها وجود دارد، بنابراین دانشمندان داده می‌توانند تقریباً از هرجای دیگر با اکسل کار کنند.

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

هزینه: نسخه آزمایشی رایگان

20. Feature Labs

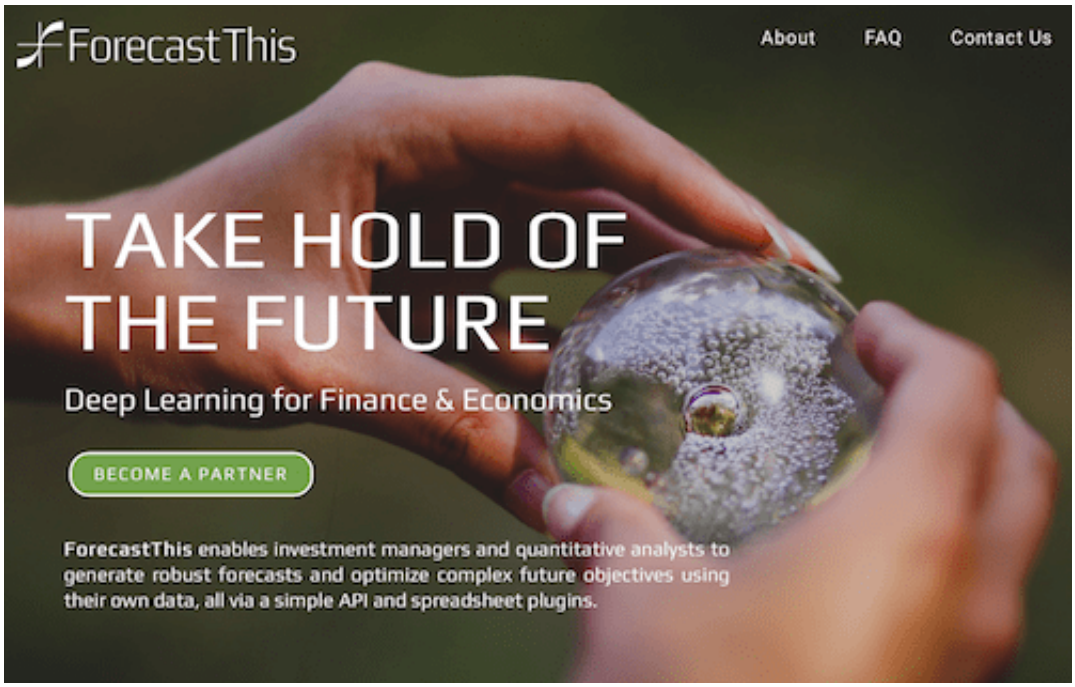


فیچرلب[۳۴] که یک راه‌حل برای پایان دادن به اطلاعات داده‌ها است، محصولات و خدمات هوشمندی را برای داده‌های شما تولید و ارائه می‌دهد. آن‌ها همچنین با دانشمندان داده همکاری می‌کنند تا به شما در تولید و استقرار محصولات، ویژگی‌ها و خدمات هوشمند کمک کنند.

هزینه: دارد

21. ForecastThis

@forecastthis



این ابزاری برای دانشمندان

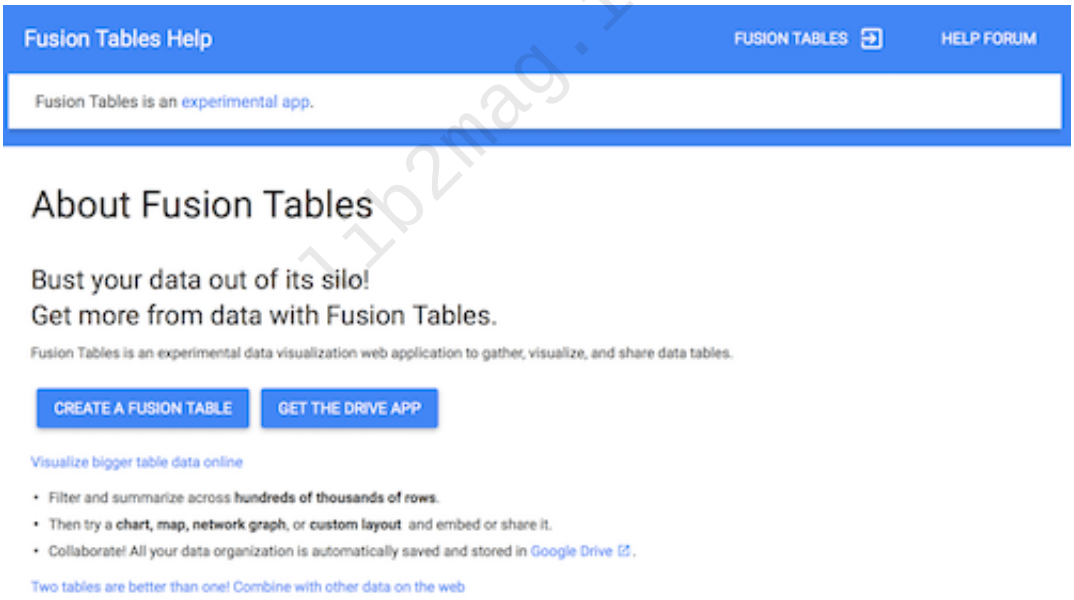
تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

داده است که انتخاب مدل پیش‌بینی را خودکار می‌کند. این شرکت تلاش می‌کند یادگیری عمیق مربوط به امور مالی و اقتصادی را از طریق امکان دادن به مدیران سرمایه‌گذاری، تحلیل‌گران کمی و دانشمندان داده در استفاده از داده‌های خود برای تولید پیش‌بینی‌های قوی و بهینه‌سازی اهداف پیچیده آینده، مورداستفاده قرار دهند.

هزینه: دارد

22. Fusion Tables

@GoogleFT



گوگل[۳۵] یک سرویس مدیریت داده مبتنی بر ابر است که بر همکاری، سهولت استفاده و مصورسازی متمرکز است. فیژنزتیبل یک برنامه آزمایشی است، یک ابزار برنامه‌نویسی وب برای دانشمندان داده است که به شما امکان می‌دهد جداول داده را جمع‌آوری، مصورسازی و به اشتراک بگذارید.

هزینه: رایگان

23. Gawk

جی‌ان‌یو[۳۶] یک سیستم

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

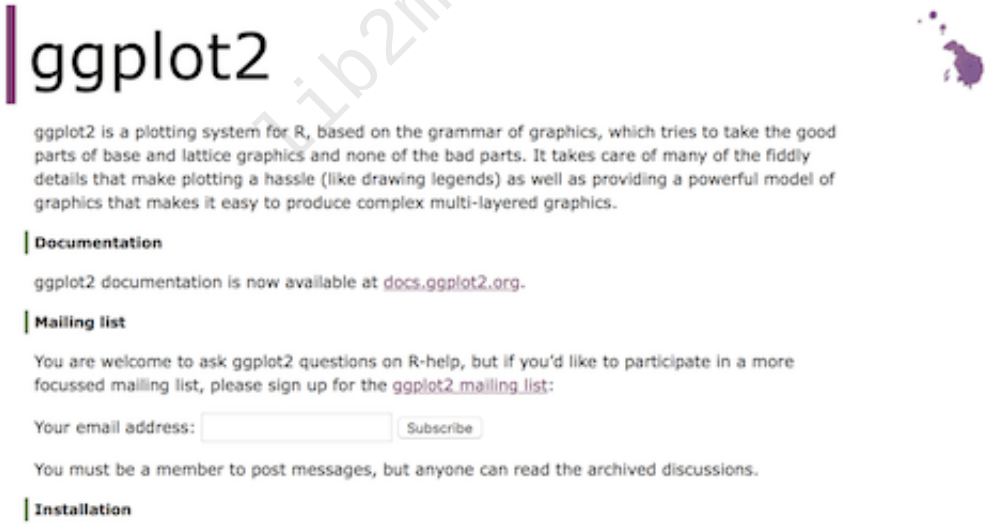
عامل است که به شما امکان می‌دهد از یک کامپیوتر بدون نرم‌افزار «که کار کردن با آن مشکل است» استفاده کنید. آن‌ها گوک[۳۷] را ایجاد کرده‌اند، یک ابزار اوک [۳۸] که یک زبان برنامه‌نویسی برای اهداف خاص را تفسیر می‌کند. گوک به کاربران این قدرت را می‌دهد تا فقط با استفاده از چند خط کد، کارهای ساده در زمینه تغییر شکل داده‌ها را مدیریت کنند.

هزینه:رایگان

24. ggplot2

@hadleywickham

@winston_chang



هادلی ویکام و وینستون

چانگ[۳۹]، جی‌جی‌پلات‌تو[۴۰] را توسعه دادند، یک سیستم نقشه‌کشی برای آر [۴۱] که بر اساس دستور زبان گرافیک است. با جی‌جی‌پلات‌تو، دانشمندان داده می‌توانند در هنگام حفظ قسمت‌های جذاب گرافیک پایه و شبکه و تولید گرافیک‌های پیچیده چند لایه، از بسیاری از دردسرهای طرح‌ریزی جلوگیری کنند.

هزینه:رایگان

25. GraphLab Create

و دانشمندان

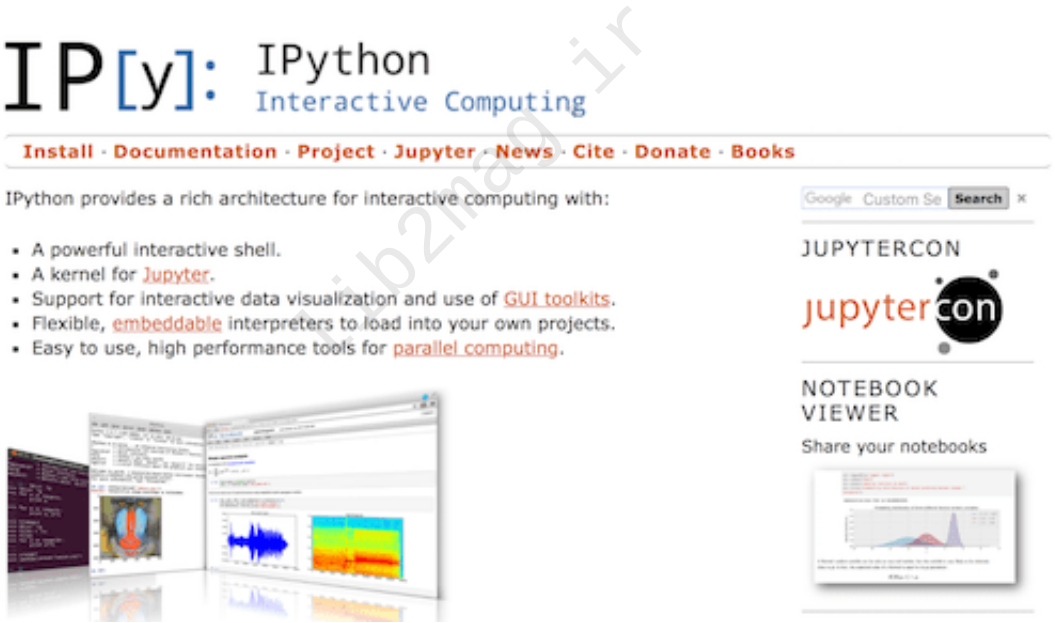
تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

توسعه‌دهندگان داده از گراف‌لب‌کریت[۴۲] برای ساخت پیشرفته‌ترین داده‌ها از طریق یادگیری ماشین استفاده می‌کنند. این ابزار مدل‌سازی یادگیری ماشین به کاربران کمک می‌کند تا برنامه‌های هوشمند را از پایتن به پایان برسانند.

هزینه:اشتراک یک ساله تجدید پذیر رایگان برای استفاده دانشگاهی

26. IPython

@IpythonDev



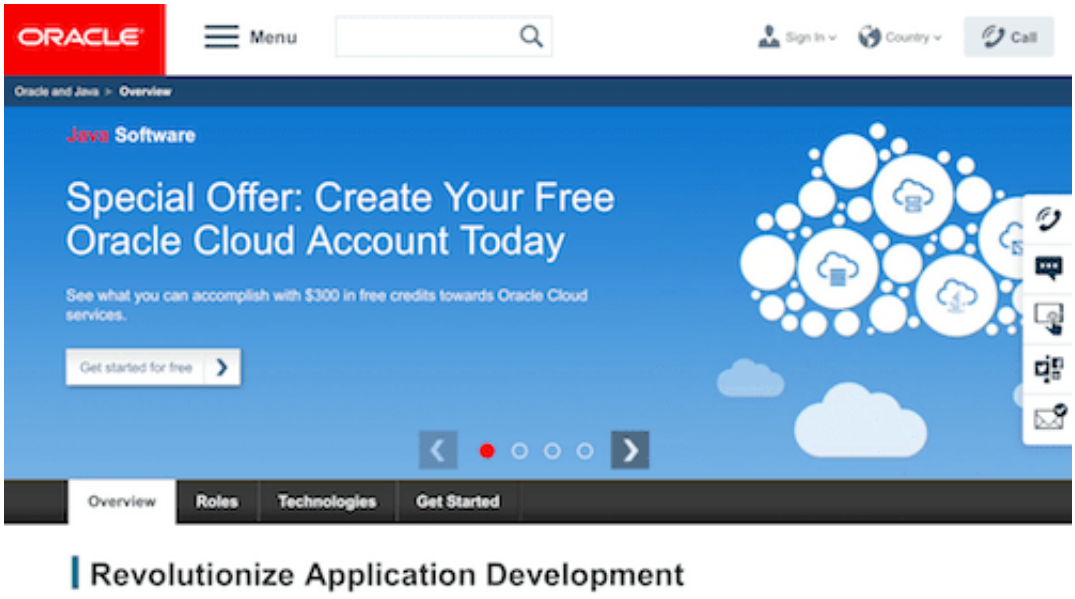
ابزارهای پایتن تعاملی یا

آی‌پایتن[۴۳]، پروژه‌ای در حال رشد با گسترش مؤلفه‌های مجهز به زبان است و معماری غنی را برای محاسبات تعاملی فراهم می‌کند. ابزاری منبع‌باز برای دانشمندان داده است، آی‌پایتن از پایتن ۲.۷ و ۳.۳ یا جدیدتر پشتیبانی می‌کند.

هزینه:رایگان

27. Java

@SW_Java



جاوا زبانی با پایگاه کاربری

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

گسترده است که به عنوان ابزاری برای دانشمندان داده ایجاد محصولات و چارچوب‌هایی شامل سیستم‌های توزیع شده، تجزیه‌وتحلیل داده‌ها و یادگیری ماشین است. اکنون جاوا به همان اندازه آر و پایتن برای علم داده مهم شناخته شده است زیرا برای کاربردهای علم داده قوی، راحت و مقیاس‌پذیر است.

هزینه: نسخه آزمایشی رایگان

28. Jupyter

@ProjectJupyter

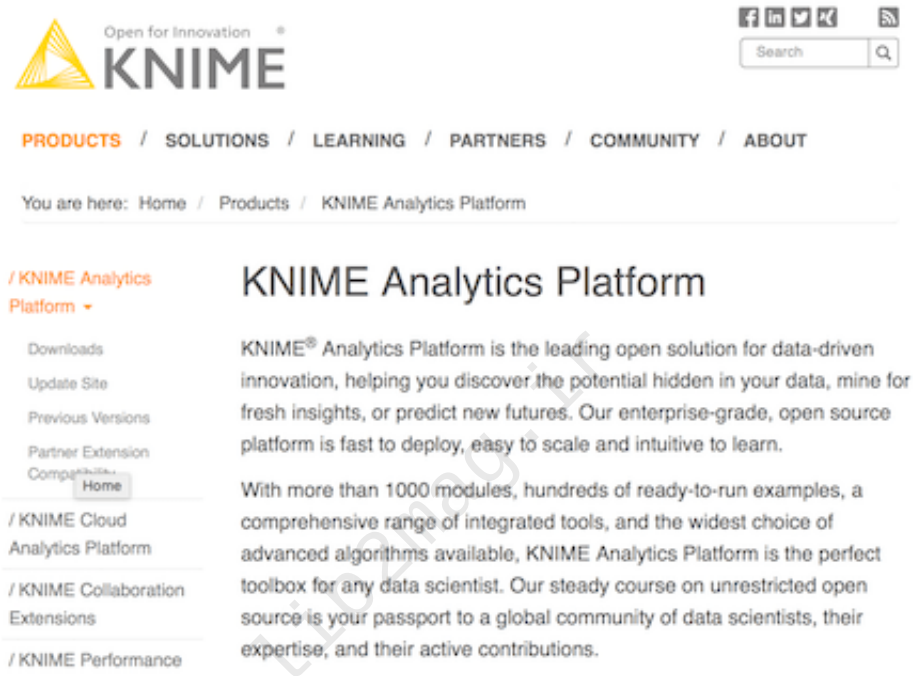
جی‌پایتر[۴۴] محیط‌های محاسباتی تعاملی چند زبانه را فراهم می‌کند. دفترياداشت[۴۵] آن، به عنوان یک برنامه وب منبع‌باز، به دانشمندان داده اجازه می‌دهد تا اسنادی حاوی کد زنده، معادلات، تجسم‌ها و متن توضیحی را ایجاد و به اشتراک بگذارند.

هزینه: رایگان

29. KNIME Analytics Platform

@knime

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.



به لطف سیستم‌عامل باز،

کی‌ان‌آی‌ام‌ای[۴۶] ابزاری برای پیمایش آزادانه داده‌های پیچیده است. پلتفرم تجزیه‌وتحلیل کی‌ان‌آی‌ام‌ای[۴۷] یک راه‌حل باز پیشرو برای نوآوری مبتنی بر داده است که به دانشمندان داده کمک می‌کند تا ارزش پنهان داده‌ها، معادن برای چشم‌انداز و پیش‌بینی آینده را کشف کنند.

هزینه: رایگان

30. Logical Glue

@logicalglue

یک برنده جایزه یادگیری ماشین جعبه سفید و پلتفرم هوش مصنوعی، لاجیکال‌گلو[۴۸] باعث افزایش بهره‌وری و سود برای سازمان‌ها می‌شود.

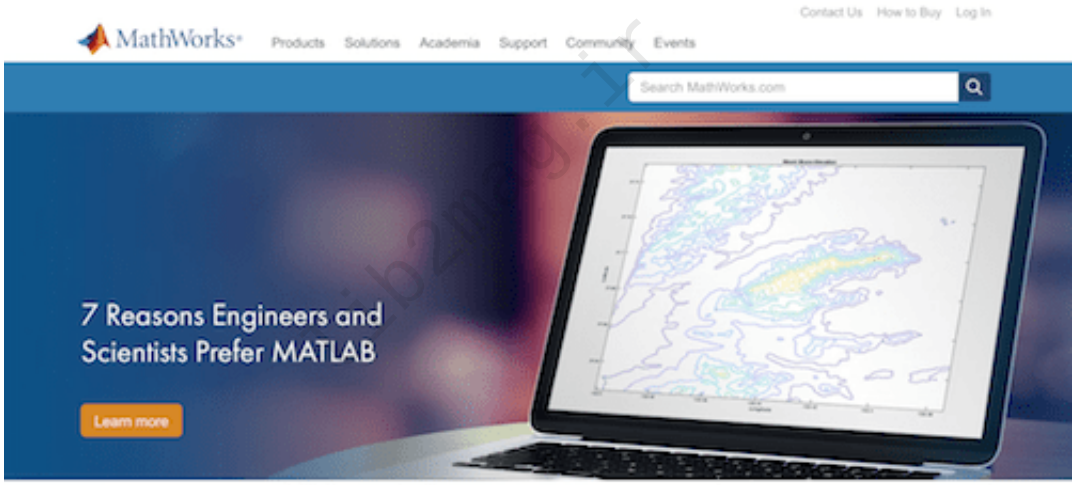
تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

دانشمندان داده این ابزار را انتخاب می‌کنند زیرا بینش شما را برای مخاطبان زنده می‌کند.

هزینه: دارد

31. MATLAB

@MATLAB

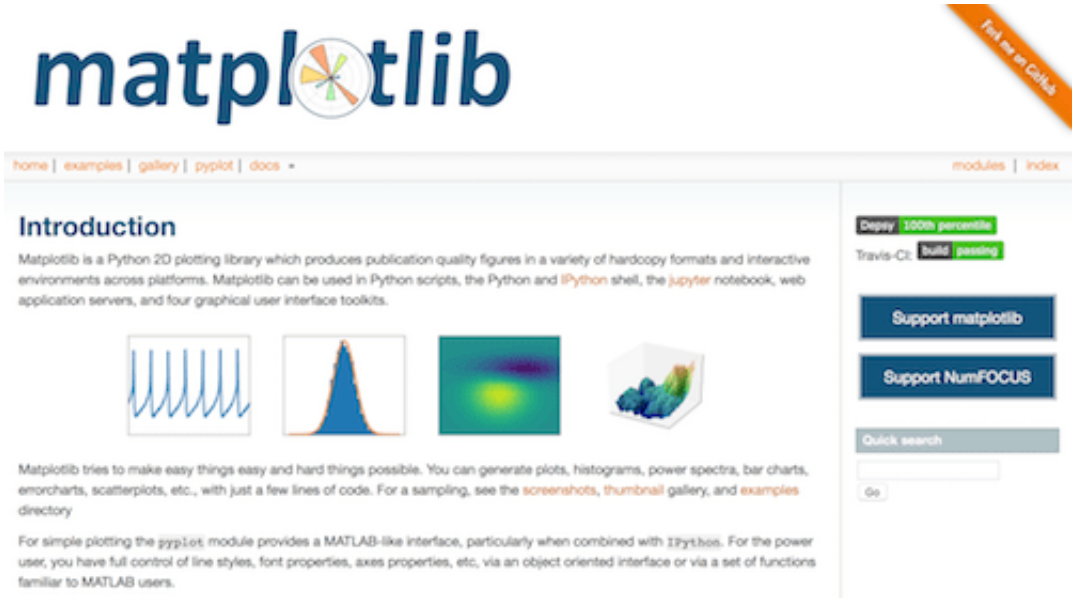


ام‌ای‌تی‌ال‌ای‌بی[۴۹] یکپارچه
سطح بالا و یک محیط تعاملی برای محاسبات عددی، مصورسازی و برنامه‌نویسی، ابزاری قدرتمند برای دانشمندان داده است. ام‌ای‌تی‌ال‌ای‌بی به عنوان زبان محاسبات فنی عمل می‌کند و برای ریاضیات، گرافیک و برنامه‌نویسی مفید است.

هزینه: دارد

32. Matplotlib

@matplotlib



متپلوت‌لیب[۵۰] یک کتابخانه نقشه‌کشی پایتن‌تری‌دی[۵۱] است که ارقام با کیفیت انتشار در انواع قالب‌های چاپی و محیط‌های تعاملی را در

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

سیستم‌عامل‌ها تولید می‌کند. دانشمندان داده از این ابزار در اسکریپت‌های پایتن، پوسته پایتن و آی‌پایتن، نوت بوک جی‌پایتر، سرورهای برنامه‌های وب و چهار جعبه‌ابزار رابط کاربر گرافیکی استفاده می‌کنند.

هزینه:رایگان

33. MLBase

@amplab

پوسی‌برکلیزای‌ام‌پی‌لب[۵۲] الگوریتم‌ها، ماشین‌ها و افراد را برای درک ابرداده‌ها ادغام می‌کند. آن‌ها همچنین ام‌ال‌بیس[۵۳] را توسعه دادند، یک پروژه منبع باز که یادگیری ماشین توزیع شده را برای دانشمندان داده آسان‌تر می‌کند.

هزینه:رایگان

34. MySQL

@MySQL

مای‌اس‌کیوال[۵۴] یکی از محبوب‌ترین پایگاه‌های اطلاعاتی منبع‌باز امروزی است. این همچنین یک ابزار محبوب برای دانشمندان داده است که

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

می‌توانند برای دسترسی به داده‌ها از پایگاه داده استفاده کنند. حتی اگر مای‌اس‌کیوال به طور معمول در برنامه‌های وب یک نرم‌افزار باشد، می‌تواند در تنظیمات مختلف مورد استفاده قرار گیرد.

هزینه:رایگان

35. Narrative Science

@narrativesci

نرتیوساینس[۵۵] به

شرکت‌ها کمک می‌کند تا با استفاده از روایت‌های خودکار و هوشمند تولید شده توسط نسل پیشرفته زبان روایی با را هاداده که نرتیوساینس فناوری از استفاده با داده دانشمندان .برسانند حداکثر به را خود های‌داده تأثیر ،(NLG) سرعت و مقیاس بی‌نظیر تفسیر و سپس تبدیل می‌کند، داده‌ها را انسان‌سازی می‌کنند.

هزینه:دارد

36. Natural Language Toolkit (NLTK)

@NLTK_org

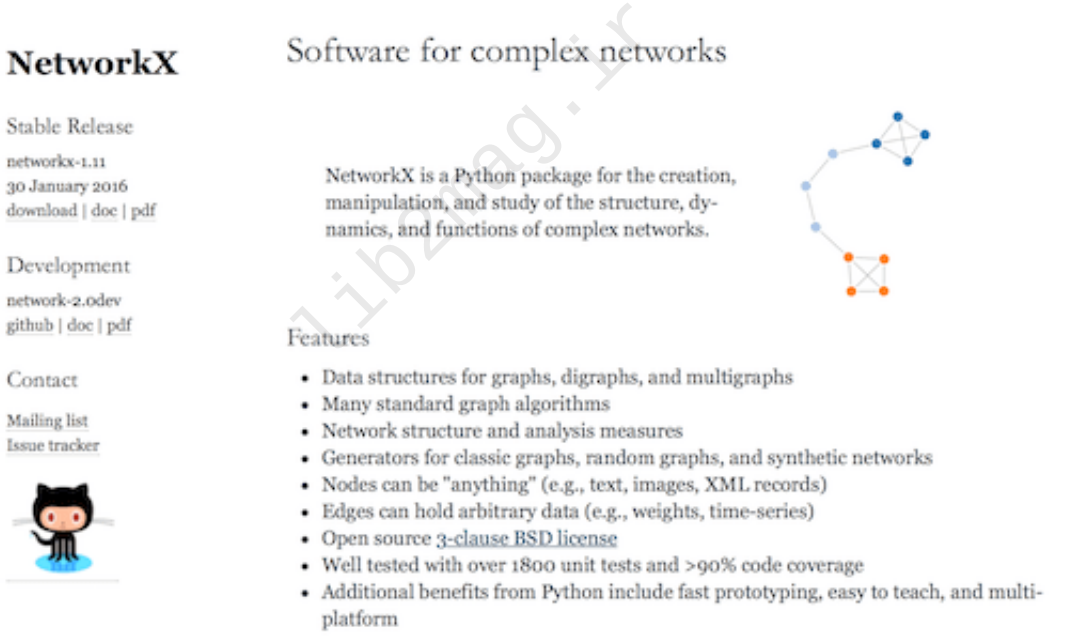
یک‌بستراصلی‌برای‌ساخت

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

برنامه‌های پایتن، جعبه‌ابزار طبیعی زبان (ان‌ال‌تی‌کی)[۵۶] ابزاری برای کار با داده‌های زبان انسان است. ان‌ال‌تی‌کی ابزاری مفید برای دانشمندان مبتدی داده و دانشجویان علوم داده است که در زبان‌شناسی محاسباتی با استفاده از پایتن کار می‌کنند.

هزینه:رایگان

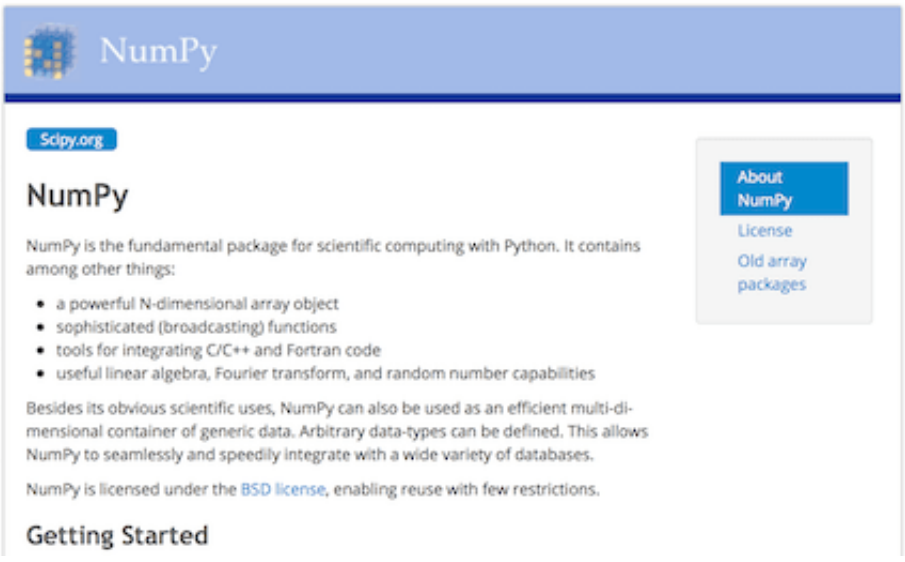
37. NetworkX



نت‌وورک‌ایکس[۵۷] یک ابزار بسته پایتن برای دانشمندان داده است. ایجاد، دست‌کاری و مطالعه ساختار، پویایی و عملکرد شبکه‌های پیچیده با نت‌وورک‌ایکس انجام می‌شود.

هزینه:رایگان

38. NumPy



یک بسته اساسی برای محاسبات علمی با پایتون، نام‌پای[۵۸] برای استفاده‌های علمی مناسب است. نام پای همچنین به عنوان یک ظرف چندبعدی

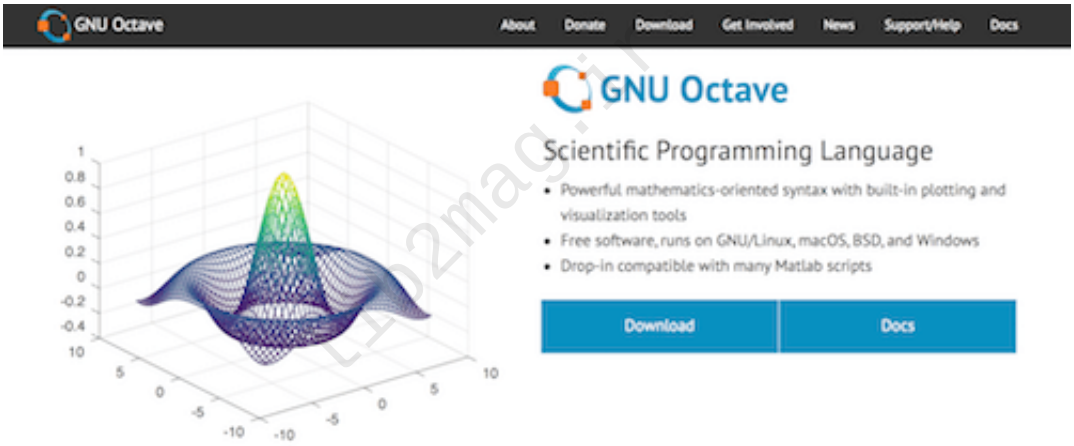
تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

داده‌های عمومی عمل می‌کند.

هزینه:رایگان

39. Octave

@GnuOctave

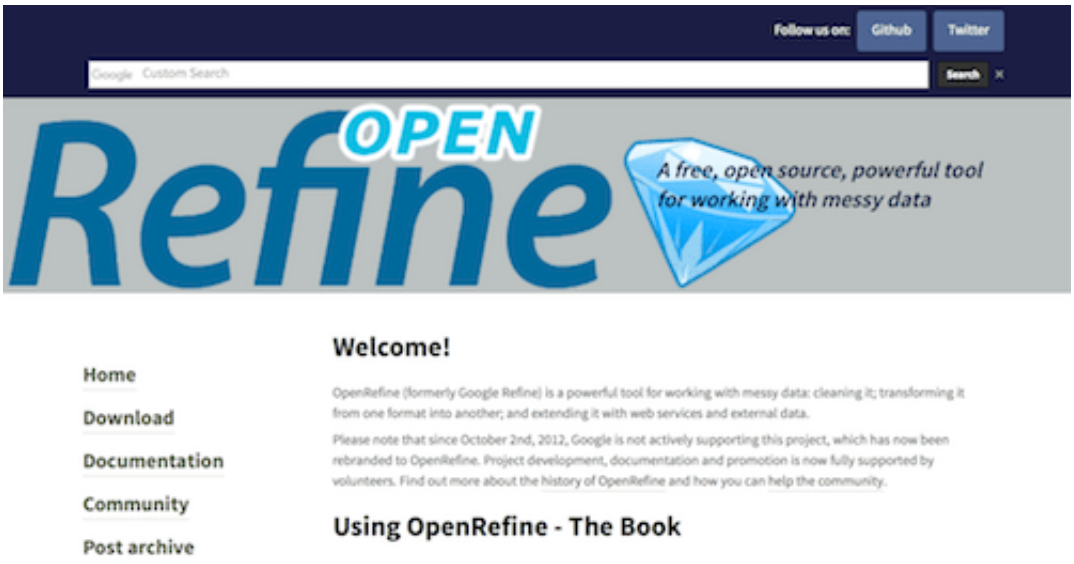


جی‌ان‌یواکتیو[۵۹] یک‌زبان برنامه‌نویسی علمی است که ابزاری مفید برای دانشمندان داده است که به دنبال حل سیستم معادلات یا مصورسازی داده‌ها با دستورات رسم سطح بالا هستند. این ابزار با ام‌ای‌تی‌ال‌ای‌بی[۶۰] سازگار است و مفسر آن را می‌توان در حالت جی‌یوآی[۶۱]، به عنوان کنسول اجرا کرد یا به عنوان بخشی از پوسته اسکریپت فراخوانی کرد.

هزینه:رایگان

40. OpenRefine

@OpenRefine



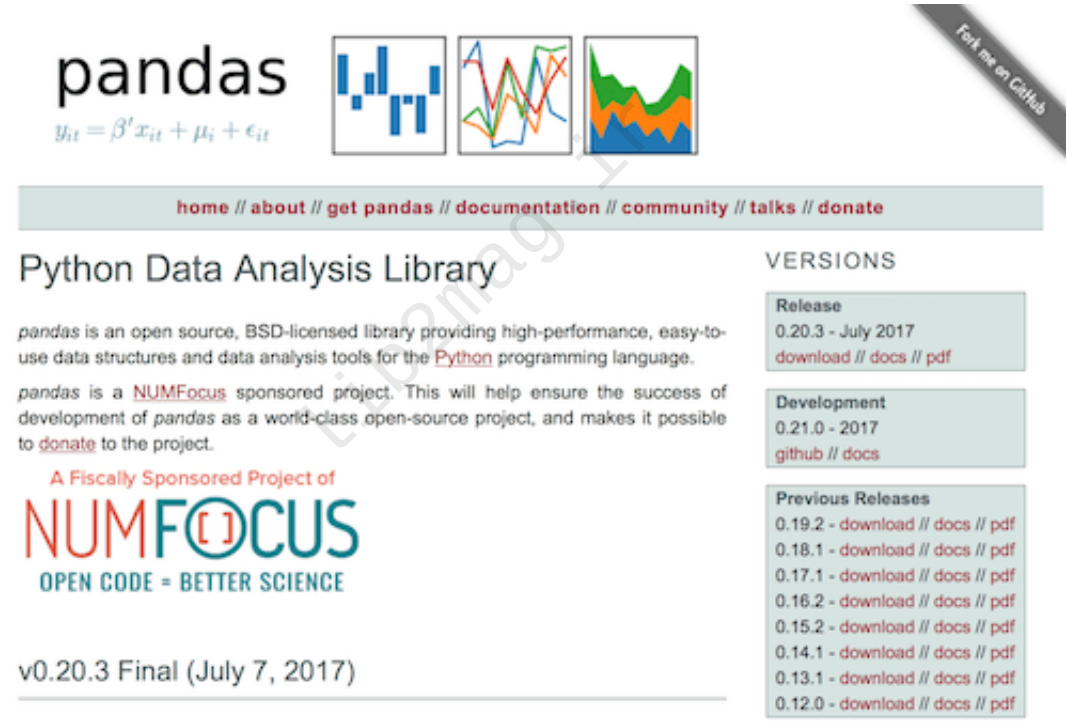
اوپن‌ریفاین[۶۲] ابزاری قدرتمند برای دانشمندان داده است که می‌خواهند داده‌ها را با سرویس‌های وب پاک، تبدیل و گسترش دهند و سپس آن‌ها را به پایگاه‌های داده پیوند دهند. قبلاً گوگل‌ریفاین[۶۳] اکنون اوپن‌ریفاین یک پروژه منبع باز است که به طور کامل

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

توسط داوطلبان پشتیبانی می‌شود.

هزینه:رایگان

41. pandas



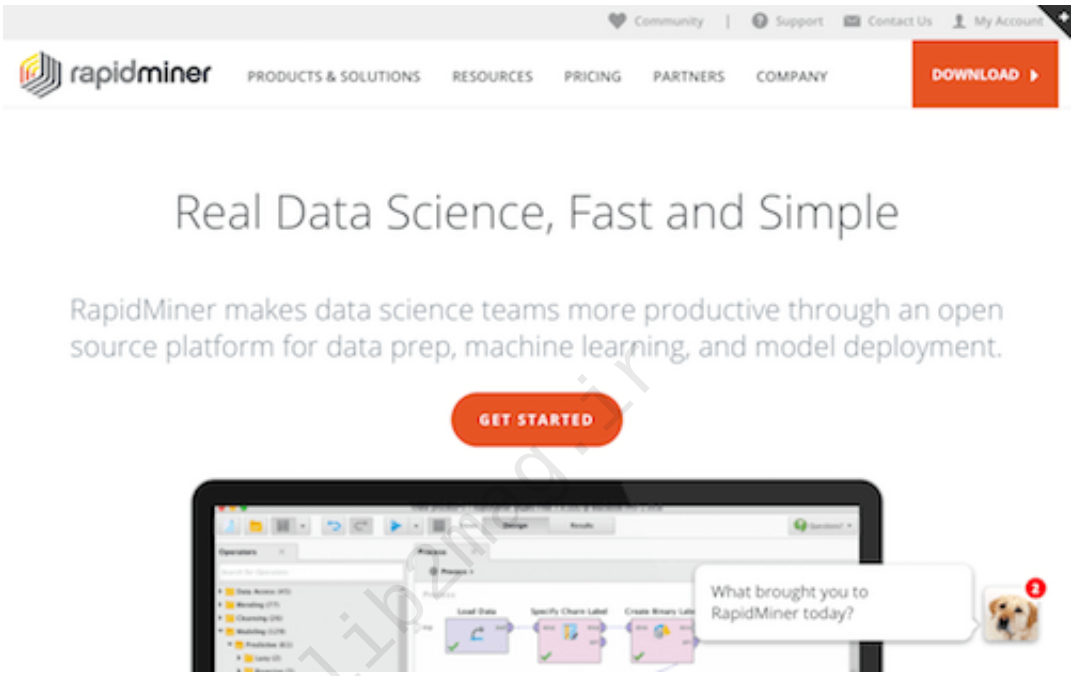
پانداس[۶۴] یک کتابخانه منبع‌باز است که ساختارهای داده‌ای با کاربرد آسان و ابزارهای تجزیه‌وتحلیل داده‌ها را برای زبان برنامه‌نویسی پایتون ارائه می‌دهد. دانشمندان داده در صورت نیاز به کتابخانه تجزیه‌وتحلیل داده‌های پایتون از این ابزار استفاده می‌کنند.

هزینه:رایگان

42. RapidMiner

@RapidMiner

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

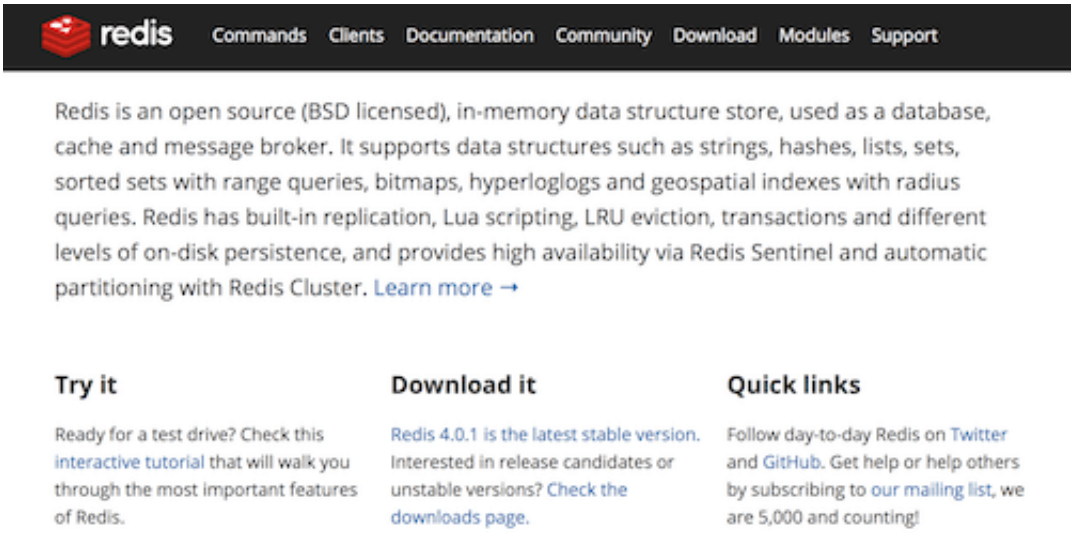


دانشمندان داده هنگامی

که از راپیدماینر[۶۵] ، یک بستر واحد برای آماده‌سازی داده‌ها، یادگیری ماشین و استقرار مدل استفاده می‌کنند، بهره‌وری بیشتری دارند. راپیدماینر ابزاری برای ساخت سریع و ساده دانش داده، یک رهبر در "Quadrant Magic Gartner 2017" برای سیستم‌های علوم داده، یک رهبر در "2017 Wave Forrester" برای تجزیه‌وتحلیل پیش‌بینی و یادگیری ماشین و عملکرد بالا در شبکه تجزیه‌وتحلیل پیش‌بینی جی‌تی‌کرود[۶۶]است.

43. Redis

@redisfeed



ردیس[۶۷] یک سرور

ساختار داده است که دانشمندان داده از آن به عنوان پایگاه داده، حافظه پنهان و کارگزار پیام استفاده می‌کنند. این منبع باز، حافظه ذخیره‌سازی ساختار داده از رشته‌ها، هش‌ها، لیست‌ها و موارد دیگر پشتیبانی می‌کند.

هزینه:رایگان

44. RStudio

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

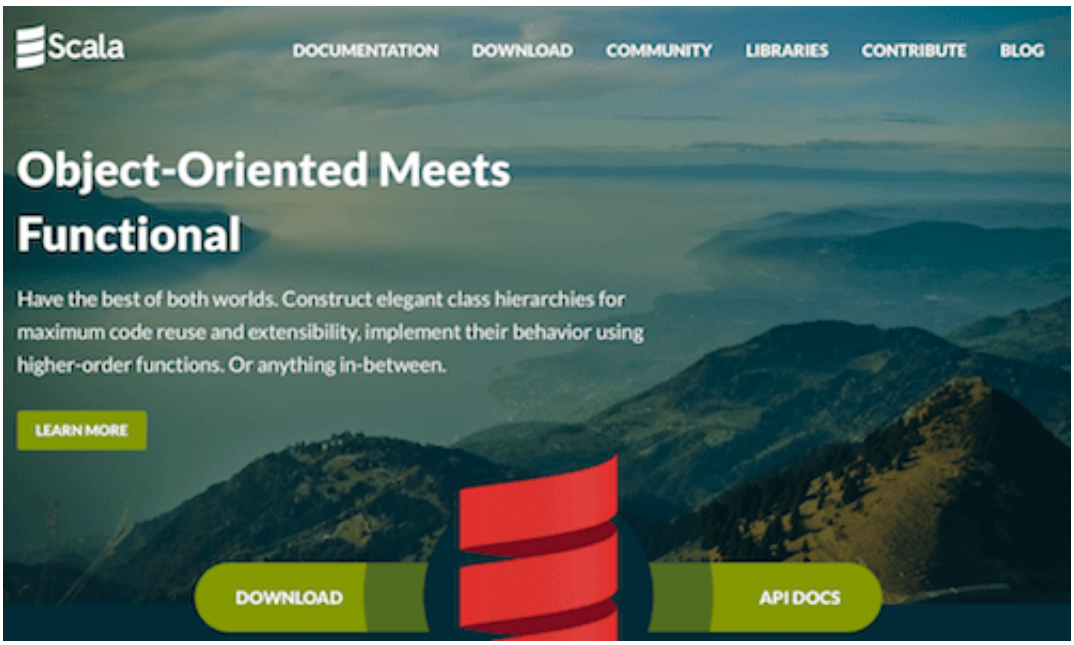
@rstudio

استودیوآر[۶۸] ابزاری برای
دانشمندان داده است که متن‌باز و آماده برای شرکت است. این نرم‌افزار حرفه‌ای برای جامعه آر، استفاده از آر را آسان‌تر می‌کند.

هزینه: نسخه منبع‌باز رایگان

Scala .45

@scala_lang



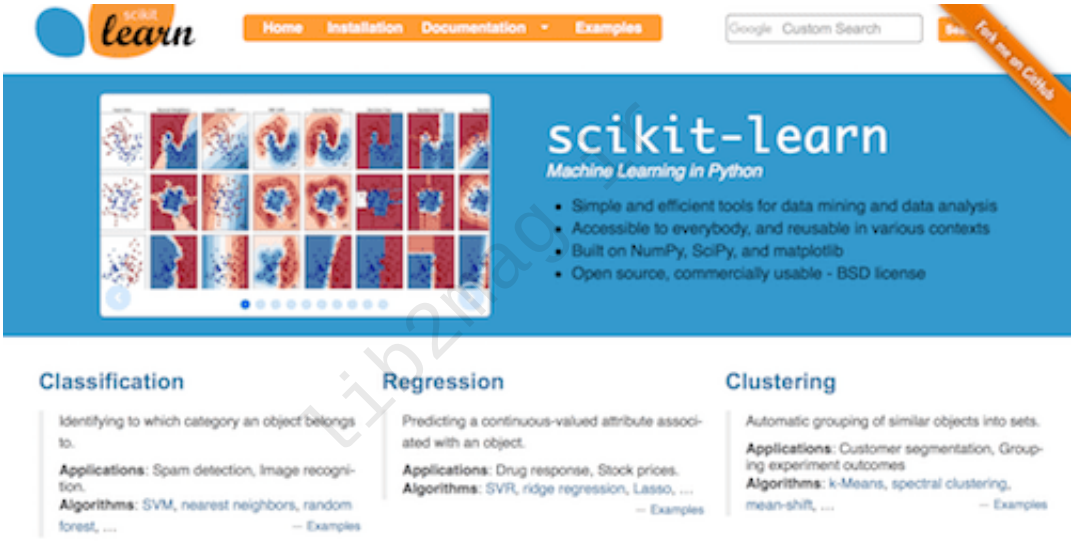
زبان‌برنامه‌نویسی‌مقیاس
ابزاری[۶۹] برای دانشمندان داده است که به دنبال ساخت سلسله مراتب برای به حداکثر رساندن استفاده مجدد و گسترش کد هستند. این ابزار همچنین به کاربران امکان می‌دهد تا رفتار سلسله مراتبی کلاس را با استفاده از توابع مرتبه بالاتر پیاده‌سازی کنند.

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

هزینه:رایگان

46. scikit-learn

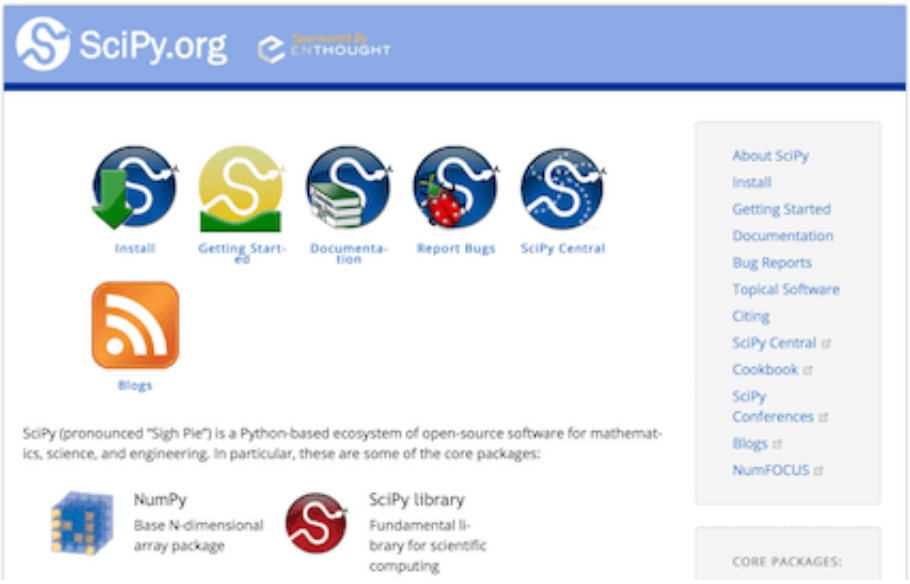
@scikit_learn



سایکیت_لرن[۷۰] برای پایتن
یک یادگیری ماشینی با کاربردهای عمومی است که به راحتی قابل استفاده است. دانشمندان داده، یادگیری سایکیت را ترجیح می‌دهند زیرا دارای ابزارهای ساده و کارآمد برای داده‌کاوی و تجزیه‌وتحلیل داده‌ها است.

هزینه:رایگان

47. SciPy



سایپی[۷۱] ، اکوسیستم نرم‌افزار منبع‌باز مبتنی بر پایتن، برای ریاضیات، علوم و کاربردهای مهندسی در نظر گرفته شده است. سایپی‌استاک[۷۲] شامل پایتن، نام‌پی، متلپ‌لوت‌لیب، پایتون، کتابخانه سایپی و موارد دیگر است.

هزینه:رایگان

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

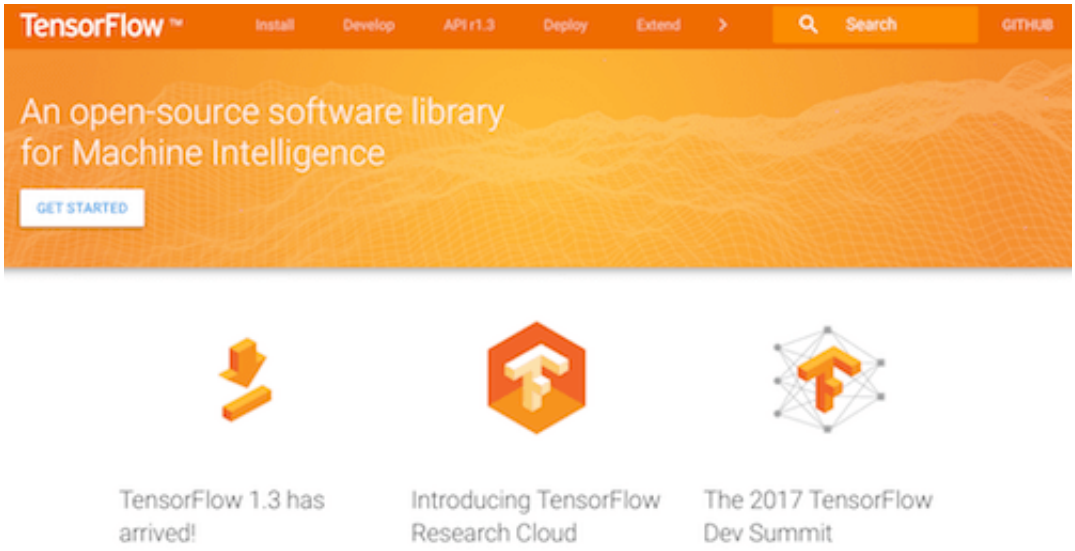
48. Shiny

شینی[۷۳] یک چارچوب
برنامه وب برای آر توسط استودیوآر است، ابزاری است که دانشمندان از آن برای تبدیل تجزیه‌وتحلیل به برنامه‌های وب تعاملی استفاده می‌کنند. شینی ابزاری ایده‌آل برای دانشمندان داده است که در زمینه توسعه وب کم‌تجربه هستند.

هزینه:دارد

49. TensorFlow

@tensorflow



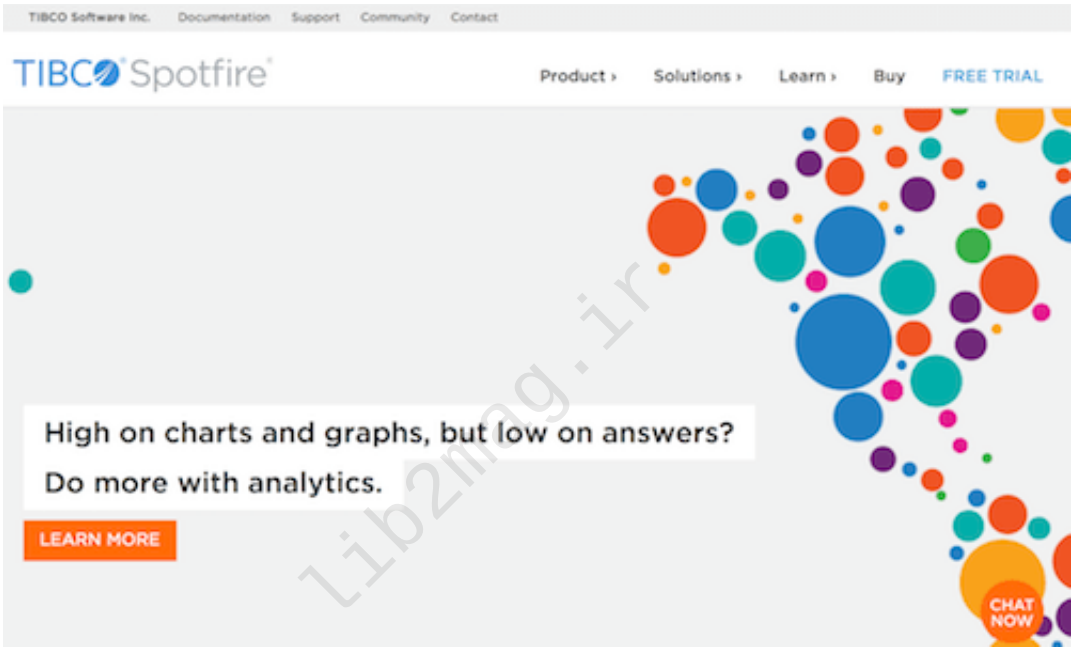
تنسورفلو[۷۴] یککتابخانه
یادگیری ماشین منبع‌باز سریع، انعطاف‌پذیر، مقیاس‌پذیر برای تحقیق و تولید است. دانشمندان داده با استفاده از نمودارهای جریان داده از تنسورفلو برای محاسبه عددی استفاده می‌کنند.

هزینه:رایگان

50. TIBCO Spotfire

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

@TIBCO



تی‌آی‌بی‌سی‌او[۷۵] **بامکان**
تصمیم‌گیری بهتر و اقدامات سریع‌تر و هوشمندانه، تجارت دیجیتال را پیش می‌برد. راه‌حل اسپاتفیر[۷۶] آن‌ها، ابزاری برای دانشمندان داده است که به کشف داده‌ها، اختلافات داده‌ها، تجزیه‌وتحلیل‌های پیش‌بینی شده و موارد دیگر می‌پردازد.

هزینه: نسخه آزمایشی رایگان

Pxyll.com BONUS .51

@pyxll

این وبلاگ دارای لیستی کامل از ابزارهای کار با پایتن و اکسل است. این وبلاگ نوشتن افزونه‌های اکسل در پایتن، خواندن و نوشتن فایل‌های اکسل و تعامل با اکسل را شامل می‌شود. این یک منبع عالی برای درک تفاوت بین همه ابزارهای مختلف پایتن/ اکسل به شکل یکجا است.

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

پانویس:

[۱] Algorithms.io
[۲] یک پلتفرم فناوری سازمانی است که شرکت‌ها را قادر می‌سازد تا از برنامه‌های در حال ظهور استفاده کنند و از چهارمین انقلاب صنعتی استفاده کنند.
[۳] LumenData
[۴] Apache Giraph
[۵] Pregel
[۶] Apache Hadoop
[۷] Apache HBase
[۸] Bigtable
[۹] HDFS
[۱۰] Apache Hive
[۱۱] SQL
[۱۲] Apache Kafka
[۱۳] Apache Mahout
[۱۴] Apache Mesos
[۱۵] Apache Pig
[۱۶] Apache Spark
[۱۷] Cassandra
[۱۸] S3
[۱۹] Apache Storm
[۲۰] RPC
[۲۱] BigML
[۲۲] Bokeh
[۲۳] Python
[۲۴] Cascading
[۲۵] Clojure
[۲۶] Mike Bostock
[۲۷] SVG
[۲۸] Canvas
[۲۹] HTML
[۳۰] DataRobot
[۳۱] DataRPM
[۳۲] Frost & Sullivan
[۳۳] Excel
[۳۴] Feature Labs
[۳۵] Google Fusion Tables
[۳۶] GNU
[۳۷] Gawk
[۳۸] awk
[۳۹] Hadley Wickham and Winston Chang
[۴۰] ggplot2
[۴۱] R
[۴۲] GraphLab Create
[۴۳] IPython
[۴۴] Jupyter
[۴۵] Notebook
[۴۶] KNIME

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

KNIME Analytics Platform [۴۷]
Logical Glue [۴۸]
MATLAB [۴۹]
Matplotlib [۵۰]
Python 2D [۵۱]
UC Berkeley's AMPLab [۵۲]
MLBase [۵۳]
MySQL [۵۴]
Narrative Science [۵۵]
NLTK [۵۶]
NetworkX [۵۷]
NumPy [۵۸]
GNU Octave [۵۹]
MATLAB [۶۰]
GUI [۶۱]
OpenRefine [۶۲]
Google Refine [۶۳]
pandas [۶۴]
RapidMiner [۶۵]
G2 Crowd [۶۶]
Redis [۶۷]
RStudio [۶۸]
Scala [۶۹]
scikit-learn [۷۰]
SciPy [۷۱]
SciPy Stack [۷۲]
Shiny [۷۳]
TensorFlow [۷۴]
TIBCO [۷۵]
Spotfire [۷۶]

lib2mag.ir

منبع:
این نوشته ترجمه مطلبی با این عنوان است:

Top Tools for Data Scientists Analytics Tools, Data Visualization Tools, Database Tools, Moreand

مترجم: علی اوچی

دانشجوی کتابداری و اطلاع رسانی پزشکی دانشگاه علوم پزشکی همدان

مشخصات استناددهی به این مقاله

نویسنده(ها): علی اوچی
عنوان مقاله: ابزارهای برتر برای دانشمندان داده: ابزارهای تجزیه‌وتحلیل، ابزار مصورسازی داده، ابزار پایگاه داده و موارد دیگر
عنوان مجله: کتابدار ۲.۰ –(عنوان لاتین: i-Kitابdar 2.0)

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.

مشخصات استناددهی به این مقاله

دوره مجله(Vol): ۷
شماره مجله(Issue): ۱
سال(Year): ۱۴۰۰
شناسه دیجیتال(DOI):
لینک کوتاه: http://lib2mag.ir/12290

lib2mag.ir

تمامی حقوق برای مجله کتابدار 2.0 محفوظ است. نقل مطالب مجله تنها با ذکر منبع و نام نویسنده مجاز است.