

من أدوات علوم البيانات الأكثر استخدامًا 14 لعام 2022- مكونات علوم البيانات الأساسية

أهم أدوات علوم البيانات

1. SAS

هو برنامج مملوك مغلق SAS. إنها إحدى أدوات علوم البيانات المصممة خصيصًا للعمليات الإحصائية الأساسية لأداء SAS لغة برمجة SAS تستخدم. تستخدمه المؤسسات الكبيرة لتحليل البيانات المصدر النمذجة الإحصائية.

تقدم. يتم استخدامه على نطاق واسع من قبل المتخصصين والشركات العاملة على برامج تجارية موثوقة والأدوات الإحصائية التي يمكنك استخدامها كعالم بيانات لنمذجة وتنظيم العديد من المكتبات SAS بياناتهم.

موثوق للغاية وتحظى بدعم قوي من الشركة ، إلا أنها باهظة الثمن ولا تستخدم إلا في SAS في حين أن يتضاءل بالمقارنة مع بعض الأدوات الأكثر حداثة والتي هي مفتوحة SAS أيضا ، الصناعات الكبيرة المصدر.

غير متوفرة في الحزمة الأساسية ويمكن SAS علاوة على ذلك ، هناك العديد من المكتبات والحزم في أن تتطلب ترقية باهظة الثمن.

2. اباتشي سبارك

هو محرك تحليلات قوي للغاية وهو أداة علوم البيانات الأكثر Spark أو ببساطة Apache Spark الدفق خصيصًا للتعامل مع معالجة الدفعات ومعالجة Spark تم تصميم. استخدامًا

يأتي مزودًا بالعديد من واجهات برمجة التطبيقات التي تسهل على علماء البيانات إجراء وصول متكرر ويمكنه Hadoop وما إلى ذلك. إنه تحسين على SQL إلى البيانات من أجل التعلم الآلي والتخزين في MapReduce أداء 100 مرة أسرع من

على العديد من واجهات برمجة تطبيقات التعلم الآلي التي يمكن أن تساعد علماء Spark يحتوي البيانات على إجراء تنبؤات قوية باستخدام البيانات المقدمة.

بشكل أفضل من منصات البيانات الضخمة الأخرى في قدرتها على التعامل مع تدفق Spark تعمل مقارنة بالأدوات التحليلية **يمكنها معالجة البيانات في الوقت الفعلي Spark** هذا يعني أن البيانات الأخرى التي تعالج البيانات التاريخية فقط على دفعات.

لكن R و Java و Python العديد من واجهات برمجة التطبيقات القابلة للبرمجة في Spark تقدم وهي **Java Virtual Machine** التي تعتمد على Scala هو مع لغة برمجة Spark أقوى اقتران لـ متعددة المنصات بطبيعتها.

حيث يتم Hadoop بكفاءة عالية في إدارة المجموعة مما يجعلها أفضل بكثير من Spark تتميز بمعالجة التطبيقات Spark إن نظام إدارة الكتلة هذا هو الذي يسمح لـ استخدام هذا الأخير فقط للتخزين بسرعة عالية.

3. BigML

يوفر بيئة واجهة مستخدم رسومية ، هي أداة أخرى لعلوم البيانات مستخدمة على نطاق واسع BigML **خوارزميات التعلم** قابلة للتفاعل بالكامل ومستندة إلى مجموعة النظراء يمكنك استخدامها لمعالجة برنامجاً قياسياً باستخدام الحوسبة السحابية لمتطلبات الصناعة BigML يوفر . **الآلي** على سبيل . من خلاله ، يمكن للشركات استخدام خوارزميات التعلم الآلي عبر أجزاء مختلفة من شركتهم . المثال ، يمكنه استخدام هذا البرنامج الواحد للتنبؤ بالمبيعات وتحليلات المخاطر وابتكار المنتجات.

يستخدم مجموعة متنوعة من خوارزميات التعلم الآلي مثل . متخصص في النمذجة التنبؤية BigML التجميع والتصنيف والتنبؤ بالسلاسل الزمنية وما إلى ذلك.

ويمكنك إنشاء حساب مجاني أو Rest APIs واجهة ويب سهلة الاستخدام باستخدام BigML يوفر يسمح بالتصورات التفاعلية للبيانات ويوفر لك القدرة على . حساب مميز بناءً على احتياجاتك من البيانات . تصدير المخططات المرئية على هاتفك المحمول أو أجهزة إنترنت الأشياء.

مع العديد من طرق التشغيل الآلي التي يمكن أن تساعدك على أتمتة BigML علاوة على ذلك ، يأتي ضبط نماذج المعلمات الفائقة وحتى أتمتة سير عمل البرامج النصية القابلة لإعادة الاستخدام.

4. D3.js

، مكتبة D3.js . بشكل أساسي كلغة برمجة نصية من جانب العميل Javascript تُستخدم باستخدام العديد من . تسمح لك بعمل تصورات تفاعلية على متصفح الويب الخاص بك Javascript ، يمكنك استخدام العديد من الوظائف لإنشاء تصور ديناميكي D3.js واجهات برمجة التطبيقات لـ . وتحليل البيانات في متصفحك .
المستندات ديناميكية من D3.js يجعل . هي استخدام انتقالات متحركة D3.js ميزة أخرى قوية في خلال السماح بالتحديثات من جانب العميل والاستخدام النشط للتغيير في البيانات لعكس المرنیات على المتصفح .

لإنشاء تصورات رائعة وعابرة ستساعدك على تنفيذ الرسوم البيانية CSS يمكنك دمج هذا مع المخصصة على صفحات الويب .

على الأجهزة القائمة بشكل عام ، يمكن أن تكون أداة مفيدة جدًا لعلماء البيانات الذين يعملون على .إنترنت الأشياء التي تتطلب تفاعلًا من جانب العميل للتصور ومعالجة البيانات .

5. ماتلاب

إنه برنامج مغلق . هي بيئة حوسبة رقمية متعددة النماذج لمعالجة المعلومات الرياضية MATLAB يستخدم .المصدر يسهل وظائف المصفوفة وتنفيذ الخوارزميات والنمذجة الإحصائية للبيانات .على نطاق واسع في العديد من التخصصات العلمية MATLAB

باستخدام . والمنطق الضبابي الشبكات العصبية لمحاكاة MATLAB في علم البيانات ، يتم استخدام أيضًا في معالجة MATLAB يستخدم . ، يمكنك إنشاء تصورات قوية MATLAB مكتبة رسومات .الصور والإشارات

هذا يجعلها أداة متعددة الاستخدامات لعلماء البيانات حيث يمكنهم معالجة جميع المشكلات ، من تنظيف .التعلم العميق الأكثر تقدمًا خوارزميات البيانات وتحليلها إلى

السهل لتطبيقات المؤسسات والأنظمة المدمجة يجعلها أداة MATLAB علاوة على ذلك ، فإن تكامل مثالية لعلوم البيانات

كما أنه يساعد في أتمتة المهام المختلفة التي تتراوح من استخراج البيانات إلى إعادة استخدام البرامج ومع ذلك ، فهو يعاني من محدودية كونه برنامجًا مملوكًا مغلق المصدر .النصية لاتخاذ القرار

6. إكسل

في الغالب لعمليات Microsoft Excel طورت. ربما تكون أداة تحليل البيانات الأكثر استخدامًا حساب جداول البيانات ، وهي تستخدم اليوم على نطاق واسع لمعالجة البيانات والتصور والحسابات المعقدة.

على الرغم من أنها كانت الأداة التقليدية لتحليل البيانات ، . أداة تحليلية قوية لعلوم البيانات Excel يعد . لا يزال يحزم لكلمة Excel إلا أن

مع العديد من الصيغ والجداول والفلاتر والمقطعات وما إلى ذلك. يمكنك أيضًا Excel يأتي برنامج ليس مخصصًا لحساب Excel في حين أن Excel إنشاء وظائف وصيغ مخصصة باستخدام برنامج الكمية الهائلة من البيانات ، إلا أنه لا يزال خيارًا مثاليًا لإنشاء تصورات قوية للبيانات وجداول بيانات

يستخدم الكثير من علماء . واستخدامه لمعالجة البيانات وتحليلها Excel بـ SQL يمكنك أيضًا ربط لتنظيف البيانات لأنه يوفر بيئة واجهة مستخدم رسومية قابلة للتفاعل لمعالجة Excel البيانات برنامج المعلومات مسبقًا بسهولة

ومع . ، أصبح من السهل الآن حساب التحليلات المعقدة Microsoft Excel لـ ToolPak مع إصدار بشكل عام ، على SAS. ذلك ، لا يزال يتضاءل بالمقارنة مع أدوات علوم البيانات الأكثر تقدمًا مثل أداة مثالية لتحليل البيانات Excel المستوى الصغير وغير المؤسسي ، يعد

7. ggplot2

أنشأ المطورون هذه الأداة لاستبدال R. اللغة برمجة عبارة عن حزمة عرض بيانات متقدمة ggplot2 وهي تستخدم أوامر قوية لإنشاء تصورات رائعة R حزمة الرسومات الأصلية لـ . إنها المكتبة الأكثر استخدامًا التي يستخدمها علماء البيانات لإنشاء تصورات من البيانات التي تم تحليلها مصممة لعلوم البيانات R ، وهي حزمة في tidyverse هي جزء من Ggplot2

أفضل بكثير من بقية تصورات البيانات هي ggplot2 إحدى الطرق التي يكون فيها ، يمكن لعلماء البيانات إنشاء تصورات مخصصة من أجل المشاركة في ggplot2 باستخدام الجماليات . سرد القصص المحسن

، يمكنك إضافة تعليق توضيحي لبياناتك في تصورات وإضافة تسميات نصية إلى ggplot2 باستخدام يمكنك أيضًا إنشاء أنماط مختلفة من نقاط البيانات وتعزيز استعصاء الرسوم البيانية الخاصة بك وما إلى ذلك. إنها أداة علوم البيانات hexbins و cartograms و choropleths الخرائط مثل الأكثر استخدامًا

8. تابلوه .8

يركز على .مليء بالرسومات القوية لعمل تصورات تفاعلية هو برنامج لتصور البيانات Tableau .الصناعات التي تعمل في مجال ذكاء الأعمال .هو قدرته على التفاعل مع قواعد البيانات ، وجداول البيانات ، ومكعبات Tableau أهم جانب في Tableau المعالجة التحليلية عبر الإنترنت) ، وما إلى ذلك إلى جانب هذه الميزات ، يمتلك (OLAP .القدرة على تصور البيانات الجغرافية ورسم خطوط الطول وخطوط العرض في الخرائط

مع مجتمع Tableau يأتي .إلى جانب التصورات ، يمكنك أيضًا استخدام أداة التحليلات لتحليل البيانات هو برنامج Tableau في حين أن .نشط ويمكنك مشاركة نتائجك على النظام الأساسي عبر الإنترنت Tableau Public .مؤسسي ، فإنه يأتي مع إصدار مجاني يسمى

9. كوكب المشتري .9

لمساعدة المطورين في صنع IPython هي أداة مفتوحة المصدر تعتمد على Project Jupyter Julia لغات متعددة مثل Jupyter يدعم .برامج مفتوحة المصدر وتجربة الحوسبة التفاعلية R. و Python و

إنها أداة تطبيق ويب تُستخدم لكتابة التعليمات البرمجية الحية والتصورات والعروض .هي أداة شائعة على نطاق واسع تم تصميمها لتلبية متطلبات علوم البيانات Jupyter .التقديمية

إنها أيضًا أداة قوية لسرد .إنها بيئة قابلة للتفاعل يمكن من خلالها لعلماء البيانات أداء جميع مسؤولياتهم .القصص حيث توجد ميزات عرض مختلفة فيها

، يمكن للمرء إجراء تنظيف البيانات والحساب الإحصائي Jupyter Notebooks باستخدام .إنه مفتوح المصدر بنسبة 100٪ وبالتالي فهو مجاني . التنبؤية نماذج تعلم الآلة والتصوير وإنشاء تعمل على السحابة وتخزن البيانات في Collaboratory عبر الإنترنت تسمى Jupyter هناك بيئة Google Drive.

10. Matplotlib

إنها الأداة الأكثر شيوعًا Python. تم تطويرها من أجل هي مكتبة للتخطيط والتصوير Matplotlib يتم استخدامه بشكل أساسي لرسم الرسوم البيانية المعقدة . لإنشاء الرسوم البيانية بالبيانات التي تم تحليلها باستخدام هذا ، يمكن للمرء إنشاء مخططات شريطية . باستخدام أسطر بسيطة من التعليمات البرمجية ورسوم بيانية ومخططات مبعثرة وما إلى ذلك

واحدة من أكثر الوحدات استخدامًا هي .على العديد من الوحدات الأساسية Matplotlib يحتوي هو أيضًا بديل مفتوح المصدر لوحدة Pyplot. مثل الواجهة MATLAB يقدم pyplot. الرسومية MATLAB.

هي الأداة المفضلة لتصورات البيانات ويستخدمها علماء البيانات على الأدوات Matplotlib .المعاصرة الأخرى.

لتوضيح تصورات البيانات أثناء هبوط مركبة Matplotlib في واقع الأمر ، استخدمت ناسا Python. إنها أيضًا أداة مثالية للمبتدئين في تعلم تصور البيانات باستخدام .الفضائية Phoenix

11. NLTK

يتعامل مع تطوير .معالجة اللغة الطبيعية باعتبارها المجال الأكثر شيوعًا في علوم البيانات برزت النماذج الإحصائية التي تساعد أجهزة الكمبيوتر على فهم لغة الإنسان. تعد هذه النماذج الإحصائية جزءًا من التعلم الآلي ويمكن من خلال العديد من خوارزمياتها مساعدة Natural مع مجموعة من المكتبات تسمى Python تأتي لغة .أجهزة الكمبيوتر في فهم اللغة الطبيعية Language Toolkit (NLTK) تم تطويرها لهذا الغرض المحدد فقط.

على نطاق واسع لتقنيات معالجة اللغة المختلفة مثل الترميز ، والاشتقاق ، والعلامات ، NLTK يستخدم يتكون من أكثر من 100 هيئة وهي عبارة عن مجموعة من البيانات لبناء .والتحليل ، والتعلم الآلي نماذج التعلم الآلي.

يحتوي على مجموعة متنوعة من التطبيقات مثل أجزاء من تمييز الكلام وتجزئة الكلمات والترجمة الآلية والتعرف على النص إلى الكلام وما إلى ذلك.

12. سكيكيت ليرن

من السهل .تُستخدم لتنفيذ خوارزميات التعلم الآلي Python هي مكتبة في لغة Scikit-Learn .والسهل تنفيذ أداة مستخدمة على نطاق واسع للتحليل وعلوم البيانات.

وهو يدعم مجموعة متنوعة من الميزات في التعلم الآلي مثل المعالجة المسبقة للبيانات ، والتصنيف ، والانحدار ، والتكتل ، وتقليل الأبعاد ، إلخ.

لذلك فهي في المواقف التي تتطلب .استخدام خوارزميات التعلم الآلي المعقدة Scikit-Learn تسهل يستخدم .نماذج أولية سريعة وهي أيضًا منصة مثالية لإجراء البحوث التي تتطلب التعلم الآلي الأساسي وما إلى ذلك Matplotlib و Numpy و SciPy مثل Python المكتبات الأساسية في العديد من

13. TensorFlow

يستخدم على نطاق واسع لخوارزميات التعلم الآلي. للتعلم الآلي أداة قياسية TensorFlow أصبح وهي مصفوفات Tensors اسم TensorFlow أطلق المطورون على. المتقدمة مثل التعلم العميق متعددة الأبعاد.

إنها مجموعة أدوات مفتوحة المصدر ومتطورة باستمرار وهي معروفة بأدائها وقدراتها الحسابية على كل من وحدات المعالجة المركزية ووحدات معالجة TensorFlow يمكن تشغيل. العالية. أكثر قوة TPU الرسومات وقد ظهر مؤخرًا على منصات.

وهذا يمنحها ميزة غير مسبقة من حيث قوة المعالجة لخوارزميات التعلم الآلي المتقدمة.

مثل لديها مجموعة متنوعة من التطبيقات Tensorflow فإن نظرًا لقدرتها العالية على المعالجة، التعرف على الكلام، وتصنيف الصور، واكتشاف الأدوية، وتوليد الصور واللغة، وما إلى ذلك. هي أداة يجب معرفتها Tensorflow بالنسبة لعلماء البيانات المتخصصين في التعلم الآلي، فإن

14. وىكا

هو برنامج تعلم آلي W aikato E nvironment for K nowledge A nalysis أو Weka يتكون. إنها مجموعة من خوارزميات التعلم الآلي المختلفة لاستخراج البيانات. Java مكتوب بلغة من العديد من أدوات التعلم الآلي مثل التصنيف والتجميع والانحدار والتصوير وإعداد البيانات Weka إنه برنامج واجهة مستخدم رسومية مفتوح المصدر يسمح بتنفيذ أسهل لخوارزميات التعلم الآلي من خلال نظام أساسي قابل للتفاعل.

يمكنك فهم طريقة عمل "التعلم الآلي" على البيانات دون الحاجة إلى كتابة سطر من التعليمات. إنه مثالي لعلماء البيانات المبتدئين في التعلم الآلي. البرمجية

.آمل أن يعجبك تفسيرنا. لذلك، كان هذا كله في أدوات علم البيانات