

كان العمل بمنأى آمن عن هجمات العدو، فإننا نعلم أن أحد أول استخدامات البيانات الجيدة التوثيق كان الإحصاء السكاني الذي أجراه البابليون عام ٣٨٠٠ قبل الميلاد. الأمثلة الأولى على استخدام البيانات ليست قاصرة، ينسب «قاموس أكسفورد الإنجليزي» أول استخدام معروف للكلمة إلى الكاهن الإنجليزي هنري هاموند خلال القرن السابع عشر، وبحلول عام ١٨٠٩، أرسى كلٌّ من جاوس ولاپلاس أُسساً رياضية للغاية للمنهجية الإحصائية الحديثة. جُمعت كمية هائلة من البيانات خلال تفشي وباء الكوليرا عام ١٨٥٤ في شارع برود بمدينة لندن، واشتهر بكونه أحد اختصاصيي الأوبئة الرواد. على سبيل المثال، إلا أن هذا لم يُحقَّق إلا نجاحاً محدوداً في تقليل كم العمل الذي يؤديه مكتب التعداد. ولكن حدثت طفرة جاءت في أوانها قبل تعداد عام ١٨٩٠ عندما استُخدمت آلة تبويب البطاقات المثقَّبة التي اخترعها هيرمان هوليريث لتصنيف البيانات ومعالجتها. باع هوليريث آتته إلى الشركة التي تطوّرت فيما بعد لتصبح شركة آي بي إم، البيانات في العصر الرقمي أو التجارب العلمية، ومن ثمّ، ظهرت الحاجة إلى علاج المشكلات الحتمية التي نتجت عن الكم الهائل من البيانات التي أصبح من السهل الوصول إليها بفضل شبكة الويب، وسنتناول أولاً كيفية التمييز بين أنواع البيانات المختلفة. أو شبه هيكلية. من النوع المكتوب يدوياً والمحفوظ في دفاتر أو في خزانات الملفات، المعلومات الضرورية لطلب سلعة ما عبر الإنترنت. ولكن الرسالة الفعلية غير الهيكلية توجد في نص الرسالة؛ ومن ثمّ يمكن تصنيفها على أنها بيانات شبه هيكلية. توجد البيانات شبه الهيكلية أيضاً في مواقع شبكات التواصل الاجتماعي التي تستخدم الوسوم حتى يمكن تحديد الرسائل (التي هي بيانات غير هيكلية) عن موضوع مُعيّن. إن التعامل مع البيانات غير الهيكلية أمر صعب؛ وشبه الهيكلية التي تُنتج كل دقيقة، طبقاً لدراسة عالمية حديثة أجرتها شركة آي بي إم، حوالي ٢/٥ إكسابايت من البيانات تُنتج كل يوم. سواءً كانت هيكليةً أو غير هيكلية، ومواقع شبكات التواصل الاجتماعي. ومن ثمّ، فإنه لا يمكن التعامل معها باستخدام أساليب تحليل البيانات التقليدية عليها. عام ٢٠١٢، تُجمَع كمية كبيرة من البيانات. وكتدريب على ذلك، كما استخدمنا أحد محركات البحث، وتحتوي هذه السجلات على معلومات مفيدة على غرار الكلمة المُستعمل عنها نفسها، وكذلك تصفُّحنا لكل موقع. تُسجَل كل نقرة نقرها في مكان ما لاستخدامها في المستقبل. جميعهم يُنتجون بيانات، تأثيرٌ كبير منذ إطلاقه عام ٢٠٠٥، ويزعم بيان صحفي حديث عن يوتيوب أن عدد مستخدميه قد تجاوز المليار مستخدم في جميع أنحاء العالم. عند التعامل مع المشكلات الصحية. وتحقيق أهداف معينة مثل خسارة الوزن أو السير لعدد محدد من الخطوات كل يوم. ولكن سننشأ حتماً مشكلاتٍ تتعلّق بالخصوصية لا بد من وضعها في الاعتبار، مثل تتبّع جميع أنشطة الموظفين على أجهزة الكمبيوتر والهواتف الذكية التي توفّرها الشركة. في خبر صحفي في يونيو عام ٢٠١٦ إلى «الملايسات الشديدة الندرة التي أحاطت بحادث التصادم». مقارنةً بحادث مميت واحد كل ٩٤ مليون ميل تتسبّب فيه القيادة العادية غير الآلية في الولايات المتحدة. أي حوالي ١٠ أضعاف حجمه عام ٢٠١٣. ويُقدَّر أنه سيُنتج إجمالي ٦٠ بيتابايت (الأصوات المُعطاة لكل صفحة ٢٥٠ بايت) من البيانات. وبدأ العمل به عام ٢٠١٨. ومنحتنا ثورة البيانات الضخمة السيارات الذكية وأجهزة المراقبة المنزلية. نتج عن القدرة على جمع البيانات إلكترونياً ظهور مجال علم البيانات المثير، تساعد مثل هذه الاحتمالات في تكريس الموارد لوضع إجراءات، العاملة في هذه المجالات ومجالات أخرى،