

أصبحت البيانات كنزا ثميناً في مجتمع اليوم ، حيث تترجم إلى معرفة مباشرة ومؤثرة. وأصبحت الشركات والمؤسسات تبذل الكثير من الأموال والجهود في سبيل الحصول على البيانات وتحليلها ومن ثم تعدل استراتيجياتها وخططها بناء على ميولات عملائها وجمهورها. إذ تتمثل عملية تحليل البيانات الضخمة في محاولة معالجة مجموعة متنوعة من البيانات للكشف عن الأنماط المخفية والارتباطات غير المعروفة واتجاهات السوق وتقضيلات العملاء وأي معلومات أخرى تساعد أصحاب الأعمال على اتخاذ قرارات مدروسة وبناءة. عملية التحليل هذه تحتاج كما ذكرنا إلى منصات معالجة متطورة. سنتعرف في هذا المقال إلى أهمية تحليل البيانات الضخمة بالنسبة للشركات والمؤسسات وإلى أبرز منصات تحليل ومعالجة البيانات الكبيرة.

أهمية تحليل البيانات الضخمة

يشير مصطلح "البيانات الضخمة" Big Data إلى مجموعات البيانات الكبيرة للغاية ، المنظمة أو غير المنظمة ، المعقدة لدرجة أنها تحتاج إلى أنظمة معالجة أكثر تطوراً من أنظمة معالجة البيانات التقليدية.

بحلول العام 2020، سيبلغ حجم البيانات التي ينتجها كل شخص في اللحظة الواحدة إلى 1.7 ميجا بايت. بينما سيصبح عدد الأجهزة المتصلة على شبكة الإنترنت أكبر من عدد سكان العالم بثلاث مرات. قسم كبير من هذه الأجهزة يعمل ضمن إطار تطبيقات إنترنت الأشياء التي سيصل حجم البيانات الناتجة عنها إلى 600ZB كل سنة. كل هذه المعطيات تبين التطور الكبير الذي تشهده البيانات من حيث الحجم والنوع والسرعة. ويعزى هذا التطور إلى عدة أسباب أهمها عمليات الرقمنة التي انتهجتها الشركات والمؤسسات في السنوات الماضية والإنتشار الواسع لشبكات التواصل الاجتماعي وتطبيقات المحادثات وتكنولوجيا إنترنت الأشياء.

لذلك يعتبر تحليل البيانات أمراً حيوياً وضرورياً لأصحاب الأعمال والشركات إذ سيقدم لهم المزايا التالية:

تقييم سمعة ومكانة الشركة

تحليل بيانات مواقع التواصل الاجتماعي تمنحك القدرة على معرفة ما يتحدث به الناس عن شركتك أو مؤسستك ومعرفة هوية المتحدثين أيضا. وهذا ما يمكنك من تقييم سمعة الشركة ومعرفة مكانتها الحقيقية.

التحليل التنافسي

هذا النوع من التحليلات يتيح لك المقارنة بين المنتجات أو الخدمات التي تقدمها وبين المنتجات والخدمات التي تقدمها الشركات المنافسة. وذلك من خلال معالجة سلوك المستخدمين ومعرفة كيفية تفاعل المستهلكين مع الأنشطة التجارية في قطاعك. كما يقدم لك تحليل البيانات الكبيرة الآني Real-Time Big Data Analytics معلومات آنية عن منافسيك فيحال تخفيضهم للأسعار مثلا أو تغييرهم للإستراتيجيات.

استهداف المستخدمين والعملاء

استقراء البيانات الضخمة على مواقع التواصل الاجتماعي والمدونات والاستطلاعات التي تجرى عبر الإنترنت وبيانات المبيعات يوفر لك قدرات هائلة على تحديد جمهورك المستهدف. إذ يمنحك معرفة دقيقة بهوية المستخدمين المهتمين بمنتجاتك أو خدماتك وبالتالي يمكنك استهدافهم بالحملات الترويجية والتسويقية.

تطوير المنتجات ومعرفة مدى رضا المستخدمين

يمكن أن تساعدك البيانات الكبيرة في فهم نظرة الآخرين لمنتجاتك وردود أفعالهم تجاه علامتك التجارية. ويسمح لك تحليل محتوى وسائل التواصل الإجتماعية بالكشف عن مشاعر زبائنك وتقييمهم لما تقدمه. وهذا مفيد جدا لمعرفة عيوب ومشاكل المنتجات ومن ثم اصلاحها وإعادة تأهيلها حتى تتال رضا العملاء والمستخدمين.

رؤية أفضل للمبيعات

مما قد يؤدي إلى تحقيق إيرادات إضافية. تخبرنا تحليلات البيانات الضخمة في الوقت الفعلي بالضبط عن حجم المبيعات، وفي حالة ما إذا رأى أحد تجار التجزئة عبر الإنترنت أن المنتج يحقق أداءً جيدًا ، يمكنه اتخاذ إجراء لمنع فقدانه أو إساءة توزيعه وبالتالي خسارة الأرباح.

خفض تكاليف الصيانة

يمكن للمؤسسات استخدام تحليلات البيانات الضخمة لتحديد المجالات التي تتفق عليها دون داعٍ أو فائدة. هذا من شأنه أن يعطيهم رؤى لمساعدتهم على اتخاذ قرارات مالية أفضل وخفض التكاليف والمصاريف.

منصات تحليل البيانات

منصة MongoDB

منصة حرة ومفتوحة المصدر تصنف ضمن فئة قواعد البيانات NoSQL. تقوم بتخزين ومعالجة البيانات في صيغة JSON. وهي أداة مفيدة جدًا لتحليل البيانات الضخمة.

منصة Cassandra

هي عبارة عن نظام لإدارة قواعد البيانات الموزعة. كما أنها مفتوحة المصدر. وقد تم انشاؤها من قبل شركة فيسبوك في أول الأمر. وهي قادرة على معالجة كمية كبيرة من البيانات الموزعة على عدة خوادم servers.

منصة Spark

هي واحدة من أكثر مشاريع Apache نشاطا. وهي منصة حوسبة عنقودية مفتوحة المصدر.

منصة Hadoop

هي حزمة برمجية مفتوحة المصدر مكتوبة بلغة جافا JAVA. وتتميز بقدرتها الكبيرة على تخزين ومعالجة البيانات الضخمة سواء المنتظمة منها أو غير المنتظمة.

منصة Google Fusion Tables

منصة ذات إمكانيات كبيرة في تحليل البيانات وتجسيدها على شكل رسوم بيانية وخرائط. ولا شك أن خرائط جوجل لها فضل كبير في جعل هذه الاداة واحدة من أهم أدوات تحليل البيانات.

منصة NodeXL

هي منصة لتحليل وتجسيد العلاقات والشبكات. وهي حرة ومفتوحة المصدر وتمتلك القدرة على تجميع البيانات من منصات التواصل الاجتماعي. مما يجعلها واحدة من أهم أدوات الإحصاء وتحليل البيانات الكبيرة.

منصة CouchDB

قاعدة بيانات مفتوحة المصدر من فئة NoSQL قادرة على تخزين ومعالجة البيانات الضخمة في صيغة JSON.

منصة Apache Hive

واحدة من أفضل منصات تحليل البيانات الضخمة.

منصة Tableau

هي أداة مجانية تمكن من التأكد من صحة الفرضيات واستكشاف البيانات بشكل سريع.

منصة Rapid Miner

هي منصة تحليل بيانات مفتوحة المصدر تساعد على تمثيل البيانات ومعالجتها إضافة إلى النشر والتقييم والتحليلات التنبؤية. توفر Rapid Miner بيئة متكاملة للتحليلات الخاصة بالأعمال business analytics والتحليلات التنبؤية والتعلم الآلي machine learning.

منصة OpenRefine

تعرف أيضا باسم Google Refine. وهي أداة فعالة للعمل على البيانات غير المنظمة. فهي قادرة على تنظيف وترتيب البيانات وتحويلها من صيغة إلى أخرى. كما أنها قادرة على تجميع البيانات من مواقع الويب المختلفة.

منصة Knote

تعد واحدة من أهم منصات تحليل البيانات الضخمة والتي تساعد على اكتشاف الإمكانات الخفية في البيانات والتنبؤ بخصائص جديدة والحصول على رؤى جديدة.

منصة Wolfram Alpha

هو محرك بحث متطور قادر على تحليل البيانات المتعلقة بموضوع البحث ومن ثم تقديم جداول ورسوم بيانية حول الأسعار، أحدث المعاملات، مقارنة الأداء وغيرها.

منصة Solver

هي أداة مضافة على برنامج Microsoft Office Excel. إنها أداة معالجة متقدمة تساعد على تحليل البيانات بشكل سريع.

خصائص ونقاط بعض منصات تحليل البيانات الضخمة

المنصة	الخصائص	النقاط
Tableau Public	*إمكانية مشاركة التحليلات والرسوم البيانية التفاعلية على مواقع الويب وشبكات التواصل الاجتماعي. *لا تحتاج إلى مهارات برمجية.	* حجم البيانات المعالجة محدود *لا تقبل سوى بيانات في صيغة Excel أو txt.
OpenRefine	*تنظيم وترتيب البيانات غير المنظمة أو الناقصة. *تحويل البيانات من صيغة إلى أخرى.	*غير مناسب لمجموعة بيانات كبيرة جدا.

	*تجميع البيانات من مواقع الويب.	
*قدرات متواضعة على تجسيد البيانات وتصويرها في شكل رسومات بيانية وغيرها....	*لا تحتاج إلى مهارات برمجية ولا إلى كتابة الاكواد ومع ذلك فهي تدعم مختلف لغات البرمجة.	KNIME
*قيود على حجم البيانات المعالجة *يحتاج تنصيبها إلى أجهزة ذات قدرات كبيرة.	*بيئة متكاملة لتحليل البيانات وعمليات التنبؤ. *ملائمة للأنشطة التجارية وإدارة الأعمال.	RapidMiner
*حجم البيانات المعالجة محدود.	*تجسيد وتحليل البيانات على الخرائط. *القدرة على معالجة صيغ مختلفة من البيانات في نفس الوقت (جداول، معلومات من مواقع الويب...*)	Google Fusion Tables
	*القدرة على تجميع، تحليل، تجسيد وتمثيل البيانات.	NodeXL