**استخدام تويتر للبيانات الضخمة**

أحد أشكال التقدم والتقنية وأحد أبرز آثارها هو تضخم المعلومات الهائل، ونستطيع القول فعلاً أن البيانات لا تنام ولا تتوقف عن النمو. حسب آخر الإحصائيات زاد عدد مستخدمي شبكة الانترنت في عام 2017 بنسبة 7.5% عن عام 2016 ليصل إلى 3.7 مليار مستخدم. الموقع الاجتماعي تويتر على سبيل المثال، يضم كماً كبيراً من المستخدمين الفاعلين على مدار الساعة، ويتولد عنهم في كل دقيقة ما يقرب 456,000 تغريده تحتوي على نصوص، صور، فيديوهات، روابط وهشتاقات، ولك أن تتخيل كم البيانات الناتج عن ذلك خلال يوم واحد فقط.

**المراحل الرئيسية لعملية تنقيب البيانات**

**المرحلة الأولى: جلب البيانات**

في هذه المرحلة يتم تحديد مصادر البيانات وكيفية دمجها ببعضها في حالة كانت من أكثر من مصدر.

 **المرحلة الثانية: تنظيف البيانات**

تعتبر هذه المرحلة من المراحل المهمة جداً والتي تتضمن تهيئة البيانات لعملية التنقيب. بعض البيانات تحتوي على قيم فارغة، وبعضها يحتوي على أخطاء، ومنها ما هو غير منطقي. في هذه المرحلة يتم التعامل مع هذه المشاكل وحلها.

**المرحلة الثالثة: اختيار البيانات**

في هذه المرحلة يتم اختيار البيانات المهمة في عملية التنقيب، ويتم استبعاد البيانات الغير مهمة، المُكررة والتي ليس لها علاقة. البيانات المكررة أو الغير مهمة ستؤثر على عملية تنقيب البيانات بالتأكيد، لذا يجب الانتباه لها.

**المرحلة الرابعة: تحويل البيانات**

في بعض الأحيان نحتاج إلى تحويل شكل البيانات أو قيمها إلى شكل اخر أو قيم أخرى بهدف جعلها مناسبة للخوارزمية المستخدمة في التنقيب، فبعض الخوارزميات مثلا تتعامل مع قيم رقمية فقط، وفي هذه الحالة سيتم إما استبعاد البيانات النصية في المرحلة الثالثة أو تحويلها لقيم رقمية في هذه المرحلة

**المرحلة الخامسة: تنقيب البيانات**

هنا تتم عملية استخدام أحد الخوارزميات الذكية (مثل شجرة القرار، الشبكات العصبية..الخ) والتي ستقوم بعملية تنقيب البيانات واستخراج المعرفة منها واستخدام المعرفة الناتجة في عملية اتخاذ القرار.

**المرحلة السادسة: التقييم**

في هذه المرحلة يتم استخدام طرق تقييم خاصة لعملية التنقيب، وهدفها قياس درجة دقة العملية التي تمت باستخدام عينات اختبارية من البيانات.

**المرحلة السابعة: المعرفة**

هنا يتم عرض/استخدام النتيجة النهائية لتنقيب البيانات في عملية اتخاذ القرار والاستفادة من المعرفة الناتجة عن ذلك.

المراحل السابقة تشكل مجتمعةً عملية تنقيب البيانات، وهي ليست بالضرورة في اتجاه واحد فقط، نستطيع بعد إتمام أي مرحلة الرجوع لأي مرحلة سابقة وذلك في حال وجدنا أن هناك خلل ما أو عدم دقة في النتائج أو حتى لأغراض التجربة، وهذا الأمر يعطي عملية تنقيب البيانات نوعاً من الديناميكية التي تسهل العمل.