داده کاوی فرایند استخراج داده های مفید یا الگو از داده خام می باشد. فرض کنید یک مقدار مشخص از داده داریم، و به دنبال یک حالت مشخص از یک رویداد درون این داده ها می باشیم. می شود انتظار داشت که این رویداد بخصوص رخ دهد، حتی اگر داده ها کاملا تصادفی باشند. تعداد رویداد ها با افزایش اندازه داده ها، افزایش می یابد. اما این روی داد ها فریبنده هستند؛ زیرا داده های تصادفی همیشه تعدادی خصوصیت غیر عادی دارند که به نظر مهم می آیند اما اینطور نیست. یک تئوری آماری به نام تصحیح بنفرونی1، یک روش آماری ارائه می دهد که می تواند اکثر رویداد های فریبنده را تشخیص دهد. بدون ورود به جزئیات آماری، ما یک روش غیر رسمی بر اساس اصل بنفرونی ارائه می کنیم، که کمک می کند به اشتباه رویداد های تصادفی را پردازش نکنیم. با فرض این که داده ها تصادفی هستند، تعداد رخ دادن رویداد مورد نظر را محاسبه می کنیم. اگر عدد به دست آمده به شکل چشمگیری از تعداد واقعی نمومه ها بیشتر باشد، آنگاه باید انتظار داشت هر چیزی که پیدا شده فریبنده باشد. البته این یک مسئله آماری ساختگی است تا یک مدرک برای پیدا کردن چیزی که به دنبال آن هستیم.

داده کاوی در اینترنت اشیاء جهت مدیریت مقدار زیاد داده هایی که توسط دستگاه های اینترنت اشیاء تولید می شوند، کاربرد دارد. داده کاوی شامل کشف و تجزیه و تحلیل دانش از مجموعه گسترده ای از داده ها است. هدف اصلی از استخراج داده، یافتن الگوهای مفید از مجموعه داده های بزرگ دریافت شده از دستگاه های اینترنت اشیاء (IoT) و سنسورها است.