

البيانات الضخمة ورهان المستقبل .. فهم الواقع الجديد الموهل في الرمزية

Big data and the future bet ..

Understanding the new reality that is deep in symbolism

أ.د/ الحاج سالم عطية

كلية علوم الإعلام والاتصال/ جامعة الجزائر 3

تاريخ الاستلام، 2022/06/14 ؛ تاريخ المراجعة، 2022/06/25 ؛ تاريخ القبول، 2022/06/26

ملخص:

ارتبطت البيانات بالعديد من التخصصات العلمية التي أصبحت تشتغل عليها بشكل لافت، و منها علوم الإعلام والاتصال التي شهدت ميلاد العديد من الأدبيات المساهمة في فهم البيانات، والبيانات الضخمة، إذ قدمت جملة من المنهجيات المتكاملة لفهم كيفية التعامل مع ثروة الثروات ووقود القرارات والاستراتيجيات في المنظمات على اختلافها. انطلاقا من ذلك سوف يتناول المقال جملة من المفاهيم، وتحديدًا تلك التي ترتبط بفهم البيانات الضخمة والاختلافات بينها وبين البيانات، وجملة الخصائص المرتبطة بها، ثم إلى أنواعها والطرق المعتمدة في تحليلها، ووصولًا إلى كيفية الاستفادة منها في دعم قرارات واستراتيجيات المؤسسات العالمية المترتبة على عرش البيانات الضخمة، وذلك من خلال عرض حالتين للدراسة وهما Netflix و Facebook وكيف استفادتا من توظيف البيانات الضخمة لزيادة تنافسيتهما وبالتالي مداخلهما.

الكلمات المفتاحية:

البيانات، البيانات الضخمة، مستقبل المعلومة، Netflix، Facebook .

Abstract:

Data has been linked to many scientific disciplines that it has become remarkably engaged in, including media and communication sciences, which witnessed the birth of many literatures contributing to the understanding of data, and big data, as it provided a set of integrated methodologies to understand how to deal with the wealth of wealth, decision fuels and strategies in organizations at their difference.

Based on that, the article will address a number of concepts, specifically those related to understanding big data and the differences between it and the data, and the set of characteristics associated with it, then to its types and methods adopted in analyzing it, and how to use it in support of the decisions and strategies of global institutions sitting on the throne of data By presenting two case studies, Netflix and Facebook, and how they benefited from employing big data to increase their competitiveness and thus their incomes.

Keywords:

Data, Big Data, The future of information, Netflix, Facebook.

مقدمة:

ارتبطت عمليتنا وضع الإستراتيجية، وصناعة القرار إلى حد كبير بالبيانات بل أن 95 % منهما هي بيانات، ومع التحولات التكنولوجية الأخيرة التي غيرت من طبيعة إنتاج البيانات وحولت هذه الأخيرة إلى طبيعة جديدة لتصبح بيانات ضخمة، تحولت معها معادلي الإستراتيجية وصناعة القرار لتعتمد بالدرجة الكلية على هذا المنتج الجديد "البيانات الضخمة"، من هذا المنطلق ظهرت هناك طرق جديدة للتعامل مع هذا النوع من البيانات، كما سمح ذلك بابتكار طرق جديدة للتحليل، وتوليد المعاني المساعدة على فهم الماضي، الحاضر و المستقبل، لقد مكن ذلك من الاعتماد المتزايد على الثروة الجديدة لزيادة التنافسية، وبالتالي زيادة الإرباح بشكل مذهل إنه التحول نحو الاقتصاد الجديد المبني على الرمزية الضخمة خصوصاً من قبل عمالقة التعاطي مع البيانات (قوقل، أبل، فيسبوك، أمازون) والمعروفة باختصار (GAFA)، لقد برهنت منهجية التعاطي مع البيانات الضخمة بكل ما تحتويه من تكنولوجيات، ومعدات، وموارد بشرية، ودكاء اصطناعي، أن المستقبل لا يمكن أن يخرج عن هيمنة من يتمكنون من البيانات الضخمة.

I. البيانات الضخمة:

تُعرف البيانات الضخمة على أنها مجموعات من مجموعات البيانات التي يكون حجمها أو سرعتها أو تنوعها كبيراً جداً من الصعب تخزينها وإدارتها ومعالجتها وتحليلها باستخدام قواعد البيانات التقليدية وأدوات معالجتها¹.

للبينات الضخمة القدرة على تشغيل الجيل التالي من التطبيقات الذكية التي تزيد من قوة البينات. تمتد تطبيقات البينات الضخمة إلى مجموعة واسعة من المجالات مثل الويب وتجارة التجزئة والتسويق والخدمات المصرفية والمالية والصناعية والرعاية الصحية والبيئية وإنترنت الأشياء².

وقد عرف معهد ماكنزي العالمي سنة 2011 البيانات الضخمة أنها أي مجموعة من البينات التي هي بحجم يفوق قدرة أدوات قواعد البيانات التقليدية من جمع، تخزين، إدارة و تحليل تلك البينات . تنشأ البينات الضخمة عن طريق كل شيء من حولنا وفي كل الأوقات.

كما تعرفها شركة (IBM) بأنها كل عملية رقمية وكل تبادل في وسائل التواصل الاجتماعي ينتج لنا البينات الضخمة، تتناقضها الأنظمة، وأجهزة الاستشعار، والأجهزة النقالة. وللبينات الضخمة مصادر متعددة في السرعة والحجم والتنوع ولكي نستخرج منفعة معنوية من البينات الضخمة نحتاج إلى معالجة مثالية، وقدرات تحليلية ومهارات.

وفقاً لتقدير شركة (IBM)، يتم إنشاء 2.5 كوينتيليون بايت من البينات يومياً.

يقدر تقرير حديث صادر عن DOMO (شركة عالمية متخصصة في التعامل مع البينات الضخمة) كمية البينات التي يتم إنشاؤها كل دقيقة على منصات الإنترنت فيما يلي بعض البينات الأساسية من التقرير³:

- يشارك مستخدمو Facebook ما يقرب من 4.16 مليون منشور من المحتوى،
- يرسل مستخدمو Twitter ما يقرب من 300.000 تغريدة،
- يحب مستخدمو Instagram ما يقرب من 1.73 مليون صورة،
- يقوم مستخدمو YouTube بتحميل 300 ساعة من محتوى الفيديو الجديد،
- يقوم مستخدمو Apple بتنزيل ما يقرب من 51.000 تطبيق،
- يجري مستخدمو سكايب ما يقرب من 110.000 مكالمة جديدة،
- تستقبل أمازون 4.300 زائر جديد،
- يأخذ ركاب Uber ما يقرب من 694 رحلة،
- ييث مشتركو Netflix ما يقرب من 77.000 ساعة من الفيديو .

II. خصائص البيانات الضخمة⁴:

1/ الحجم: البيانات الضخمة هي شكل من أشكال البيانات التي يكون حجمها كبيراً جداً بحيث لا يتناسب مع جهاز واحد، وبالتالي فإن الأدوات والأطر المتخصصة مطلوبة لتخزين العملية وتحليل هذه البيانات. على سبيل المثال تعالج تطبيقات الوسائط الاجتماعية مليارات الرسائل اليومية ، ويمكن للأنظمة الصناعية وأنظمة الطاقة أن تولد تيرابايت من بيانات المستشعر كل يوم، ويمكن للتطبيقات المختلفة معالجة ملايين المعاملات في اليوم ، وما إلى ذلك. فأحجام البيانات التي يتم إنشاؤها بواسطة تكنولوجيا المعلومات الحديثة والصناعية والرعاية الصحية و انترنت الأشياء، وأنظمة أخرى ينمو بشكل كبير، مدفوعاً بانخفاض تكاليف تخزين البيانات ومعالجتها والحاجة إلى استخراج رؤى قيمة من البيانات لتحسين العمليات التجارية والكفاءة والخدمة للمستهلكين. على الرغم من عدم وجود عتبة ثابتة لحجم البيانات التي يجب اعتبارها بيانات ضخمة ، إلا أنه عادةً ما يتم استخدام مصطلح البيانات الضخمة للبيانات التي يصعب تخزينها وإدارتها ومعالجتها باستخدام قواعد البيانات التقليدية وبني معالجة البيانات .

2/ السرعة: تشير سرعة البيانات إلى مدى سرعة إنشاء البيانات. يمكن أن تصل البيانات التي تم إنشاؤها بواسطة مصادر معينة بسرعات عالية جداً ،على سبيل المثال، بيانات الوسائط الاجتماعية أو بيانات الاستشعار. السرعة هي خاصية مهمة أخرى للبيانات الضخمة والسبب الرئيسي للنمو الأسّي للبيانات. تؤدي السرعة العالية لتوليد البيانات إلى أن يصبح حجم البيانات المتراكمة كبيراً جداً في فترة زمنية قصيرة. يمكن أن يكون لبعض التطبيقات مواعيد نهائية صارمة لتحليل البيانات (مثل التداول أو كشف الاحتيال عبر الإنترنت) وتحتاج البيانات إلى التحليل في الوقت الفعلي.

الأدوات المتخصصة مطلوبة لإدخال مثل هذه البيانات عالية السرعة في البنية التحتية للبيانات الضخمة وتحليلها البيانات في

الوقت الفعلي

3/ التنوع: يشير التنوع إلى أشكال البيانات. حيث تأتي البيانات الضخمة بأشكال مختلفة فمنها المنظمة، وغير المنظمة، وشبه المنظمة، إضافة إلى التنوع من حيث شكلها إذ يمكن أن تكون في أشكال مختلفة، فمنها البيانات النصية والصورة والصوت والفيديو وبيانات أجهزة الاستشعار. يجب أن تكون أنظمة البيانات الضخمة مرنة بما يكفي للتعامل مع مثل هذه المجموعة المتنوعة من البيانات

4/ الموثوقية: تشير إلى الصدق وإلى مدى دقة البيانات. فلاستخراج قيمة من البيانات ، يجب تنظيفها لإزالة الضوضاء. يمكن للتطبيقات القائمة على البيانات الضخمة جني فوائد منها عندما تكون هذه الأخيرة مفيدة ودقيقة. لذلك ، تعد عملية تنقية البيانات الضخمة أمراً مهماً بحيث يمكن تصنيفها من البيانات غير الصحيحة أو الخاطئة.

5/ القيمة: تشير قيمة البيانات إلى فائدة البيانات للغرض المقصود. الهدف النهائي لأي نظام لتحليل البيانات الضخمة هو استخراج قيمة من البيانات. ترتبط قيمة البيانات أيضاً بصحة البيانات أو دقتها. بالنسبة لبعض التطبيقات، تعتمد القيمة أيضاً على مدى سرعة قدرتنا على معالجة البيانات.

III. أنواع البيانات الضخمة ومصادرها:

في أي تطبيق أو نظام أساسي لتحليل البيانات الضخمة، قبل معالجة البيانات وتحليلها ، يجب التقاطها من مصادر البيانات الأولية إلى أنظمة وأطر عمل البيانات الضخمة.

تنقسم البيانات الضخمة إلى بيانات ضخمة منظمة أو مهيكلة وهي تشكل حوالي 10% من مجموع البيانات الضخمة في حين أن هناك بيانات ضخمة غير منظمة أو غير مهيكلة وهي التي تشكل الحجم الأكبر منها⁵.

1/ البيانات المنظمة: هي البيانات المخزنة في حقول قواعد البيانات والتي يميزها إمكانية البحث فيها وتحليلها، كما يمكن إدارتها وفق أهداف محددة.

2/ البيانات غير المنظمة: هي كل ما لا يمكن تصنيفه بسهولة كالصور والرسوم البيانية ومقاطع الأصوات والأغاني والفيديو، والنقرات على المواقع، صفحات الويب، ملفات pdf، وسائل البريد الإلكتروني، تغريدات التويتر، ومنشورات الفيسبوك، ورسائل الدردشة... الخ .

ورغم أن هذه الأنواع من الملفات لها هيكل داخلي خاص بها، إلا أنها تعتبر غير منظمة لأن بياناتها لا تتسق في أعمدة موحدة ومناسبة لقاعدة بيانات .

3/ البيانات الهجينة: وهي البيانات التي تجمع بين النوعين السابقين

تتضمن بعض أمثلة مصادر البيانات الضخمة الخام ما يلي⁶:

- **السجلات:** السجلات التي تم إنشاؤها بواسطة تطبيقات الويب و الخوادم والتي يمكن استخدامها لمراقبة الأداء،
- **بيانات المعاملات:** بيانات المعاملات الناتجة عن تطبيقات مثل التجارة الإلكترونية والخدمات المصرفية والمالية،
- **وسائل التواصل الاجتماعي:** البيانات الناتجة عن منصات التواصل الاجتماعي،
- **قواعد البيانات:** البيانات المهيكلة الموجودة في قواعد البيانات العلائقية،
- **بيانات الاستشعار:** بيانات الاستشعار التي تم إنشاؤها بواسطة أنظمة إنترنت الأشياء (IoT)،
- **بيانات النقر على الويب :** البيانات التي يتم إنشاؤها بواسطة تطبيقات الويب والتي يمكن استخدامها لتحليل أنماط تصفح المستخدمين،
- **بيانات المراقبة:** بيانات الاستشعار والصورة والفيديو التي تم إنشاؤها بواسطة أنظمة المراقبة،
- **بيانات الرعاية الصحية:** بيانات الرعاية الصحية التي تم إنشاؤها بواسطة السجل الصحي الإلكتروني (EHR) وتطبيقات الرعاية الصحية الأخرى،
- **بيانات الشبكة:** بيانات الشبكة التي تم إنشاؤها بواسطة أجهزة الشبكة مثل أجهزة التوجيه وجدران الحماية.

IV. تحليل البيانات الضخمة:

1/ مفهوم التحليل:

مصطلح واسع يشمل العمليات والتقنيات والأطر والخوارزميات لاستخراج رؤى ذات مغزى من البيانات. البيانات الخام في حد ذاتها ليس لها معنى حتى يتم وضعها في سياقها ومعالجتها لتتحول إلى معلومات مفيدة. فالتحليل هو عملية استخراج وإنتاج المعلومات من البيانات الأولية عن طريق تصنيفها ومعالجتها وتصنيفها ووضعها في سياقها. ثم يتم تنظيم هذه المعلومات التي تم الحصول عليها وهيكلتها لاستنتاج المعرفة حول النظام و/أو مستخدميه، وبيئته، وعملياته والتقدم نحو أهدافه، مما يجعل الأنظمة أكثر ذكاءً وأكثر فعالية⁷. يعتمد اختيار التقنيات والخوارزميات وأطر التحليل على أهداف التطبيق. على سبيل المثال ، قد تكون أهداف مهمة التحليل:

- التنبؤ بشيء ما (على سبيل المثال ما إذا كانت المعاملة احتيالية أم لا، أو ما إذا كانت ستمطر في يوم معين، أو ما إذا كان الوبم حميداً أو خبيثاً ...)،
- إيجاد أنماط معينة في البيانات (على سبيل المثال العثور على أكثر 10 أيام برودة في العام، والعثور على الصفحات الأكثر زيارة على موقع ويب معين، أو العثور على المشاهير الأكثر بحثاً في عام معين ...).
- إيجاد العلاقات بين البيانات (على سبيل المثال العثور على مقالات إخبارية مماثلة، أو العثور على مرضى مشاهير في نظام السجلات الصحية الإلكترونية، أو العثور على المنتجات ذات الصلة على موقع التجارة الإلكترونية، أو العثور على صور مماثلة، أو إيجاد علاقة بين الأخبار وأسعار الأسهم ...).

2/ أنواع التحليل:

2- 1/ التحليل الوصفي:

تشمل التحليلات الوصفية تحليل البيانات السابقة لتقديمها في شكل موجز يمكن تفسيره بسهولة. تهدف التحليلات الوصفية إلى الإجابة - ماذا حدث؟

جزء كبير من التحليلات التي يتم إجراؤها اليوم هي التحليلات الوصفية من خلال استخدام وظائف الإحصاء مثل الأعداد، الحد الأقصى، الحد الأدنى، المتوسط، النسبة المئوية ...، وتساعد هذه الإحصائيات في وصف الأنماط في البيانات وتقديم البيانات في شكل موجز، على سبيل المثال حساب العدد الإجمالي للإعجاب لمنشور معين أو حساب متوسط هطول الأمطار الشهري أو إيجاد متوسط عدد الزوار شهرياً على موقع ويب ...، فالتحليلات الوصفية مفيدة في تلخيص البيانات⁸.

2- 2/ تحليل التشخيصي:

تشمل التحليلات التشخيصية تحليل البيانات السابقة لتشخيص أسباب حدوث أحداث معينة. تهدف التحليلات التشخيصية إلى الإجابة - لماذا حدث ذلك؟

ومثال ذلك نظام يجمع ويحلل بيانات المستشعر من الأجهزة لمراقبة صحتها والتنبؤ بالأعطال الممكنة. يمكن أن توفر التحليلات التشخيصية مزيداً من الأفكار حول سبب حدوث خطأ معين بناءً على الأنماط في بيانات المستشعر للأعطال السابقة⁹.

2- 3/ التحليل التنبؤي:

تشمل التحليلات التنبؤية التنبؤ بحدوث حدث أو النتيجة المحتملة لحدث ما أو التنبؤ بالقيم المستقبلية باستخدام نماذج التنبؤ. تهدف التحليلات التنبؤية إلى الإجابة - ما الذي قد يحدث على الأرجح؟

على سبيل المثال، يمكن استخدام التحليلات التنبؤية للتنبؤ بوقوع عطل ما في الجهاز، والتنبؤ بما إذا كان الورم حميداً أم خبيثاً، أو التنبؤ بحدوث الطوارئ الطبيعية (أحداث مثل حرائق الغابات أو فيضانات الأنهار) أو التنبؤ بمستويات التلوث. يتم إجراء التحليلات التنبؤية باستخدام النماذج التنبؤية التي يتم اختبارها بواسطة البيانات الموجودة. تتعلم هذه النماذج الأنماط والاتجاهات من البيانات الموجودة والتنبؤ بحدوث حدث أو النتيجة المحتملة لنماذج تصنيف الحدث) أو أرقام التنبؤ (نماذج الانحدار). تعتمد دقة نماذج التنبؤ على جودة وحجم البيانات الحالية المتاحة لاختبار النماذج، بحيث يمكن تعلم جميع الأنماط والاتجاهات في البيانات الحالية بدقة¹⁰.

2- /4 التحليل التقديري:

بينما تستخدم التحليلات التنبؤية نماذج التنبؤ للتنبؤ بالنتيجة المحتملة لحدث ما، تستخدم التحليلات التقديرية نماذج تنبؤ متعددة للتنبؤ بالنتائج المختلفة وأفضل مسار للعمل لكل نتيجة. تهدف التحليلات التقديرية إلى الإجابة - ما الذي يمكننا فعله لتحقيق ذلك؟

يمكن أن تتنبأ التحليلات التقديرية بالنتائج المحتملة بناءً على الاختيار الحالي للإجراءات. يمكننا اعتبار التحليلات التقديرية نوعاً من التحليلات تستخدم نماذج تنبؤ مختلفة لمدخلات مختلفة. تصف التحليلات التقديرية الإجراءات أو الخيار الأفضل لمتابعة الخيارات المتاحة. على سبيل المثال، يمكن استخدام التحليلات التقديرية لوصف أفضل دواء لعلاج المريض بناءً على نتائج الأدوية المختلفة لمريض مشابهين¹¹.

V. مراحل استغلال البيانات الضخمة:

1/ جمع البيانات:

جمع البيانات هو الخطوة الأولى لأي تحليل. قبل أن يتم تحليل البيانات، يجب جمع البيانات واستيعابها في مكس بيانات ضخمة. يعتمد اختيار الأدوات والأطر لجمع البيانات على مصدر البيانات ونوع البيانات التي يتم استيعابها. لجمع البيانات، يمكن استخدام أنواع مختلفة من الموصلات مثل أطر عمل مراسلة النشر، وقوائم انتظار الرسائل، وموصلات مصدر المصدر، وموصلات قاعدة البيانات والموصلات المخصصة¹².

2/ تحضير البيانات:

غالبًا ما تكون البيانات قذرة ويمكن أن تحتوي على العديد من المشكلات التي يجب حلها قبل معالجتها، مثل السجلات الفاسدة والقيم المفقودة والتكرارات والاختصارات غير المتسقة والوحدات غير المتسقة والأخطاء الإملائية والتهجئة غير الصحيحة والتنسيق غير الصحيح. تتضمن خطوة إعداد البيانات مهامًا مختلفة مثل تنقية البيانات، أو تعديلها، وإزالة الازدواجية، والتطبيع، وأخذ العينات، والتصفية. يقوم تنظيف البيانات باكتشاف وحل مشكلات مثل السجلات الفاسدة، والسجلات ذات القيم المفقودة، والسجلات ذات التنسيق السيئ، على سبيل المثال. يتعامل التغيير في البيانات مع تحويل البيانات من تنسيق خام إلى تنسيق آخر. على سبيل المثال، عندما نجمع السجلات كملفات نصية أولية تشكل مصادر مختلفة، فقد نواجه تناقضات في فواصل الحقول المستخدمة في ملفات مختلفة. قد يستخدم بعض الملفات الفاصلة كفاصل للحقول، بينما قد يستخدم البعض الآخر علامة التبويب كفاصل للحقول. يعمل تحضير البيانات على حل هذه التناقضات عن طريق تحليل البيانات الأولية من مصادر مختلفة وتحويلها إلى تنسيق واحد ثابت. تكون التسوية مطلوبة عندما تستخدم البيانات من مصادر مختلفة ووحدات أو مقاييس مختلفة أو لها اختصارات مختلفة لنفس الشيء. على سبيل المثال، قد

تحتوي بيانات الطقس التي أبلغت عنها بعض المحطات على درجة حرارة بمقياس مئوية بينما قد تستخدم البيانات من محطات أخرى مقياس "فهرنهايت". قد تكون التصفية وأخذ العينات مفيدة عندما نريد معالجة البيانات التي تلي قواعد معينة فقط. يمكن أن تكون التصفية مفيدة أيضًا لرفض السجلات التالفة ذات القيم غير الصحيحة أو خارج النطاق¹³.

3/ اختيار نوع التحليل:

تتمثل الخطوة التالية في تدفق التحليل في تحديد نوع التحليل المناسب. -لقد قمنا بوصف أنواع التحليل فيما سبق- فهناك خيارات مختلفة لأنواع التحليل والخوارزميات الشائعة لكل نوع من أنواع التحليل.

4/ طبيعة التحليل:

مع تحديد أنواع التحليل، فإن الخطوة التالية هي تحديد طبيعة التحليل، والذي يمكن أن يكون إما دفعة، أو في الوقت الفعلي، أو بشكل تفاعلي. يعتمد اختيار طبيعة التحليل على متطلبات التطبيق. إذا طلب تطبيقك تحديث النتائج بعد فترات زمنية قصيرة (كل بضع ثوانٍ)، فسيتم اختيار وضع التحليل في الوقت الفعلي. أما إذا كان تطبيقك يتطلب فقط إنشاء النتائج وتحديثها على نطاقات زمنية أكبر (على سبيل المثال يوميًا أو شهريًا)، فيمكن استخدام وضع الدفعات. بينما إذا كان تطبيقك يتطلب مرونة في الاستعلام عن البيانات عند الطلب، فإن الوضع التفاعلي مفيد¹⁴.

5/ العرض التصوري للبيانات:

يعتمد اختيار أدوات التصوير وقواعد البيانات التي تخدم وأطر الويب على متطلبات التطبيق. يمكن أن تكون المرئيات ثابتة أو ديناميكية أو تفاعلية. يتم استخدام الصور الثابتة عندما يكون لديك نتائج التحليل مخزنة في قاعدة بيانات معروضة وتريد ببساطة عرض النتائج. ومع ذلك، إذا طلب تطبيقك تحديث النتائج بانتظام، فستحتاج إلى صور ديناميكية (مع عناصر واجهة مستخدم مباشرة أو مخططات أو أجهزة قياس). إذا كنت تريد أن يقبل التطبيق الخاص بك مدخلات من المستخدم ويعرض النتائج، فستحتاج إلى صور تفاعلية¹⁵.

VI. دراسات حالة:

1/ دراسة حالة Netflix:

يقال إن خدمة بث الأفلام والتلفزيون Netflix تمثل ثلث حركة مرور الإنترنت في وقت الذروة في الولايات المتحدة، وتضم الخدمة الآن 65 مليون عضو في أكثر من 50 دولة يتمتعون بأكثر من 100 مليون ساعة من البرامج التلفزيونية والأفلام يوميًا. يتم جمع البيانات من هؤلاء الملايين من المشتركين ومراقبتها في محاولة لفهم عادات المشاهدة لدينا. لكن بيانات Netflix ليست "كبيرة" بالمعنى الحرفي. إن الجمع بين هذه البيانات والتقنيات التحليلية المتطورة هو ما يجعل Netflix شركة بيانات ضخمة حقيقية¹⁶.

قال كاتب السيناريو الأسطوري في هوليوود ويليام جولدلمان: "لا أحد، لا أحد - ليس الآن، ولا أبدًا - يعرف ما يمكن أن ينجح أو لا ينجح في شبك التذاكر. كان يتحدث قبل وصول الإنترنت والبيانات الضخمة، ومنذ ذلك الحين، صممت Netflix على إثبات خطئه من خلال بناء نشاط تجاري حول التنبؤ بالضبط بما سنستمتع بمشاهدته¹⁷.

يكفي إلقاء نظرة سريعة على صفحة وظائف Netflix لإعطائك فكرة عن مدى جدية التعامل مع البيانات والتحليلات. يتم تعيين المتخصصين للانضمام إلى فرق ماهرة على وجه التحديد في تطبيق المهارات التحليلية على مجالات عمل معينة: التحليل المتخصص، وتحليل الرسائل، وتحليل تسليم المحتوى، وتحليل الجهاز. . . والقائمة تطول. ومع ذلك، على الرغم من استخدام البيانات الضخمة في كل جانب من جوانب أعمال Netflix، إلا أن هدفهم المقدس دائمًا هو التنبؤ بما سيستمع العملاء بمشاهدته. تحليل البيانات الضخمة هي الوقود الذي يطلق "محركات التوصية" المصممة لخدمة هذا الغرض.

بدأت الجهود هنا في عام 2006، عندما كانت الشركة لا تزال في الأساس تعمل في مجال إرسال البريد عن طريق "DVD". لقد أطلقوا جائزة Netflix، حيث قدموا مليون دولار للمجموعة التي يمكنها التوصل إلى أفضل خوارزمية للتنبؤ بكيفية تقييم عملائهم لفيلم بناءً على تقييماتهم السابقة. تم الإعلان أخيرًا عن الفائزة في المسابقة في عام 2009، وعلى الرغم من مراجعة الخوارزميات وإضافتها باستمرار، إلا أن المبادئ لا تزال عنصرًا رئيسيًا في محرك التوصية.

في البداية، كان المحللون مقيدون بسبب نقص المعلومات التي لديهم عن عملائهم (كانت أربع نقاط بيانات فقط: معرف العميل ومعرف الفيلم والتصنيف وتاريخ مشاهدة الفيلم متاحة للتحليل)، وبمجرد أن أصبح التدفق هو طريقة التسليم الأساسية، أصبح الوصول إلى العديد من نقاط البيانات الجديدة من عملائها. مكنت هذه البيانات الجديدة Netflix من بناء نماذج للتنبؤ بحالة العاصفة المثالية للعملاء الذين يتم إفادتهم باستمرار بأفلام سيستمعون بها. والتي من المرجح أن تسعد العملاء وتجعلهم يجددون اشتراكاتهم.

عنصر مركزي آخر في محاولة Netflix لمنحنا أفلامًا سنستمتع بها هو وضع العلامات. حيث تدفع الشركة للأشخاص مقابل مشاهدة الأفلام، ثم تأخذ بالحسبان توصيفهم لهذه الأفلام.

سيقترحون عليك بعد ذلك مشاهدة المنتجات الأخرى التي تم وضع علامة عليها بشكل مشابه لتلك التي استمتعت بها. هذا هو المكان الذي تأتي منه "الاقتراحات" غير العادية في بعض الأحيان. في الواقع، حددت Netflix بشكل فعال ما يقرب من 80.000 "نوع صغير" جديد من الأفلام بناءً على عادات المشاهدة¹⁸.

تحركت Netflix في الآونة الأخيرة نحو وضع نفسها كمنشئ محتوى، وليس مجرد طريقة توزيع لاستوديوهات الأفلام والشبكات الأخرى. كانت إستراتيجيتهم هنا أيضًا مدفوعة بقوة ببياناتهم - والتي أظهرت أن مشتركهم لديهم شهية للمحتوى الذي أخرجه ديفيد فينشر وبطولة كيفن سيسي. بعد المزايدة للحصول على حقوق House of Cards، كانوا واثقين جدًا من أنها تناسب نموذجهم التنبؤي لـ "العرض التلفزيوني المثالي" لدرجة أنهم خالفوا العرف لإنتاج نموذجي والتأليف الفوري لمسلسل من موسمين يتألفان من 26 حلقة. مع أن كل جانب من جوانب الإنتاج كان تحت رقابة بيانات Netflix بما فيها مجموعة الألوان المستخدمة في صورة الغلاف للمسلسل التي تم تحديدها لجذب المشاهدين¹⁹.

المقياس النهائي الذي تأمل Netflix في تحسينه هو عدد الساعات التي يقضيها العملاء في استخدام خدماتهم. لا تحتاج حقًا إلى إحصاءات لإخبارك بأن المشاهدين الذين لا يقضون وقتًا طويلاً في استخدام الخدمة من المحتمل أن يشعروا أنهم لا يحصلون على قيمة مقابل المال من اشتراكهم، وبالتالي قد يلغوا اشتراكهم.

ولتحقيق هذه الغاية، تتم مراقبة الطريقة التي تؤثر بها العوامل المختلفة على "جودة التجربة" عن كثب ويتم تصميم النماذج لاستكشاف كيفية تأثير ذلك على سلوك المستخدم. من خلال جمع بيانات المستخدم النهائي حول كيفية تأثير الموقع الفعلي للمحتوى على تجربة المشاهد، يمكن إجراء حسابات حول وضع البيانات لضمان وجود خدمة مثالية لأكثر عدد ممكن من المنازل.

1-1/ البيانات المستعملة:

يتم تغذية خوارزميات التوصيات وقرارات المحتوى ببيانات حول العناوين التي يشاهدها العملاء، ووقت مشاهدة الأفلام في اليوم، والوقت الذي يقضيه اختيار الأفلام، وعدد مرات إيقاف التشغيل (إما من قبل المستخدم أو بسبب قيود الشبكة) والتصنيفات المقدمة. من أجل تحليل جودة التجربة، تجمع Netflix بيانات عن التأخيرات الناتجة عن التخزين المؤقت (معدل إعادة التخزين المؤقت) ومعدل البت (الذي يؤثر على جودة الصورة)، بالإضافة إلى موقع العميل.

2-1/ النتائج المحققة:

تُظهر رسالة Netflix إلى المساهمين في أبريل 2015 أن إستراتيجية البيانات الضخمة كانت تؤتي ثمارها. لقد أضافوا 4.9 مليون مشترك جديد في الربع الأول من عام 2015، مقارنة بأربعة ملايين في نفس الفترة من عام 2014. وقد أرجعت Netflix الكثير من هذا النجاح إلى "المحتوى المتطور باستمرار"، بما في ذلك House of Cards و Orange is the New Black. هذا المحتوى الأصلي يقود إلى اكتساب أعضاء جدد والاحتفاظ بالعملاء السابقين²⁰.

في الواقع، شارك 90٪ من أعضاء Netflix في هذا المحتوى الأصلي. من الواضح أن قدرتهم على التنبؤ بما سيستمتع به المشاهدون هي جزء كبير من هذا النجاح. وماذا عن مقياسهم النهائي: كم عدد الساعات التي يقضيها العملاء في استخدام الخدمة؟ حسنًا، في الربع الأول من عام 2015 وحده، بث أعضاء Netflix 10 مليارات ساعة من المحتوى. إذا استمرت إستراتيجية البيانات الضخمة لشركة Netflix في التطور، فسيتم دفع هذا الرقم إلى الزيادة²¹.

VII. دراسة حالة Facebook:

فايسبوك Facebook، مع 1.5 مليار مستخدم نشط شهريًا (إحصائيات 2016)، لديه إمكانية الوصول إلى بيانات عن المستخدمين أكثر بكثير من أي شخص آخر. كما يمكن من خلال Facebook أيضا تتبع زيارات صفحة الويب الخاصة بنا واستنتاج الكثير من عادات التصفح لدينا، غالبًا ما يتمتع Facebook بإمكانية الوصول الكامل إلى البيانات الديموغرافية المباشرة المتعلقة بنا مثل أين نعيش ونعمل ونلعب وعدد الأصدقاء لدينا وما نفعله في أوقات فراغنا والأفلام والكتب والموسيقى المعينين الذين نحبهم²².

يمكن لناشري الكتب، على سبيل المثال، الدفع ل Facebook لوضع إعلاناتهم أمام مليون شخص يحبون الكتب المماثلة، ومطابقة الملفات الديموغرافية لعملائهم. تُستخدم البيانات التي يتم جمعها من قبل المستخدمين أثناء تصفحهم Facebook لمطابقتها مع الشركات التي تقدم منتجات وخدمات من المحتمل أن يهتموا بها إحصائيًا. يحتفظ Facebook بلا شك بوحدة من أكبر وأشمل قواعد بيانات المعلومات الشخصية التي تم جمعها على الإطلاق، ويتوسع كل ثانية من كل يوم

بالإضافة إلى كونه نظام أساسي لمشاركة الرسائل، يعد Facebook أيضًا منصة لتشغيل البرامج. تم إنشاء أكثر من نصف مليون تطبيق ل Facebook حتى الآن، يستفيد معظمه من الوصول إلى Facebook عبر واجهات برمجة التطبيقات (واجهات برامج التطبيق) الشاملة لبيانات المستخدم. تجمع هذه التطبيقات بدورها بيانات حول كيفية استخدامها التي يستخدمها مطوروها لتوجيه الإعلانات إلى عملائهم.

يتوسع Facebook أيضًا عن طريق شراء شركات وخدمات أخرى وإضافة بياناتها إلى بياناتها الخاصة. في السنوات الأخيرة، استحوذت الشركة على خدمات Instagram و WhatsApp، ووضع المزيد من البيانات حول كيفية مشاركة الصور والرسائل الفورية تحت تصرفهم.

2-1/ البيانات المستعملة:

يقوم Facebook، جنبًا إلى جنب مع مستخدميه، بإنشاء بياناته الخاصة. يقوم المستخدمون بتحميل 2.5 مليون قطعة من المحتوى كل دقيقة. يتم تحليل هذا المحتوى بحثًا عن أدلة عنا يمكن استخدامها لتقسيمنا بحسب حاجة المعلنين.

بالإضافة إلى ذلك، فهم يتفاعلون مع محتوى الأشخاص الآخرين وكذلك البيانات المخزنة في قواعد بيانات Facebook الخاصة، والتي تشمل قوائم الأعمال وقواعد بيانات الأفلام والموسيقى والكتب والعروض التلفزيونية. عندما "نعجب" بالمحتوى ونشاركه، فإنه يتعلم المزيد عنا²³.

2-2/ النتائج المحققة:

أدى أسلوب Facebook المتمثل في الاستفادة من ثروته الهائلة من بيانات المستهلك لبيع المساحات الإعلانية إلى حصولهم على حصة 24٪ من سوق الإعلانات المصورة عبر الإنترنت في الولايات المتحدة في عام 2014، وتحقيق 5.3 مليار دولار من العائدات من مبيعات الإعلانات. وبحلول عام 2017، كان من المتوقع أن تصل هذه النسبة إلى 27٪، بقيمة تزيد عن 10 مليار دولار²⁴.

خلاصة:

يبدو جليا أن رهان المستقبل لن يرتبط بالمعلومات المتجانسة التي نشتغل عليها اليوم بقصد إنتاج معاني ودلالات قادرة على توجيه السلوك، لقد تغيرت طرق التعاطي مع الواقع باعتبار أن هذا الأخير نفسه قد تغير، فلم تعد الطرق التقليدية قادرة على فك شفراته، وبالتالي فإن توجيه السلوك مستقبلا يدعو إلى الاجتهاد من أجل التحكم في ضخامة الرموز المنتجة يوميا من قبل كل ما هو موجود حولنا، لقد أضحت كل الأشياء قادرة على إنتاج بيانات لا يمكن التغافل عنها إذا ما أردنا التصرف بعقلانية مستقبلا.

الهوامش والإحالات:

- ¹Arshdeep Bahga & Vijay Madisetti , Big Data Analytics: A Hands-On Approach
- Wileys 2019 p.25.
- ² I bid.p.25.
- ³ I bid.p.25.
- ⁴ Sridhar Alla, Big Data Analytics with Hadoop 3, Packt Publishing,2018,p.44.
- ⁵ Arshdeep Bahga & Vijay Madisetti, op,cit,p..
- ⁶ Arshdeep Bahga & Vijay Madisetti, op,cit,p.39.
- ⁷ Peter Ghavami, Big Data Analytics Methods , Deutsche Nationalbibliothek, 2nd edition, 2020, p.13.
- ⁸ Bart Baesens. Analytics in a big data world : the essential guide to data science and its applications, Wiley, 2014, p .85.
- ⁹ Arshdeep Bahga & Vijay Madisetti, op,cit,p.35.
- ¹⁰ Bart Baesens, op,cit,p.35..
- ¹¹ Arshdeep Bahga & Vijay Madisetti, op,cit,p.35.
- ¹². I bid.p.36.
- ¹³ I bid.p.36.
- ¹⁴ I bid.p.36.
- ¹⁵ I bid.p.36.
- ¹⁶ Bernard Marr, Big Data In Practice How 45 Successful Companies Used Big Data Analytics To Deliver Extraordinary Results, 2016, p.17.
- ¹⁷ I bid.p.17.
- ¹⁸ I bid.p.22.
- ¹⁹ I bid.p.19.
- ²⁰ I bid.p.20.
- ²¹ I bid.p.20.
- ²² I bid.p. 69.
- ²³ I bid.p.70.
- ²⁴ I bid.p.71.