Syrian Arab Republic
Ministry of Higher Education
Syrian Virtual University

الجمهورية العربية السورية
وزارة التعليم العالي
الجامعة الافتراضية السورية

أثر تحليلات البيانات الضخمة في كفاءة الحملة التسويقية

دراسة حالة شركة سيريتل

The impact of Big Data analytics on the efficiency of marketing campaign

Case study: SyriaTel

بحث مقدم لنيل درجة ماجستير إدارة الأعمال التخصصي MBA

إعداد الطالب

ميلاد حبيب بدران

إشراف الدكتور

مضر دخان

إهداء

إلى من مهد لي طريق العلم وكافح لينير لي دربي .. والدي الحبيب

إلى القلب الناصع بالبياض والعطاء اللامتناهي .. والدتي الحنونة

إلى خير من أشدد به أزري وأشركهم في أمري سندي وعزوتي .. إخوتي

شكر وتقدير

يطيب علي عرفاناً بالجميل أن أتقدم بجزيل الشكر والعرفان إلى كل من وقف إلى جانبي خلال فترة دراستي هذه، وأخص بالذكر أستاذي الكريم المشرف على الدراسة الدكتور مضر دخان الذي قدم لي الدعم والتوجيه والإرشاد حتى أبصر هذا العمل النور

وأتقدم بوافر التقدير وعظيم الامتنان للجنة المناقشة الأفاضل الذين شرفوني بمناقشة المشروع وعلى دورهم الكبير في إثراءه بملاحظاتهم وتوجيهاتهم.

كما أتقدم بالشكر الجزيل إلى كل من ساهم من قريب أو بعيد في هذه الدراسة .

المحتويات

المقدمة: 7

1.1– الدراسات السابقة: 7

2.1– مشكلة البحث (Research problem): 12

3.1– أهداف البحث (Research Objectives): 13

4.1– نموذج البحث: (Search Model): 13

5.1– فرضيات البحث: (Research Hypothesis): 14

6.1– أهمية البحث: (Research Importance): 14

7.1– أدوات جمع البيانات: 15

8.1– أساليب تحليل البيانات الإحصائية: 15

9.1– مجتمع وعينة البحث(Research Population and sample) : 16

10.1– حدود البحث (Research Limit): 16

الفصل الثاني: الإطار النظري للبحث 17

البيانات الضخمة وأهميتها: 17

1.2– تعاريف: 17

2.2– دورة حياة البيانات الضخمة: 17

3.2– خصائص البيانات الضخمة: 18

4.2– سلسلة القيمة للبيانات الضخمة: 21

5.2– تحديات البيانات الضخمة: 23

6.2– أدوات تحليل البيانات الضخمة: التقنيات والأساليب: 25

7.2– الأمن والخصوصية في البيانات الضخمة: 29

8.2– دور Hadoop في البيانات الضخمة: 32

9.2– تأثير تحليلات البيانات الضخمة على الحملة التسويقية في شركات الاتصالات: 33

10.2– الخاتمة: 37

الفصل الثالث: القسم العملي 38

1.3– أنواع البيانات ومصادرها في شركة سيريتل: 38

2.3– تأثير تحليلات البيانات الضخمة على الحملة التسويقية في شركة SyiaTel: 39

3.3– مناقشة الفرضيات من معطيات الشركة: 40

1.3.3–الفرضية الأولى: يوجد تأثير لسرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة: 40
1- ما هي ميزات تحليلات البيانات الضخمة وكيف تتم عملية تخزين الملفات باستخدام منصة البيانات الكبيرة: 41
2- كيف أثر تطبيق تقنية البيانات الضخمة في سرعة تحليل البيانات وانعكاس ذلك على الحملة التسويقية في الشركة؟ 43
3- ماهي النتائج المباشرة لسرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية؟ 43
2.3.3–يوجد تأثير لمعالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة: 44
2- ماهي النتائج قبل تطبيق التقنية وبعد تطبيق التقنية؟ 46
3- كيف رفع ذلك من كفاءة الحملة التسويقية وماهي النتائج؟ 46
4- ماهي النتائج المباشرة لاستخراج سمات أساسية من البيانات في الرفع من كفاءة الحملة التسويقية؟ 46
3.3.3–يوجد تأثير لانخفاض تكلفة تطبيق هذه التقنية مقارنة بتقنيات أخرى على كفاءة الحملة التسويقية في الشركة محل الدراسة: 47
1- ما هي تكلفة تطبيق تقنية البيانات الضخمة في تحليل البيانات؟ 48
2- ما هو الفرق بين تكلفة تطبيق تقنية البيانات الضخمة مقارنة بالطرق التقليدية في تحليل البيانات؟ 48
كيف رفع ذلك من كفاءة الحملة التسويقية؟ 48
4.3– اختبار الفرضيات من منظور الزبائن: 49
توصيف المتغيرات الديمغرافية للاستبيان الذي أجراه الباحث على زبائن الشركة: 49
اختبار ثبات الاستبانة: 50
1.4.3–الفرضية الأولى: يوجد تأثير معنوي لسرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة: 51
الاحصاءات الوصفية للمتغير الأول "تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية": 51
مقارنة نتيجة الفرضية الأولى بين منظور الشركة ومنظور والزبائن: 53
2.4.3–يوجد تأثير لمعالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة: 54
الإحصاءات الوصفية للمتغير الثاني "تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية". 54
مقارنة نتيجة الفرضية الثانية بين منظور الشركة ومنظور والزبائن: 56
النتائج: 58
التوصيات: 60
المراجع: 62
الملحق (1): 65
الملحق (2): 66

قائمة الأشكال:

الشكل 1: نموذج الدراسة 13
الشكل 2: توقع نمو حجم البيانات حسب السنة بـ زيتا بايات 19
الشكل 3: نمو تنوع البيانات حسب السنوات 19
الشكل 4: سلسلة القيمة للبيانات الكبيرة 22
الشكل 5: تقنيات وأدوات تحليل البيانات الكبيرة 29
الشكل 6: مقارنة قدرة هادوب مع قواعد البيانات التّقليديّة 32
الشكل 7: مقارنة بين طرق حفظ البيانات في شركة SyriaTel 42
الشكل 8: تحليل شبكة التواصل الاجتماعي باستخدام البيانات الكبيرة 45
الشكل 9: إضافة سمات تابعة لـ 9 أشهر مع سمات الشبكة الاجتماعية باستخدام منصة البيانات الكبيرة 46

قائمة الجداول:

جدول (1): أوجه الاختلاف بين دراسة الباحث والدراسات السابقة 11
جدول (2): توزيع أفراد العينة حسب الجنس 49
جدول (3): توزيع أفراد العينة حسب العمر 49
جدول (4): توزيع أفراد العينة حسب العمر 50
جدول (5): معامل ألفا كرونباخ لقياس ثبات الاستبانة 51
جدول (6): تّوصيف اّلمتغير اّلمستقل اّلأول: تأثير سرعة التعامل مع البيانات في الرفع من كفاءة الحملة التسويقية 51
جدول (7): "متوسط اجابات أفراد العينة على محور "تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية" 52
جدول (8): نتائج اختبار الفرضية الأولى الخاصة بمحور تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية 53
جدول (9): توصيف المتغير المستقل الثاني: تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية 54
جدول (10): "متوسط اجابات أفراد العينة على محور "تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية" 55
جدول (11): نتائج اختبار الفرضية الثانية الخاصة بمحور تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية 56

الفصل الأول الإطار العام للبحث

المقدمة:

تعتبر ظاهرة "البيانات الضخمة " موجودة الآن في كل قطاع من قطاعات الاقتصاد العالمي. غالبًا ما ترتبط أطر التعاون المعاصرة بكمية هائلة ومتنامية من أنواع متعددة من البيانات، والتي تختلف من حيث الأهمية والموضوعية والأهمية. قد تتراوح المعرفة المستخرجة من الآراء الفردية إلى الممارسات المقبولة على نطاق واسع. تواجه أعمال اليوم تحديات ليس فقط في إدارة البيانات ولكن في تحليل البيانات الضخمة ، الأمر الذي يتطلب مناهج جديدة للحصول على رؤى من محتويات مفصلة للغاية ومضمنة وغنية في السياق.

أصبح قطاع الاتصالات منذ عام 1990 واحداً من المجالات الرئيسة للتّنمية في الدّول الصناعية، وقد صاحبت هذه الأهمية زيادة ملحوظة في الدّراسات المنشورة، وبشكلٍ أكثر تحديداً فيما يتعلّق بمجال استراتيجيات التّسويق. تقوم شركات الاتصالات باستثمار موارد كبيرة لتوفير منتج أو خدمة متميّزة عن المنافسين، وذلك لغرض اكتساب زبائن جدد، ومع ذلك فإنّ التطوير المستمرّ للمنتج نفسه غير كاف، لذلك تتبنّى العديد من الشركات استراتيجيات جديدة في العمل للبقاء في السوق التنافسيّة.

يندرج هذا العمل ضمن إطار علوم البيانات الضخمة وتحليلها الذي يهدف إلى استخلاص المعرفة انطلاقاً من تحليل البيانات التي من الصعب تخزينها، إدارتها وتحليلها باستخدام التقنيات التقليدية للبرمجيات وقواعد البيانات، حيث تعتمد عملية التحليل عل تكنولوجيا البيانات الضخمة لاستخلاص سمات محددة للزبائن تستطيع الشركة من خلالها تحديد توجهات ورغبات الزبائن للتوجه إليهم بحملات تسويقية تتناسب وتلك الرغبات والتوجهات.

1.1- الدراسات السابقة:

1- دراسة Ahmad et al (2019) بعنوان:

" Customer churn prediction in telecom using machine learning in big data platform"

"توقع تسرب الزبائن من شركات الاتصالات باستخدام تعلم الآلة في بيئة بيانات كبيرة"

قامت هذه الدراسة بتطوير نموذج للتنبؤ بتسرب الزبائن يساعد مشغلي الاتصالات على التنبؤ بالزبائن الذين من المرجح أن يتسربوا إلى شركة أخرى. يستخدم النموذج الذي تم تطويره في هذا العمل تقنيات تعلم الآلة على منصة البيانات الضخمة ويبني طريقة جديدة لميزات الهندسة والاختيار. كما وضحت الدراسة نقطة رئيسية أخرى هي استخدام شبكة العملاء الاجتماعية في نموذج التنبؤ من خلال استخراج ميزات تحليل الشبكة الاجتماعية (SNA).. تم إعداد النموذج واختباره

ضمن بيئة Spark من خلال العمل على مجموعة بيانات كبيرة تم إنشاؤها عن طريق تحويل البيانات الخام الضخمة المقدمة من شركة SyriaTel للاتصالات. احتوت مجموعة البيانات على جميع بيانات العملاء على مدار 9 أشهر، وتم استخدامها لتدريب واختبار وتقييم النظام في SyriaTel. جرب النموذج أربع خوارزميات: شجرة القرار، والغابات العشوائية ، وشجرة الماكينات المعززة للتدرج "GBM" ، وتعزيز التدرج الشديد "XGBOOST". ومع ذلك، تم الحصول على أفضل النتائج من خلال تطبيق خوارزمية XGBOOST. تم استخدام هذه الخوارزمية للتصنيف في هذا النموذج التنبئي لتسرب الزبائن.

2- دراسة Janssen et al (2017) بعنوان:

"Factors influencing big data decision-making quality"

"العوامل المؤثرة في جودة صنع القرار في البيانات الكبيرة"

حيث قام جانسن وآخرون بإجراء دراسة حول تأثير البيانات الضخمة على اتخاذ القرار ، أظهر هذا البحث الذي هو عبارة عن دراسة حالة أن قناعة المنظمات بأهمية الاستفادة من البيانات الضخمة هي عملية تطورية يؤدي فيها التفهم التدريجي لإمكانات البيانات الضخمة والقيمة المضافة لها، وروتينية عمليات جمع وتخزين وتحليل البيانات الضخمة دورا حاسما في الإفادة منها.

3- دراسة (Bughin, 2016) بعنوان:

"Reaping the benefits of big data in telecom"

"جني فوائد البيانات الضخمة في شركات الاتصالات"

قامت الدراسة بجمع حالات استخدام البيانات الضخمة لعينة تمثيلية من شركات الاتصالات في جميع أنحاء العالم، ولوحظ توزيعًا واسعًا ومتساويًا لعائدات البيانات الضخمة ، مع قيام عدد قليل من الشركات بالإبلاغ عن تأثير كبير من نطاق واسع من شركات الاتصالات ذات عائدات المحدود. استخدمت الدراسة نموذج مشترك للتبني والعودة إلى التبني ، ووجدت أن انحراف التوزيع ينشأ من عدد قليل من شركات الاتصالات القادرة على اتباع الممارسات الإدارية والتنظيمية للبيانات الضخمة الرئيسية. ووجدت أيضًا أن عوائد البيانات الضخمة تظهر وفورات في نطاقها ، وتقلص العوائد إلى الحجم ، في حين أن قدرات البيانات الضخمة مكملة لاستثمارات رؤوس الأموال الكبيرة. تهدف هذه الدراسة إلى فهم كيفية اعتماد تكنولوجيات البيانات الضخمة ، وكيف يتم الحصول على عوائد منها لصناعة الاتصالات ، وذلك باستخدام مسح سري على مشغلي الاتصالات في جميع أنحاء العالم.

4- دراسة Stryk (2015) بعنوان:

" HOW DO ORGANIZATIONS PREPARE AND CLEAN BIG DATA TO ACHIEVE BETTER DATA "GOVERNANCE? A DELPHI STUDY

"كيف تقوم المنظمات بتحضير ومسح البيانات الضخمة لتحقيق أفضل حوكمة بيانات؟ دراسة دلفي"

فقد تناولت موضوع البيانات الضخمة من جهة التعريف بأفضل ممارسات المسؤولين عن معالجة البيانات الضخمة التي تم الاطلاع عليها وآليات تحليلها واستخراج المنافع منها، وهي من الدراسات التنبئية الاستشرافية حيث اعتمدت على منهج دلفي وهو منهج تفاعلي يعتمد على التواصل مع مجموعة من الخبراء أو لجنة من الخبراء من أجل استطلاع آراءهم حول موضوع معين في عدة جولات بواسطة أداة الاستبانة في الغالب، وهو ما تم بالفعل في هذه الدراسة التي استخدمت هذا المنهج للتعرف على كيفية قيام المنظمات بإعداد البيانات وتنظيمها ليتم استخدامها بشكل مفيد في إدارة المنظمة من خلال مساعدة المنظمات على استخراج معلومات أكثر فعالية من البيانات الضخمة والحصول على قيمة مضافة لها، وبينت الدراسة أن البيانات الضخمة تمثل مشكلة للعديد من المنظمات في العديد من المجالات والتخصصات وذلك لأن تقنية معالجة البيانات الحالية لا تساعد على جعل البيانات الضخمة بيانات فعالة ومفيدة، وبينت الدراسة أن السبب الرئيسي الذي أدى بالمنظمات للقيام بمعالجة البيانات الضخمة وتحليلها والاستثمار في ذلك كان لأجل تقليل المخاطر. التنظيمية والتقنية وتحقيق أرباح أعلى من هذه البيانات، كما توصلت الدراسة إلى أن تنظيم البيانات وتحليلها يساعد في تقديم معلومات مفيدة تساند المسؤولين في اتخاذ القرار وضمان مستقبل قوي للمنظمة.

5- دراسة Lau Turner & Karacapilidis (2014) بعنوان:

" :equirements for big data analytics supporting decision making A sensemaking perspective "

" متطلبات تحليلات البيانات الضخمة التي تدعم اتخاذ القرار: منظور الاستشعار عن بعد "

تأتي للعمل على تعزيز الفهم لمجالات التكامل بين الذكاء البشري والذكاء الآلي او الصناعي في عمليات تحليل البيانات الضخمة ، في محاولة لاستكشاف نماذج تفاعلية حساسة لتحليل البيانات الضخمة ، من خلا تطوير بنية أنطولوجية مفاهيمية عامة كوسيلة لرسم خارطة مفاهيمية لتحليل البيانات وتحديد دور كل من الذكاء الصناعي والذكاء البشري في هذه العلاقة التكاملية .وأجريت دراستان حالة متناقضتان للاستخدام في العالم الحقيقي لاختبار قابلية تطبيق الهيكل المقترح لتطوير منصة داعمة لتحليل البيانات الضخمة .وقد أدى التأمل في هذه النتيجة إلى زيادة الفهم لمدى تعقيد وإمكانات النماذج الفردية والتعاونية في مجال تحليل البيانات الضخمة لتحليلات البيانات.

6- دراسة Chen & Zhang (2014) بعنوان:

"Data-intensive applications, challenges, techniques and technologies: A survey on Big Data "

" التطبيقات والتحديات والأساليب والتقنيات للبيانات الكثيفة: دراسة استقصائية عن البيانات الضخمة "

وتناولت الدراسة موضوع التحديات والتقنيات المستخدمة في إدارة البيانات الضخمة ، حيث أكدت الدراسة على أن مسألة البيانات الضخمة أو الضخمة صارت قضية مؤرقة وأضحت محل اهتمام الكثير من الباحثين وصناع القرار في القطاعات الحكومية وفي الشركات بشكل أكبر، وتشير الدراسة إلى التسارع الكبير جدا في نمو البيانات إلى درجة تشكيل بيانات ضخمة جدا، وأن البيانات الضخمة سببت متاعب كبيرة للبشر لصعوبة الاستفادة منها وعدم القدرة على التحكم فيها بوضعها الذي هي عليه، واستنتجت الدراسة أن البيانات الضخمة تحوي على الكثير من الكنوز والمنافع الغير مستغلة والتي لم يتم الكشف عنها عبر تحليلها واستخلاص النتائج منها فالبيانات الضخمة تختزن الكثير من المنافع والمعارف القيمة المحتملة والمفيدة، وهدفت الدراسة للتعريف بالبيانات الضخمة ، وأهم التطبيقات المستخدمة في تخزينها بما يساهم في التعريف بتلك التطبيقات لنشر ثقافة العمل على تحليل البيانات الضخمة ، كم بينت الدراسة أهم الفرص والتحديات عندما تتم معالجة البيانات الضخمة وآليات وطرق التعامل مع مشاكل البيانات والمنهجيات الأساسية للتعامل مع طوفان البيانات، مثل الحوسبة الحبيبية، والحوسبة السحابية، والحوسبة المستوحاة من الحيوية التفاعلية.

7- دراسة Whitworth& Suthaharan(2013) بعنوان:

"Security problems and challenges in a machine learning-based hybrid big data processing network systems "

" مشاكل وتحديات الأمان في أنظمة شبكة معالجة البيانات الضخمة الهجينة القائمة على تعلم الآلة"

وفيما يتعلق بمدى الإفادة من قدرات الحاسب الآلي في التعامل مع البيانات الضخمة فقد تحدثت هذه دراسة عن إشكالية تعاملات الحاسب الآلي مع البيانات الضخمة في القطاعات التعليمية، حيث بينت الدراسة أن «البيانات الضخمة »تشكل تحديات لنظم الحاسب الآلي التقليدية المستخدمة في التعليم خصوصا عمليات استخراج وتحليل البيانات الهامة، وهدفت الدراسة إلى المساعدة في محاولة الوصول إلى حلول جديدة ومبتكرة لمعالجة البيانات من أجل التغلب على العقبات التي تفرضها البيانات الضخمة مثل كمية البيانات الخاصة بنمو الأجيال ومراحلهم الدراسية، وتوصلت الدراسة إلى وجود بعض الحلول التي يمكن تبنيها والعمل بها، وتشمل إدخال أنظمة التخزين السحابية الهجينة للمساعدة في تخزين ومعالجة البيانات، إلا أن هذا النوع من الحلول لا تقبل به جميع المنظمات وخاصة التي تمتلك بيانات حساسة فهي قلقة من إمكانية حدوث مشاكل ذات علاقة بأمن البيانات خاصة أن هذه البيانات ستنتقل من منطقة التخزين الداخلية في المنظمة إلى منطقة تخزين سحابية يشترك فيها العديد من الأفراد والجهات فهي منطقة تخزين عامة، وتوصلت الدراسة التي تعد من الدراسات الاستشرافية إلى تبني خيارين لمواجهة هذه المشكلة، الأول :تصنيف البيانات بحسب موضوعاتها أو درجة أهميتها كآلية لمعالجة مكامن القلق من المشاكل الأمنية للبيانات القادمة من مصادر خارجية، وبهذه الطريقة يمكن السماح بتخزين

البيانات الأقل أهمية والغير حساسة في أماكن التخزين السحابية التي تعتبر خارجية وهي طريقة محدودة المنفعة إذ أنها تقدم حل لجزء من البيانات وليس لها جميعا، والخيار الثاني وهو الأقدر على حل مشكلة البيانات الضخمة للمنظمة بشكل شامل يتمثل في تخصيص مستودع بيانات سحابي رئيسي يتم تشفيره وتخزن به بيانات المنظمة بعدما تتم معالجتها وتخفيف المخاطر التي يمكن أن تنتج عن تخزينها بصيغتها الأولية إذ أن مستودعات البيانات السحابية العامة توفر امكانية تحليل البيانات الضخمة عبر الطوبولوجيا المنطقية للشبكة ومهمتها تخطيط تدفق البيانات ونقلها من وإلى منطقة التخزين السحابية بطريقة آمنة لأحجام ضخمة من البيانات.

التعقيب على الدراسات السابقة:

تشير الدراسات السابقة إلى أهمية استثمار البيانات الضخمة بعد تنظيمها وتحليلها لمساندة متخذ القرار، وأن ذلك يتم عند توفر الكفاءات البشرية المؤهلة، والأنظمة الخبيرة لتكتمل عملية تحليل البيانات الضخمة بطريقة مثالية ولكي يمكن تقديم مخرجات صحيحة، مع الأخذ في الحسبان مسألة تخزين البيانات الضخمة من خلال مستودعات الحوسبة السحابية التي توفر مساحات تخزين كبيرة مع أهمية ان يتوفر الحد الكافي من الإجراءات الأمنية المناسبة التي تضمن المحافظة على خصوصية البيانات لتكون بذلك مصدر فاعلية وقوة.

جدول (1): أوجه الاختلاف بين دراسة الباحث والدراسات السابقة

الدراسة السابقة	أوجه الاختلاف
دراسة (Ahmad et al., 2019)	دراسة الباحث اعتمدت المنهج الوصفي النوعي، بينما الدراسة السابقة استخدمت منهج وصفي تجريبي
دراسة (Janssen et al., 2017)	بيئة التطبيق: حيث أن الدراسة السابقة أجريت على هيئة الضرائب الهولندية. الدراسة سابقة نوعية بينما دراسة الباحث نوعية وصفية
دراسة (Bughin, 2016)	أجريت الدراسة السابقة على مجموعة من شركات الاتصالات حول العالم واستخدمت نموذج محدد لهذا الغرض بينما دراسة الباحث اعتمدت على

	بيانات الشركة لاستخلاص النتائج
دراسة (**Stryk, 2015**)	أجريت الدراسة على القطاع الطبي واعتمدت على الاستبيان كأسلوب لجمع البيانات بينما اعتمدت الدراسة الحالية على أسلوب المقابلات والاستبيان لجمع المعلومات وأجريت دراسة الباحث على قطاع الاتصالات
دراسة (**Lau Turner & Karacapilidis, 2014**)	اعتمدت الدراسة على الاسلوب التجريبي واعتمدت دراستي حالة، اعتمدت دراسة الباحث على الأسلوب الوصفي
دراسة (**Chen & Zhang, 2014**)	أجريت الدراسة على القطاع الحكومي واعتمدت الاستبيانات كأسلوب لجمع البيانات
دراسة (**Whitworth& Suthaharan, 2013**)	أجريت الدراسة على القطاع الحكومي بينما دراسة الباحث أجريت على قطاع الاتصالات

2.1- مشكلة البحث (Research problem):

تمكنت الشركة منذ فترة طويلة من الوصول إلى أجزاء كبيرة من البيانات من خلال قاعدة كبيرة من مشتركيها الذين يقومون بالاتصال يوميًا بشبكتهم وخدماتهم. فمن خلال توسع أعمال الشركة لتشمل النطاق العريض ، أصبحت شركات الاتصالات تستحوذ على المزيد والمزيد من حجم البيانات (يقوم المستهلكون بإجراء المزيد من المكالمات والاتصال أكثر وأكثر بالإنترنت)، ويستفيدون من مجموعة أكبر من المصادر (الاستخدام الكبير لتطبيقات الإنترنت ذات النطاق العريض المتعددة) أدى ذلك إلى السرعة العالية في توليد البيانات وبدأت طرق التحليل التقليدية للبيانات تفقد فعاليتها نتيجة الزيادة الهائلة في البيانات وبسبب الطبيعة الزمانية للبينات التي باتت تحتاج إلى تقنيات أسرع لتحليلها للاستفادة منها في الوقت الفعلي، اتجهت الشركة لتطبيق تقنية البيانات الضخمة في تحليل البيانات لما لها من اهمية عالية فهي تقدم ميزة تنافسية عالية للشركة اذا ما تم الاستفادة منها بالشكل الأمثل ومعالجتها لأنها تقدم فهما أعمق لعملائها ومتطلباتهم ويساعد ذلك على اتخاذ القرارات المناسبة و الملائمة داخل الشركة بطريقة أكثر فعالية و ذلك بناء على المعلومات المستخرجة من قواعد بيانات العملاء وبالتالي زيادة الكفاءة والربح وتقليل الخسائر.

بناء على الدراسات السابقة التي اطلع عليها الباحث وعلى الزيارة الميدانية التي قام بها، مكن تلخيص مشكلة البحث من خلال طرح التساؤل الرئيسي التالي:

ماهي أهم ميزات تحليلات البيانات الضخمة (Big Data) التي تؤثر على الرفع من كفاءة الحملة التسويقية في شركة (SyriaTel)؟

3.1- أهداف البحث (Research Objectives):

يسعى الباحث من خلال هذه الدراسة إلى تحقيق الأهداف التالية: توصف الوضع الحالي:

1- دراسة التقانات المستخدمة في البيانات الضخمة (Big Data)

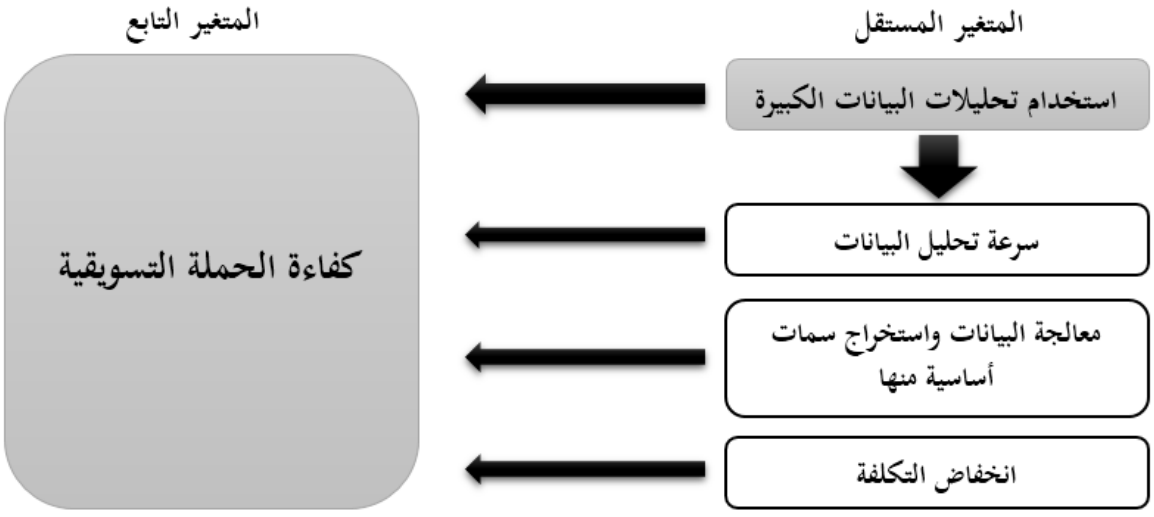
2- دراسة أثر تحليلات البيانات الضخمة على الحملة التسويقية في شركة SyriaTel من خلال:

- تحديد تأثير سرعة تحليل البيانات على كفاءة الحملة التسويقية في الشركة محل الدراسة.
- تحديد تأثير القدرة على معالجة البيانات واستخراج سمات أساسية منها على كفاءة الحملة التسويقية في الشركة محل الدراسة.
- تحديد تأثير انخفاض تكلفة تطبيق هذه التقنية مقارنة بتقنيات أخرى على كفاءة الحملة التسويقية في الشركة محل الدراسة؟

4.1- نموذج البحث: (Search Model):

لتحقيق غرض الدراسة والوصول إلى أهدافها المحددة سوف يعتمد الباحث على النموذج التالي للتعرف فيما إذا كان هناك علاقة بين المتغيرات المستقلة والمتغير التابع وفقاً لما هو مبين في الشكل:

الشكل 1: نموذج الدراسة



الشكل من إعداد الباحث

5.1- فرضيات البحث: (Research Hypothesis):

الفرضية الرئيسية:

يوجد تأثير معنوي لتحليلات البيانات الضخمة في الرفع من كفاءة الحملة التسويقة.

ومنها تتفرع الفرضيات الآتية:

1- يوجد تأثير معنوي لسرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة.
2- يوجد تأثير معنوي لمعالجة البيانات واستخراج سمات أساسية منها على الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة.
3- يوجد تأثير معنوي لانخفاض تكلفة تطبيق هذه التقنية مقارنة بتقنيات أخرى على الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة.

6.1- أهمية البحث: (Research Importance):

- ✓ تنبع أهمية هذه الدراسة من فهم كيفية اعتماد تكنولوجيات البيانات الضخمة واختبار آثار تطبيقها في الرفع من فعالية الحملة التسويقية في شركة سيرتيل وما يترتب على ذلك من عوائد مرضية ومكانة سوقية عالية وميزة تنافسية على منافستها الوحيدة في سوق الاتصالات السورية شركة MTN.
- ✓ تأتي أهمية هذه الدراسة من حيث النتائج التي يمكن أن تقدمها وكيفية الاستفادة منها في المنظمات الأخرى في الوقت التي تعيش فيه المنظمات على المستوى العالمي والمحلي إلى حد ما تحولاً نحو الاعتماد على تقنية البيانات الضخمة في المجال التسويقي، حيث يؤكد Ahmad et al (2019) أن سمات الزبائن المستخرجة باستخدام منصة البيانات الضخمة تساعد على تحديد الزبائن المتوقع تسربهم من شركة سيرتيل مما يمكن خبراء التسويق في الشركة بالقيام بخطوات فعالة للحفاظ على هؤلاء الزباء من خلال توجيه عروض تسويقية مناسبة لهذه الفئة. وبحسب (Al-Zuabi, et al, 2019) إن فهم الزبائن بشكل عميق وتمييز أنماط سلوكهم وتفضيلاتهم هو ثروة كبيرة تهدف شركة سيرتيل للوصول إليها وهي ركيزة أساسية في الحصول على ولاء الزبون، حيث تمكننا هذا المعرفة من تقديم ما يحتاج الزبون أو يفضل دون أن يطلب ذلك وبشكل صريح. ومن هنا نجد أن خصائص الزبون التي يمكن استخراجها واستنباطها من الأنشطة التي يقوم بها كخصائص الجنس والعمر تعتبر ذات أهمية كبرة بالنسبة للشركات، بالإضافة إلى الخصائص الأخرى التي تعبر عن سلوك المستخدم والتي يتم تحديدها حسب مجال التحليل المطلوب. حيث أن الخصائص المستنبطة للزبون يمكن الاستفادة منها في العملية التسويقية من خلال استهداف فئات محددة من الزبائن التي تتمتع بصفات متشابهة
- ✓ بما أن شركة الاتصالات تمتلك بيانات ذات طابع رقمي فإنها تعتبر بيئة ملائمة لتطبيق تقنية البيانات الضخمة واستخدام ميزاتها في التحليل.

✓ لم يقع بين يدي الباحث أي أبحاث منجزة في هذا المجال على مستوى الجمهورية العربية السورية من حيث تطبيق تقنية البيانات الضخمة في المجال التسويقي.

7.1- أدوات جمع البيانات:

اعتمد الباحث على كل من المصادر الأولية والمصادر الثانوية في جمع البيانات:

- **البيانات الثانوية:**

تم الاعتماد في جمع البيانات الثانوية على الأبحاث والدراسات السابقة وما كتب في المجلات العلمية ودوريات الأبحاث عن هذا الموضوع.

- **البيانات الأساسية:**

تم جمع البيانات الأولية من خلال المقابلات التي أجراها الباحث مع الموظفين وأصحاب القرار في قسم البيانات الضخمة في شركة سيرتيل كما قام بالاطلاع المباشر على بيانات الشركة لاستخلاص المعلومات، ومن خلال الاستبيان حيث قام الباحث اعتماداً على دراسات سابقة بتكييف استبيان يتناسب مع متطلبات الدراسة وتوجيهه إلى عينة من زبائن الشركة وتحليله إحصائياً وصمم على أساس مقياس ليكرت الخماسي.

8.1- أساليب تحليل البيانات الإحصائية:

تم استخلاص النتائج باستخدام برنامج تحليل البيانات (spss 20) من خلال استخدام الأساليب الإحصائية التالية:

- معامل (ألفا كرونباخ): لقياس مدى ثبات أداة القياس.
- الإحصائيات الوصفية وذلك لعرض ولذلك لعرض خصائص أفراد العينة ووصف إجاباتهم من خلال استخدام ما يلي:

– النسب المئوية: سيتم استخدامها لقياس التوزيعات التكرارية النسبية لخصائص أفراد العينة وإجاباتهم على عبارات الاستبانة.

– الوسط الحسابي: سيستخدم لإبراز مقاييس النزعة المركزية لقياس متوسط إجابات المبحوثين على أسئلة الاستبانة.

– الانحراف المعياري: سيتم استخدامه كأحد مقاييس التشتت لقياس الانحراف في إجابات أفراد العينة عن وسطها الحسابي.

- الإحصائيات التحليلية: لاختبار الفرضيات باستخدام اختبار ستودنيت (One-Sample-Test) لاختبار وجود فرق بين المتوسط المحسوب ومتوسط الحياد.

9.1- مجتمع وعينة البحث (Research Population and sample) :

مجتمع البحث: موظفي قسم البيانات الضخمة وزبائن شركة سيرتيل البالغ عددهم (9) مليون.

عينة البحث: أصحاب القرار وموظفي قسم البيانات الضخمة وعددهم (6) وعينة منتقاة من زبائن شركة سيرتيل وعددهم (41) موزعين على المحافظات التالية: (دمشق – حمص – حماة – طرطوس – اللاذقية)

10.1- حدود البحث (Research Limit):

الحدود الزمانية: تم إجراء الدراسة وتوزيع الاستبيان وتحليله من الشهر (6/1 إلى 8/15) من عام 2019.

الفصل الثاني: الإطار النظري للبحث

البيانات الضخمة وأهميتها:

1.2- تعاريف:

لا تزال البيانات الضخمة حقلاً خصباً ذا حدود غير واضحة المعالم سمحت بالتدفق الضخم للآراء والأفكار، مما أسهم في ظهور العديد من المفاهيم والتعريفات ضمن تخصصات مختلفة اقتصادياً واجتماعياً وبتضمين تقني تناوله المحللون التقنيون والممارسين ضمن مجالات وحقول عدة، ومن هذا المنطلق سيعرض الباحث مفهوم البيانات الضخمة وتعريفه من خلال بالآتي:

يرى Nasainarabic (2016) أن البيانات الضخمة هي مجموعات كبيرة جداً من البيانات التي يمكن تحليلها للكشف عن أنماط واتجاهات وارتباطات ومن ثم الحصول على رؤى منها وصولاً إلى التنفيذ العملي لها، يركز التعريف السابق على عمليات التحليل والمعالجة فركز على البيانات الضخمة من حيث أنها مجموعة بيانات مرتبة بشكل جيد ومفهوم بالرغم من ضخامتها مع وجود مستويات معروفة من عدم الوضوح، كنتيجة لتعدد عمليات ومصادر الحصول عليها، فضلاً عن حالة المزاوجة بين مختلف الأدوات والتطبيقات وأنظمة التشغيل وصيغ البيانات، ركز التعريف السابق على الأدوات والتقنيات. فيما يعرف Grobelnik (2012) البيانات الضخمة بأنها مصطلح يصف الكمية الضخمة من البيانات (الحجم) والتي تكون معقدة ومختلفة (التنوع) وسريعة تتطلب تقنيات خاصة في إدارتها وتخزينها وتحليلها.

وإذا ما أردنا أن نصل إلى تعريف شامل للبيانات الضخمة فهو كما عرفته شركة (IBM) "تنشأ البيانات الضخمة عن طريق كل شيء من حولنا وفي كل الأوقات كل عملية رقمية وكل تباد في وسائل التواصل الاجتماعي ينتج لنا البيانات الكبيرة، تتناقلها الأنظمة، وأجهزة الاستشعار، والأجهزة النقالة، البيانات الضخمة لها مصادر متعددة في السرعة والحجم والتنوع ولكي نستخرج معرفة من البيانات الضخمة نحتاج إلى معالجة سريعة، وقدرات تحليلية، ومهارات."

2.2- دورة حياة البيانات الضخمة:

يركز الكثير من الباحثين للأسف فقط على خطوة التحليل / النمذجة – فبالرغم من أهمية هذه الخطوة، فهي ذات فائدة قليلة دون المراحل الأخرى من مراحل تحليل البيانات. على سبيل المثال، يجب أن نتناول مسألة ما هي البيانات التي يجب تسجيلها من منظور أن البيانات ذات قيمة، وربما بطرق لا يمكننا توقعها بالكامل، وتطوير طرق لاشتقاق القيمة من البيانات التي يتم التقاطها بشكل غير كامل وغير مكتمل. القيام بذلك يثير الحاجة إلى تتبع الأصل ومعالجة عدم اليقين والخطأ. كمثال آخر، عندما يتم تمثيل نفس البيانات بأسلوب متكرر ومتداخل، فإنه يتيح لنا استخدام التقنيات الإحصائية للتغلب على التحديات. مثل تكامل البيانات واستخراج الكيان / العلاقة. من المحتمل أن يكون هذا هو المفتاح لاستغلال البيانات

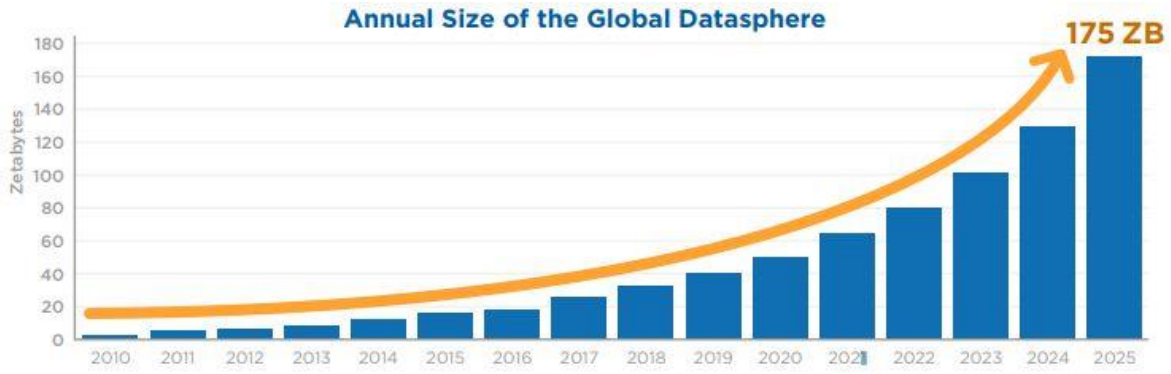
المستمدة من مصادر متعددة بنجاح (على سبيل المثال، التجارب ذات الصلة التي كشفت عنها المختبرات المختلفة، وبيانات حركة المرور الجماعية، وبيانات حول مجال معين مثل الترفيه، أزيلت من مواقع انترنيت مختلفة). هذه المواضيع ضرورية للنجاح، ونادراً ما يتم ذكرها في نفس سياق البيانات الكبيرة. حتى في مرحلة التحليل، التي حظيت باهتمام كبير، هناك تعقيدات أسيء فهمها في سياق المجموعات متعددة المستأجرين حيث تعمل برامج العديد من المستخدمين بشكل متزامن (Moothry, et al, 2014).

3.2– خصائص البيانات الضخمة:

يشار إلى خصائص البيانات الضخمة بثلاثة أبعاد أو ما يسمى بـ 3 Vs:

- **حجم البيانات (Volume)**: إلى أن الحجم هو مصطلح نسبي –من المحتمل أن يكون لدى بعض المؤسسات الأصغر حجمًا مجرد غيغابايت أو تيرابايت من تخزين البيانات مقارنةً بـ petabytes أو exabytes من البيانات التي تمتلكها الشركات العالمية الكبرى. سيستمر نمو حجم البيانات، بغض النظر عن حجم المؤسسة. تميل الشركات بشكل طبيعي لتخزين البيانات من جميع الأنواع: البيانات المالية والبيانات الطبية والبيانات البيئية وهلم جر. تقع العديد من مجموعات البيانات الخاصة بهذه الشركات في نطاق تيرابايت اليوم، ولكن سرعان ما يمكن أن تصل إلى وحدات petabytes أو حتى وحدات exabytes (2014) Moothry, et al. كما أن حجم البيانات يدل على التعامل مع المقاييس الضخمة للبيانات ضمن معالجة البيانات (مثل سلاسل التوريد العالمية، التحليل المالي العالمي، مصادم هادرون الكبير Cavanillas et al (2016, p.30). كما وضح McNamara (2010) عملية نمو حجم البيانات، حيث أنه في عام 2010 تمّ إنشاء 1.2 zettabytes من المعلومات الرّقميّة، وهذا يساوي كمّيّة التّخزين في 75 مليار هاتف آبل محمولٍ بذاكرة ذات سعة 16 جيجا بايت، أي ما يكفي من أجهزة آبل لملء مساحة ملعب ويمبلي بالكامل 41 مرة، أي ما يكفي لإعطاء كل امرأة ورجل وطفل على الأرض أكثر من 7 أجهزة آبل.

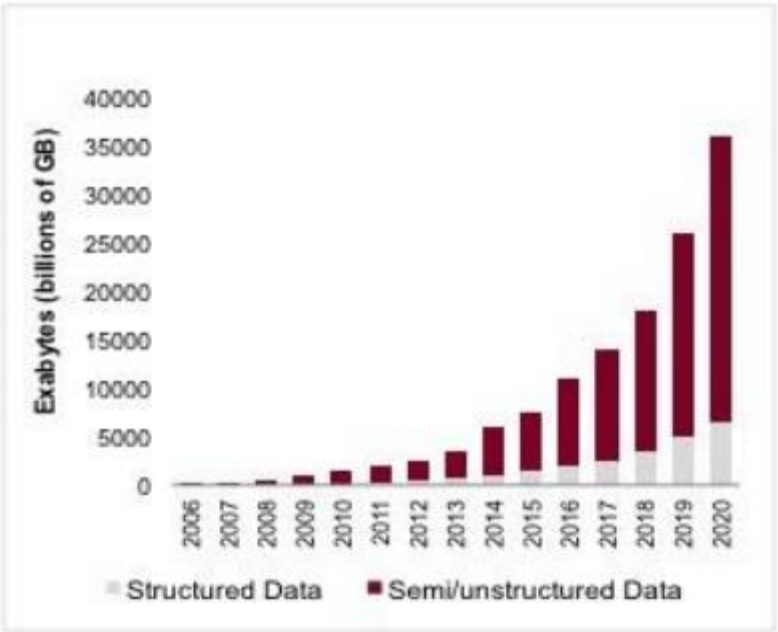
الشكل 2: توقع نمو حجم البيانات حسب السنة بـ زيتا بايات



Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

- **تنوع البيانات** (**Variety**): يمكن أن تأتي البيانات من مجموعة متنوعة من المصادر (عادة داخلية وخارجية للمؤسسة) ومن أنواع مختلفة. ومع انفجار أجهزة الاستشعار والأجهزة الذكية والشبكات الاجتماعية، أصبحت البيانات في المؤسسات معقدة لأنها لا تتضمن فقط البيانات العلائقية التقليدية المهيكلة بل البيانات شبه المهيكلة والغير مهيكلة Moothry, et al (2014). بخلاف البيانات المهيكلة النّموذجيّة، تحتوي البيانات الضخمة على نصوص، وصوت، وصور، ومقاطع فيديو، والعديد من البيانات غير المهيكلة ونصف المهيكلة، والتي تتوفر في العديد من الأشكال التناظرية والرقمية. من منظور التحليل، فإن تنوع البيانات هو أكبر تحد لاستخدام البيانات بفعالية، ويعتقد بعض الباحثين أن القدرة على التعامل مع تنوع البيانات هو المفتاح لتحليل البيانات الضخمة Sircar et al (2013)

الشكل 3: نمو تنوع البيانات حسب السنوات



Sircar Anindya, Rastogi Gaurau, Kochikar V P, Joshi Raj, Ranganath M, Towers Simon, Goparaju Subu, 2013, Infosys, **Big Data : Challenges and Opportunities**

- **السّرعة (Velocity)**: أصبحت الأعمال ديناميكية ومتقلبة. لذلك يجب معالجة التغييرات الناشئة بسبب الأحداث غير المتوقعة في الوقت المناسب من أجل تجنب الخسارة في العمل. تجد الشركات صعوبة بالغة في مواكبة سرعة البيانات هذه، فاتخاذ القرارات المثلى بسرعة، والمعالجة في وقت أقصر يكون عن طريق أنظمة إدارة البيانات التشغيلية المؤمنة Sircar et al (2013). لقد أدى هذا التحدي إلى ظهور مناطق سيل معالجة البيانات، المنطق التدفقي، وسيل استخراج البيانات للتعامل مع ارتفاع كميات البيانات الوارد من الخوادم Cavanillas et al (2016, p.30). كما أن سرعة البيانات من حيث تواتر توليدها وتسليمها هي أيضا سمة من سمات البيانات الكبيرة. عادة ما يأخذ الفهم التقليدي للسرعة في الاعتبار مدى سرعة وصول البيانات وتخزينها، ومدى سرعة استرجاعها. في سياق البيانات الكبيرة، يجب أيضًا تطبيق السرعة على البيانات المتحركة: السرعة التي تتدفق بها البيانات. إن تدفقات البيانات المختلفة وزيادة نشر شبكة الاستشعار أدت إلى تدفق مستمر للبيانات بوتيرة جعلت من المستحيل على الأنظمة التقليدية التعامل معها (Moothry, et al, 2014).

هناك أيضاً أبعاد أخرى لا تقل أهمية عن الأبعاد الثلاثة السابقة:

- **الصحة، المطابقة مع الحقائق (veracity)**: تترجم قدرة الكشف / المعالجة الملحة عبر الإنترنت. يجب اكتشاف البيانات المقاسة أو المجمعة من العمليات أو الأنظمة العملية في الوقت المناسب (قبل أي تلف أو تلاعب محتمل أو تقادم في البيانات) لضمان الثقة الكافية. على الرغم من أن البيانات الأولية التي تم الحصول عليها من مصادر أصلية يتم غسلها (تنقيتها)، أي بمعالجة مسبقة أو تصفية قبل التطبيق الحقيقي في معظم الحالات لتجنب البيانات غير ذات الصلة بشكل واضح، فإن حجمها الكبير يمكن أن يؤثر بشكل هائل على صحة البيانات الضخمة لأنه كلما زاد حجمها كلما زادت صعوبة "تنقيتها" بشكل كافٍ. بالنظر إلى صحة البيانات التي تم تنقيتها، فإن المشكلة الرئيسية التالية ليست بالضرورة الحجم الكبير، ولكن القدرة على المعالجة عبر الإنترنت. هذا هو العامل المحدد في السرعة في التطبيقات الصناعية. المزايا التي جلبتها تقنيات الشبكات المتقدمة في السنوات الأخيرة لها بعض ما يخفف من مشكلة السرعة (SHEN & kaynak, 2017).
- التباين (Variability): نظرًا لأن جميع مصادر البيانات في أنظمة البيانات الضخمة لا تنشئ تدفقات البيانات بنفس السرعة وبنوعية واحدة. لذلك، تتيح خاصية التباين معالجة المشكلات ذات الصلة. على سبيل المثال، توفير الموارد المرنة وفقًا لمتطلبات أنظمة البيانات الكبيرة (2016) Rehman et all.

- **القيمة (value)**: وهي درجة الاستفادة من هذه البيانات في المجالات الصناعية والبحثية، تحدد خاصية القيمة فائدة أنظمة البيانات الضخمة وسهولة استخدامها. تميل هذه الخاصية أكثر نحو نتائج تحليلات البيانات وعمليات معالجة البيانات وتتناسب مباشرة مع Vs5 الأخرى في أنظمة البيانات الضخمة Rehman et all (2016). قيمة البيانات تقيس مدى فائدة البيانات في اتخاذ القرارات. وقد لوحظ أن "الغرض من الحوسبة هو البصيرة وليس الأرقام". علم البيانات استكشافي ومفيد في التعرف على البيانات، لكن "العلم التحليلي" يشمل القدرة التنبئية للبيانات الضخمة (Kaisler, et al, 2013).

4.2– سلسلة القيمة للبيانات الضخمة:

تم استخدام سلاسل القيمة كأداة لدعم القرار لنمذجة سلسلة الأنشطة التي تؤديها المنظمة من أجل تقديم منتج أو خدمة ذات قيمة إلى السوق. تصنف سلسلة القيمة أنشطة القيمة المضافة العامة للمنظمة مما يسمح لها بفهمها وتحسينها. تتكون سلسلة القيمة من سلسلة من النظم الفرعية لكل منها مدخلات وعمليات تحويل ومخرجات. كأداة تحليلية، يمكن تطبيق سلسلة القيمة على أنظمة المعلومات لفهم عملية إنشاء القيمة لتقنيات البيانات (Curry et al, 2014).

تحدد هذه السلسلة الأنشطة الرئيسية عالية المستوى التالية: Cavanillas et al (2016, p.31,32).

- الحصول على البيانات (Data Acquisition): هي عملية جمع وتصفية وتنظيف البيانات قبل وضعها في مستودع بيانات أو أي وسيلة تخزين أخرى يمكن إجراء تحليل البيانات عليها. يعد الحصول على البيانات أحد أكبر التحديات للبيانات الضخمة من حيث متطلبات البنية التحتية. يجب أن تتوفر البنية التحتية اللازمة لدعم الحصول على البيانات الضخمة، زمن انتقال منخفض يمكن التنبؤ به عند التقاط البيانات وتنفيذ الاستعلامات التي تكون قادرة على التعامل مع أحجام معاملات كبيرة جدا، وغالبا في بيئة موزعة، وتدعم هياكل بيانات مرنة وديناميكية.
- تحليل البيانات (Data Analysis): تهتم هذه المرحلة بجعل البيانات الخام المكتسبة القابلة للاستخدام في صنع القرار وكذلك للاستخدام في المجال المحدد. يتضمن تحليل البيانات استكشاف البيانات وتحويلها ونمذجتها بهدف إبراز البيانات ذات الصلة وتوليف واستخراج المعلومات المخفية المفيدة ذات الإمكانات العالية من وجهة نظر العمل. تشمل المجالات ذات الصلة التنقيب عن البيانات وذكاء الأعمال وتعلم الآلة.
- معالجة البيانات (data Curation): هي الإدارة الفعالة للبيانات على مدار دورة حياتها للتأكد من أنها تقابل متطلبات جودة البيانات اللازمة للاستخدام الفعال. يمكن تصنيف عمليات تنظيم البيانات في أنشطة مختلفة مثل إنشاء المحتوى، والاختيار، والتصنيف، والتحول، والتحقق من الصحة، والمحافظة على البيانات. يتم تنظيم البيانات

من قبل المنسقين الخبراء المسؤولين عن تحسين الوصول إلى البيانات وجودتها. يتحمل أمناء البيانات (المعروفون أيضًا باسم المنسقين العلميين أو محرري البيانات) مسؤولية التأكد من أن البيانات جديرة بالثقة وقابلة للاكتشاف ويمكن الوصول إليها وقابلة لإعادة الاستخدام وتناسب الغرض منها.

- تخزين البيانات (Data Storage): هو الاستمرارية وإدارة البيانات بطريقة قابلة للتطوير التي تلبي احتياجات التطبيقات التي تتطلب الوصول السريع إلى البيانات. تعد أنظمة إدارة قواعد البيانات العلائقية (RDBMS) هي الحل الرئيسي والفريد تقريبًا لنموذج التخزين منذ ما يقرب من 40 عامًا. بينما ACID (عدم قابلية التجزئة، والاتساق، والعزل، والمتانة) هي الخصائص التي تضمن معاملات قاعدة البيانات وتفتقر إلى المرونة فيما يتعلق بتغييرات المخطط والأداء والتسامح مع الخطأ عند نمو وحدات تخزين البيانات والتعقيد، مما يجعلها غير مناسبة لسيناريوهات البيانات الكبيرة. تم تصميم تقنيات NoSQL مع مراعاة قابلية التوسع وتقديم مجموعة واسعة من الحلول القائمة على البديل.
- استخدام البيانات (Data Usage): يغطي أنشطة الأعمال التي تعتمد على البيانات والتي تحتاج إلى الوصول إلى البيانات وتحليلها والأدوات اللازمة لدمج تحليل البيانات في نشاط الأعمال. يمكن أن يؤدي استخدام البيانات في اتخاذ القرارات المتعلقة بالأعمال إلى تعزيز القدرة التنافسية من خلال خفض التكاليف أو زيادة القيمة المضافة أو أي معلمة أخرى يمكن قياسها مقابل معايير الأداء الحالية.

الشكل 4: سلسلة القيمة للبيانات الكبيرة

كما تم شرحها في:

Curry, E., Ngonga, A., Domingue, J., Freitas, A., Strohbach, M., Becker, T., et al. (2014). D2.2.2. Final version of the technical white paper. Public deliverable of the EU-Project BIG

5.2– تحديات البيانات الضخمة:

للبيانات الضخمة أيضاً العديد من التحديات JAGADISH et al (2014):

- **استحواذ البيانات:** لا تنشأ البيانات الضخمة من فراغ: إنه سجل لبعض الأنشطة الأساسية المهمة. على سبيل المثال، النظر في قدرتنا على إحساس العالم من حولنا ومراقبته، بدءًا من معدل ضربات القلب للمواطن المسن، إلى وجود السموم في الهواء الذي نتنفسه، إلى سجلات نشاط المستخدم على موقع ويب أو سجلات الأحداث في نظام برنامج ما. يمكن لأجهزة الاستشعار والمحاكاة والتجارب العلمية إنتاج كميات كبيرة من البيانات خلال اليوم. على سبيل المثال، سينتج عن تليسكوب مصفوفة الكيلومتر المربع المخطط ما يصل إلى مليون تيرابايت من البيانات الخام يوميًا. يمكن تصفية الكثير من هذه البيانات وضغطها دون المساس بقدرتنا على التفكير في النشاط الأساسي محل الاهتمام. يتمثل أحد التحديات في تعريف عوامل التصفية "على الإنترنت" بطريقة لا يتم فيها تجاهل المعلومات المفيدة، نظرًا لأن البيانات الخام غالبًا ما تكون ضخمة جدًا لدرجة لا تسمح حتى بخيار تخزينها بالكامل. علاوة على ذلك، غالبًا ما يمثل تحميل مجموعات البيانات الضخمة تحديًا، خاصةً عند دمجها مع التصفية عبر الإنترنت وتقليل البيانات، ونحن بحاجة إلى تقنيات اختزال تزايدي فعالة. قد لا تكون هذه كافية للعديد من التطبيقات، لذلك يجب تصميم المعالجة الفعالة في الوقت الحقيقي. وبحسب Chen (2016) تظهر في هذه المرحلة التحديات التي تمتاز بها البيانات الكبيرة، حيث أن تحصيل البيانات يتطلب بناء تسلسل للبيانات data pipeline وكفاءة عالية وقابلية للتوسعة والقدرة عل التعامل مع الأشكال المختلفة ضمن هذا النظام .
- **تنقية المعلومات:** في كثير من الأحيان، لن تكون المعلومات التي يتم جمعها بتنسيق جاهز للتحليل. يعد فهم مصادر الخطأ هذه ووضع نماذج لها بمثابة خطوة أولى نحو تطوير تقنيات تنظيف البيانات. لسوء الحظ، فإن الكثير من هذا يعتمد على مصدر البيانات والتطبيق المعتمد.

بحسب Banumathi & Aloysius (2017) تعد في الواقع هذه المرحلة من أكثر المراحل كلفة من حيث الجهد والوقت ، وتتطلب فهم أعمق للبيانات حيث أن جودة النموذج النهائي تعتمد بشكل أساسي عل نتيجة ومخرجات هذه المرحلة، حيث أن تنقية البيانات تتم من خلال الحذف، الاستبدال ، مثل القيم المفقودة أو المغلوطة من خلال عدة منهجيات للتنبؤ بالقيمة، أو الاعتماد عل توزيع البيانات أو غيرها من الطرق.

- **تكامل البيانات، التجميع، والتمثيل:** غالبًا ما يتطلب التحليل الفعال على نطاق واسع جمع البيانات غير المتجانسة من مصادر متعددة. على سبيل المثال، يستفاد من الحصول على العرض الصحي بزاوية 360 درجة

للمريض (أو السكان) من دمج وتحليل السجل الصحي الطبي إلى جانب البيانات البيئية المتوفرة عبر الإنترنت ومن ثم مع قراءات من أنواع متعددة من العدادات (على سبيل المثال، عدادات الجلوكوز والقلب، التسارع). تساعد مجموعة من أدوات تحويل البيانات وتكاملها محلل البيانات على حل عدم التجانس في بنية البيانات ودلالاتها. قرار عدم التجانس هذا يؤدي إلى بيانات متكاملة موحدة يمكن تفسيرها داخل المجتمع، لأنها تتلاءم مع خطط القياس واحتياجات التحليل. ومع ذلك، غالبًا ما تكون تكلفة الاندماج الكامل هائلة وتتغير احتياجات التحليل بسرعة، لذا توفر تقنيات التكامل الحديثة "الدفع الفوري والليونة "، حيث تقوم بالكثير من هذا العمل سريعًا لدعم استكشاف الإعلان المخصص. من الجدير بالملاحظة أن التوافر الهائل للبيانات على الإنترنت، إلى جانب أدوات التكامل والتحليل التي تتيح إنتاج البيانات المشتقة، يؤدي إلى نوع آخر من انتشار البيانات، والذي لا يمثل مشكلة في حجم البيانات فحسب، بل مشكلة في تتبع مصدر هذه البيانات المشتقة. حتى بالنسبة للتحليلات الأكثر بساطة التي تعتمد على مجموعة بيانات واحدة فقط، عادة ما يكون هناك العديد من الطرق البديلة لتخزين البيانات نفسها، حيث يتضمن كل بديل مفاضلات معينة.

- **النمذجة والتحليل**: تعتبر طرق الاستعلام عن البيانات الضخمة واستخراجها مختلفة بشكل أساسي عن التحليل الإحصائي التقليدي على العينات الصغيرة. البيانات الضخمة غالبًا ما تكون صاخبة وديناميكية وغير متجانسة وغير مترابطة وغير جديرة بالثقة. ومع ذلك، يمكن أن تكون البيانات الضخمة أكثر قيمة من العينات الصغيرة نظرًا لأن الإحصائيات العامة التي يتم الحصول عليها من الأنماط المتكررة وتحليل الارتباط تغلب عادة عليها التقلبات الفردية وغالبا ما تكشف عن أنماط ومعارف خفية أكثر موثوقية. في الواقع، مع الرعاية الإحصائية المناسبة، يمكن استخدام التحليلات التقريبية للحصول على نتائج جيدة دون أن تطغى عليها وحدة التخزين. حسب Sircar et al (2013) تحليل البيانات الضخمة هي ممارسة النّمذجة التّحليلية بالاعتماد بشكل كبير على الرياضيات، مما يتطلب الأدوات المناسبة والأشخاص المهرة في هذا المجال، وكذلك فالبيانات الضخمة تتطلب أيضاً أدوات عالية الكفاءة لتمثيل البيانات، لأن الأدوات التقليدية عادةً ما تكون مصممة لكمية صغيرة من البيانات.
- **التفسير**: في النهاية، يتعين على صانع القرار، المزود بنتيجة التحليل، تفسير هذه النتائج. عادة، هذا ينطوي على فحص جميع الافتراضات المقدمة وإعادة تتبع التحليل. علاوة على ذلك، هناك العديد من مصادر الخطأ المحتملة: يمكن أن تحتوي أنظمة الكمبيوتر على أخطاء، والنماذج دائمًا ما تكون مفترضة، ويمكن أن تستند النتائج إلى بيانات خاطئة. لكل هذه الأسباب، المستخدم غير المسؤول سوف يحيل الخطأ لنظام الكمبيوتر. بدلاً من ذلك، يجب أن يحاول فهم النتائج التي ينتجها الكمبيوتر والتحقق منها. كما يجب عليه أن يسهل على نظام الكمبيوتر القيام بذلك. هذا يمثل تحديا كبيرًا بالنسبة للبيانات الضخمة نظرًا لتعقيدها. باختصار،

نادراً ما يكون تقديم النتائج فقط كافياً. بدلا من ذلك، يجب أن يوفر للمستخدمين القدرة على تفسير النتائج التحليلية التي تم الحصول عليها وتكرار التحليل بافتراضات أو معلمات أو مجموعات بيانات مختلفة لدعم عملية التفكير الإنساني والظروف الاجتماعية بشكل أفضل وغالبًا ما تكون النتيجة الصافية للتفسير هي صياغة الآراء التي تشرح البيانات الأساسية بشكل أساسي.

6.2- أدوات تحليل البيانات الضخمة: التقنيات والأساليب:

لاستخلاص القيمة من البيانات الكبيرة، نحتاج إلى تطوير تقنيات وأساليب جديدة لتحليلها. حتى الآن، يطور العلماء مجموعة واسعة من التقنيات والأساليب لاستخلاص البيانات الضخمة وتحليلها وتصورها. ومع ذلك، فهي بعيدة عن تلبية مجموعة متنوعة من الاحتياجات. هذه التقنيات والأساليب تتقاطع مع عدد من الضوابط، بما في ذلك علوم الكمبيوتر، والاقتصاد، والرياضيات، والإحصاءات وغيرها من الخبرات. هناك حاجة إلى طرق متعددة التخصصات لاكتشاف المعلومات القيمة من البيانات الكبيرة. سنناقش التقنيات والأساليب الحالية لاستغلال التطبيقات كثيفة البيانات.

نحن بحاجة إلى أدوات (منصات) لفهم البيانات الكبيرة. تركز الأدوات الحالية على ثلاثة أقسام، وهي أدوات معالجة الدُفعات وأدوات معالجة الدفق وأدوات التحليل التفاعلية. تعتمد معظم أدوات معالجة الدُفعات على بنية Apache Hadoop الأساسية، مثل Mahout وDryad وهذا الأخير هو ضروري في التحليل في الوقت الحقيقي لتطبيقات البيانات الدفقية وتطبيقات معالجة الدفعات. يعالج التحليل التفاعلي البيانات في بيئة تفاعلية، مما يسمح للمستخدمين بإجراء تحليلهم للبيانات الخاصة بهم. المستخدم متصل مباشرة بالكمبيوتر، وبالتالي يمكنه التفاعل معه في الوقت الفعلي. يمكن مراجعة البيانات ومقارنتها وتحليلها بتنسيق جدولي أو رسومي أو كليهما في نفس الوقت. يعد Google Dremel و Apache Drill من منصات البيانات الضخمة استنادًا إلى التحليل التفاعلي، حيث يعد Apache Drill نظام برمجي مفتوح المصدر يدعم التطبيقات الموزعة لكثافة البيانات من أجل التحليل التفاعلي لمجموعات البيانات الكبيرة. Drill هو الأصدار المفتوح المصدر لنظام Dremel من Google والذي يتوفر كخدمة بنية تحتية تسمى Google BigQuery. أحد أهداف التصميم المعلنة بشكل صريح هو أن Drill قادر على توسيع نطاقه إلى 10000 خادم أو أكثر وأن يكون قادراً على معالجة عدد كبير من الـ petabytes وتريليونات السجلات في ثوان. (C.L. Philip Chen & C.-Y. Zhang,2014).

أساليب تحليل البيانات الضخمة:

تحتاج البيانات الضخمة إلى تقنيات غير عادية لمعالجة حجم كبير من البيانات بكفاءة خلال أوقات تشغيل محدودة. يتم تشغيل تقنيات البيانات الضخمة بواسطة تطبيقات محددة. على سبيل المثال، تطبق Wal-Mart تعلم الآلة والتقنيات

الإحصائية لاستكشاف الأنماط من حجمها الكبير من بيانات المعاملات. يمكن أن تنتج هذه الأنماط قدرة تنافسية أعلى في استراتيجيات التسعير والحملات الإعلانية. تتبنى شركة تاوباو (شركة صينية مثل eBay) تقنيات كبيرة لاستخراج بيانات الدفق على بيانات تصفح المستخدمين المسجلة على موقعها على الإنترنت، وتستغل قدرًا كبيرًا من المعلومات القيمة لدعم عملية صنع القرار.

تتضمن تقنيات البيانات الضخمة عددًا من التخصصات، بما في ذلك الإحصاء، واستخراج البيانات، وتعلم الآلة، والشبكات العصبية، وتحليل الشبكات الاجتماعية، ومعالجة الإشارات، والتعرف على الأنماط، وطرق التحسين، وأساليب التصور. هناك العديد من التقنيات المحددة في هذه التخصصات، وتتداخل مع بعضها البعض كل ساعة (C.L. Philip Chen & C.Y. Zhang,2014).

- **طرق التحسين (Optimization Methods)**: تم تطبيقها لحل المشكلات الكمية في العديد من المجالات، مثل الفيزياء والبيولوجيا والهندسة والاقتصاد. يعتبر التحسين العشوائي، بما في ذلك البرمجة الجينية والبرمجة التطورية، تقنيات تحسين مفيدة ومحددة مستوحاة من عملية الطبيعة. ومع ذلك، غالبا ما يكون لديهم تعقيد كبير في الذاكرة واستهلاك الوقت.
- **الإحصاء (Statistics)**: هو علم جمع وتنظيم وتفسير البيانات. تستخدم الأساليب الإحصائية لاستغلال العلاقات والعلاقات السببية بين الأهداف المختلفة. يتم توفير الأوصاف العددية أيضا عن طريق الإحصاءات. ومع ذلك، عادةً ما لا تكون التقنيات الإحصائية القياسية مناسبة تمامًا لإدارة البيانات الضخمة ، وقد اقترح العديد من الباحثين امتدادات للتقنيات الكلاسيكية أو طرق جديدة تمامًا (Ciaccio et al, 2012).
- **التنقيب في البيانات (Data mining)**: استخراج البيانات هو مجموعة من التقنيات لاستخراج المعلومات القيمة (الأنماط) من البيانات، بما في ذلك تحليل المجموعات والتصنيف والانحدار وتعلم قواعد الارتباط. أنها تنطوي على أساليب تعلم الآلة والإحصاءات. يعد التنقيب عن البيانات الضخمة أكثر صعوبة مقارنة بخوارزميات التنقيب عن البيانات التقليدية. إذا أخذنا التجميع كمثال، فإن الطريقة الطبيعية لتجميع البيانات الضخمة تتمثل في تمديد الأساليب الحالية (مثل التجميع الهرمي وK-Mean وFuzzy CMean) حتى يتمكنوا من التعامل مع أعباء العمل الضخمة (Zhou et al,2013, 99). تعتمد معظم الإضافات عادة على تحليل كمية معينة من عينات البيانات الكبيرة، وتختلف في كيفية استخدام النتائج المستندة إلى العينة لاشتقاق قسم للبيانات الإجمالية.
- **تعلم الآلة (Machine learning)**: تعلم الآلة هو موضوع مهم للذكاء الصنعي يهدف إلى تصميم خوارزميات تسمح لأجهزة الكمبيوتر بتطوير سلوكيات تستند إلى بيانات تجريبية. السمة الأكثر وضوحا لتعلم

الآلة هي اكتشاف المعرفة واتخاذ القرارات الذكية تلقائيا. عندما يتعلق الأمر بالبيانات الضخمة، نحتاج إلى توسيع نطاق خوارزميات تعلم الآلة، كل من التعلم الخاضع للإشراف والتعلم غير الخاضع للإشراف، للتعامل معها. أصبح التعلم العميق بالآلات بمثابة واجهة بحثية جديدة في مجال الذكاء الصنعي (Yu & Deng, 2011).

بعض خوارزميات تعلم الآلة المستخدمة في منصة البيانات الضخمة في شركة سيرتيل: Ahmad et al (2019)

1- **خوارزمية الغابة العشوائية Random Forest**: عبارة عن مجموعة من أشجار القرار التفرعية، وكل منها يتم تعليمه على عينة عشوائية من السمات وذلك من أجل تخفيف خطر التعلم الزائد.

2- **خوارزمية شجرة القرار Decision Tree**: تعتبر خوارزميّة شجرة القرار واحدةً من الخوارزميّات المشهورة في تعلم الآلة، وهي مكوّنة من مجموعة من العقد والوصلات، وكل عقدة يمكن أن تعبر عن سمة من السمات، وكل وصلة تعبر عن قيمة ما لهذه السمة، أما عقد الأوراق فإنها تعبر عن الهدف المتنبأ به وهو حالة التسرب في هذا البحث. تكمن شهرة هذه الخوارزميّة بسبب سهولة تفسيرها وكونها قادرةً على التعامل مع السمات الفئوية ويمكن توسيعها لتتعامل مع أكثر من صنفين، بالنسبة لخوارزمية شجرة القرار فإن لها بارامتراً مهماً هو عمق الشجر.

3- **خوارزمية Adboost**: هو نوع من خوارزميات التعلم الجمعية، وتعتبر أولى الخوارزميات الشجرية التسلسلية، حيث يتم توظيف العديد من المتعلمين لبناء خوارزمية تعلم أقوى، عن طريق اختيار خوارزمية أساسية (مثل أشجار القرار) وتحسينها بشكل تكراري عن طريق حساب الأمثلة المصنفة بشكل غير صحيح في مجموعة التدريب.

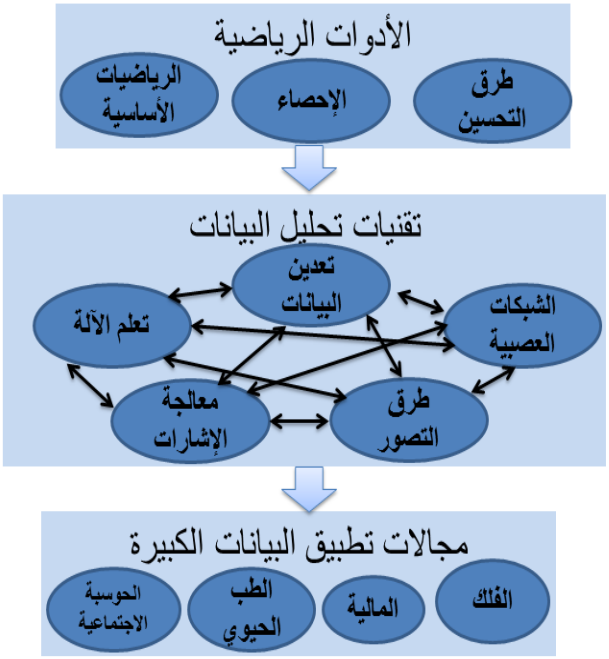
4- **خوارزمية شجرة التعزيز التدرّجي Gradient Boosting Machine Tree "GBM"**: تتشابه خوارزمية GBM وظيفياً مع خوارزمية الغابة العشوائية حيث تنشئ مجموعة أشجار، ويستخدم التوزيع العشوائي أثناء إنشاء الشجرة، ومع ذلك فإن الغابة العشوائية تبني الأشجار بالتوازي وتصوت على التنبؤ، في حين أن GBM تخلق سلسلة من الأشجار، وتقوم بتحسين كل شجرة حسب التّدريب الذي تم على الشجرة التي قبلها. حيث تعتمد على توليد العديد من أشجار التوقع الضعيفة ثم تدمجهم أو تحسنهم من خلال نموذج قوي.

5- **خوارزمية XGBoost**: وهي خوارزمية مطورة عن خوارزمية شجرة التعزيز التدرجي تستخدم نموذج أكثر تنظيماً للتحكم في مشكلة التعلم الزائد، الأمر الذي يمنحها أداءً أفضل. تم تطوير هذه الخوارزمية

في عام 2016 حيث فازت بالمركز الأول في مسابقة كاغل Kaggle لتحليل البيانات وتم نشر ورقة بحثية بطريقة عملها.

- **التنقيب في البيانات المرئية (Visualization Approaches)**: هي التقنيات المستخدمة لإنشاء الجداول والصور والرسوم البيانية وغيرها من طرق العرض البديهية لفهم البيانات Simoff (2008 et al). التصور الكبير للبيانات ليس بهذه السهولة مثل مجموعات البيانات الصغيرة النسبية التقليدية بسبب التعقيد في Vs3 (أبعاد البيانات الكبيرة) Geng et al (2012). لقد نشأت بالفعل إضافات أساليب التصور التقليدية، لكنها بعيدة عن أن تكون كافية. عندما يتعلق الأمر بتصور البيانات على نطاق واسع، يستخدم العديد من الباحثين استخراج المعالم والنمذجة الهندسية لتقليل حجم البيانات بشكل كبير قبل عرض البيانات الفعلي. لتفسير البيانات عن كثب وبشكل حدسي، يحاول بعض الباحثين تشغيل برنامج دفعي لتقديم البيانات بأعلى دقة ممكنة بطريقة موازية. يعد اختيار التمثيل الصحيح للبيانات أمرًا مهما أيضًا عندما نحاول تصور البيانات الكبيرة.
- **تحليل الشبكات الاجتماعية (Social Network Analysis)**: الذي ظهر كطريقة أساسية في علم الاجتماع الحديث، يستعرض العلاقات الاجتماعية من حيث نظرية الشبكة، ويتكون من العقد والروابط. وقد اكتسب أيضًا عددًا كبيرًا من المتابعين في الأنثروبولوجيا، وعلم الأحياء، ودراسات الاتصال، والاقتصاد، والجغرافيا، والتاريخ، وعلوم المعلومات، والدراسات التنظيمية، وعلم النفس الاجتماعي، ودراسات التنمية، وعلم اللغة الاجتماعي، وهو متاح الآن بشكل شائع كأداة للمستهلكين. واحدة من العقبات الرئيسية المتعلقة بتحليل الشبكات الاجتماعية هو اتساع البيانات الكبيرة. تحليل شبكة تتكون من ملايين أو مليارات الكائنات المتصلة عادة ما تكون مكلفة حسابيًا.

الشكل 5: تقنيات وأدوات تحليل البيانات



C.L. Philip Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Inform. Sci. (2014),

7.2– الأمن والخصوصية في البيانات الضخمة:

بالنسبة للتسويق والبحث، تستخدم العديد من الشركات البيانات الضخمة ، ولكن قد لا تمتلك الأصول الأساسية خاصة من منظور الأمان. إذا حدث خرق أمني للبيانات الكبيرة، فسيؤدي ذلك إلى تداعيات قانونية أكثر خطورة وإضراراً في السمعة أكثر من الوقت الحالي. في هذا العصر الجديد، تستخدم العديد من الشركات التكنولوجيا لتخزين وتحليل الـ petabytes للبيانات الخاصة بشركتهم، وعملهم وعملائهم. نتيجة لذلك، يصبح تصنيف البيانات أكثر أهمية. لجعل البيانات الضخمة آمنة، يجب أن تكون التقنيات مثل التشفير وتسجيل الدخول واكتشاف أساليب مخترقي الشبكات ضرورية. في العديد من المؤسسات، يعد نشر البيانات الضخمة للكشف عن الاحتيال جذابًا ومفيدًا للغاية.

يجب أن يتم حل التحدي المتمثل في اكتشاف ومنع التهديدات المتقدمة والمتطفلين الضارين، وذلك باستخدام تحليل أنماط البيانات الكبيرة. تساعد هذه التقنيات في اكتشاف التهديدات في المراحل المبكرة باستخدام تحليل أنماط أكثر تطوراً وتحليل مصادر بيانات متعددة.

ليس فقط الأمن ولكن أيضًا تحديات خصوصية البيانات الموجودة في الصناعات والمؤسسات الفيدرالية. مع ازدياد استخدام البيانات الضخمة في الأعمال التجارية، تصارع العديد من الشركات مشاكل الخصوصية. خصوصية البيانات هي مسؤولية، وبالتالي يجب أن تكون الشركات في حالة دفاعية عن الخصوصية. ولكن على عكس الأمان، ينبغي اعتبار الخصوصية أحد الأصول، وبالتالي تصبح نقطة بيع لكل من العملاء وأصحاب المصلحة الآخرين (Moothry, et al, 2014).

تحديات الأمن والخصوصية في البيانات الضخمة:

الخصوصية والأمان هي تحديات معروفة في البيانات الضخمة . نشرت مجموعة عمل CSA للبيانات الضخمة قائمة بأفضل 10 تحديات لأمن البيانات الضخمة والخصوصية Mora et al (2012). فيما يلي خمسة من تلك التحديات التي تعتبر مهمة للغاية لتخزين البيانات الضخمة (Cavanillas et al, 2016, pp. 127-128-129):

1. ممارسات الأمان الأفضل لمخازن البيانات غير العلائقية:

تشبه تهديدات الأمان لقواعد بيانات NoSQL التهديدات على قواعد البيانات التقليدية RDBMS بالتالي يجب تطبيق أفضل ممارسات الأمان (Winder, 2012). ومع ذلك، فإن العديد من تدابير الأمان التي يتم تنفيذها افتراضيًا ضمن RDBMS التقليدية مفقودة في قواعد بيانات NoSQL (Okman et al, 2011) وتشمل هذه التدابير تشفير البيانات الحساسة، وصناديق العمليات، والتحقق من صحة المدخلات، والتوثيق القوي للمستخدم. يوصي بعض داعمي NoSQL باستخدام قواعد البيانات في بيئة موثوقة مع عدم وجود تدابير أمان أو مصادقات إضافية مطبقة. ومع ذلك، فإن هذا النهج بالكاد معقول عند نقل تخزين البيانات الضخمة إلى التخزين السحابي. يحظى أمان قواعد بيانات NoSQL بمزيد من الاهتمام من قِبل الباحثين في مجال الأمن والمخترقين على حد سواء، وسيستمر تحسين الأمن مع نضوج السوق. على سبيل المثال، هناك مبادرات لتوفير قدرات التحكم في الوصول لقواعد بيانات NoSQL استنادًا إلى وحدات مصادقة Kerberos. (Winder, 2012).

2. تأمين تخزين البيانات وسجلات المعاملات:

تنشأ تحديات أمان خاصة لتخزين البيانات بسبب توزيع البيانات. من خلال ميزة الضبط التلقائي، يمنح المشغلون التحكم في تخزين البيانات للخوارزميات من أجل تقليل التكاليف. يجب حساب مكان البيانات وحركات الطبقة والتغييرات بواسطة سجل المعاملات. يجب تصميم الاستراتيجيات التلقائية بعناية لمنع نقل البيانات الحساسة إلى مستويات أقل أمانًا وبالتالي أقل تكلفة، يجب أن تكون آليات المراقبة والتسجيل قائمة من أجل الحصول على رؤية واضحة لتخزين البيانات وحركة البيانات (Mora et al. 2012).

3. التحكم في الوصول القسري والاتصال الآمن:

اليوم، غالبًا ما يتم تخزين البيانات بدون تشفير، ويعتمد التحكم في الوصول فقط على تطبيق يشبه البوابة. ومع ذلك، لا ينبغي الوصول إلى البيانات إلا عن طريق الكيانات المرخص لها بضمانات التشفير –وكذلك في التخزين

والإرسال. لهذه الأغراض، هناك حاجة إلى آليات تشفير جديدة توفر الوظائف المطلوبة بطريقة فعالة وقابلة للتطوير. بينما يبدأ مزودو خدمات التخزين السحابي في تقديم التشفير ومفتاح التشفير، يجب إنشاء المواد وتخزينها لدى العميل وعدم تسليمها مطلقًا إلى المزود السحابي.

4. تحديات الأمن والخصوصية للتحكم في الوصول الدقيق:

يمثل تنوع البيانات تحديًا كبيرًا نظرًا لمتطلبات الأمان المتنوعة على قدم المساواة، مثل القيود القانونية وسياسات الخصوصية وسياسات الشركة الأخرى. هناك حاجة إلى آليات التحكم في الوصول الدقيق لضمان الامتثال لهذه المتطلبات.

5. مصدر البيانات:

سلامة وتاريخ كائنات البيانات داخل سلاسل القيمة أمر بالغ الأهمية. الأصل التقليدي يحكم الملكية والاستخدام. ومع البيانات الكبيرة، سيزداد تعقيد بيانات تعريف المصدر Glavic (2014). بذلت جهود أولية لدمج الأصل في النظام البيئي للبيانات الضخمة (Ikeda et al. 2011; Sherif et al. 2013) ومع ذلك، يتطلب المصدر الآمن ضمانات سلامة وسرية بيانات الأصل بجميع أشكال تخزين البيانات الكبيرة ويبقى تحديا مفتوحا. علاوة على ذلك، فإن تحليل الرسوم البيانية للأصول الضخمة جدًا مكثف من الناحية الحسابية ويتطلب خوارزميات سريعة.

تحديات الخصوصية في تخزين البيانات الضخمة Cavanillas et al (2016, p.129). أظهر الباحثون (Acquisti and Gross, 2009) أن تحليل البيانات الضخمة للبيانات المتاحة للجمهور يمكن استغلاله لتخمين رقم الضمان الاجتماعي للشخص. تقوم بعض المنتجات بتشفير حقول البيانات بشكل انتقائي لإنشاء هوية مجهولة، وفقًا لامتيازات الوصول. قد تكون إخفاء الهوية وإلغاء تحديد البيانات غير كافية لأن الكمية الضخمة من البيانات قد تسمح بإعادة التحديد. دعى Bollier & Firestone (2010) إلى الشفافية في التعامل مع البيانات والخوارزميات كصفقة جديدة بشأن البيانات الضخمة Wu & Guo (2013) لتمكين المستخدم النهائي كمالك للبيانات. لا يشتمل كلا الخيارين فقط على شفافية المؤسسة، ولكن أيضًا على الأدوات التقنية مثل الأمان والخصوصية حسب تصميم ونتائج مشروع الاتحاد الأوروبي EEXCESS EU FP7 project (Hasan et al. 2013).

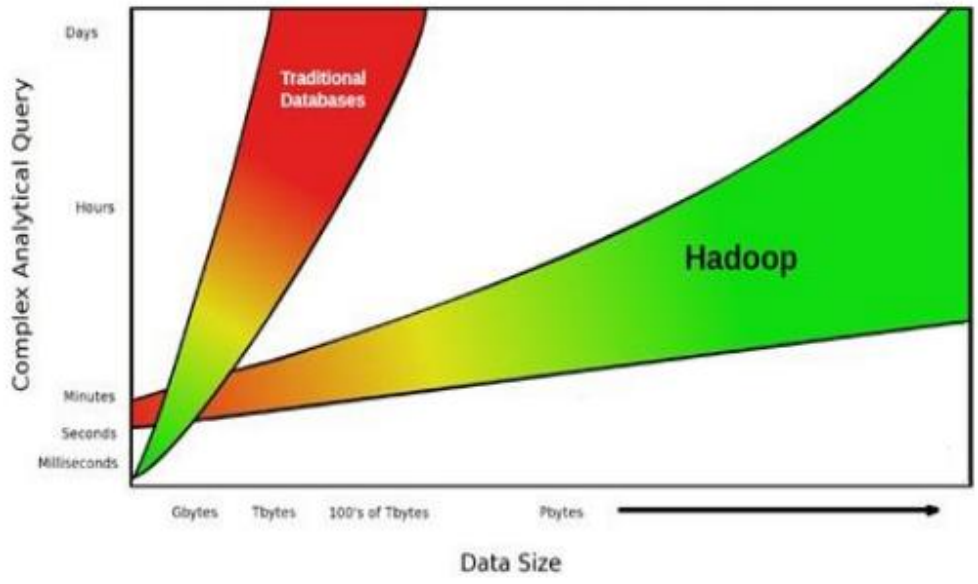
8.2– دور Hadoop في البيانات الضخمة:

Hadoop هو نظام معالجة دفعي لمجموعة من العقد التي توفر الأسس لمعظم الأنشطة التحليلية للبيانات الضخمة لأنه يجمع بين مجموعتين من الوظائف الأكثر حاجة للتعامل مع مجموعات البيانات الضخمة غير المهيكلة ، نظام الملفات الموزعة ومعالجة Map/Reduce. هو مشروع من مؤسسة Apache Software Foundation مكتوب بلغة جافا لدعم التطبيقات الموزعة على البيانات (Padhy,2013).

Hadoop يعمل بشكل جيد في معالجة كمية كبيرة من البيانات في نفس الوقت. يوفر آلية تقسيم عامة لتوزيع عبء العمل التجميعي عبر أجهزة مختلفة. ومع ذلك، تم تصميم Hadoop لمعالجة دفعة واحدة. يعتبر محرك متعدد الأغراض ولكنه ليس محركًا في الوقت الفعلي وليس وعالي الأداء، نظرًا لوجود معدل كبير من الكمون في تطبيقاته Chen (2014).

قواعد البيانات التقليدية جيدة في الاستعلامات التّفاعلية السريعة للغاية على مجموعات البيانات المعتدلة الحجم والصغيرة، ولكن عادةً ما تنفد ببطء، ويمكن أن تستغرق وقتاً طويلاً عندما تصبح البيانات ذات حجم كبير. ومن ناحية أخرى، فإن النظام الشامل هادوب بمعالجته الموزعة، ذو أداء رديء عند الاستعلام على البيانات الصغيرة، ولكنها مثالية لأعباء الأعمال الضخمة التي تأخذ كمية كبيرة من البيانات والاستعلامات الأكثر تعقيداً (*Sircar* et al, 2013)، وهذا ما يوضحه الشكل:

الشكل 6: مقارنة قدرة هادوب مع قواعد البيانات التّقليديّة



Sircar Anindya, Rastogi Gaurau, Kochikar V P, Joshi Raj, Ranganath M, Towers Simon, Goparaju Subu, 2013, Infosys, **Big Data : Challenges and Opportunities**

خصائص Hadoop:

إنّ نظام هادوب هو نظام مفتوح المصدر من توزيعة أباتشي ويتميّز بما يلي Sircar et al (2013)

1. قابلية التوسع، وهذه إحدى الخصائص الهامة لهادوب التي تجعلها مناسبة لتحليلات البيانات الضخمة ويسمح ذلك بإضافة عقد جديدة حسب الحاجة، دون اعتبار لتنسيقات البيانات أو كيفية تحميل البيانات.
2. نظام مرن، فلا يحتاج إلى مخططات للبيانات للعمل عليها، ويمكن أن يتعامل مع أي نوع أو تنسيق للبيانات حيث تفشل قواعد البيانات التقليدية عادة في تحميل بيانات غير مهيكلة.
3. هو نظام متسامح مع الأخطاء، فعندما تفشل إحدى العقد، يقوم بتوجيه المعالجة إلى موقع آخر للبيانات مستفيداً من خاصية تعددية البيانات، فتبقى العملية قيد التشغيل.
4. نظام معالجة سريع، حيث إنه يقوم بتشغيل العمليات بالتوازي على العقد الموزعة.

9.2- تأثير تحليلات البيانات الضخمة على الحملة التسويقية في شركات الاتصالات:

1- تأثير سرعة تحليل البيانات (chen,2016):

قبل عقدين من الزمن، كان مستودع البيانات على نطاق تيرابايت يُعتبر شائعًا (وصل مستودع بيانات Walmart إلى 1 تيرابايت في عام 1992). في الوقت الحاضر مجموعات البيانات في نطاق تيرابايت موجودة في كل مكان. نمت بيانات الاتصالات بشكل كبير منذ عصر النطاق العريض والجيل الثالث. سيستمر هذا الاتجاه مع تطور الشبكات (البصرية، G4 LTE /، 5G) ، مما يتيح للمستخدمين الوصول إلى المحتويات المتزايدة والمساهمة ومشاركتها على الإنترنت (Ericsson Mobility Report, 2016). تواجه صناعة الاتصالات كمية غير مسبوقة من البيانات، حيث تتدفق عند سرعة لم يسبق لها مثيل من قبل. كان تحليل البيانات ذات المقياس الصغير وتدفق البيانات يمثل تحديًا من قبل، إن لم يكن مستحيلًا، لأن تكاليف الأجهزة والبرامج اللازمة للتعامل مع البيانات كانت باهظة. علاوة على ذلك، نظرًا لأن عائد الاستثمار في تحليلات البيانات يصعب أحيانًا قياسه (على سبيل المثال، ملكية العميل) أو يستغرق وقتًا طويلاً لتحقيقه، فقد كان عادةً ما يتم تثبيط الإنفاق الكبير.

اليوم، قلل اتجاه استخدام الأجهزة مفتوحة المصدر والسلع الأساسية في منصة البيانات الضخمة من القلق أعلاه إلى حد ما. تحتوي العديد من المصادر المفتوحة إما على إصدار مجتمع مجاني أو إصدار مؤسسة مع رسوم خدمة / ترخيص معتدلة. تعد مشاريع [Apache] واحدة من أبرز الأمثلة. إن التحول من القرص المشترك غير المكلف، SMP (المعالجات المتعددة المتماثلة) إلى بنية لا تعتمد على مشاركة، سلعة تعمل أيضًا على خفض سعر التذكرة في لعبة البيانات الكبيرة.

يمتلك المشغلون الآن أدوات ميسورة التكلفة لاستكشاف التحليلات الجديدة والمعقدة مع البيانات الضخمة. من منظور التنوع والحجم والسرعة، يواجه مشغلو الاتصالات، مثل الشركات في جميع القطاعات الأخرى، التحديين الرئيسيين التاليين:

- إبرة في كومة قش –كيفية الكشف عن الارتباط والرؤى القابلة للتنفيذ من مساحة البيانات عالية الأبعاد؟ تتمثل المهمة الأكثر تحديا والأكثر إثارة في أي مشروع لتحليل البيانات في تحديد الارتباطات بين المتغيرات (أو الميزات) وكشف العلاقة بين المتغيرات والقياس المتوقع (التسميات). فقط عندما يتم الكشف عن هذه القرارات، يمكن اتخاذ قرارات تجارية ذات معنى. تصبح هذه المهمة أكثر صعوبة اليوم بسبب المصادر المتنوعة للبيانات المتاحة التي قد تؤدي إلى مئات أو حتى آلاف المتغيرات.
- تكامل البيانات وجودتها –تكون النتائج جيدة بقدر جودة البيانات. على عكس ما يعتقد معظم الناس، عادة ما يستغرق تكامل البيانات وتطهيرها الجزء الأكبر من الوقت في أي مشروع لتحليل البيانات. قد تتضمن المهام استخراج البيانات التقليدية وتحويلها وتحميلها (ETL) وتسوية البيانات. هذا الأخير يشير إلى مهمة حل التعارضات من مصادر بيانات متعددة، على المستويات التخطيطية والنمذجية والدلالية. في حين أن الأدوات التي يمكن أن تساعد في أتمتة عملية ETL تعد وفيرة، لا تزال تسوية البيانات تشكل جهدا شاقا للغاية. من المعتقد بشكل عام أن أكثر من 70 في المئة من وقت علماء البيانات يقضونه في "تدوين البيانات" بدلاً من تحليل البيانات فعليًا.

وخير مثال على ذلك هو تحليل تسرب الزبائن. معدل التذبذب هو النسبة المئوية للمشتركين الذين ينهون الخدمة مع الشركة. يمكن أن يتراوح متوسط معدل التسرب الشهري من 1% إلى 5% (http://www.statista.com.). للمشغلين الكبار التي يمكن أن تعني خسارتهم خسارة مئات الملايين من الإيرادات. ونظرًا لأن تكلفة الحصول على مشترك جديد أعلى بعدة مرات من الاحتفاظ بمشترك حالي، يعتمد المشغلون على تحليل التسرب لتحديد الهوية والاحتفاظ بالعملاء القيمة الذين سيغادرون على الأرجح.

مع تقدم وتوافر تقنيات البيانات الكبيرة، أصبح بإمكان المشغلين الآن جمع المزيد من البيانات الكاملة تقريبًا عن تجربة المستخدم وسلوكه. يمكنهم بعد ذلك بناء نموذج تنبؤ قائم على السلوك لتنبيههم عندما يكون المشترك على وشك المغادرة. على سبيل المثال، يجد المشغلون أن المشترك الذي يكون عقده على وشك الانتهاء وقد بحث في مواقع ويب المنافسين، سيكون لديه احتمال كبير للاتصال بإلغاء الخدمة (http://www.teradata.com.).

تقدم السرعة في تحليل البيانات الضخمة لمشغلي الاتصالات فرصة حقيقية لالتقاط رؤية أكثر شمولًا لعملياتهم وعملائهم، ولتعزيز جهودهم الابتكارية. تتمثل إحدى القوى المحركة الرئيسية في النمو الهائل للبيانات الناتجة من الأجهزة المحمولة والمتصلة

ومحتويات الوسائط الاجتماعية. تشمل التطبيقات المهمة التي يمكن أن تستفيد من البيانات الضخمة تحليل جودة التجربة لإطلاق الحملات التسويقية سواء للحفاظ على الزبائن الحاليين أو اكتساب زبائن جدد في وقت قصير مقارنة مع الطرق التقليدية في تحليل البيانات.

2- تأثير معالجة البيانات واستخراج سمات أساسية منها (https://medium.com/):

في عالم الاتصالات المكثف اليوم، تنتج شبكات التواصل الاجتماعي والأجهزة المتصلة وسلوكيات العملاء والبوابات الحكومية وسجلات بيانات المكالمات ومعلومات الفوترة وما إلى ذلك كمية هائلة من البيانات. من الصعب على مشغلي الاتصالات التعامل مع هذه الزيادة الضخمة في أحجام البيانات. ومع ذلك، يمكن تحويل هذا التحدي إلى فرصة باستخدام تقنيات البيانات الضخمة وتحليلات البيانات الضخمة بكفاءة.

الكم الهائل من البيانات عند التقاطها بحكمة وتحليلها بشكل احترافي يمكن أن تكشف عن رؤى قوية. توفر البيانات الضخمة والتحليلات المتقدمة لشركات الاتصالات الأدوات والتقنيات لتسخير ودمج مصادر جديدة وأنواع جديدة من البيانات في أحجام أكبر وفي الوقت الفعلي.

يمكن أن تساعد تحليلات البيانات المشغلين على تحسين القيمة الإجمالية لأعمالهم فيما يتعلق بتحسين الخدمة ورضا العملاء والإيرادات. الحالات التي تستفيد فيها شركات الاتصالات من تحليلات البيانات الضخمة بشكل كبير:

رؤى العملاء المحسنة:

بمساعدة قدرات تحليل البيانات الضخمة، يمكن لشركات الاتصالات تحويل بيانات هائلة منظمة وغير منظمة إلى رؤى عملية للعملاء. البيانات الضخمة التي يولدها العملاء، مع التحليلات الصحيحة، تمكن شركات الاتصالات من تطوير ملفات تعريف العملاء المخصّصة 360 وتأسيس مؤشرات الأداء الرئيسية التي تركز على العملاء وتطوير عروض تسويقية أكثر استهدافًا. بفضل بنيات البيانات المتقدمة، يمكن للمشغلين أيضًا تخزين أنواع جديدة من البيانات والاحتفاظ بها لفترة أطول والانضمام إلى مجموعات البيانات المتنوعة معًا لاكتساب رؤى جديدة.

فيما يلي بعض مكونات بيانات العميل التي تؤخذ في الاعتبار من أجل أن تصل شركات الاتصالات إلى رؤى العملاء المحسنة:

- بيانات العميل: معرف العميل، MSISDN، الديموغرافية، الخدمات المستخدمة، نمط الإنفاق، خطة الاستخدام.
- بيانات الجهاز: العلامة التجارية، النموذج، السلسلة، التطبيقات، التكنولوجيا المستخدمة، سجل الجهاز

- بيانات الاستخدام: CDR ، متوسط الإيرادات ، VAS (خدمات القيمة المضافة) ، بيانات استخدام الإنترنت عبر الهاتف النقال (عناوين URL ، الوقت المستغرق ، نوع المحتوى ، التنزيلات)
- بيانات الموقع: الموقع الحالي، الموقع الأكثر زيارة، بيانات التجوال، استخدام خدمات الموقع، إلخ.

بما أن شركة الاتصالات تخدم عدد كبير من العملاء. من خلال تحليل بيانات العملاء بشكل صحيح، يمكنهم أيضًا إنشاء شرائح صغيرة، تمكنهم من تقسيم العملاء إلى مجموعات متشابهة. سيسمح ذلك لمشغلي الاتصالات بتخصيص نهجهم لتلبية احتياجات كل مجموعة من العملاء، وتحديد العملاء الأكثر قيمة –المكون الحاسم للغاية من أجل اتخاذ القرارات الاستراتيجية –وإنشاء حملات تسويقية تناسبهم. فمثلا:

1- يمكنهم التعرف على العملاء المخلصين الذين لديهم قيمة محتملة عالية مدى الحياة مع تجزئة قيمة العملاء تقوم الشركة ببناء حملات تسويقية مستهدفة وتقلل من معدلات تسرب الزبائن.
2- تحديد العملاء ذوي القيمة العالية الذين هم أكثر عرضة لتكرار أنماط الشراء مع التحليل التنبئي.
3- تحديد العملاء المحتملين حيث يمكنهم ضبط الوصول المستهدف وخفض التكاليف على قاعدة العملاء غير المستهدفة.
4- تقديم منتجات مخصصة لكل شريحة وفقًا لاحتياجات العملاء، والسلوك، والديموغرافيا، وبيانات الجهاز، إلخ.

3- تأثير انخفاض تكلفة استخدام البيانات الضخمة:

1- **مرونة أعلى ومساحات تخزين أقل** Ahmad et al (2019): تم تطبيق العديد من المحاولات السابقة باستخدام نظام Warehouse Data (مستودع البيانات) في إطلاق حملات تسويقية للتخفيض من تسرب الزبائن في شركة سيريتل. جمع مستودع البيانات نوعًا من بيانات الاتصالات مثل بيانات الفواتير والمكالمات / الرسائل القصيرة / الإنترنت والشكاوى. تم تطبيق تقنيات استخراج (تعدين) البيانات في قمة نظام Warehouse، لكن فشل النموذج في إعطاء نتائج عالية باستخدام هذه البيانات. في المقابل، تم تجاهل مصادر البيانات الضخمة في الحجم بسبب التعقيد في التعامل معها. لم يتمكن مستودع البيانات من الحصول على هذه الكمية الضخمة من البيانات وتخزينها ومعالجتها في نفس الوقت. بالإضافة إلى ذلك، كانت مصادر البيانات من أنواع مختلفة، وكان جمعها في مستودع البيانات عملية صعبة ومكلفة للغاية، لذا فإن إضافة ميزات جديدة لخوارزميات تعدين البيانات تتطلب وقتًا طويلًا، وقدرة معالجة عالية، والمزيد من سعة التخزين. من ناحية أخرى، تتم كل هذه العمليات الصعبة في مستودع البيانات بسهولة باستخدام المعالجة الموزعة التي توفرها منصة البيانات الكبيرة.

2- **رضا العملاء:** تقوم شركات الاتصالات بتجميع كمية كبيرة من البيانات بسببها يمكنها بسهولة إنشاء عرض 360 درجة لعملائها باستخدام بياناتهم الخاصة، وهي بيانات الاتصال والبيانات الجغرافية وبيانات استخدام الإنترنت والبيانات العامة من الشبكات الاجتماعية. إن العرض بزاوية 360 درجة سيساعد الشركات على إنشاء تجارب مخصصة للغاية وستكون الشركات قادرة على تقديم خدمات أفضل مع هدف محدد.

10.2- الخاتمة:

تم في هذا الفصل تغطية أهم الجوانب التي تتعلق بالبيانات الضخمة بدءاً من دورة حياة البيانات وخصائصها، انتقالاً إلى سلسلة القيمة، بعد ذلك تم شرح أهم التحديات التي تواجه البيانات الضخمة وتقنيات وأساليب تحليلها بالإضافة إلى تحديات الأمن والخصوصية، تم أيضا شرح المنصة المستخدمة في تحليل البيانات الضخمة (Hadoop). أما القسم الأخير فقد خصص لشرح تأثير تحليلات البيانات الضخمة على الحملة التسويقية في شركات الاتصالات، حيث اتضح أن سرعة تحليل البيانات الضخمة تؤثر على لإطلاق الحملات التسويقية سواء للحفاظ على الزبائن الحاليين أو اكتساب زبائن جدد في وقت قصير مقارنة مع الطرق التقليدية في تحليل البيانات. كما اتضح أن استخلاص سمات أساسية من البيانات يسمح لمشغلي الاتصالات بتلبية احتياجات كل مجموعة من العملاء، وتحديد العملاء الأكثر قيمة وإنشاء حملات تسويقية تناسبهم، أخيراً فيما يخص التكلفة فقد اتضح وجود فرق كبير عند تطبيق تقنية البيانات الضخمة من حيث السعات التخزينية الأقل والسرعة العالية في التحليل مقارنة بالطرق التقليدية.

الفصل الثالث: القسم العملي

قام الباحث في هذا الفصل بمناقشة المعطيات المتعلقة بالفرضيتين الأولى والثانية من منظور الشركة: من خلال إجراء مقابلات مع أصحاب القرار في قسم البيانات الضخمة (Big Data) في شركة SyriaTel وعددهم (6) ومن خلال الاطلاع المباشر على قاعدة البيانات في الشركة لاستخلاص البيانات المطلوبة المتعلقة بالـ (Big Data).

وقام باختبارهما من منظور الزبائن: من خلال استبيان قام الباحث بتوزيعه على عينة من زبائن الشركة والتي بلغت (41) موزعين على خمس محافظات (دمشق – حمص – حماه – طرطوس – اللاذقية) طرح من خلالها الأسئلة التي تتعلق بمحاور البحث الأساسية حيث قام بتفريغ الاستبيان وتحليل النتائج باستخدام برنامج التحليل الإحصائي (spss) ومقارنتها مع النتائج التي حصل عليها من المرحلة الأولى.

قام الباحث بعد ذلك بمقارنة النتائج التي حصل عليها من المنظورين السابقين وصل من خلالها إلى مجموعة من النتائج.

ناقش الباحث المعطيات المتعلقة بالفرضية الثالثة من منظور الشركة فقط وذلك لأن التكاليف تتعلق بالبيانات الداخلية للشركة ولا يمكن اختبارها عن طريق الزبائن.

1.3– أنواع البيانات ومصادرها في شركة سيريتل:

في مجال الاتّصالات، تستهدف معظم جهود البيانات الضخمة في وقت مبكّر تحليل البيانات الحاليّة مثل CDRs وبيانات الشبكة وبيانات الزبائن. يُعرف أكثر من نصف خبراء التشغيل البيانات الداخلية كمصدر أساسيٍّ للبيانات الضخمة داخل مؤسّساتهم. هناك ثلاثة أنواعٍ من البيانات في مجال الاتّصالات من وجهة نظر المشغّل، وهي بيانات الزّبائن، وبيانات الشّبكة وسجلّ تفاصيل المكالمات CDRs.

1– بيانات الزبائن

عادةً ما يكون لمشغلي الاتصالات الخليوية الملايين من الزبائن، ويتطلّب هذا الحفاظ على قاعدة بيانات المتعلقة بهؤلاء الزبائن، على سبيل المثال: يحتفظ المشغل ببيانات حساب الزّبون التي تتضمن على سبيل المثال: الاسم والعنوان وبيانات الفوترة بما في ذلك طرق الدّفع وتفاصيل الدّفع، وتتضمن بيانات الزبون أيضاً البيانات المتعلقة بالاتصال وقطع الاتصال وإعادة الاتصال والمعلومات التّفصيلية للخدمات التي يستخدمها. يقوم المشغّل أيضاً بحفظ حركة البيانات، مثل البيانات التي تحدد هوية المرسل والمتلقي وبيانات التوجيه والتتبع عبر الإنترنت لاتّصالات الزّبون. تحتوي بيانات الزبون على بيانات مهيكلة وشبه مهيكلة وغير مهيكلة. تنشأ بيانات الملف الشخصي للزبون من أنظمة الشبكة مثل Home Location Register (HLR)، أو إدارة علاقات الزبائن (CRM).

2- بيانات الشّبكة

هي البيانات التي يتم إنشاؤها بواسطة عناصر الشبكة. تقريباً جميع معدات شبكة الاتصالات قادرة على توليد رسائل الخطأ وحالة التشغيل، الأمر الذي يؤدي إلى كمية كبيرة من بيانات الشّبكة، وتحتوي هذه البيانات على الطابع الزمني، وهي سلسلة تحدد بشكل فريد المكونات المادية أو البرامج التي تنشئ الرسالة، والرماز الذي يشرح سبب إنشاء الرسالة. تحتوي بيانات الشبكة أيضاً على بيانات مهيكلة وغير مهيكلة. يمكن تصنيف الإنذارات أو رموز الأخطاء على أنها بيانات مهيكلة، ولكن رسالة الخطأ إن لم تكن محددة مسبقاً، فهي بيانات غير مهيكلة.

3- سجلّ تفاصيل المكالمة

في كل مرة يتم فيها إجراء مكالمة أو إرسال رسالة أو استخدام للإنترنت أو القيام بأي عملية على شبكة الاتّصالات، يتم حفظ المعلومات الوصفية لها كسجل تفاصيل المكالمات (CDR). ويجدر الذكر بأن عدد سجلاّت تفاصيل المكالمات التي يتم إنشاؤها وتخزينها في قاعدة بيانات مشغل الخليوي كبيرة جداً.

وهنا يمكن أن نرى بوضوح أهمية استخدام منهجية البيانات الضخمة في مجال الاتصالات بسبب التنوع الكبير في البيانات وطبيعتها والحجم الكبير وتوليدها في وقت صغير جداً، ولذلك فإن تحضير هذه البيانات ليتم استخدامها لحل أي مشكلة، سيكون عائقاً كبيراً إذا ما استخدمنا بيئة بيانات كبيرةٍ بسبب صعوبة معالجتها وتحويلها.

2.3- تأثير تحليلات البيانات الضخمة على الحملة التسويقية في شركة SyiaTel:

- حسب Ahmad et al (2019) تم استخدام تقنية البيانات الضخمة في الشركة لبناء نموذج لتوقع تسرب الزبائن يساعد مشغلي الاتصالات على التنبؤ بالعملاء الذين هم على الأرجح عرضة للتسرب إلى شركة أخرى، استطاعت الشركة التعامل مع كمية بيانات أكبر تابعة لفترة زمنية أطول تمتد لمدة 9 أشهر حتى تم الاكتشاف إلى إن المدة المثلى لكشف توقع تسرب الزبائن هي 6 أشهر، واستطاعت الوصول إلى سمات أكثر تعقيدا من السمات التي كان يتم التعامل معها بالطريقة التقليدية وهي السمات المولدة من شبكة التواصل الاجتماعي، حيث أن شبكة التواصل الاجتماعي هي عبارة عن شبكة رسم بيانية كبيرة جداً تحتوي على جميع مكالمات العملاء على سبيل المثال (FaceBook) فهي نوع من أنواع البيانات التي يصعب التعامل معها بالطريقة التقليدية لذلك تم استخدام بيئة البيانات الضخمة حيث يمكن التعامل مع بيانات أكثر تعقيداً وطبيعة مختلفة واستخدام سمات أكثر غناء من النموذج المستخدم بالشركة وبالتالي تحسين نتائج توقع تسرب الزبائن في الشركة، كل ذلك يمكن الشكرة من ابتكار حملات تسويقية موجهة إلى مجموعة محددة من الزبائن المتوقع تسربهم من الشركة لتتمكن من الحفاظ عليهم.

- وحسب Al-Zuabi et al (2019) تم استخدام تقنية البيانات الضخمة في شركة اتصالات SyriaTel لتوقع عمر وجنس المستخدم، من حيث أن مستخدم الشبكة الحقيقي ليس بالضرورة أن يكون المالك الفعلي للخط. تم الاستفادة من تقنية البيانات في حساب وتحليل ما يقارب 1400 خاصية لكل زبون في مدة زمنية قدرها 3 أشهر وهو رقم كبير جداً، حيث يصعب على النموذج الموجود في الشركة تحليل هذه الكمية من البيانات واستخراج السمات المطلوبة من قاعدة البيانات لإنجاز العمل المطلوب. تلعب السمات الديموغرافية للعميل، مثل الجنس والعمر، دورًا أساسيًا قد يمكّن الشركات من تعزيز خدماتها واستهداف العميل المناسب في الوقت والمكان المناسبين، تم استخدام سجلات تفاصيل المكالمات (CDR) وإدارة علاقات العملاء (CRM) وبيانات الفوترة كمصدر للبيانات لتحليل سلوك عميل الاتصالات وأنواع مختلفة من خوارزميات تعلم الآلة لتزويد الحملات التسويقية بمعطيات أكثر دقة حول الخصائص الديموغرافية للعملاء.

3.3– مناقشة الفرضيات من معطيات الشركة:

1.3.3– الفرضية الأولى: يوجد تأثير لسرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة:

قام الباحث بطرح مجموعة من الأسئلة الفرعية تتعلق بالمحور الأول "تأثير سرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية" خلال المقابلة التي أجراها مع موظفي قسم البيانات الضخمة في شركة سيرتيل وهي:

1– في أي عام تم تطبيق تقنية البيانات الضخمة في الشركة؟

2– لماذا تم تطبيقها؟

3– كم استمرت الدراسة التجريبية قبل تطبيقها فعلياً؟

4– ما هي أهم ميزات تطبيقها؟

5– ما هي الطرق التقليدية لتخزين البيانات؟

6– ما هي طرق تخزين البيانات باستخدام منصة البيانات الكبيرة؟

7– كيف أثرت منصة البيانات الضخمة على سرعة تحليل البيانات؟

8– ما هو واقع الحملات التسويقية قبل تطبيق تقنية البيانات الكبيرة؟

9– كيف استفاد قسم التسويق من نتائج تحليل البيانات باستخدام منصة البيانات الكبيرة؟

10– كيف يتم إطلاق الحملات التسويقية باستخدام النتائج السابقة؟

بعد طرح الأسئلة السابقة قام الباحث بجمع الإجابات وتنسيقها مع المعلومات التي استخلصها من الاطلاع المباشر على قاعدة البيانات ثم قام بصياغة ما توصل إليه ضمن التساؤلات الرئيسية التالية:

1– ما هي ميزات تحليلات البيانات الضخمة وكيف تتم عملية تخزين الملفات باستخدام منصة البيانات الكبيرة:

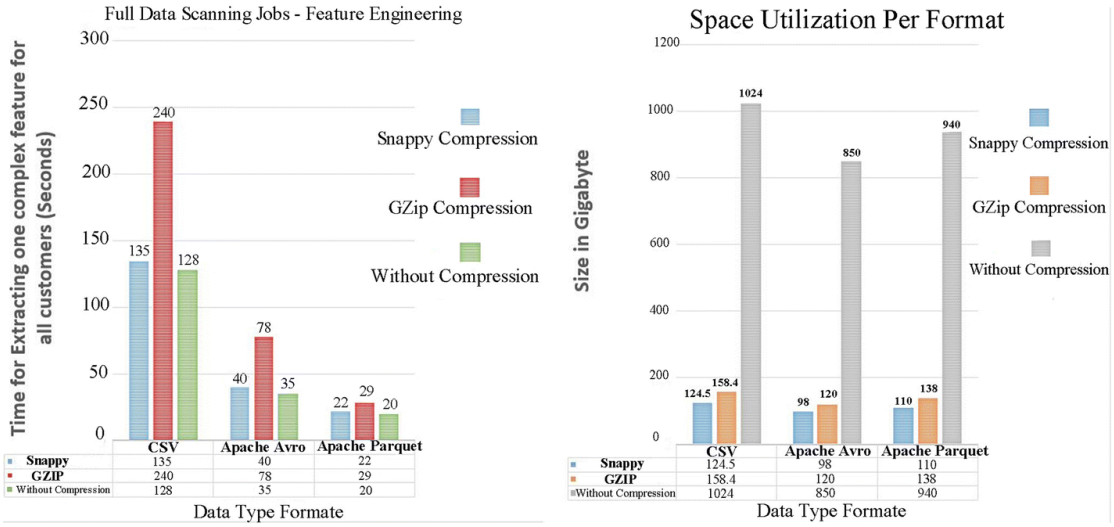
تم تطبيق تقنيات البيانات الضخمة في بداية عام 2019 بعد دراسة تجريبية استمرت لمدة عام ونصف تقريبا، قبل هذه الفترة كان حجم البيانات المولدة كبيراً جداً يصل تقريبا ل 50 غيغا بايت بالثانية، لم تستطع الشركة تخزين كل البيانات على مخدمات الشركة لسرعة وصول هذه البيانات وتنوعها وحجومها الكبيرة، فكانت تقتصر على تخزين المهم منها والتي يكمن حجمها ب 5 غيغا بالساعة.

بعد استخدام البيانات الكبيرة، غدت الشركة قادرة على تخزين هذه الحجوم الضخمة واستخراج البيانات التي تحتاجها من أصل 50 غيغا بايت في الثانية فكان حجم البيانات بعد التصفية واستخدام المهم منها حوالي 10 غيغا بايت بالثانية، وتم اهمال الباقي لعدم أهميته أو بسبب أمور لها علاقة بالتكرار.

تم أيضا الوصول لطرق جديدة لحفظ هذه البيانات فتمت مقارنة ثلاث طرق لحفظ هذه البيانات المخزنة باستخدام ثلاث طرق ضغط، فكان أهم انجاز تم تقديمه هو بإمكانية الاستعلام عن البيانات من ملفات مضغوطة من صيغة معينة دون تخزينها في بيئة بيانات تقليدية بسرعة كبيرة جداً.

في الشكل التالي (7) نعرض مقارنة لهذه التجربة مقدمة من الشركة على كتلة بيانات حجمها 1 تيرا ونلاحظ الفرق في التخزين وسرعة الأداء أثناء التنفيذ باستخدام تقنيات البيانات الكبيرة.

الشكل 7: مقارنة بين طرق حفظ البيانات في شركة SyriaTel



المصدر: قسم البينات الضخمة في شركة SyriaTel

إنّ Apache Parquet عبارة عن ملفات تسلسل بياناتٍ موجّهةٍ بالعمود تستخدم في حالات تحليل البيانات. تتضمّن التّحسينات الإضافيّة الفهرسة والضّغط المطبّق على سلسلةٍ من القيم من نفس الأعمدة ممّا يعطي نسب ضغطٍ جيّدةٍ جدّاً. وعند تخزين البيانات على نظام ملفات هادوب الموزع HDFS في تنسيق Parquet أعطى فعّاليّةً عاليةً في حالة التّحليل.

يوضح الشكل السابق عملية عليه مقارنة الحجم وسرعة الأداء بين تطبيق الطرق التقليدية وتقنية البيانات الضخمة لملفّ بياناتٍ يبلغ حجمه حوالي 1000غيغا بايت مخزن بصيغةٍ نصيّةٍ.

من خلال تحويل بيانات CSV إلى تنسيق أعمدة Parquetثمّ ضغطه وتقسيمه، تم توفير حجوم التّخزين والحصول على أداءٍ أفضل.

- نلاحظ من الشكل أنه تم توفير في الحجم كما هو موضح بالجانب الأيمن فكان نتيجة ضغط ملف CSV بالطريقة التقليدية هي 124.5غيغا بايت.

  بينما باستخدام تنسيق Parquet فقد وصل حجم الملف المضغوط إلى 110 غيغا بايت.
- الجانب الأيسر من الشكل يوضح سرعة فك ضغط ملف CSV بالطريقة التقليدية وهي 135 ثانية.

  بينما باستخدام تنسيق Parquet فقد وصلت سرعة فك الضغط إلى 22 ثانية وهو فارق كبير جداً.

2- كيف أثر تطبيق تقنية البيانات الضخمة في سرعة تحليل البيانات وانعكاس ذلك على الحملة التسويقية في الشركة؟

قدمت تقنيات البيانات الضخمة تسهيلات كبيرة في مجال سرعة التحليل، ففي بعض التطبيقات يحتاج قسم التسويق لإرسال عروض مباشرة للزبون في الزمن الحقيقي كمثال على ذلك: عند دخوله منطقة معينة أو استهلاك قيمة معينة فوق حد ما.

تكمن المشكلة الأساسية التي عالجتها تطبيقات البيانات الضخمة في المعالجة في الزمن الحقيقي، حيث أن الوقت اللازم لمعالجة ملايين السجلات الخاصة بالمكالمات والرسائل والانترنت يحتاج إلى ساعات باستخدام الأساليب التقليدية لسبب أن هذه البيانات ستمر بعدّة مراحل قبل أن تتاح للاستخدام من قبل فريق قسم التسويق، مثل مرحلة جلب البيانات وتحويلها وتحميلها على قواعد بيانات خاصة ومن ثم فهرستها ليتم أخيرا معالجتها.

أما في بيئة البيانات الضخمة فأن المعالجة تتم بنفس الوقت الذي يتم فيه جلب هذه البيانات بغض النظر عن حجمها أو طريقة تخزينها وتمثيلها، حيث تعتبر العروض المقدمة من الشركة على بيئة بيانات كبيرة بعيدة عدّة ثوانٍ من الحدث الذي سبب حدوثها، بينما تحتاج الطرق التقليدية لمعدل ساعة ونصف من الوقت لإصدار نفس العرض، وبالتالي أخذ القرار في الوقت غير المناسب.

كل هذه الأسباب أعطت الشركة القدرة على طرح عروض في الزمن الحقيقي سواء بالنسبة لسلوك الاستهلاك أو مكان وجود الزبائن.

3- ماهي النتائج المباشرة لسرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية؟

❖ **الحالة الأولى:**

طرح العروض بالوقت الحقيقي (real Time offers): يتم حساب متوسط عدد دقائق المكالمات اليومية للمستخدم على مدار شهر ونصف عن طريق تحليل سجلات تفاصيل المكالمات (CDR)، فعند تجاوز أي مستخدم لمتوسط عدد الدقائق باليوم (Average call per day) يتم باللحظة ذاتها طرح عرض للمستخدم يحتوي على نفس عدد الدقائق (المتوسط اليومي) ولكن بتكلفة مخفضة قد تصل إلى النصف. أي إذا كان معدل الاستخدام اليومي للمستخدم هو 100 ليرة سورية أو (10) دقائق، عند تجاوز المستخدم هذا الحد يتم طرح عرض مباشر لـ (10) دقائق أخرى بتكلفة 50 ليرة سورية.

المشكلة الأساسية التي كانت تواجه الشركة سابقاً أن الحصول على هذه المعلومات من سجل تفاصيل المكالمات (CDR) كانت تستغرق ما يقارب (3) ساعات وذلك لأنها كانت تخزن في قاعدة بيانات علائقية صعبة المعالجة

بالتالي أدى ذلك إلى التأخير. المشكلة في هذا التأخير هي إذا كانت المكالمة الأخيرة للمستخدم التي تخطى بها المعدل اليومي هي بعد الساعة (9) مساءً، فمدة (3) ساعات تأخير ستؤدي إلى طرح العرض في اليوم التالي وهذه تعتبر مشكلة. بينما باستخدام تقنية البيانات الضخمة تمكنت الشركة من المعالجة خلال (30) ثانية من آخر مكالمة يقوم بها المستخدم.

❖ **الحالة الثانية:**

طرح العروض على حدود المحافظات: عندما يسافر المستخدم إلى محافظة أخرى يتلقى عرض بلحظة دخوله إلى هذه المحافظة، وهو عبارة عن عرض ترحيبي لوصول المستخدم إلى المحافظة الوجهة. مثلاً مسافر وصل إلى دمشق قادماً من اللاذقية، يتلقى عرض لحظة دخوله لحدود محافظة دمشق. قامت الشركة بإجراء تجربة على مجموعة من المستخدمين وكانت نسبة الاستجابة لهذا العرض إلى ما يقارب (70%).

لتوليد هذا النوع من العروض تتم المعالجة في (ram memory on time) ذواكر الرام، ويعتمد على (patch processing) إي معالجة سابقة لكمية كبيرة من البيانات التاريخية (historical data) ذات حجوم كبيرة جداً عائدة إلى شهر تقريباً من الزمن. والتي باستخدام تقنية البيانات الضخمة تحولت إلى معالجة بالوقت الحقيقي.

2.3.3-يوجد تأثير لمعالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة:

قام الباحث بطرح مجموعة من الأسئلة الفرعية التي تتعلق بالمحور الثاني "تأثير معالجة البيانات واستخراج سمات أساسية منها على الرفع من كفاءة الحملة التسويقية" خلال المقابلة التي أجراها مع موظفي قسم البيانات الضخمة في شركة سيرتيل وهي:

1. كيف حسنت منصة البيانات الضخمة من تحليل الشبكة الاجتماعية؟
2. ما هي عوائق تحليل الشبكة الاجتماعية قبل استخدام منصة البيانات الكبيرة؟
3. كيف يتم استخراج سمات جديدة من الشبكة الاجتماعية باستخدام منصة البيانات الكبيرة؟
4. ما هي نتائج استخراج السمات وتحليلها؟
5. كيف انعكس ذلك على الحملة التسويقية؟

بعد طرح الأسئلة السابقة قام الباحث بجمع الإجابات وتنسيقها مع المعلومات التي استخلصها من الاطلاع المباشر على قاعدة البيانات ثم قام بصياغة ما توصل إليه ضمن التساؤلات الرئيسية التالية:

1- كيف يتم استخلاص سمات جديدة من شبكة التواصل الاجتماعية باستخدام البيانات الكبيرة؟

عمدت الشركة على استخدام تقنيات تعلم الآلة وذلك من أجل تقوية حملاتها التسويقية أو من أجل زيادة أرباحها، من أهم هذا المشاريع كان توقع تسرب الزبائن وتحديد جنس وعمر الزبون، حيث أن المشكلة الأكبر كانت تكمن في معرفة من هو الزبون الأكثر احتمالية لترك الشركة والانتقال للشركة المنافسة، حيث أن هذه المشكلة تخسر الشركة الكثير من الأموال وتضر سمعتها وانتشارها.

كان للبيانات الضخمة أثر مهم جدا في هذا التطبيق حيث أن حجم البيانات المستخدم كان أكبر بكثير وكانت المصادر التي تم الاعتماد عليها لاستخراج سمات النموذج التنبئي أكثر بكثير مما سبق في المحاولات السابقة، حيث تمكنت الشركة من العمل على سمات لها علاقة بشبكة التواصل الاجتماعي المبنية من الاتصالات والرسائل والمبينة في الشكل (8)، والتي تعبر واحدة من أكثر الشبكات تعقيدا وصعوبة في المعالجة، وهي أحد نتائج استخدام البيانات الكبيرة.

الشكل 8: تحليل شبكة التواصل الاجتماعي باستخدام البيانات الكبيرة

المصدر: قسم البينات الضخمة في شركة SyriaTel

يبين الشكل تحليل الشبكة الاجتماعية يوم بإجراء تحويل البيانات وإعدادها لتلخيص الاتصالات بين كل عميلين وإنشاء رسم بياني للشبكة الاجتماعية استنادًا إلى بيانات CDR التي تم التقاطها خلال أربعة أشهر.

يتكون الرسم البياني للشبكة الاجتماعية من العقد والحواف:

- العقد: تمثل عدد المشتركين.

- الحواف: تمثل التفاعلات بين المشتركين (المكالمات والرسائل القصيرة ورسائل الوسائط المتعددة).

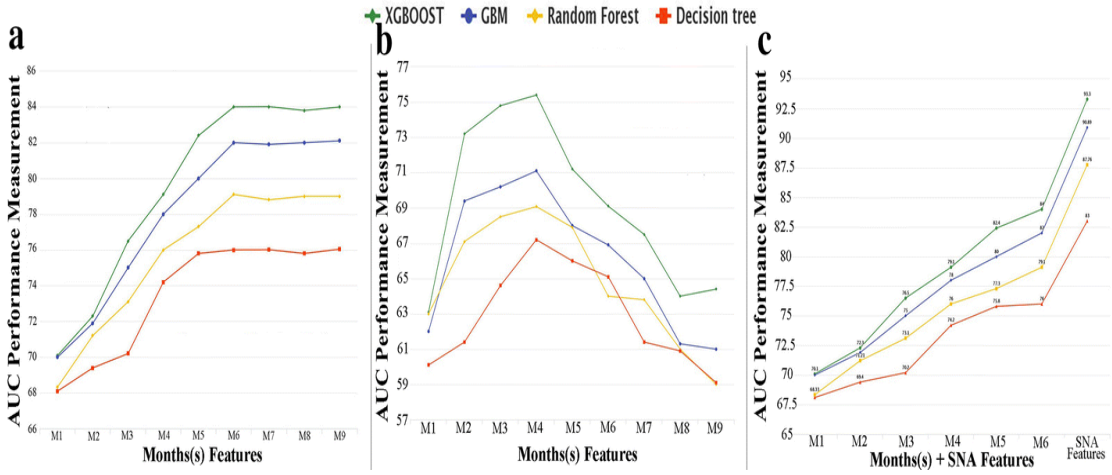
تم استخراج ما يقارب عشرين سمة من السمات الاجتماعية أدت في النهاية لزيادة دقة النموذج التنبئي المستخدم بنسبة 10% تقريبا وهي نسبة عالية جدا ساعدت في تحسين دقة الحملة التسويقية التي تهدف لاستهداف الزبائن المحتمل تسربهم.

2- ماهي النتائج قبل تطبيق التقنية وبعد تطبيق التقنية؟

لقد كانت التقنيات القديمة تعطي نتائج مقبولة ولكن غير مرضية حيث أن قلة البيانات المستخدمة، وصعوبة تحليل وتجميع بيانات تابعة لفترة زمنية طويلة كانت عائقا في نجاح أي نموذج قيد التطوير.

في الشكل (9) نعرض نتيجة إضافة سمات تابعة لفترة زمنية كبيرة وصلت ل 9 أشهر للبيانات التحليلية ودمج سمات من الشبكة الاجتماعية لفترة وصلت ل ستة أشهر، حيث تم توليدها باستخدام منصة البيانات الكبيرة، وهذا ما كان صعباً للغاية ومضيعة للوقت في البيئات السابقة.

الشكل 9: إضافة سمات تابعة لـ 9 أشهر مع سمات الشبكة الاجتماعية باستخدام منصة البيانات الكبيرة



المصدر: قسم البيانات الضخمة في شركة SyriaTel

3- كيف رفع ذلك من كفاءة الحملة التسويقية وماهي النتائج؟

زاد هذا الاستخدام دقة النموذج في المراحل التجريبية الأولى ليصل إلى 84% وبعد إضافة سمات الشبكة الاجتماعية، زادت حوالي 10% لتصل لـ93.3% وهي أعلى دقة وصلت لها الشركة حتى الأن.

4- ماهي النتائج المباشرة لاستخراج سمات أساسية من البيانات في الرفع من كفاءة الحملة التسويقية؟

❖ **طرح عروض للإناث في محافظة حماة:** بداية تم طرح التساؤل التالي من قسم التسويق على قسم الـ Big data، من هم الإناث المتواجدين في محافظة حماه؟ بالحالة التقليدية الإجابة على هذا السؤال تكون عن طريق طلبات التسجيل

أي المستخدم الأساسي صاحب الخط كما هو مسجل في الطب، ولكن فعلياً ليس بالضرورة أن يكون صاحب الخط هو المستخدم الفعلي، وهذه الحالات منتشرة بكثرة فالمستخدم قد يمتلك عدة شرائح موزعة على أولاده أقاربه أو غير ذلك. وهنا يصبح التفريق صعب جدا ما بين المستخدم الفعلي (user) والمالك الحقيق (owner) بالتالي يجب أن يكون هناك طريقة للتفرقة بين المستخدم والمالك وما هو عمره.

واجه قسم التسويق مشكلتين للقيام بالعرض السابق: أولاً من هم الذين يعيشون في حماة وثانياً ما هو جنس المستخدم. لحل هذه المشكلة قامت الشكرة بتصميم نظام توقع الجنس (Gender prediction) الذي يستخدم لغات تعلم الآلة (machine learning). هنا ظهرت سمات إضافية مكنت الشركة من زيادة دقة النموذج المستخدم. حيث تكمنت من تحديد (700 إلى 800) ألف أنثى من القاطنين في حماة.

قامت الشركة بعد ذلك بإجراء اختبار للتأكد من العينة السابقة فكانت دقة العينة بنسبة (94%) صحيحة (ذكور أو إثاث).

تم ذلك باستخدام سمات مستخلصة من منصة البيانات الضخمة (Big data platform)، من هذه السمات (حركة المستخدم خلال اليوم والمسافات التي يقطعها، معدل المكالمات باليوم، مع من يتكلم، كم عدد الأصدقاء الذين يتكلم معهم) بالتالي تم تحقيق دقة عالية لتحديد جنس العينة وإذا كان المستخدم قاطن فعلياً في حماة، رفع ذلك من كفاءة الحملة التسويقية من حيث الدقة الضخمة في استهداف العينة مثلاً في حال توجيه عروض للذكور فقط أو الإناث فقط أو توجيه عروض للقاطنين الفعليين في حماة وهذا ما تم تطبيقه فعلياً وتم توجيه عروض على هذا الأساس.

3.3.3-يوجد تأثير لانخفاض تكلفة تطبيق هذه التقنية مقارنة بتقنيات أخرى على كفاءة الحملة التسويقية في الشركة محل الدراسة:

قام الباحث بطرح مجموعة من الأسئلة الفرعية التي تتعلق بالمحور الثاني "تأثير انخفاض تكلفة تطبيق تحليلات البيانات الضخمة على الرفع من كفاءة الحملة التسويقية خلال المقابلة " التي أجراها مع موظفي قسم البيانات الضخمة في شركة سيرتيل وهي:

1- ما هي تكلفة تطبيق تقنية البينات الضخمة في الشركة؟

2- كيف يتم حساب التكلفة؟

3- ما هي تكاليف الطرق التقليدية

4- كيف انعكست تلك التكلفة على الشركة؟

5- كيف انعكست تلك التكلفة على قسم التسويق؟

6– بعد طرح الأسئلة السابقة قام الباحث بجمع الإجابات وتنسيقها مع المعلومات التي استخلصها الباحث من الاطلاع المباشر على قاعدة البيانات ضمن الأسئلة الرئيسية التالية:

1– ما هي تكلفة تطبيق تقنية البيانات الضخمة في تحليل البيانات؟

إن تكلفة البيانات الضخمة كانت قليلة جدا حيث تم استخدام مخدمات منهية العمر كانت قيد الاتلاف، حيث ان مواصفات المخدم الواحد غير مناسبة للاستخدام في أي تطبيق مفرد ولكن تقنيات البيانات الضخمة تعتمد على استثمار الموارد الحسابية بطريقة موزعة وبالتالي تم نصب حوالي عشرين مخدم ضعيف الإمكانات ليقوموا بمعالجة البيانات على التوازي بكفاءة عالية.

2– ما هو الفرق بين تكلفة تطبيق تقنية البيانات الضخمة مقارنة بالطرق التقليدية في تحليل البيانات؟

الطرق القديمة كانت مكلفة جداً لأن زيادة أداء المخدم الواحد بالطرق التقليدية كان أيضا مكلفاً جداً، حيث أن شراء ذاكرة عشوائية الوصول بحجم 1 تيرا مكلف جدا وقد يصل لوحده لمبلغ 30 ألف دولار، وهو نفس المبلغ الذي كلف الشركة تقريبا لبناء منصة بيانات كبيرة مكونة من عشر عقد (مخدمات) بذاكرة نصف تيرا وبقدرة معالجة عالية تصل ل 300 نواة معالجة.

كيف رفع ذلك من كفاءة الحملة التسويقية؟

1- أثر ذلك بطريقة غير مباشرة من حيث انخفاض تكاليف تخزين البيانات وتحليلها وبالتالي انخفاض تكلفة العتاد المستخدم وانعكاس ذلك على التكاليف العامة للشركة وكفاءة معالجة البيانات وسرعتها بالتالي رفع كفاءة الحملة التسويقية أي أن العلاقة عكسية ما بين انخفاض تكلفة استخدام البيانات الضخمة ورفع كفاءة الحملة التسويقية.

2- يمكن للحملات التسويقية التي تركز على الجماهير أن تستهدف العميل في الموقع الصحيح وفي الوقت المناسب بالرسالة الصحيحة. هنا يمكن أن نقول أن البيانات الضخمة تقدم مساعدة كبيرة للشركة عن طريق خفض التكلفة بالتالي زيادة حصتها في السوق من خلال التنفيذ السليم لترويج المبيعات واستجابة العملاء. يمكن تقديم المنتجات أو الخدمات المخصصة الجديدة للعملاء، مما يقلل تكاليف العملاء ويزيد من رضاهم.

4.3- اختبار الفرضيات من منظور الزبائن:

توصيف المتغيرات الديمغرافية للاستبيان الذي أجراه الباحث على زبائن الشركة:

- **الجنس:**

جدول (2): توزيع أفراد العينة حسب الجنس

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ذكر	25	58.5	58.5	58.5
	أنثى	17	41.5	41.5	100.0
	Total	41	100.0	100.0	

المصدر: الجدول من إعداد الباحث بناءً على نتائج على نتائج التحليل الإحصائي باستخدام برنامج (spss)

يبين الجدول السابق أن نسبة (61%) من أفراد العينة هي من الذكور، وأن نسبة (39%) من أفراد العينة هي من الإناث.

- **العمر:**

جدول (3): توزيع أفراد العينة حسب العمر

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	أقل من (25) سنة	12	29.3	29.3	29.3
	بين (25) و(34) سنة	28	68.3	68.3	97.6
	بين (35) و(44) سنة	1	2.4	2.4	100.0
	Total	41	100.0	100.0	

المصدر: الجدول من إعداد الباحث بناءً على نتائج على نتائج التحليل الإحصائي باستخدام برنامج (spss)

يبين الجدول السابق أن نسبة (29.3%) من أفراد العينة هم من دون (25) سنة وبلغ عددهم (12)، أن نسبة (68.3%) من أفراد العينة تتراوح أعمارهم بين (25) و(34) سنة وبلغ عددهم (28)، وأن نسبة (2.4%) من أفراد العينة تتراوح أعمارهم بين (35) و(44) سنة وبلغ عددهم (1)، هذا يعني أن النسبة الأكبر من أفراد العينة هم الذين تتراوح أعمارهم بين (25) و(34).

- **العمل:**

جدول (4): توزيع أفراد العينة حسب العمر

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	طالب	17	41.5	41.5	41.5
	مدرس	6	14.6	14.6	56.1
	موظف	13	31.7	31.7	87.8
	غير ذلك	5	12.2	12.2	100.0
	Total	41	100.0	100.0	

المصدر: الجدول من إعداد الباحث بناءً على نتائج على نتائج التحليل الإحصائي باستخدام برنامج (spss)

يبين الجدول السابق أن نسبة (41.5%) من أفراد العينة هم من الطلاب، وأن نسبة (14.6%) من أفراد العينة هم من المدرسين، وأن نسبة (31.7%) من أفراد العينة هم من الموظفين، وأن نسبة (12.2%) من أفراد العينة هم غير ما سبق. يوضح الجدول أن النسبة الأكبر من أفراد العينة هم من الطلاب.

اختبار ثبات الاستبانة:

" ثبات الاستبانة "فيقصد بها أن تعطي هذه الاستبانة نفس النتيجة فيما لو تم إعادة توزيعيا أكثر من مرة تحت نفس الظروف والشروط، أو يمكن القول بعبارة أخرى أن ثبات الاستبانة يعني الاستقرار في نتائج الاستبانة وعدم تغييرها بشكل كبير فيما لو تم إعادة توزيعيا على أفراد العينة عدة مرات خلال فترات زمنية معينة.

وبشكل عام يمكن قياس ثبات الاستبانة بثلاث طرق هي :الاختبار واعادة الاختبار، التجزئة النصفية، ومعامل ثبات ألفا كرونباخ .قام الباحث بتطبيق الطريقة الثالثة لقياس ثبات الاستبانة وهي حساب معامل ثبات ألفا كرونباخ. يتم حساب معامل ثبات ألفا كرونباخ عن طريق برنامج (SPSS)، وعادة ما تكون قيمته مفضلة إذا زادت عن (0.70) ، حيث تزداد قيمة هذا المعامل كلما ازدت عبارات الاستبانة، مما يدل أن الاستبانة تشمل كل تفاصيل البحث، كما تزداد قيمة هذا المعامل أيضا كلما كانت إجابات أفراد العينة متباينة وغير متجانسة.

وللتأكد من ثبات الاستبانة قام الباحث بحساب معامل ثبات ألفا كرونباخ لكل محور من محاور الاستبانة من جهة ولجميع فقرات الاستبانة من جهة ثانية وكانت النتائج كما هي مبينة في الجدول الآتي:

جدول (5): معامل ألفا كرونباخ لقياس ثبات الاستبانة

م	المحور	معامل ألفا كرونباخ
1	تأثير سرعة التعامل مع البيانات في الرفع من كفاءة الحملة التسويقية	0.699
2	تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية	0.728
3	المحورين السابقين معاً	0.818

المصدر: الجدول من إعداد الباحث بناءً على نتائج على نتائج التحليل الإحصائي باستخدام برنامج (spss)

نلاحظ أن قيمة معامل ألفا كرونباخ للمحاور يتراوح بين (0.699) و(0.728). أما قيمة هذا المعامل للمحورين معاً فقد بلغت (0.818).

يمكن القول أن الاستبانة تتمتع بدرجة جيدة جداً من الصدق والثبات، مما يعني أنها قابلة للتوزيع على أفراد العينة وصالحة للحصول على بيانات مطلوبة.

1.4.3-الفرضية الأولى: يوجد تأثير معنوي لسرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة:

الاحصاءات الوصفية للمتغير الأول "تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية":

لتوصيف متغيرات البحث قام الباحث بتوصيف بيانات الدراسة حسب متوسطات إجابة كل مجموعة من العبارات، إضافة إلى حساب الانحراف المعياري ومعامل الاختلاف الذي يعطى بالعلاقة الآتية:

$$cv = \frac{\alpha}{\mathrm{m}} . 100$$

يبين الجدول الآتي قيمة الحد الأدنى والحد الأعلى والمتوسط الحسابي والانحراف المعياري ومعامل الاختلاف لإجابات أفراد العينة على العبارات الخاصة بمحور **"تأثير سرعة التعامل مع البيانات في الرفع من كفاءة الحملة التسويقية"**، وهو الذي يمثل المتغير المستقل الأول:

جدول (6): تّوصيف اّلمتغير اّلمستقل اّلأول: تأثير سرعة التعامل مع البيانات في الرفع من كفاءة الحملة التسويقية

	N	Minimum	Maximum	Mean	Std. Deviation	C.V%
أتلقى عروض تتناسب مع معدل صرفي اليومي (مكالمات – رسائل – انترنيت)	41	1	5	3.3415	1.01513	30.37

أتلقى عروض تتزامن مع انتهاء استهلاك سابق (في الوقت المناسب)	41	1	5	3.3415	.93834	28.08
أتلقى عروض تتزامن مع انتهاء عرض سابق	41	1	5	3.3171	.90662	27.33
أتلقى عروض تناسب توقيت عملي (ليلاً – نهاراً)	41	1	5	3.3171	.98588	29.72
أتلقى عروض تتزامن مع وجودي في منطقة جغرافية ما	41	1	5	3.3415	.96462	28.86
العروض التي أتلقاها تعد مغرية	41	1	5	3.3415	1.17494	35.16
أستجيب بسرعة للعروض التي أتلقاها	41	1	5	3.2439	.83007	25.58
Valid N (listwise)	41			3.3206		

المصدر: الجدول من إعداد الباحث بناءً على نتائج على نتائج التحليل الإحصائي باستخدام برنامج (spss)

يتضح من الجدول السابق أن المتوسط الحسابي لإجابات أفراد العينة على العبارات الخاصة بمحور " تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية" قد بلغ (3.3206)، وهذا يعني أن إجابات أفراد العينة فيما يتعلق بالمحور السابق كان أقرب إلى الموافق، مما قد يعطي مؤشر إلى وجود تأثير لسرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية.
كما يتضح من الجدول السابق انخفاض قيم معامل الاختلاف نسبياً لأغلب عبارات المحور السابق، الأمر الذي يدل على تجانس إجابات أفراد العينة ووحدة رأيهم حول مضمون عبارات هذا المحور. فالقيمة الضخمة لمعامل الاختلاف تدل على التشتت واختلاف الرأي وعدم التجانس في إجابات أفراد العينة حول مضمون العبارة، بعكس القيمة الصغيرة التي تدل على الاتفاق والتجانس ووحدة الرأي.
لاختبار هذه الفرضية قام الباحث بحساب متوسط إجابات أفراد العينة على محور تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية" ومن ثم اختبار وجود فرق بين المتوسط المحسوب ومتوسط الحياد باستخدام اختبار ستودنيت (One-Sample-Test)، وهذا ما يبينه الجدول الآتي:

جدول (7): "متوسط اجابات أفراد العينة على محور "تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية"

	N	Mean	Std. Deviation	Std. Error Mean
متوسط اجابات أفراد العينة على محور "تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية	41	3.3206	.97366	.15206

المصدر: الجدول من إعداد الباحث بناءً على نتائج على نتائج التحليل الإحصائي باستخدام برنامج (spss)

يبين الجدول التالي أن متوسط إجابات أفراد العينة على المحور الأول قد بلغ (3.3206)، وهو ما يعني أن إجابات أفراد العينة تميل باتجاه الموافقة ما يدل على وجود تأثير لسرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة.

ولكن بما أن حساب المتوسط الحسابي ومعرفة قيمته وميله باتجاه الموافقة من عدمها يعد شرطاً لازماً ولكنه غير كاف فقد قام الباحث باختبار وجود فرق بين قيمته وقيمة متوسط الحياد (3) في مقياس ليكرت المستخدم. وهو ما يبينه الجدول التالي:

One-Sample Test

جدول (8): نتائج اختبار الفرضية الأولى الخاصة بمحور تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية

	Test Value = 3					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
متوسط إجابات أفراد العينة على محور "تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية"	2.113	40	.044	.32056	.0132	.6279

المصدر: الجدول من إعداد الباحث بناءً على نتائج على نتائج التحليل الإحصائي باستخدام برنامج (spss)

يتضح من الجدول السابق أن القيمة الاحتمالية (Sig. = 0.044) أغر من مستوى الدلالة ($\alpha = 0.05$) وهو ما يعني قبول الفرضية البديلة الأولى التي تقول: يوجد تأثير معنوي لسرعة تحليل البيانات على الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة.

مقارنة نتيجة الفرضية الأولى بين منظور الشركة ومنظور والزبائن:

يتضح من النتائج التي تم الوصول إليها من خلال تحليل البيانات الداخلية للشركة أن هناك فوارق كبيرة بين سرعة تحليل البيانات باستخدام منصة البيانات الضخمة مقارنة مع الطرق التقليدية، من حيث كمية السجلات الهائلة التي كانت تتطلب ساعات من التحليل بالطرق التقليدية والتي تحولت إلى تحليل في الزمن الفعلي، وهو ما أثر بشكل مباشر على كفاءة الحملة التسويقية ويتضح ذلك من خلال العروض التي تم تقديمها للزبائن بناء على هذه النتائج والمتمثلة بعروض الزمن الحقيقي، وهو ما يمكن تأكيده بناء على نتائج الاستبيان الذي أجراء الباحث على عينة من زبائن الشركة، فيما اتضح للباحث من خلال نتائج الاستبيان أن التأثير موجود ولكنه ليس كبير حيث بلغ معدل الإجابات (0.32056) وهي نتيجة تميل للموافقة ولكنها أقرب للحياد، مما يدل أنه لا يزال هناك قصور في طرح العروض التسويقية ويعود السبب في ذلك إلى أن تحليلات

البيانات الضخمة بدأ استخدامها في المجال التسويقي منذ بداية عام (2019) ولم تشمل العروض حتى الآن كافة الزبائن مما انعكس على نتائج تحليل الاستبيان الذي أجراء الباحث، كما أن التحيز بالإجابات من قبل بعض الزبائن ضد الشركة بناء على أسباب وقناعات مسبقة قد يكون سبباً أيضاً. كما اتضح للباحث أن متوسط اجابات الزبائن على السؤال المتعلق بتلقي عروض تتزامن مع انتهاء عرض سابق والسؤال المتعلق بتلقي عروض تتناسب مع توقيت العمل كان الأقل حيث بلغ (3.3171) وهو دليل على وجود ضعف في تقديم عروض مناسبة بالنسبة للنقطتين السابقتين بناء على إجابات الزبائن.

2.4.3-يوجد تأثير لمعالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة:

الإحصاءات الوصفية للمتغير الثاني "تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية".

يبين الجدول الآتي قيمة الحد الأدنى والحد الأعلى والمتوسط الحسابي والانحراف المعياري ومعامل الاختلاف لإجابات أفراد العينة على العبارات الخاصة بمحور " **تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية**"، وهو الذي يمثل المتغير المستقل الثاني:

جدول (9): توصيف المتغير المستقل الثاني: تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية

	N	Minimum	Maximum	Mean	Std. Deviation	C.V%
أتلقى عروض تناسب جنسي (ذكر – أنثى)	41	1	5	3.3902	.91864	27.09
اتلقى عروض تناسب سني	41	1	4	3.4146	.99939	29.26
أتلقى عروض تناسب طبيعة عملي (طالب – موظف – ...)	41	2	5	3.4390	1.00122	29.11
أتلقى عروض تتناسب مع المنطقة الجغرافية التي أتواجد بها	41	1	5	3.3415	.91131	27.27
العروض التي أتلقاها تعد مغرية	41	1	5	3.3415	1.17494	35.16
أستجيب بسرعة للعروض التي أتلقاها	41	1	5	3.2439	.83007	25.58
Valid N (listwise)	41			3.3618		

المصدر: الجدول من إعداد الباحث بناءً على نتائج على نتائج التحليل الإحصائي باستخدام برنامج (spss)

يتضح من الجدول السابق أن المتوسط الحسابي لإجابات أفراد العينة على العبارات الخاصة بمحور " **تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية** " قد بلغ (3.3292)، وهذا يعني أن إجابات أفراد العينة فيما يتعلق بالمحور السابق كان أقرب إلى الموافق، مما قد يعطي مؤشر إلى وجود تأثير لمعالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية.

كما يتضح من الجدول السابق انخفاض قيم معامل الاختلاف نسبياً لأغلب عبارات المحور السابق، الأمر الذي يدل على تجانس إجابات أفراد العينة ووحدة رأيهم حول مضمون عبارات هذا المحور. فالقيمة الضخمة لمعامل الاختلاف تدل على التشتت واختلاف الرأي وعدم التجانس في إجابات أفراد العينة حول مضمون العبارة، بعكس القيمة الصغيرة التي تدل على الاتفاق والتجانس ووحدة الرأي.

لاختبار هذه الفرضية قام الباحث بحساب متوسط إجابات أفراد العينة على محور "تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية" ومن ثم اختبار وجود فرق بين المتوسط المحسوب ومتوسط الحياد باستخدام اختبار ستودنيت (One-Sample-Test)، وهذا ما يبينه الجدول الآتي:

جدول (10): "متوسط اجابات أفراد العينة على محور "تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية"

	N	Mean	Std. Deviation	Std. Error Mean
متوسط اجابات أفراد العينة على محور "تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية"	41	3.3618	.97260	.15189

يبين الجدول التالي أن متوسط إجابات أفراد العينة على المحور الأول قد بلغ (3.3618)، وهو ما يعني أن إجابات أفراد العينة تميل باتجاه الموافقة ما يدل على وجود تأثير لسرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة.

ولكن بما أن حساب المتوسط الحسابي ومعرفة قيمته وميله باتجاه الموافقة من عدمها يعد شرطاً لازماً ولكنه غير كاف فقد قام الباحث باختبار وجود فرق بين قيمته وقيمة متوسط الحياد (3) في مقياس ليكرت المستخدم. وهو ما يبينه الجدول التالي:

One-Sample Test

جدول (11): نتائج اختبار الفرضية الثانية الخاصة بمحور تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية

	Test Value = 3					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
متوسط إجابات أفراد العينة على محور " تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية "	2.388	40	.031	.36179	.0548	.6688

يتضح من الجدول السابق أن القيمة الاحتمالية (Sig. = 0.031) أغر من مستوى الدلالة ($\alpha = 0.05$) وهو ما يعني قبول الفرضية البديلة الثانية التي تقول: "يوجد تأثير معنوي لمعالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية في الشركة محل الدراسة".

مقارنة نتيجة الفرضية الثانية بين منظور الشركة ومنظور والزبائن:

يتضح من النتائج التي تم الوصول إليها من خلال البيانات الداخلية للشركة أن تحليل الشبكة الاجتماعية مكن الشركة من استخراج سمات إضافية تمكنت من خلالها من تحديد عمر وجنس الزبائن الفعليين والتي فتحت أبواب أمام الشركة لطرح عروض للزبائن في هذا الإطار كما تمكنت من خلال هذا التحليل من تحديد الزبائن المحتمل تسربهم من الشركة وامكانية طرح عروض خاصة لهذه الفئة من الزبائن، وهذا ما أثر بشكل مباشر على كفاءة الحملة التسويقية وانعكس من خلال العروض التي تم تقديمها للزبائن بناء على هذه النتائج والمتمثلة بالعروض المقدمة للإناث في محافظة حماه والعروض المتعلقة بالمقيمين ضمن المحافظة، وهو ما يمكن تأكيده بناء على نتائج الاستبيان الذي أجراء الباحث على عينة من زبائن الشركة والذين بلغ معدل إجاباتهم (0.36197) وهو معدل يميل إلى الموافقة وكاف لإثبات الفرضية ولكن لا يمكن اعتباره معدل كبير، ولازال

يدل على بعض القصور في طرح العروض التسويقية ويعود السبب في ذلك إلى أن تحليلات البيانات الضخمة بدأ استخدامها في المجال التسويقي منذ بداية عام (2019) ولم تشمل العروض حتى الآن كافة الزبائن مما انعكس على نتائج تحليل الاستبيان الذي أجراء الباحث، إضافة إلى أن اختلاف عينة الباحث والتي تعد عشوائية عن العينة المستهدفة من قبل الشركة يعد سبباً آخر، كما اتضح للباحث أن عدد اجابات الزبائن على السؤال المتعلق بتلقي عروض تتناسب مع المنطقة الجغرافية كان الأقل حيث بلغ (3.3415) وهو دليل على وجود ضعف في تقديم عروض تناسب النقطة السابقة.

النتائج:

1- أثرت سرعة تحليل البيانات باستخدام منصة البيانات الضخمة بشكل مباشر على الحملة التسويقية في الشركة وذلك من خلال إتاحة امكانية تحليل ملايين السجلات وتقديم النتائج لقسم التسويق في الزمن الحقيقي بالتالي طرح العروض التسويقية في الزمن الحقيقي بالنسبة لسلوك الاستهلاك أو مكان وجود الزبائن.

2- تمكنت الشركة من خلال منصة البيانات الضخمة من توليد سمات إضافية للمستخدمين وكانت من أهم نتائجها توقع تسرب الزبائن من الشركة كما عالجت مشكلة رئيسية بالتعرف على عمر وجنس المستخدم الحقيقي بدقة عالية من خلال تحليل سلوكهم واستخراج سمات تعبر عن هذه السلوكيات وهو ما يمثل ركيزة أساسية لمعظم الحملات التسويقية.

3- تمكنت الشركة من خلال منصة البيانات الضخمة من بناء نظام متكامل يحقق جميع مراحل دورة حياة البيانات من تحصيل البيانات مروراً بالمعالجة والتحليل والتخزين وحتى واجهات الويب التي تؤمن سهولة الاستخدام للمختصين في عمل الحملات التسويقية.

4- ساهمت منصة البيانات الضخمة في تخفيض تكلفة استخدام العتاد الصلب بشكل كبير وذلك بسبب اعتمادها على استثمار الموارد الحسابية بطريقة موزعة بالتالي امكانية استخدام مخدمات منهية العمر مما أدى إلى تخفيض تكاليف الشركة إلى أضعاف التكلفة الفعلية. كما أثر ذلك على كفاءة الحملة التسويقية بطريقة غير مباشرة من خلال العلاقة العكسية التي تنطوي على ميزات تطبيق تحليلات البيانات الضخمة .

5- تم الحصول على نتائج مقبولة من تحليل الاستبيان الموزع على الزبائن فيما يخص المحور الأول "تأثير سرعة تحليل البيانات في الرفع من كفاءة الحملة التسويقية" ولكن اتضح من خلال النتائج وجود قصور في تقديم العروض بشكل عام والعروض التي تتعلق بوقت العمل بشكل خاص.

6- تم الحصول على نتائج مقبولة أيضاً من تحليل الاستبيان الموزع على الزبائن فيما يخص المحور الثاني "تأثير معالجة البيانات واستخراج سمات أساسية منها في الرفع من كفاءة الحملة التسويقية" فيما اتضح للباحث وجود قصور في العروض التي تتعلق بهذا المحور عموماً والعروض التي تتعلق بمكان الإقامة بشكل خاص.

7- ويعود السبب في النتائج التي توصل إليها الباحث من خلال الاستبيان الموجه إلى الزبائن إلى أن العروض لم تشمل جميع عملاء الشركة أي أنها لازالت توجه إلى فئات من الزبائن بالإضافة إلى طرح بعض العروض التي تعد ربحية على المستوى القريب بالنسبة للشركة كالعروض التي تطرح في المناطق التي تحتوي عدد قليل من

الزبائن (شراء خطوط جديدة بعروض مغرية) بقصد الاستحواذ على عدد من الزبائن والتي تشكل أثر رجعي على الشركة من حيث أن نتائج التحليل قد تكون مضللة.

التوصيات:

1- زيادة البحث في مجال تحليل شبكات التواصل الاجتماعي لمعرفة تأثير وأهمية كل زبون وتأثره بالوسط المحيط به وتحديد فيما إذا كان هناك إمكانية لزيادة دقة النتائج.

2- امكانية الاعتماد على السمات المستخرجة باستخدام منصة البيانات الضخمة لبناء نظام لتوليد العروض بشكل آلي بحيث يقوم بالاستفادة من سمات الزبون التي تميز سلوكه وجنسه وفئته العمرية لإعطائه العرض المناسب له.

3- يجب على الشركة التوسع في نتائج تحليلات البيانات الضخمة لتحسين العروض المقدمة للزبائن والتركيز على توقيت عمل الزبائن وأماكن الإقامة الفعلية واستجابة الزبائن للعروض للرفع من كفاءة الحملة التسويقية الحالية.

4- يجب استخدام نتائج تحليلات البيانات الضخمة في توقع الأوقات من خلال البيانات الواردة من أبراج الاتصال التي تتعرض فيها الشبكة لأكبر كمية ازدحام ثم العمل على توجيه الحزم لتفادي حصول تصادمات ضمن الشبكة.

5- يجب أن تستفيد الشركة أيضاً من نتائج هذه التحليلات في تحديد العملاء الأكثر عرضة لمواجهة مشاكل في عمليات الدفع وإيجاد الحلول المناسبة لتحسين استرداد المدفوعات.

6- تستطيع الشركة من خلال معالجة البيانات الموزعة جغرافياً في الزمن الحقيقي باستخدام منصة البيانات الضخمة من فهم المشاكل الخدمية بشكل أفضل وكيفية تأثيرها على المستخدم، كما يمكن مراقبة تغيير أنماط استخدام الشبكة بسرعة وإعادة تكوين دعم الشبكة لمعالجة الأحداث القصيرة الأجل مثل الاستخدام المكثف بالقرب من استاد أثناء حدث رياضي.

7- بدء العمل على استخدام نتائج تحليلات البيانات الضخمة للتنبؤ باحتياجات العملاء والعمل على علاج الخلل الذي يتسبب به معدل الاستجابة المنخفض أو الرديء لدى شركة الاتصالات. فإذا كان بالإمكان التنبؤ باحتياجات العملاء، لكن من دون استجابة تلبي هذه الاحتياجات فإن التنبؤ سيظل مجرد وظيفة لا نفع لها ويتطلب إحداث التغيير بين العملاء إحداث تحول في جهود التسويق والانتقال من تتبع توجهات السوق إلى الحرص على تلبية الاحتياجات الفردية.

8- تستطيع الشركة من الاستفادة من ميزة البيانات الضخمة مع التحليلات المتراكبة في التسارع إلى تحديد فرص أعمال جديدة وزيادة نمو منظومتها لمصلحة العملاء حيث يُظهر اللجوء إلى خدمات جديدة من جهات خارجية حتى عند تقديمها إلى جمهور صغير تمحوراً حول العميل ينم عن عقلية مبتكرة.

9- وضع نتائج البحث أمام مختلف المنظمات التي تعتمد في عملها على تحليل البيانات من مصادر مختلفة والتي قد تتجه إلى تطبيق هذا النوع من التقنيات للاستفادة منها كالجهات الحكومية ومؤسسات الأعمال حيث يمكن أن تستفيد من هذه التحليلات في زياد أرباحها وتعزيز وضعها التنافسي من خلال معرفة رغبات الزبائن وميولهم وتوجهاتهم الاجتماعية والنفسية بالتالي توفير منتجات وخدمات بناء على تلك الرغبات، بذلك تضمن رضا العملاء بعيداً عن التخمين والمجازفة.

المراجع:

[1] Banumathi, S & Aloysius, A. (2017)"PREDICTIVE ANALYTICS CONCEPTS IN BIG DATA-A SURVEY", **International Journal of Advanced Research in Computer Science**, p. 8.

[2] Bughin, J. (2016) "Reaping the benefits of big data in telecom", **journal of big data**, PP. 14-15-16.

[3] J.M. Cavanillas, E. Curry, and W. Wahlster. (2016**) New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe**: Springer International Publishing, ISBN 9783319215686.

[4] Chen Tianqi, Guestrin Carlos (2016) XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, San Francisco, California, USA.

[5] Ciaccio, A.D. & Coli, M. & Ibanez, A. & Miguel, J. (2012) "Advanced Statistical Methods for the Analysis of Large Data-Sets", **Springer**.

[6] Chen, C.P. & Zhang, C. Y. (2014) "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", **Information Sciences**, pp. 275-314-347.

[7] Curry, E., Ngonga, A., Domingue, J., Freitas, A., Strohbach, M., Becker, T., et al. (2014) ". Big Data Technical Working Groups, Final version of the technical white paper. Public deliverable of the EU-Project BIG, **CSA**.

[8] Glavic, B., Rabl, T., Poess, M., Baru, C. & Jacobsen, H.A. (2014) "Big data provenance: Challenges and implications for benchmarking. Specifying big data benchmarks in Berlin ", **Springer**, pp. 72–80.

[9] Grobelnik, M. (2012). Big data tutorial. Jozef Stefan Institute, Ljubljana, Slovenia. Stavanger.

[10] Geng Bo, Li Yangxi, Tao Dacheng, Wang Meng, Zha Zheng-Jun, Xu Chao. (2012) "Parallel lasso for large-scale video concept detection", **IEEE Trans**. Multimedia
14 (1) PP. 55–65.

[11] JAGADISH, H.V & GEHRKE, J & LABRINIDIS, A & PAPAKONSTANTINOU, Y & PATEL, J.M & RAMAKRISHNAN, R AND SHAHABI, C (2014) " Exploring the inherent technical challenges in realizing the potential of Big Data ", **COMMUNICATIONS OF THE ACM,** VOL. 57, NO. 7, p. 89.

[12] Janssen, M., van der Voort, H. & Wahyudi, A.(2017)" Factors influencing big data decision-making quality", **Journal of Business Research**, pp. 338-345.

[13] Lau, L., yang-Turner, F. & Karacapilidis, N. (2014) "Requirements for big data analytics supporting decision making: A sensemaking perspective. In: Karacapilidis, N, (ed.) Mastering Data-Intensive Collaboration and Decision Making", **Springer Science & Business Media**, pp. 49 - 70. Springer, Cham.

[14] Moorthy, M., Baby, R., Senthamaraiselvi, S. (2014) "An Analysis for Big Data and its Technologies", **International Journal of Computer Science Engineering and Technology,** Vol 4, pp. 412-416-418.

[15] Mora, A. C., Chen, Y., Fuchs, A., Lane, A., Lu, R., Manadhata, P. et al. (2012). "Top Ten Big Data Security and Privacy Challenges", **Cloud Security Alliance (CSA).**

[16] Rehman, H.R. Liew, C.S. Abbas, A. Jayaraman, P.P. War, T.Y. Khan, S.U (2016) "Big Data Reduction Methods ", **Springer**, P. 267.

[17] Simoff, S & Bohlen, M.H &. Mazeika, A. (2008) "Visual Data Mining: Theory, Techniques and Tools for Visual Analytics", **Springer**.

[18] Sircar Anindya. Rastogi Gaurau. Kochikar, V.P. Joshi Raj. Ranganath. M, Towers Simon, Goparaju Subu, (2013), **Big Data Challenges and Opportunities**: Infosys.

[19] Stryk, B. (2015). How do organizations prepare and clean Big Data to achieve better data governance? A Delphi Study (doctoral dissertation, Capella University).

[20] Yu Dong, Deng Li. (2011) "Deep learning and its applications to signal and information processing", **IEEE Signal Process. Mag**. 28 (1) pp. 145–154.

[21] Yin, shen. Kaynak, okyay. (2015) " Big Data for Modern Industry: Challenges and Trends ", **Proceedings of the IEEE**, Vol. 103, p. (145).

[22] Whitworth, J. N. (2013). Applying hybrid cloud systems to solve challenges posed by the big data problem. The University of North Carolina at Greensboro, ProQuest Dissertations Publishing.

[23] Zhou, J & Chen, P. & Chen, L. & Li, H.X & Zhao, W. (2013) "A collaborative fuzzy clustering algorithm in distributed network environments", IEEE Trans. **Fuzzy Syst**. PP. 99.

مواقع الانترنت:

[1] **Nasainarabic. (2016)** (big data) ما هي البيانات الضخمة, www.nasainarabic.net. Access date 12/6/2019.

[2] **Statista. (2015)** Average Monthly Churn Rate for Wireless Carriers, www.statista.com. Access date 13/6/2019.

[3] **Teradata. (2017)** Featured Big Data products, Database management software, www.teradata.com. Access date 14/6/2019.

[4] **Techworld. (2010)** How big is a zettabyte?, www.techworld.com. Access date 11/6/2019.

الملحق (1):

الأسئلة الموجهة إلى موظفي قسم البيانات (Big Data) في شركة SyriaTel عند إجراء المقابلة.

1- في أي عام بدأ تطبيق تقنية البيانات الضخمة في الشركة؟

2- لماذا تم تطبيقها؟

3- كم استمرت الدراسة التجريبية قبل تطبيقها فعلياً؟

4- ما هي أهم ميزات تطبيقها؟

5- ما هي الطرق التقليدية لتخزين البيانات؟

6- ما هي طرق تخزين البيانات باستخدام منصة البيانات الكبيرة؟

7- كيف أثرت منصة البيانات الضخمة على سرعة تحليل البيانات؟

8- ما هو واقع الحملات التسويقية قبل تطبيق تقنية البيانات الكبيرة؟

9- كيف استفاد قسم التسويق من نتائج تحليل البيانات باستخدام منصة البيانات الكبيرة؟

10- كيف يتم إطلاق الحملات التسويقية باستخدام النتائج السابقة؟

11- كيف حسنت منصة البيانات الضخمة من تحليل الشبكة الاجتماعية؟

12- ما هي عوائق تحليل الشبكة الاجتماعية قبل استخدام منصة البيانات الكبيرة؟

13- كيف يتم استخراج سمات جديدة من الشبكة الاجتماعية باستخدام منصة البيانات الكبيرة؟

14- ما هي نتائج استخراج السمات وتحليلها؟

15- كيف انعكس ذلك على الحملة التسويقية؟

16- ما هي تكلفة تطبيق تقنية البينات الضخمة في الشركة؟

17- كيف يتم حساب التكلفة؟

18- ما هي تكاليف الطرق التقليدية

19- كيف انعكست تلك التكلفة على الشركة؟

20- كيف انعكست تلك التكلفة على قسم التسويق؟

الملحق (2):

الاستبيان الذي تم توزيعه على زبائن شركة SyriaTel:

ملاحظة: تم تطبيق تقنية البيانات الضخمة منذ بداية عام (2019) وهو ما تم التنويه إليه للزبائن قبل الإجابة على أسئلة الاستبيان من قبل الزبائن.

المعلومات الشخصية		
الجنس		ذكر
		أنثى
العمر		أقل من (25) سنة
		من (25) إلى (34) سنة
		من (35) إلى (44) سنة
		من (45) إلى (54) سنة
		أكثر من (55) سنة
العمل		طالب
		مدرس
		موظف
		غير ذلك

البيان	غير موافق بشدة 1	غير موافق 2	محايد 3	موافق 4	موافق بشدة 5
أتلقى عروض تتناسب مع معدل صرفي اليومي (مكالمات – رسائل – انترنيت)					
أتلقى عروض تتزامن مع انتهاء استهلاك سابق (في الوقت المناسب)					
أتلقى عروض تتزامن مع انتهاء عرض سابق					
أتلقى عروض تناسب توقيت عملي (ليلاً – نهاراً)					
أتلقى عروض تتزامن مع وجودي في منطقة جغرافية ما					
أتلقى عروض تناسب جنسي (ذكر – أنثى)					
اتلقى عروض تناسب سني					
أتلقى عروض تناسب طبيعة عملي (طالب – موظف – ...)					
أتلقى عروض تتناسب مع المنطقة الجغرافية التي أتواجد بها					
العروض التي أتلقاها تعد مغرية					
أستجيب للعروض التي أتلقاها					