

تنقيب بيانات وسائل التواصل الإجتماعى واستخداماته فى البحوث الإعلامية ... تحليل المشاعر نموذجاً

د. طارق الخليفى

قسم الإعلام - كلية الآداب - جامعة الإسكندرية

ملخص الدراسة :

تأتى هذه الدراسة فى إطار منهجية التكامل المعرفى بين علوم الاتصال والإعلام وغيره من التخصصات المعرفية الأخرى و يتحدد موضوعها فى تنقيب البيانات وتحليل المشاعر فى وسائل التواصل الإجتماعى بوصفه مجالاً بحثياً حديثاً لم يحظ بعد بالاهتمام الكافى من جانب باحثى الاتصال وهى دراسة وصفية تحليلية تهدف الى التأسيس النظرى لذلك الاتجاه البحثى و التعريف بأهم المفاهيم المستخدمة فيه، وأبرز الطرق والآليات المستخدمة فى تنقيب بيانات وسائل التواصل الاجتماعى ومناقشة أهم القضايا المنهجية والأخلاقية المرتبطة بمجال تحليل المشاعر، والدراسة الحالية بمثابة تحليل كیفى للتراث النظرى من خلال



عينة غرضية من الأبحاث والدراسات والمقالات المنشورة التى تيسر للباحث الحصول عليها من بنك المعرفة المصرى ومصادر أخرى .

إن أبحاث الاتصال المحوسبة وتحديدًا تنقيب بيانات وسائل التواصل الاجتماعى، يقع فى مفترق طرق بين تخصصات مختلفة أهمها تعلم الآلة ومعالجة اللغة الطبيعية وتحليل الشبكات الاجتماعية وهو يعد بكثير من الفرص وي طرح مجموعة من التحديات فى الوقت ذاته ينبغى استكشافها بحثًا وتحليلًا ونقدًا.

الكلمات الدالة :

البيانات الكبيرة **Big Data** ، وسائل التواصل الاجتماعى **Social Media**

تنقيب البيانات **Data Mining** ، تحليل المشاعر **sentiment analysis**

استخراج الرأى **Opinion mining** ، الإعلام الجديد **New media**



المقدمة :

ظهرت وسائل التواصل الاجتماعي Social media في إطار ما يعرف بالإعلام الجديد New media بوصفها منصات إلكترونية أو قنوات تتيح للناس التواصل ومشاركة المحتوى ، كما وأنها تسهل التفاعلات المتعددة بين المستخدمين.

لقد شكلت وسائل التواصل الاجتماعي تحدياً لوسائل الإعلام التقليدية القائمة بما أفرزته من قضايا وما أثارته من إشكاليات ولم تقتصر هذه التحديات على تغيير الإعلام إنتاجاً واستهلاكاً وتأثيراً بل امتدت لتشمل الدراسات الإعلامية على مستوى المفاهيم والمناهج والنظريات.

و بعد أن حررت الهواتف النقالة وسائل التواصل الاجتماعي من قيود الزمان والمكان ورافقت المستخدمين سواء في منازلهم أو اجازاتهم أو أثناء حضورهم لإحدى الفعاليات ، فإن الخبرات والظواهر والأحداث التي يمر بها الناس في هذه الأيام تتحول بشكل فوري تقريباً إلى محتوى في وسائل التواصل الاجتماعي، تلك الوسائل التي أصبحت أداة لمشاركة خبرات الحياة اليومية .

إن نظرة سريعة على معدلات انتشار الإنترنت عالمياً ، وعلى النمو المتواصل في استخدام منصات التواصل الاجتماعي يعطى مؤشراً على التغير الذي حدث في بيئة الدراسات الإعلامية ، حيث تشير الاحصائيات إلى أنه، في حين كان إجمالي عدد من يستخدمون الإنترنت عام ٢٠١٧ نحو ٣,٨ مليار شخص من جميع أنحاء العالم (Salem, 2017,p:5) فقد ارتفع عدد المستخدمين الى ٤,٢ مليار في اكتوبر عام ٢٠١٨ م. ولقد كشفت أحدث مجموعة من تقارير Global Digital عن زيادة اعداد مستخدمي الإنترنت بمعدل يزيد عن مليون مستخدم جديد يومياً حتى أنه مع حلول شهر



يناير من عام ٢٠١٩م أصبح هناك ٤.٣٩ مليار مستخدم للإنترنت و بزيادة قدرها ٣٦٦ مليون مستخدم جديد وبنسبة زيادة تعادل ٩% عن التوقيت ذاته من عام ٢٠١٨م. (kemp, 2019)

و من حيث استخدام وسائل التواصل الاجتماعي ، أشار تقرير الإعلام الاجتماعي للعام الماضي (We are social 2018) أنه في شهر سبتمبر كان ثمة ٣,٤ مليار شخص يستخدمون وسائل التواصل الاجتماعي من مختلف أنحاء العالم وبزيادة ١٠% مقارنة بذات التاريخ من عام ٢٠١٧م وأن الفيس بوك لا يزال الأكثر شعبية بين منصات التواصل الاجتماعي ويمثل اثنين من كل ثلاثة يستخدمون مختلف وسائل التواصل الاجتماعي عبر العالم. وبمجيء شهر يناير من عام ٢٠١٩م زادت أعداد مستخدمي وسائل التواصل الاجتماعي عالمياً لتصل الى ما يقرب من ٣.٥ مليار ، لتبلغ نسبة الانتشار عالمياً نسبة ٤٥ في المائة. وبحسب Global Web Index هناك زيادة في وقت استخدام شبكات التواصل حيث يبلغ متوسط وقت الإستخدام الآن ساعتين و١٦ دقيقة يومياً على هذه المنصات الاجتماعية وهو ما يعادل ثلث إجمالي وقت استخدام الإنترنت تقريباً ، ولا زال موقع Facebook يحتفظ بأعلى ترتيب للمنصات في أوائل عام ٢٠١٩ حيث زاد عدد المستخدمين النشطين بأكثر من ٤ في المائة خلال الأشهر الثلاثة الماضية ليصل إلى ٨٩٥ مليون مستخدم نشط حول العالم (kemp, 2019)

إن أحد نتائج استخدام وسائل التواصل الاجتماعي على نطاق واسع ، توافر قدر كبير جداً من البيانات من جميع الأنواع ، نصية وبصرية وعاطفية وحميمية وخاصة (Kennedy, 2012,p:435) ، فمع وجود مئات الملايين من الناس يقضون ساعات لا تحصى على هذه الوسائل ليتشاركوا ويتباطوا ويتواصلوا ويتفاعلوا معاً فإن ذلك أوجد بيانات يولدها المستخدمون بمعدلات غير مسبوقه (*) الى الحد الذي يمكن



أن نقول معه أننا نعيش عصر البيانات الكبيرة Age of Big Data. لقد أصبحت وسائل الإعلام الاجتماعية أحد المصادر الجديدة والفريدة التي توفر قدراً ضخماً من البيانات تعد مصدراً ثرياً يتيح فرصاً لامثيل لها وإمكانات كبيرة للبحث والتطوير. (Zafarani, &Abbasi, 2014 p: xi)

و من هذا المنطلق نشأ نوع جديد من أنواع الأعمال وهو تنقيب بيانات وسائل التواصل الاجتماعي . ويمثل تحليل المشاعر Sentiment Analyses أحد أنواع تنقيب البيانات Data Mining الذى يستخدم مجموعة من التقنيات لتحديد المشاعر تجاه موضوعات معينة يتم التعبير عنها داخل وسائل التواصل الاجتماعي وذلك بهدف التوصل الى الشعور العام تجاه هذه الموضوعات.

و يعد هذا المجال البحثى (تنقيب البيانات وتحليل المشاعر) مجالاً متعدد التخصصات Interdisciplinary Field يرتبط بفروع معرفية مختلفة مثل معالجة اللغة الطبيعية Natural language Processing وتنقيب النص واللغويات الحاسوبية Text Mining & computational Linguistics ، ومن ثم يشمل تقنيات من علوم الحاسب الآلى واستخراج البيانات والتعلم الآلى، وتحليل الشبكات الاجتماعية وعلم الاجتماع والإثنوجرافى Ethnography ، وعلم الإحصاء والرياضيات.

تنقسم الدراسة الحالية الى ستة أجزاء نعرض فى الجزء الأول للإطار المنهجى الذى يشتمل على المفاهيم والاستراتيجية المنهجية والأهمية والأهداف والتساؤلات، ويحتوى الجزء الثانى على تأصيل تاريخى للموضوع منذ البدايات الأولى لدراسات تحليل المشاعر وحتى توسع الدراسات والنظر الى المجال بوصفه حقلاً معرفياً ناشئاً ينتمى الى العلوم الاجتماعية المحوسبة ، يلى ذلك وفى الجزء الثالث توضيح لمستويات ومراحل تنقيب البيانات وتقنياته، ويتناول الجزء الرابع



الاتجاهات البحثية الرئيسية لدراسات تحليل المشاعر ، ويختص الجزء الخامس بعرض رؤية نقدية لدراسات تنقيب البيانات وتحليل المشاعر وتختتم الدراسة بالجزء السادس الذي يمثل رؤية مستقبلية شاملاً على أهم النتائج والتوصيات.

أولاً : الإطار المنهجي للدراسة:

١- مفاهيم الدراسة:

وسائل التواصل الاجتماعي Social Media .

على الرغم من أن التعريف لا يزال نوعاً ما في حالة تغير ، إلا أن مصطلح Social Media أو وسائل التواصل الاجتماعي يعرف على نطاق واسع بأنه منصات إلكترونية Electronic Platforms أو قنوات تتيح للناس التواصل وتبادل المحتوى كما وأنها تسهل إقامة مجموعة متنوعة من العلاقات عبر الانترنت (Pendergast,t.& Pendergast,s 2000 , p :619)

يصف ذلك المصطلح أنواع تطبيقات شبكة الويب (www) مثل المدونات Blogs والمدونات الصغيرة Microblogs مثل تويتر Twitter، ومواقع الشبكات الاجتماعية Social Network Sites ، أو منصات تشارك الفيديو والصور والملفات أو الويكي Wikis (Fuchs S ch .2014 , p: 33)

إن وسائل التواصل الاجتماعي وبحسب ما جاء في تقرير الإعلام الاجتماعي العربي " هي تطبيقات قائمة على الانترنت تسهل إنشاء شبكات منظمة من الأفراد أو الكيانات أو الأشياء الافتراضية المرتبطة على أساس الخصائص

الاجتماعية ، وهي تسهل التفاعلات المتعددة بما فى ذلك إنشاء والمحتوى ومشاركته من قبل المستخدمين ". (Salem, 2017,p:8)

البيانات الكبيرة Big Data

إن البيانات الكبيرة هي مصادر البيانات ذات الحجم والسرعة الكبيرة . وهي تمثل مجموعة متنوعة من البيانات التي تتطلب أدوات وطرق جديدة لاستخراجها وتنظيمها وإدارتها ومعالجتها بطريقة فعالة . ويعرفها الاتحاد الدولي للاتصالات بأنها نموذج لتمكين وجمع وتخزين وإدارة وعرض مجموعات واسعة من البيانات ذات الخصائص غير المتجانسة . (Salem, 2017,p:55)

لقد استخدم مصطلح "البيانات الكبيرة " لأول مرة فى منتصف التسعينيات للإشارة إلى معالجة مجموعات البيانات الضخمة وتحليلها، تلك البيانات التي تميزت بثلاثة خصائص تتمثل في ضخامة الحجم ، إذ تتكون من تيرا بايت terabytes أو بيتابايت petabytes من البيانات ، والسرعة العالية حيث يتم إنشاؤها في الوقت الحقيقي ، والتنوع فى الأنماط أو الأشكال ، فضلاً عن كونها مهيكلة أو غير مهيكلة في طبيعتها، وفي الآونة الأخيرة ، تم إضافة خصائص أخرى لتكون شاملة في نطاقها تشمل مجموعات أو أنظمة كاملة . (Kitchin, 2016, p:27)

إن مصطلح البيانات الكبيرة Big Data له معنى نسبي و يميل للإشارة الى مجموعات من البيانات تكبر و تكبر بمرور الوقت، و المصطلح يستخدم فى سياق علوم الحاسبات للإشارة إلى مجموعات البيانات الكبيرة جدا التى تختلف عن المجموعات الصغيرة من البيانات و بالتالى تتطلب أدوات مختلفة لمعالجتها و لاكتشاف الأنماط داخلها و هي تسمح بنوعية من التحليلات يستحيل إجراؤها على نطاق ضيق. (Mahrt & Scharkow, 2013,p:22)



تنقيب البيانات

يستخدم مصطلح تنقيب البيانات (DT) Data mining ، أو اكتشاف المعرفة من البيانات (KDD) knowledge discovery from data أو اكتشاف المعرفة وتنقيب البيانات (KDDM) knowledge discovery and data mining كمصطلح شامل يصف العديد من الأنشطة والتقنيات لاستخراج المعلومات من البيانات واقتراح أنماط في قواعد بيانات كبيرة للغاية ، والتي تشمل على تجمعات ضخمة، ومستودعات للبيانات data warehouses والتحليل الإحصائي وتقنيات التعلم الاستنتاجي deductive learning techniques . (Harvey, 2014,p:349)

و يشير جيمس واتسون (James Watson & Anne Hill 2012,p:71) الى أن تنقيب البيانات هو "مجموعة من التقنيات الآلية لاستخراج المعلومات المخفية أو التي لم تكن معروفة مسبقاً من قواعد البيانات الكبيرة " إن عمليات تنقيب البيانات الناجحة تجعل من الممكن استخراج معلومات ذات قيمة أو الكشف عن الأنماط والعلاقات ومن ثم استخدام هذه المعلومات الجديدة لاتخاذ قرارات الأعمال القائمة على المعرفة الاستباقية .

تحليل المشاعر Sentiment Analysis

تحليل المشاعر والمعروف أيضاً باسم استخراج الرأى Opinion Mining هو المهمة الأساسية لتصنيف نص أساسى آلياً أو تلقائياً ، وتحديد ما إذا كان اتجاهه أو ميله إيجابياً أو سلبياً أو محايداً . (Oliveira, Bermejo, & Santos, 2017,p:4) وتحليل المشاعر هو تحليل للنص يهدف إلى تحديد اتجاه أو موقف المتكلم أو الكاتب فيما يتعلق ببعض الموضوعات أو قطبية السياق الكلى لجزء من النص، ويهتم تحليل المشاعر أو استخراج الرأى بتطوير البرامج

لاستخراج آراء المستخدمين تلقائياً حول المنتجات أو الكيانات *entities* الأخرى من النص وعادةً ما تكون من مصادر عبر الإنترنت . وإن كان تحليل المشاعر يمكن أن يختلف عن استخراج الرأي إذ يشتمل على البرامج التي تستخرج المشاعر من النص لأغراض أخرى مثل تقدير الحالة العاطفية لمؤلف النص ذاته. (Thelwall, 2017,p:344-346)

ويختلف تحليل المشاعر أيضاً عن تحليل العواطف *Emotion Analysis* وتصنيفها . فالمشاعر تشير الى آراء الأشخاص حول موضوع ما (فيلم أو منتج جديد مثلاً) ويكون إما إيجابياً أو سلبياً ، بينما تحليل العواطف يصف شعور الشخص الأكثر عمقاً من تحليل المشاعر وهو أكثر انفتاحاً للتأويل (مثل العواطف المتعلقة بالحزن والسعادة والغضب) وللتمييز بين العواطف والمشاعر نقول أن عواطف "الحزن" و"الغضب" تعد أمثلة للعواطف المعبرة عن مشاعر سلبية عموماً ولكن من الواضح أن كلا من الحزن والغضب تمثل عواطف مختلفة. (Abdullah, et.al,2018,p:1)

و تتمثل الوظيفة الرئيسية لتحليل المشاعر في تحديد ما يفكر فيه الناس، واستخلاص وجهات نظرهم حول موضوعات معينة باستخدام التكنولوجيا ، إذ يقلل تحليل المشاعر بشكل كبير الحاجة الى مسح أعداد كبيرة من المستندات لاستخلاص الآراء . إن تحليل المشاعر باختصار عبارة عن تقنية تسعى الى مواجهة التحدي المتمثل في تحديد معلومات ذاتية واستخراجها من كميات كبيرة من البيانات غير المنظمة من خلال الجمع بين تقنيات استخراج البيانات وتعلم الآلة ومعالجة اللغات الطبيعية واسترجاع المعلومات وإدارة المعرفة. (Oliveira, Bermejo, & Santos, 2016,p:4) ويشير استخراج الرأي إلى مهمة إيجاد آراء الأشخاص حول



كيانات محددة باستخدام اللغويات الحاسوبية ومعالجة اللغات الطبيعية لتحديد المعلومات الذاتية واستخلاصها من النصوص المصدر ، ويهدف تحليل المشاعر الى تحديد قطبية الفرد فيما يتعلق بالحكم أو الحالة العاطفية ، ويستخدم للتحليل الآلي للنصوص والموارد التقييمية وغالبا ما ينظر الى استخراج الرأى وتحليل المشاعر على أنهما مترادفان ويعبران عن معنى متبادل ايضا.(Zhu, Y., Kim, M.C. & Chen, C. 2017)

٢ - الإستراتيجية المنهجية :

نوعية الدراسة: الدراسة الحالية هي دراسة وصفية تحليلية لا تنتمي الى ما يسمى الدراسات الكمية للعلم ، ونظراً لحدائثة الموضوع وعدم وجود تراكم معرفى فى الدراسات الاتصالية فيما يختص بتقيب بيانات وسائل التواصل الاجتماعى وتحليل المشاعر كان من المناسب من وجهة نظر الباحث اختيار عينة غرضية من الدراسات الأجنبية تفى بأهداف الدراسة وتساعد على استكشاف الجهود البحثية الناشئة فى تلك النوعية من الدراسات_ سواء ما يتعلق منها بتوضيح طبيعة ذلك المجال منذ نشأته وأهم مصطلحاته وباحثيه أو مداخله التقنية ومراحله الإجرائية _ مع اختيار أهم الدراسات التى توضح كيف يمكن أن يسهم تقيب البيانات فى إثراء حقل دراسات الاتصال مع ملاحظة أن اختيار الباحث لعينة الدراسات خضع لعدة معايير أهمها أن تكون الدراسة حديثة فى الفترة من عام ٢٠٠٠ وحتى ٢٠١٨ م _ يوجد هنا استثناء فى دراسة واحدة عام ٢٠٠٨ _ وأن تحتوى فى كلماتها الدالة على مصطلحات تقيب البيانات ، تحليل المشاعر، وسائل التواصل الاجتماعى أو أى منهم وتم البحث فى قواعد البيانات باللغة الانجليزية .

صعوبات واجهت الباحث

كان من المتصور في بداية الدراسة الاكتفاء بالدراسات التي يمكن الحصول عليها من بنك المعرفة المصري غير أنه وفي ضوء بعض الصعوبات البحثية المتعلقة بأن بنك المعرفة لا يغطي الاشتراك في بعض قواعد البيانات العلمية (مثل SAGE Reference) وبعض الدوريات في تخصص الاتصال التي تحتوى على مراجع وثيقة الصلة بالموضوع تم الاستعانة بقواعد البيانات الرقمية الخاصة بالمكتبة الرقمية السعودية وأيضاً الاشتراك في بوابة الباحثين (<https://www.researchgate.net>) المخصصة لتبادل الأبحاث الحديثة ، كما تم استخدام كل من محرك بحث جوجل العلمى والعام .

أهمية الدراسة:

تأتى هذه الدراسة فى إطار التعريف بأحد الاتجاهات البحثية الحديثة فى دراسات الإعلام الجديد وتحديدًا ذلك الاتجاه الذى يندرج ضمن دراسات تحليل الشبكات الاجتماعية ويركز على المحتوى الذى ينتجه مستخدمو شبكات التواصل الاجتماعى مستخدماً تقنيات وتطبيقات الحوسبة ومعالجة اللغة الطبيعية للتنقيب فى هذا المحتوى واستخراج الآراء والمشاعر المتضمنه فيه.

اتضح للباحث من خلال مسح التراث والأدبيات المنشورة حول الموضوع أن هناك جهوداً بحثية نشطة للغاية فى مجال تنقيب البيانات وتحليل المشاعر وهى تتوزع فى مجالات معرفية متنوعة أبرزها يتركز فى علوم الحاسبات ومعالجة اللغة الطبيعية وتطبيقات تعلم الآلة والذكاء الإصطناعى ، وفى مقابل ذلك هناك محاولات بحثية من جانب العلوم الاجتماعية - وإن كانت بدرجة أقل- للاستفادة من هذه التطورات فى



أبحاثها. و يكاد يكون حقل الدراسات الإعلامية هو الاقل مواكبة لهذه التطورات ويصل إلى حد الندرة فيما لو كانت دراسات إعلامية عربية. ويشكل ذلك فجوة علمية تتطلب العمل في هذا الاتجاه .

تتطلق الدراسة من ضرورة التأكيد على التكامل المعرفي بين دراسات الإعلام وغيره من حقول معرفية إذ يتيح ذلك المجال الجديد- بل يحتم - ضرورة التعاون بين الباحثين في تخصصات مختلفة والعمل على موضوعات بحثية مشتركة تفيد من خبراتهم جميعا كل في مجاله. إن أبحاث الاتصال المحوسبة تعد بكثير من الفرص وتطرح مجموعة من التحديات في الوقت ذاته ينبغي استكشافها بحثاً وتحليلاً ونقداً.

الأهداف

يتمثل الهدف الرئيس للدراسة الحالية في التعريف بمجال تنقيب البيانات وتحليل المشاعر بوصفه مجالاً بحثياً حديثاً لم يحظ بعد بالاهتمام الكافي من جانب باحثي الاتصال . وثمة عدة اهداف فرعية أهمها التأصيل النظري لذلك المجال، والكشف عن بنيته الفكرية وخصائصه المعرفية ، وما يطرحه من قضايا يمثل بعضها فرصاً بحثية إلى جانب ما يفرضه من تحديات منهجية وأخلاقية .

التساؤلات

- ١- متى نشأت أبحاث تنقيب البيانات وتحليل المشاعر؟
- ٢- ما أهم المفاهيم التي ترتبط بذلك المجال البحثي؟

٣- ما مراحل تنقيب بيانات وسائل التواصل الاجتماعي وما مستوياته وأبرز

تقنياته؟

٤- ما الخصائص المعرفية لأبحاث تحليل المشاعر ؟

٥- كيف يمكن الاستفادة من تنقيب البيانات في أبحاث الاتصال؟

٦- ما الاتجاهات البحثية الرئيسية لأبحاث تنقيب البيانات وتحليل المشاعر؟

٧- ما أهم القضايا المنهجية والتحديات الأخلاقية في تنقيب بيانات وسائل التواصل

الاجتماعي؟

ثانيا : تنقيب البيانات وتحليل المشاعر : تأصيل تاريخي.

١-الإرهاصات الأولية :

في عام ٢٠٠٨ م وفي إحدى المقالات المبكرة حول تنقيب البيانات أشار بو بانج Bo Pang إلى عدم وجود أي مظهر من مظاهر الاتفاق العالمي على المصطلحات في عالم مراقبة وتحليل وسائل الإعلام الاجتماعية social media monitoring and analysis، فقد وُصِفَ ذلك المجال الناشئ بمصطلحات عدة من "مراقبة العلامات التجارية brand monitoring"، و"مراقبة الطنين buzz 'monitoring' و الأنثروبولوجيا على الإنترنت" إلى "تحليلات المؤثرات السوقية 'market influence analytics'" و"تنقيب المحادثة conversation mining" و"ذكاء المستهلك عبر الإنترنت." ولم يكن هناك مصطلحات موحدة لذلك الحقل المعرفي الناشئ. (Pang & Lee, 2008,p:5)

كان التنقيب عن البيانات ، كحقل ناشئ ، رائداً في البحث والتطور في اعتماده على الأساليب والخوارزميات التي تعالج كميات هائلة من البيانات وتسهم في حل مشاكل العالم الحقيقي. ويسعى الباحثون المنقبون في البيانات إلى استخراج معلومات ذات



مغزى من مجموعة بيانات لا تبدو واضحة ولا يمكن الحصول عليها بسهولة.
(Barbier & Liu, 2011,p:328)

وتتنوع المصطلحات التي قد يجدها القارئ المهتم في الأوراق البحثية في ذلك المجال ، مثل المعالجة الحاسوبية computational treatment للرأي opinion ، والمشاعر, sentiment ، والذاتية في النص. subjectivity in text . والذى أصبح يُعرف باسم استخراج الرأي وتحليل المشاعر opinion mining, sentiment analysis. أو تحليل الذاتية subjectivity analysis . وقد تم استخدام تعبير مراجعة الآراء واستخراج التقييم أو استخلاص التوصيات أيضا review mining and appraisal extraction ، وهناك أيضاً بعض الارتباطات بالحوسبة العاطفية affective computing ، التي تتضمن تمكين أجهزة الكمبيوتر من التعرف على المشاعر وفهم كيفية التعبير عنها . وتعكس كثرة المصطلحات الاختلافات في الدلالات التي تحملها هذه المصطلحات ، سواء في الاستخدام الأصلي للخطاب العام أم في الاستخدامات التي تطورت فيما بعد في الأدبيات التقنية لكثير من المجتمعات .
(Pang & Lee, 2008,p:5)

ترجع بداية ظهور مصطلح Opinion Mining الى ورقة بحثية كتبها ديف Kushal dave وآخرون تم نشرها في وقائع مؤتمر الشبكة العنكبوتية العالمية لعام ٢٠٠٣ ؛ قد يفسر مكان النشر شعبية مصطلح استخراج الرأي داخل المجتمعات المرتبطة بشدة ببحث الويب أو استرجاع المعلومات . و وفقاً لـ Dave "ستقوم أداة التنقيب المثالية ، بمعالجة مجموعة من نتائج البحث لعنصر معين ، وتوليد قائمة بسمات المنتج (الجودة ، والميزات ، وما إلى ذلك) وتكوين آراء حول كل منها (ضعيف ، مختلط ، جيد) كان ذلك الوصف مناسباً لكثير من الأبحاث اللاحقة التي



نظرت الى استخراج الرأي على أنه استخراج الأحكام التقييمية الخاصة بجوانب مختلفة لعناصر معينة وتحليلها. ومؤخراً اكتسب هذا المفهوم تفسيراً أوسع ليشمل أنواعاً مختلفة من التحليل التقييمي للنص. (Pang & Lee, 2008, p: 6)

ويمثل تاريخ عبارة "تحليل المشاعر" تعبير "استخراج الرأي" في بعض الجوانب، حيث يظهر مصطلح "المشاعر" "sentiment" المستخدم في الإشارة إلى التحليل التلقائي للنص التقييمي وتتبع الأحكام التنبؤية الواردة فيه. وفي الوقت الحاضر يتم تفسير المصطلح بشكل أوسع ليعني المعالجة الحاسوبية للرأي، والمشاعر، والذاتية في النص. وهكذا، عندما يتم تطبيق التفسيرات الواسعة، يشير "تحليل المشاعر" و"استخراج الرأي" إلى ذات المجال الدراسي. (Pang & Lee, 2008, p: 6)

٢- التوجه التجارى المبكر لدراسات تنقيب البيانات.

كان مجال الأعمال التجارية هو الأسبق في الاهتمام بدراسات تنقيب البيانات، مدفوعاً بهدف استخراج بيانات معينة ذات أهمية بالغة لتطبيقات الأعمال من خلال تحليل الأنشطة التجارية التاريخية المخزنة على شكل بيانات ثابتة في مستودع قواعد البيانات وذلك لتحقيق عدة أهداف، منها الكشف عن أنماط واتجاهات خفية أو معلومات تجارية إستراتيجية لم تكن معروفة من قبل والأمثلة على ذلك هو ما تستخدمه الشركات لاستخراج بيانات تشمل إجراءات تحليلات للسوق لتحديد حزم المنتجات الجديدة ومنع خسارة أو فقدان العملاء واكتساب عملاء جدد. فمن خلال برامج تنقيب البيانات وخوارزميات (*) التعرف على الأنماط يمكن أيضاً التنبؤ وبمستويات دقة عالية بمدى ولاء العملاء وتحديد اتجاهات بيع المنتجات وتطوير الحملات التسويقية. Library, W. (E,2017)



و في السنوات الأخيرة كان هناك زيادة في استخدام تحليل المشاعر من جانب شركات العلامات التجارية Branding Marketing Companies وشركات التسويق والإعلانات ، وقد نشأ تحليل المشاعر في سياق ما يقال أنه يتميز بالافتراض في رسائل الشركات ورسائل التسويق وانخفاض نسبة الثقة في الإعلانات في مقابل ازدياد الثقة في توصيات الأقران ومحادثاتهم . إن الشركات حريصة على معرفة ما يقال عن منتجاتها في وسائل التواصل الاجتماعي والتعرف على طبيعة المشاعر التي يتم التعبير عنها. و لا تقتصر استخدامات تحليل المشاعر على الشركات التجارية فقط بل لجأ كثير من السياسيين والعلماء الأكاديميين والمؤسسات الإعلامية والجمعيات الخيرية إلى توظيف تحليل المشاعر لخدمة أهدافهم الخاصة.(Kennedy, 2012, p:435)

٣- تحليل المشاعر: التوسع خارج مجال الأعمال التجارية.

بالنظر الى أن تحليل المشاعر Sentiment Analysis هو ظاهرة حديثة إلى حد ما فإنه ليس من المستغرب قلة عدد الدراسات التي أجريت بشأنه داخل أي تخصص أكاديمي . (Kennedy, 2012, p:436)) إن الإمكانيات البحثية التي ينطوي عليها تحليل المشاعر والقضايا المنبثقة عنها تجعلنا نتفق مع وجهة النظر القائلة بأنه لا ينبغي أن يحجب التوجه التجاري المبكر لمشروعات تحليل المشاعر المحوسبة الإمكانيات الأوسع لهذه الطريقة بالنسبة للباحثين ، وأنه قد حان الوقت كما يرى أوبال Uppal لاستغلال التقدم الذي حدث في استخراج الرأي من أجل الكشف عن المشاعر في مجموعة متنوعة من السياقات ولأغراض اجتماعية في المقام الأول بدلاً من الأهداف التجارية (Driscoll, 2015, p:864)



٤- تنقيب البيانات وتحليل المشاعر بوصفه حقلاً أكاديمياً

قبل منتصف التسعينيات كان من النادر استخدام مصطلح "تنقيب البيانات data mining" في منشورات العلوم الاجتماعية. ومنذ ذلك الحين ، حدثت زيادة سريعة في أعداد المنشورات التي تذكر هذا المصطلح . و من بين جميع المنشورات المفهرسة في Scopus ، وهي واحدة من أكبر قواعد البيانات الببليوغرافية المتاحة ، استخدم ٢١٣٥ منشوراً في العلوم الاجتماعية مصطلح "data mining" بين عامي ١٩٩٥ و ٢٠٠٤. وبالمقارنة ، و بعد مرور عشر سنوات ارتفع عدد المنشورات ليصل إلى ٢٩,٩٧٠ منشوراً (بين عامي ٢٠٠٥ و ٢٠١٤) في معظم التخصصات ، بما في ذلك علوم الحياة والصحة والعلوم الفيزيائية ، ثم شهد "استخراج البيانات" نمواً آخر في السنوات الأخيرة بسبب انتشار وسائل الإعلام الاجتماعية وتعميم الجيل الثاني من شبكة الويب Web 2.0 (Baram-Tsabari, Segev, & Sharon, 2017 , p: 93)

وعلى الرغم مما يمثله ذلك من أهمية في مجال دراسات الاعلام "كانت معظم المنشورات التي أشارت لمصطلح تنقيب البيانات تدرج ضمن تخصص أو قطاع الأعمال أو الإدارة . والتي تشمل المجال الفرعي لنظم المعلومات والتكنولوجيا . وفي الواقع كانت حوالي ٦٠% من جميع منشورات العلوم الاجتماعية التي تستخدم مصطلح تنقيب البيانات في الفترة ما بين ١٩٩٤ وحتى ٢٠١٤ متصلة بالأعمال التجارية والإدارة وبالمقارنة كانت نسبة ٤ % فقط من منشورات العلوم الاجتماعية متصلة بعلم النفس. (Baram-Tsabari, Segev, & Sharon, 2017, p:94) ومع هذا النمو في اعداد الأبحاث والمنشورات في مجال العلوم الاجتماعية ، يظل هناك فجوة واضحة في أبحاث الاتصال التي أجريت حول تحليل المشاعر باستخدام تقنيات تنقيب البيانات.



٥ - الخصائص المعرفية لأبحاث تحليل المشاعر.

نتيجة لتزايد استخدام البيانات الكبيرة في المجال الأكاديمي كان ثمة آراء بأن ذلك سيؤدي إلى تغير طبيعة إجراء الأبحاث بل أن ذلك يمثل " ثورة حقيقية " تتعلق بفهمنا للعالم خاصة وأنها تؤدي الى بزوغ أو نشأة نماذج بحثية جديدة عبر التخصصات ، والتي سميت بالعلوم الموجهة بالبيانات data-driven science ، والعلوم الاجتماعية المحوسبة computational social science والعلوم الإنسانية الرقمية digital humanities . وترتكز بصفة خاصة على المدى الذي تسهم من خلاله بيانات وسائل التواصل الاجتماعي ، إلى جانب البيانات الاجتماعية الكبيرة الأخرى، في إمكانية وجود نوع مختلف من العلوم الاجتماعية. (Kitchin, 2016, p:27)

وتختلف البيانات الناتجة عن شبكات التواصل الاجتماعي _ والتي يتم تحليلها _ سواء في نوعيتها أو طبيعتها عن تلك التي تنتجها دراسات العلوم الاجتماعية. وبالنظر إلى منصة " فيس بوك" مثلاً سنجد تياراً من البيانات أو سيلاً متدفقاً باستمرار من المعلومات الثرية والمفيدة للغاية عن الناس ، وحياتهم وتفاعلاتهم ، وما يحدث في مجتمعات مختلفة حول العالم. وينطبق الشيء ذاته على تويتر و، واتس آب، وسناب شات Snapchat ، و، فور اسكوير Foursquare وغيرها من منصات وسائل التواصل الاجتماعي. وفي حين ينحصر عدد المستجيبين في معظم دراسات العلوم الاجتماعية _ سواء في استطلاعات الرأي السياسية ، أو المقابلات ، أو المجموعات البؤرية _ في حدود ١٠ أو ١٠٠ ونادراً ما يتجاوز ١٠٠٠ ، و يتم جمع البيانات لفترة زمنية واحدة في الوقت المناسب (الذي عادة ما يكون على مدى أسبوعين أو أشهر) ، وتكون البيانات محدودة التنوع ، نجد أنه بالمقارنة مع بيانات وسائل التواصل يصبح الفرق أكثر حدة. (Kitchin, 2016, p:29)

تعد أبحاث تحليل المشاعر حقلاً متعدد التخصصات Interdisciplinary

Field ينمو بسرعة وهو في مفترق طرق بين تخصصات مختلفة تشمل تقنيات من علوم الحاسب الآلي واستخراج البيانات وتعلم الآلة، وتحليل الشبكات الاجتماعية وعلم الاجتماع والاثنوجرافي Ethnography ، وعلم الإحصاء لذا فهو عميق الجذور في علوم الحاسب الآلي والعلوم الاجتماعية (Zafarani, & Abbasi, 2014, p:21) يرتبط تحليل المشاعر بمجالات بحثية مختلفة مثل معالجة اللغة الطبيعية Natural Language Processing و تنقيب النص واللغويات الحاسوبية Text Mining & computational Linguistics . (Aldayel & Azmi, 2016, p:785) .

في حين تعتبر بحوث الاتصال بحثاً بينية Interdisciplinary فإن وسائل الإعلام الرقمية قد تكون أكثر من ذلك فكل تخصص يستدعي نظرياته وأساليبه وطرقه البحثية الخاصة في دراسته للظواهر. لقد دفعت الأشكال المتنوعة من الاتصالات الحاسوبية _ التي تم تطويرها مؤخراً _ الباحثين في العلوم الإنسانية الى اكتشاف الأساليب الكمية لمجموعة البيانات الضخمة أو الكبيرة في حين اكتشف علماء المعلومات مزايا التحليلات الكيفية كما أن المداخل التي طورت من أجل تحليل البيانات التي تولدت _ وأصبحت متاحة من خلال استخدام الناس لوسائل الإعلام الرقمية وهياكل البيانات الناتجة عن ذلك والتي غالباً ما تكون مختلفة عن تلك التي يتم دراستها في حقل او تخصص معين_ عملت على تشجيع وإثراء الابتكار المنهجي) Mahrt & (Scharrow, 2013, p:20)

و يتوقع من الباحثين في هذا المجال الناشئ أن يكون لديهم معرفة في مجالات مختلفة مثل تنقيب النصوص Text mining وتحليل الشبكات Social Network



Zafarani, Information retrieval وعلم استرجاع المعلومات Analysis (&Abbasi, 2014, p: 5)

إن أبحاث تنقيب البيانات ذات طبيعة استكشافية exploratory nature، إذ أن نتائج عملية التنقيب غالباً ما يُتوقع أن تكون "مثيرة للاهتمام" أو "غير عادية" أو "غير متوقعة"، أو كما كان يشار إليها في البداية أو سابقاً بأنها "غير معروفة". فالشخص الذي يقوم بعملية تنقيب البيانات، هو شخص لديه معرفة متخصصة يمكن أن تساعد في تقييم الأنماط الناشئة (أو المكتشفة) ووضعها في سياقها. وفي هذا الصدد يمكن التمييز بين نوعين من تنقيب البيانات الأول: تحليل الافتراض، ويسمى التحقق "verification"، على غرار ما هو موجود في كثير من الأحيان في التحليل الكمي التقليدي، والثاني: نمط اكتشاف آلي مستقل بذاته autonomous, automated pattern-finding يسمى، الاكتشاف. وهذا النوع الأخير هو الأكثر شيوعاً في عمليات تنقيب البيانات وعندما يتم العثور على كيانات أو هياكل في مجموعات البيانات، يتم تقييم بعض أو كل منها بحسب معناها أو قيمتها من قبل خبير المجال. domain expert. (Baram-Tsabari, Segev, & Sharon, 2017, p:95)

ثالثاً: تنقيب بيانات وسائل التواصل الاجتماعي: مستوياتها ومراحلها وتقنياتها.

١ - المستويات:

يمكن أن يكون النص المراد استخراج المشاعر منه مستنداً كاملاً، أو جملة واحدة أو جزءاً من جملة، يشتمل على بيانات تعريف مضمنة (ما وراء بيانات) embedded metadata. أو بدونها. (Yusof, N. F., Lin, C., & He, ، p:2385) (Y. 2018) ولذلك يمكن إجراء تحليل المشاعر على عدة مستويات مختلفة سواء على مستوى المستندات document-level sentiment analysis أو الجملة

aspect-، sentence-level sentiment analysis أو الكلمة أو مستوى الجانب ،- sentiment analysis level أو الميزة .

٢ - مراحل عملية تنقيب البيانات :

تتضمن عملية استخراج الرأي في وسائل الإعلام الاجتماعية استخدام تقنيات محوسبة لتجميع واستخلاص المشاعر ومواقف المؤلفين في محتوى هذه الوسائط الاجتماعية ثم تحليلها وعرضها بطريقة سهلة الفهم. ويمكن إيجاز مراحل هذه العملية كما يلي : (Kim & Jeong, 2015, p:397-399)

المرحلة الأولى: الاتصال بقنوات وسائل التواصل الاجتماعي المستهدفة وجمع البيانات منها
تتضمن هذه المرحلة اختيار القنوات الإعلامية والاجتماعية المستهدفة وتجميع البيانات منها. وعلى سبيل المثال توفر بعض المواقع مثل Twitter و Facebook واجهة برمجة تطبيقات مفتوحة للوصول إلى بياناتها وجمعها. في حين أن موقع ياهو مثلًا Yahoo.com لا يدعم ذلك ، ولكن يمكن للمحلل استخدام أدوات وتقنيات محرك البحث مثل برامج زاحف الويب(*) للقيام بذلك أو يمكن شراء مثل هذه البيانات من مزودي بيانات وسائل التواصل الاجتماعي ، ونظراً لأن أساليب جمع البيانات تختلف وفقاً لنوع الوسائط الاجتماعية ، فإن طبيعة البيانات المجمعة تختلف أيضاً. فقد يتم تكديس بعض بيانات الوسائط الاجتماعية كملفات مخزنة في سجل بينما يتم إدارة البيانات الأخرى في قاعدة بيانات علائقية. ومع ذلك ، فهذه البيانات هي في الأساس عبارة عن نص كبير، وغير منظم يتم إنشاؤه من جانب المستخدمين ، بالإضافة إلى ذلك ، تمتلئ وسائل التواصل الاجتماعي بوضاء كبيرة مثل الإعلانات والرموز التعبيرية مما يمكن أن يؤدي إلى تشوه نتائج التنقيب. لذلك ، يجب معالجة البيانات الكبيرة التي يتم جمعها لاستخراج ما يفيد منها لإجراء تحليل ذي معنى.



المرحلة الثانية: تأهيل أو تهيئة البيانات التي تم جمعها باستخدام معالجة اللغة الطبيعية

بعد تجميع مجموعة البيانات غير المهيكلة ، يجب القيام ببعض الإجراءات المهمة تتمثل في المعالجة الأولية أو المسبقة لتحليل البيانات وذلك باستخدام معالجة اللغة الطبيعية وهي تقنية حاسوبية تعمل على تحليل نص اللغة الطبيعية وفهمه وتفسيره وتقديمه من أجل التحليل اللغوي. في هذه المرحلة ، تكون معالجة اللغة الطبيعية مسئولة عن أنشطة المعالجة المسبقة: تحليل الجمل ، وإزالة الأحرف المعطلة ، وميزات الاستخراج ، ووضع علامات على أحرف معينة. على سبيل المثال ، يتم التخلص من الأحرف التي لا معنى لها مثل علامات النص المترابط html وعلامات الترقيم والأرقام والرموز التعبيرية. بالإضافة إلى ذلك ، فإن إزالة الكلمات ، مثل حروف الجر ، والضمائر ، وبعض الكلمات التي يتم تعريفها بأنها لا قيمة لها ، يتم إزالتها في عملية التطهير هذه. ثم يتم تحويل البيانات المؤهلة التالية إلى تنسيق بيانات قابل للتحليل مثل البيانات العلانية أو البيانات المنظمة. يتضمن التنسيق نصاً تعويضياً ومعلومات التعريف مثل تاريخ الإنشاء واسم المؤلف والأعداد والتعليقات والمفضلات. بعد ذلك يتم تخزين البيانات المؤهلة كهيكل إطار بيانات مدمج مع بيانات التعريف. بالإضافة إلى ذلك ، يمكن إنشاء موارد مثل معجم يختص بمجالات معينة مثل قاموس للمشاعر لتحسين دقة عملية استخراج الرأي.

المرحلة الثالثة: تطبيق تحليلات استخراج الرأي (تحليل المشاعر) على البيانات المؤهلة

تُطبق هذه المرحلة مجموعة من التحليلات على المعلومات اعتماداً على الغرض من التحليل، ويمكن للمحلل والباحث اختيار أدوات التنقيب المناسبة. على سبيل المثال ، يتم استخدام تحليل المشاعر لتقييم سمعة المنتجات والخدمات والشركات ، ومعرفة استجابات العملاء للأنشطة التسويقية. وإذا تم إنشاء موارد لغوية خاصة بمجال

معين مثل المعاجم و استخدامها في هذه المرحلة ، يمكن للمحلل أن يتوقع نتائج أكثر شمولية وواقعية حيث يمكن لنتائج تحليل المشاعر المُصنَّفة من قبل المُعاجم الخاصة بمجال معين توفير خريطة مفصلة حول الصفات أو السمعة سواء السلبية أم الإيجابية.

المرحلة الرابعة: عرض نتائج تحليل المشاعر بطريقة سهلة ومفهومة.

تتمثل المرحلة الأخيرة والختامية في عرض نتائج التحليل وهى نتائج استخراج الرأي باستخدام مخرجات مرئية مثل الرسوم البيانية ، والغرض من هذه المرحلة هو جعل النتائج بسيطة وواضحة وغير معقدة بحيث يمكن استخدامها في صنع القرارات. على سبيل المثال يمكن لخريطة حرارة المشاعر أن تكشف آراء العملاء الإيجابية أو السلبية باستخدام لونين متباينين مثل الأحمر والأخضر ؛ ويمكن لخريطة شجرة التكافؤ تقديم تصور مباشر باستخدام كل من الحجم والمشاعر في تصنيف هرمي، ويمكن لسحابة الكلمات الكشف عن تردد الكلمات أو الجمل في النصوص الخاضعة للتحليل .

٣- طرق وتقنيات تحليل المشاعر

هناك ثلاث مداخل أو طرق مختلفة لتحليل المشاعر هى نهج تعلم الآلة والطريقة القائمة على المعجم والنهج المختلط

نهج تعلم الآلة: **Machine learning approach** :

تعتمد طريقة تعلم الآلة على مجموعة من البيانات لإنشاء أو تصميم نماذج تستخدم فى تصنيف البيانات الجديدة أو غير المعروفة بدقة.و يعد هذا المدخل هو الأسلوب الأكثر استخداماً فى دراسات تحليل المشاعر، نظراً لمعدلات الدقة العالية التى يحققها، ومع ذلك فإنه يعتمد على المجال ويحتاج الى مجموعة بيانات مُعلّمة وينطوى



ذلك على كثير من الجهد في بيئة مثل الشبكات الاجتماعية التي تحتوي على كثير من المجالات. (Ramírez-Tinoco, Alor.et. Al, 2018,p:271) وهذا النهج ينقسم الى ثلاث تصنيفات فرعية فقد يكون خاضعاً للإشراف ويعتمد على خوارزمية وتسمية لكل مثال ، أو قد يكون غير خاضع للإشراف ولا يعتمد على شيء وأخيراً قد يكون خاضعاً لشبه إشراف.

تتطلب تقنية تعلم الآلة في حالة التعلم تحت الإشراف القيام بعملية التصنيف اليدوي لعينات من قاعدة البيانات manual sorting of database samples لتكوين مجموعة التدريب (training set) بهدف بناء نماذج تستند إلى نص تكون بمثابة مقاييس parameters في تحليل البيانات الآلي (Oliveira, Bermejo, & Santos, 2016,p:4)

النهج القائم على المعجم The lexicon-based approach:

و يتم بأسلوبين الأول هو استخدام القواميس التي تحتوي على مجموعة واسعة من المصطلحات أو العبارات لاستنتاج المشاعر المنقولة في النص . والثاني يعتمد على موارد تحوي مجموعة من الكلمات (Ramírez-Tinoco, Alor.et. Al, 2018,p:264) ويمكن أن نسمى ذلك النهج أسلوب التوجيه الدلالي الذي يعتمد على تطبيق أو استخدام مجموعة كلمات تم إنشاؤها سابقاً تحتوي على مصطلحات موجبة وسالبة ، وبالتالي ، يتم التحديد التلقائي automated identification للقضية باستخدام تكرار هذه المصطلحات في النص ، دون الحاجة إلى نموذج و قاعدة بيانات مصنفة سابقاً (Oliveira, Bermejo, & Santos, 2016,p:4)

إن استخدام قاموس الكلمات أو (n-grams) الذي قد يكون موجباً أو سالباً يحدد وزن الكلمات الرئيسية وفقاً لتكرارها أو ترددها ويتنبأ بقطيبتها (إيجابية أم سلبية) غير

أن هذا النهج المعتمد على المعجم وحده لا يكون فعالاً دائماً في تحديد المشاعر بدقة نظراً لأن النصوص تحوى بعض الإختصارات ، وبعض الكلمات العامية ، والرموز والعناصر غير التقليدية الأخرى، مما يجعل من الصعب وجود قاموس مُحدّث باستمرار من أجل تحديد المشاعر ولهذا السبب يلجأ الباحثين في دراساتهم إلى نهج تعلم الآلة (Aston, Liddle, & Hu, 2014, p:11)

النهج المختلط The hybrid approach : وهو يجمع بين أساليب النهج القائم على المعجم وبين تقنيات التعلم الآلي من أجل الاستفادة من خصائص كل منهما.
برامج وتقنيات تنقيب البيانات وتحليل المشاعر.

يوجد الكثير من البرامج التي يمكن استخدامها في تنقيب البيانات سواء لأغراض تجارية في المحل الأول أو لأغراض بحثية أكاديمية وهي متنوعة ومختلفة الإمكانيات وبعضها يناسب مرحلة معينة من مراحل التنقيب مما يعنى ضرورة الاستعانة بالبرنامج أو التقنية المناسبة بحسب الهدف وبحسب المرحلة التي وصل إليها البحث. فمثلا يمكن استخدام برنامج ما لاستخلاص البيانات من قواعد البيانات ثم استخدام تقنية اخرى مساعدة في عملية تهيئة البيانات أو المعالجة المسبقة وتقنية ثالثة في تحسين أداء عملية التصنيف قبل أن تتم عملية استخراج الآراء وفق خوارزميات معينة يمكن تعديلها وتكييفها بما يناسب كل حالة وفي مجال محدد ، ونعرض هنا لبعض النماذج من البرامج غير التجارية وكذلك أدوات من الأدوات المتاحة على الإنترنت والمختصة بتحليل المشاعر .



برنامج Orange :

يعد برنامج Orange من أدوات تنقيب البيانات وهو يعتبر بيئة برمجية شاملة مفتوحة المصدر توفر أدوات وتقنيات تنقيب البيانات وتعلم الآلة. وقد تم تطوير البرنامج في مختبر المعلوماتية الحيوية بكلية علوم الكمبيوتر والمعلومات، جامعة ليوبليانا، سلوفينيا.

و البرنامج مجاني، يمكن إعادة تعديله تحت شروط رخصة GNO المشاع الإبداعي _ كما نشرتها مؤسسة البرمجيات المجانية_ بحيث يكون مفيداً لكل فئات المستخدمين من العلماء والباحثين حول العالم. ولكن بدون أي ضمانات، بما فيها ضمان الملاءمة لغرض معين. و هو أحد أدوات تنقيب البيانات ذات الطابع الأكاديمي غير التجاري، ويعتبر من أدوات تنقيب البيانات التي لا تتطلب كتابة لغات برمجة.

برنامج Weka:

تم تصميم برنامج ويكا من قبل جامعة Waikato في نيوزيلندا وهو برنامج خاص بخوارزميات التنقيب عن البيانات وتعلم الآلة ، مكتوب بلغة الجافا ، ويعمل تحت رخصة GPL ، ومتوافق مع عدة بيئات حوسبية مثل (جنو، لينوكس ، ويندوز، ماك) كما يحتوي هذا البرنامج على أدوات قادرة على التعامل مع البيانات في مراحل ما قبل المعالجة والتصنيف والتجميع وقواعد الارتباط والعرض المرئي visualization للنتائج (M, 2009) والبرنامج عبارة عن حزمة برمجية مفتوحة المصدر تساعد على تحليل ، وإستخراج البيانات . وهذه الخوارزميات يمكن تطبيقها بسهولة على مجموعة من البيانات إما بشكل مباشر من خلال واجهة برنامج WEKA أو من خلال استدعاؤها عن طريق (Java code) بإستخدام classes الخاصة بها، وذلك من خلال تحميل المكتبة الخاصة بـ WEKA (العربي، ٢٠١٧)

SocialMention (SM): هي خدمة ويب مجانية أو واحدة من محركات البحث في وسائل التواصل الاجتماعي تسمح لمستخدميها بالبحث عن المنشورات posts المتنوعة وتتبع وقياس وتحليل ما يقوله الأشخاص عن الشركات أو المنتجات أو العلامات التجارية في مختلف الشبكات الاجتماعية على الويب ، بما في ذلك: Facebook و Twitter و YouTube و Digg وغيرها . وهي منصة تحليل تُنتج عدداً من الإحصاءات الأساسية في مجال تحليل المشاعر وهي تدعم التعليقات والتحليلات باللغة العربية (Al-Kabi, Alsmadi, et.al,2018:975)

SentiStrength: عبارة عن أداة أو خدمة ويب ومحرك بحث عبر وسائل التواصل الاجتماعي تم تطويره في جامعة ولفرهامبتون ولفرهامبتون Wolverhampton يسمح لمستخدميه بالبحث في شبكات اجتماعية مختلفة. وتم استخدامها في البداية عام ٢٠١٠ لإستخراج المشاعر من النصوص باللغة الإنجليزية غير الرسمية . وقد تم تطويرها للتعامل أيضاً مع لغات أخرى مثل: العربية ، الألمانية ، الإسبانية ، الروسية ، التركية، الفرنسية ، الإيطالية والهولندية وعدة لغات أخرى ، وتعد أداة SentiStrength قادرة على القياس العددي لقوة المشاعر النصية الإيجابية والسلبية وذلك بتعيين عدد صحيح (موجب أو سالب) لكل شعور في النطاق من (-٥ إلى ٥). ويمكنه أيضاً التعامل مع التعبيرات غير الرسمية للمشاعر -975 p: (Al-Kabi, Alsmadi, et.al, 2018 977)

مهام تحليل المشاعر: تتمثل المهام الأساسية لتحليل المشاعر في الكشف عن الذاتية في النصوص حيث يتم تصنيف النصوص على أنها تحتوي على تعبيرات ذاتية (غير موضوعية) عن المشاعر ، والكشف عن قطبية النصوص فتصنف النصوص على أنها إما إيجابية أم سلبية بشكل عام وكشف قوة المشاعر حيث تصنف النصوص



بحسب القوة الكلية للمشاعر الإيجابية أو السلبية ، وتحديد قطبيتها. وأخيراً كشف العاطفة Emotion detection حيث تصنف النصوص بحسب العاطفة السائدة (على سبيل المثال ، حزين ، غاضب...) (Thelwall, 2017 ,p:344)

وباختصار يهدف تحليل المشاعر إلى فهم المعلومات الذاتية مثل الآراء والمواقف والمشاعر المعبر عنها في النص. وتشمل مهام تحليل المشاعر على سبيل المثال لا الحصر ما يلي (Yusof, N. F., Lin, C., & He, Y. 2018)

• تصنيف المشاعر: الذي يصنف جزءاً معيناً من النص على أنه إيجابي أو سلبي أو محايد.

• استرجاع الرأي الذي يسترد الآراء ذات الصلة بموضوع أو استعلام معين.

• تلخيص الرأي الذي يلخص الآراء على مصادر نصية متعددة نحو موضوع معين

• تحديد هوية صاحب الرأي Opinion holder identification الذي يحدد من يعبر عن رأي محدد.

• تتبع ديناميات الموضوع أو المشاعر: تهدف إلى تتبع المشاعر وتغييرات الموضوع مع مرور الوقت.

• الكشف عن البريد المزعج: Opinion spam detection الذي يحدد الآراء المزيفة أو الكاذبة.

• التنبؤ: الذي يتنبأ سلوكيات الناس ، واتجاهات السوق ، ونتائج الانتخابات السياسية ، وما إلى ذلك ، بناءً على الآراء أو المشاعر المعبر عنها في المحتويات على الإنترنت



لمزيد من التوضيح عن كيفية القيام بالمهام الرئيسية لتحديد الكيانات فى النص و استخراج المشاعر المتعلقة بها أنظر الملحق المرفق فى نهاية الدراسة

رابعاً : الاتجاهات البحثية الرئيسية لأبحاث تنقيب البيانات وتحليل المشاعر:

تكشف مراجعة الأدبيات العلمية لمجال تنقيب البيانات عامة وتنقيب وسائل التواصل الاجتماعي خاصة عن اتجاهات رئيسة لهذا المجال قام الباحث بتصنيفها الى ستة اتجاهات رئيسية و لقد استفاد الباحث من التصنيف المذكور فى (Baram - Tsabari, Segev, & Sharon, 2017) الذى صنف الدراسات الى ثلاث فئات استناداً إلى مجموعات البيانات التى تدرسها، الفئة الأولى تختص بتنقيب بيانات وسائل الإعلام السائدة ، و الثانية : تهتم بالمحتوى الذى يُنشره المستخدمون، و الثالثة: الدراسات التى تفحص أنشطة المستخدم وآثاره الرقمية على المواقع المختلفة . و يختلف التصنيف الذى سنعرضه هنا فى إضافة معيار آخر للتصنيف يتمثل فى الهدف الأساسى للدراسات التى سوف نعرض لها كأمثلة على الجهود البحثية المعاصرة ، لتصبح الفئات ستة اتجاهات رئيسية بحيث يشتمل التصنيف على الجهود المبكرة فى أبحاث تنقيب البيانات و التى جاء معظمها فى تخصصات بعيدة عن العلوم الإنسانية و هو (الاتجاه الأول) و كذلك على المحاولات البحثية الحديثة سواء تلك التى جمعت بين تنقيب البيانات و الأبحاث التقليدية القائمة على المسوح أو الاستبيانات بهدف المقارنة بينهما (الاتجاه الخامس) ، أو تلك التى حاولت تقديم منظور نقدي(الاتجاه السادس) يراجع التراث النظرى و يكشف عن المثالب المنهجية و التحديات الأخلاقية لأبحاث تنقيب البيانات و تحليل المشاعر. لذا يستند التصنيف المقترح فى دراستنا الحالية الى معيارين أساسيين هما : طبيعة مجموعات البيانات التى تقوم الدراسة بتنقيبها ، و الهدف من الدراسة، و فيما يلى أهم الاتجاهات البحثية:



الاتجاه الأول : الدراسات الخاصة بتطوير تقنيات التنقيب واختبارها وتحسين مستوى دقتها.

نشأت هذه النوعية من الدراسات مبكراً وركزت فى مجالات علوم الهندسة والكمبيوتر أو أنظمة الحاسبات والمجالات المرتبطة بها وبخاصة التطبيقات متعددة التخصصات لعلوم الحاسب واهتمت بالبحث فى الجوانب اللغوية والخواريزمية فى تنقيب البيانات واستخراج الرأى بهدف تحديد اتجاهات النصوص كما غلب عليها المفاهيم الرياضية

أول الدراسات التى نعرض لها دراسة شين وكيم (Zhu, Y., Kim, M.C. & Chen, C. 2017) التى اهتمت بالبناء الفكرى لأبحاث تنقيب البيانات ونشرت فى دورية خاصة بعلوم المعلوماتية وقام فيها فريق البحث بفحص ٥٩٥٩ سجلاً ببلوجرافياً بهدف تحديد المشهد الفكرى والإتجاهات الناشئة والتطورات الأخيرة فى أبحاث استخراج الرأى واعتمدت الدراسة على طريقة البحث بالموضوع وفحصت اتجاهات الإقتباس على مستوى النطاق أو التخصصات العلمية لاكتشاف شبكات مشاركة الأبحاث والوثائق والمقالات البارزة فى مجال تنقيب البيانات وتمثلت أهم نتائجها فى الآتى:

- ركزت الاهتمامات البحثية فى التعرف على المعلومات الذاتية واستخراجها من النصوص وكان لمفاهيم مثل تنقيب النص Text mining واستخراج الرأى Opinion mining دوراً ملحوظاً فى نشر الأفكار الخاصة بالمجال الناشئ
- انتقلت نتائج المنشورات المبكرة الصادرة عن الهندسة الكهربائية والذكاء الاصطناعى واللغويات وعلوم المعلومات الى مجال استخراج الرأى وتم استخدام الاساليب الفنية وتنقيب النصوص وتعلم الآلة ومعالجة اللغات الطبيعية



بشكل متكرر للقيام بمهام التنقيب و تركزت الابحاث فى معظمها على اقتراح تقنيات محوسبة لقياس توجهات مراجعات المنتجات بطريقة ثنائية (إيجابى أو سلبى)

- كانت فئة التطبيقات متعددة التخصصات ذات تأثير أكبر على ظهور وتطوير أفكار تنقيب البيانات ونشرها فى حين كان عدد قليل من التخصصات مثل علوم الكمبيوتر واللغويات والهندسة وعلم النفس وعلم المعلومات هى التى قادت أبحاث استخراج الراي .
- ومؤخراً او فى الاونة الاخيرة حظى استخراج الرأى باهتمام كبير فى بحوث العمليات والإدارة.

وتتركز الادبيات الآن على التطبيقات العملية لتحليل المشاعر فى مجال التسويق وشبكات التواصل الاجتماعى والدراسات الحديثة تبحث بشكل أعمق فى النصوص باستخدام مجموعة متنوعة من التقنيات ووحدات التحليل مثل الوثيقة والجملة والميزة أو الجانب وذلك ضمن اهداف تتعلق بالتنبؤ بالقيمة السوقية (Zhu, Y., Kim, M.C. & Chen, C. 2017)

و حرفياً أجريت مئات الدراسات فى إطار هذا الاتجاه تركزت جهودها فى تطوير التقنيات الخاصة بالتنقيب للنصوص الانجليزية فى البداية ثم امتدت بعد ذلك لتشمل لغات اخرى منها الصينية والعربية وغيرها ، ومؤخرا بدأت الدراسات فى تناول تحليل المشاعر بلهجات مختلفة أيضاً ونقدم هنا فقط نماذج من الجهود البحثية الحديثة

فى سياق دراسات علم المعلومات التى اهتمت بتحليل المشاعر حاولت هيفاء الدايل تطوير منهجية هجينة تجمع بين المصنف القائم على المعجم والمصنف القائم على تعلم الآلة لتقدم نهج هجين يجمع بين التوجه الدلالي وتقنيات تعلم الآلة وذلك



بهدف تنقيب الشبكات الاجتماعية لفهم الخطاب المتعلق بكيفية نشر الأفكار وفهم ورصد تأثيرات القضايا الاجتماعية على آراء الناس وأفكارهم تجاه نشر الأفكار في سياق المجتمع السعودي (Aldayel & Azmi, 2016)

في دراسة تنتمي الى علوم الكمبيوتر حاولت رفع مستوى الدقة في تحليل المشاعر واستعرضت نماذج من الدراسات التي تعتمد على مجموعات من المصنفين البشريين لمحتوى وسائل التواصل الاجتماعي للتشفير والتكويد وللحفاظ على مستوى دقة مقبول وهو ٨٩%، اقترح الباحثين تحليلاً آلياً للمشاعر يعتمد على نمط بناء الجملة Morphological sentence pattern mood بدلاً من المصنفين البشريين ليرفع مستوى دقة التحليل لأعلى من ٩١.٢ وتم تطبيق ذلك على مراجعات الأفلام سواء في يوتيوب أو تويتر. (Al-Kabi , Al-Ayyoub, & Alsmadi, 2016)

و حاول هادي في دراسته (Hadi. 2015) تقديم مورد لتصنيف الكلمات باللغة العربية ومعجم لتصنيف قطبية الرأي لبيانات وسائل التواصل الاجتماعي على موقع تويتر وتم جمع ٣٧٠٠ تغريدة عربية بأطوال مختلف من أرشيف تويتر واعتمدت الدراسة على خوارزمية نيف بايسين (Naïve Bayesian (NB) و خوارزمية الجار الأقرب (K Nearest Neighbour (KNN) وآلات دعم المتجهات (SVMs) Support Vector Machines وأشارت النتائج التجريبية في مجموعة تويتر باللغة العربية إلى أن طريقة التعلم في آلات دعم المتجهات (SVM) تفوقت على الطريقتين الأخرين ، وانتهت الدراسة باقتراح لعمل مستقبلي يختص بتطوير طريقة جديدة لتحليل المشاعر في لهجات الكلمات العربية ، لأن المغردين لا يلتزمون بكتابة التغريدات حسب اللغة القياسية العربية وكل منهم يكتب بلهجته الخاصة لذلك من المهم بناء موارد لكل لهجة عربية، لتصنيف التغريدات العربية بشكل صحيح قدر الإمكان.



و هذا ما قام به شكرى (Shoukry & Rafea, 2015) في دراسة استهدفت تحسين مقاييس أداء تحليل المشاعر باللهجة المصرية على مستوى الجملة من خلال اقتراح منهج هجين يجمع بين كل من منهج التعلم الآلي وباستخدام آلات دعم المتجهات ومنهج التوجه الدلالي. وتم اقتراح نهج هجين لتصنيف المشاعر لتغريدات اللهجات المصرية وأثبت النهج المقترح فعاليته في تصنيف المشاعر على مستوى الجملة أيضاً .

و في الواقع إن الجهود البحثية في هذه النوعية من الدراسات مستمرة ، و معظمها يقوم على التعاون البحثي بين تخصصات مختلفة. ومن هذه الدراسات ، دراسة حديثة (Al-Kabi, Alsmadi, et.al, 2018) قام خلالها فريق بحثي من أربعة جامعات هي جامعة الزرقاء Zarqa University وجامعة العلم والتكنولوجيا بالأردن وجامعة New Haven بالولايات المتحدة الأمريكية وجامعة الملك خالد بالمملكة العربية السعودية بدراسة استهدفت تقييم أداتين من الأدوات المتاحة على الإنترنت لتحليل المشاعر وتدعمان اللغة العربية وهما SocialMention و SentiStrength وأعطت النتائج دليلاً قوياً على أن أداة SentiStrength غير قادرة على تحديد القطبية الصحيحة للآراء المكتوبة بالكلمات العربية العامية وكانت تصنفها في فئة محايدة بسبب عدم القدرة على ترجمة الكلمات العربية العامية الى الإنجليزية ، وكذلك فإن الأداة الثانية SocialMention فشلت في تحديد قطبية الآراء وكانت فاعليتها أقل من ٥٠ % وشجعت محدودية فعالية الأداة، الفريق على بناء ١٠ معاجم عربية يدوية من خلال مجموعة مرجعية مكونة من ٣٠١٥ من الآراء العربية عبارة عن التعليقات النصية على Facebook باللهجات العربية (المصرية ، بلاد ما بين النهرين ، المشرق ، وشبه الجزيرة العربية). وتم تصميم أداة جديدة لتحليل المشاعر



لل كلمات العربية بالعامية تسمى تحديد هوية قطبية الآراء العربية Arabic Opinions Polarity Identification (AOPI)

ولا يزال هناك جهود ضخمة للغاية ودراسات أحدث لتقيب بيانات الويب الدلالي و تحليل المشاعر المقارن أيضاً. و في دراسة أخرى (Ramírez-Tinoco, et.al 2018) في مجال النظم الخبيرة واتجاهات الحوسبة وتطبيقاتها في هندسة البرمجيات قام راميرز وآخرون بمراجعة ٢٢٩ ورقة بحثية حول استخدام مقاربات تحليل المشاعر في الشبكات الاجتماعية بهدف توجيه الباحثين لاختيار النهج الأنسب وفقاً لنوعية دراستهم وتزويدهم بالتوجهات المستقبلية للأبحاث في هذا المجال وايضاً لتكون دراستهم دليلاً للمهنيين المهتمين بتطوير الأدوات والخوارزميات والنهج والأطر المناسبة واتضح من توزيع الأبحاث بحسب المنصة أن تويتر هو المهيمن كشبكة اجتماعية في دراسات تحليل المشاعر وعلى الرغم من أن واجهة برمجة التطبيقات تعد الأسهل والأسرع في تحليل المشاعر إلا أنها نادراً ما تستخدم في الأوراق البحثية حيث استخدمت بنسبة ٩.٦٨% فقط في حين اختارت نسبة ٩٠.٣٢% من الأبحاث بدائل أخرى مثل استخدام بعض خوارزميات التصنيف الموجودة من خلال أدوات التتقيب عن البيانات لتطوير خوارزميات خاصة . وأن نهج تعلم الآلة كان الأكثر استخداماً وبنسبة ٥٩% تقريباً ويقال أنه في الممارسة يحقق نسبة دقة أعلى في التصنيف ولا يتطلب موارد معجمية يصعب في بعض الأحيان الحصول عليها أو إنشاؤها ومع ذلك فإنه يحتاج الى إنشاء مجموعة بيانات معلمة تتطوى على كثير من الجهد في ضوء بيئة مثل وسائل التواصل التي تحتوى على كثير من المجالات في حين حصل النهج القائم على المعجم على نسبة ٣٠% تقريباً وحصل النهج الهجين على النسبة الباقية ١٠% تقريباً لأنه يتطلب جهداً أكبر من المجالات الأخرى وأخيراً ركزت

الأعمال البحثية على القطبية العاطفية وتركت جوانب أخرى مثل قوة العواطف ومسألة استخدام صيغة النفي في النصوص.

الاتجاه الثاني: الدراسات الخاصة بتلقيب بيانات وسائل الإعلام السائدة

يعد زيادة حجم العينة أحد المزايا المهمة للدراسات التي تستهدف تلقيب بيانات محتوى وسائل الإعلام الرئيسية أو السائدة و يلاحظ هنا أن حجم البيانات أكبر بكثير من الأبحاث التقليدية " حيث يشكل ثراء الأشكال الجديدة من وسائل الإعلام مجالاً أكثر تعقيداً ويتجاوز الأشكال التقليدية لأبحاث وسائل الإعلام التي تقوم باستقصاء أو استبيان يطبق على ٣٠٠ مفردة حول نقاش سياسي مثلاً أو جمع ٤٠ ساعة من النشرات ولمدة أسبوع وهي تصميمات بحثية مصممة لتحليل مجموعات محددة من البيانات في مرحلة واحدة حيث هناك نقاط بداية ونهاية واضحة لإنتاج ونشر المحتوى في حين يختلف نمط دراسات تلقيب البيانات بسبب أن هناك تدفق مستمر للمعلومات و كمثل على ذلك قام أحد الباحثين في الإعلام الجديد بتحليل المواقع الاخبارية وفحص ١٦٠٠٠ مقالة مختلفة على هذه المواقع إضافة الى الروابط المتصلة بهذه المقالات على

فيس بوك وتويتر وبينترست (Nevius , 2017)

و في دراسة أخرى قام أحمد الراوي (Al-Rawi, 2017) بدراسة حول قناتي الجزيرة والعربية من خلال دراسة موقع القناتين على الفيس بوك وقام بفحص ١١٦٨٥ تقريراً إخبارياً منشوراً على الصفحتين بهدف تقييم المشاعر العامة والنقضيات الإخبارية وتصنيف المحتوى المنشور الى فئات مثل السياسة الداخلية والنزاع ، السياسة الإقليمية والدولية ، علوم وثقافة وفنون ، الاقتصاد والمال ، رياضة وتسلية ... وهكذا وحاول التعرف على نوعية النص الإخباري الذي يحظى باهتمام الجمهور من خلال تعليقاتهم على القصص الإخبارية وقياس المشاعر تجاه هذه



القصص الإخبارية وتوصل الى أن قناة الجزيرة كانت الأكثر نشاطاً على الفيس بوك من حيث تفاعل الجمهور و أن أغلب المشاعر المسيطرة على كلا القناتين كانت تتماشى في الغالب مع تغطية القناتين لأحداث معينة.

و في دراسة ثالثة قامت سيجيف Segeve وفي إطار مشروع بحثي متعدد الجنسيات شمل ٣٢ دولة بفحص أكثر من مليون عنصر إخباري من ٣٥ موقعا إخباريا حيث سهل توافر أرشيفات الاخبار على الإنترنت القيام بزيادة حجم البيانات المدروسة على هذا النحو ، وبالمثل لم يكتف Seaefer بزيادة الحجم بل قام بزيادة الفترة الزمنية ليحلل في دراسة مقارنة محفوظات الصحف الأمريكية والألمانية على مدى ٥٠ عاما (Baram-Tsabari, Segev, & Sharon, 2017)

الاتجاه الثالث: الدراسات الخاصة بالتنقيب في المحتوى الذي يُنشئه المستخدمون.

ينصب التركيز الرئيسي هنا على تنقيب بيانات المحتوى الذي ينشئه المستخدم ميزة دراسة الأنماط البنائية للاتصال باستخدام تحليل الشبكات Data Mining of User-Generated Content network analysis و لدى هذه الفئة من الدراسات و مثال ذلك تحليل شبكة المناقشات بين محرري موسوعة ويكيبيديا أو تحليل شبكة تدفق الاتصال في تويتر. (Baram-Tsabari, Segev, & Sharon, 2017).

في دراسة ميك ثيلوال التي أجريت لتحليل المشاعر المرتبطة بثلاثين حدثاً من الأحداث الشعبية لمدة شهر على تغريدات تويتر باللغة الانجليزية ومحاولة تقييم ما إذا كانت الأحداث الشعبية ترتبط عادة مع زيادة قوة المشاعر ، أعطت النتائج دليلاً قوياً على أن الاحداث الشعبية ترتبط عادة مع قوة المشاعر السلبية ويبدو أن بعض الأحداث الإيجابية مثل توزيع جوائز الأوسكار قادرة على توليد مشاعر سلبية متزايدة في ردود الفعل تجاهها وخلصت الدراسة الى أن الأحداث المهمة في تويتر ترتبط بزيادة في



معنويات المشاعر السلبية ومع بعض الاستثناءات غالباً ما تكون المشاعر السلبية مفتاح الأحداث الشعبية في تويتر وأن متوسط التغير المفاجيء في المشاعر من السلبية الى الايجابية كان بنسبة بسيطة تتراوح من واحد الى ستة بالمائة. (Thelwall, Buckley, & Paltoglou, 2010)

و في دراسة لتحليل المشاعر جمعت بين التحليلات الكمية والكيفية أجريت على عينة عشوائية قوامها ٤٥٨٦ تغريدة تم اختيارها من إجمالي ٤٥٨٦٠ تغريدة تمثل تعليقات على صحيفة El pais (صحيفة يومية إسبانية متاحة رقمياً) اهتمت بالكشف عن مشاعر الأشخاص العاديين أو القراء تجاه الكلمات الإسبانية من أصول عربية متعلقة بالإسلام، وأسهمت نتائج الدراسة في تعميق المناقشات حول تأثير اللغة العربية على مفردات اللغة الإسبانية وأشارت الى وجود شعور إيجابي عموماً تجاه كثير من الكلمات الإسبانية ذات الأصل العربي. (Mostafa & Nebot, 2017).

و في دراسة حديثة (Giaxoglou, Döveling, & Pitsillides, 2017) بعنوان " من العواطف الى المشاركة الشبكية نشرت في دورية الإذاعة ووسائل الإعلام الإلكترونية ، أشارت الى أن العواطف وثيقة الصلة بأبحاث وممارسات الاتصال عبر الأنترنت وأنها تسهم في إعادة تشكيل الحياة الاجتماعية ومن خلال دراسة صفحات النصب التذكاري على فيس بوك Facebook memorial page والمشاركين في ما يسمى بالحداد الشبكي Networked mourning ممن يشاركون مشاعرهم بالخسارة علناً أو شبه علناً مما يسمح برؤية ما كان ينظر اليه سابقاً على أنه تجربة عاطفية خاصة وحميمية . وقدمت الدراسة رؤى متعددة التخصصات ومشاركة بين الثقافات حول تقاسم الخسارة والحداد عبر الإنترنت مما يعكس أو يقلل أو يببالغ ويضخم من واقعة الوفاة وأشارت الى قدرة وسائل التواصل الاجتماعي على تمديد الهويات العامة



حتى بعد وفاة أصحابها وكشفت عن تعقيد ممارسات مشاركة العواطف عبر هذه الوسائل الأمر الذي أعاد التفاوض حول المجال العام والمجال الخاص وهذا يعني أن الآراء والتعاريف الحالية للعاطفة ليست كافية وتحتاج إلى أن تُستكمل بفهم السياقات الشبكية وبعبارة أخرى ، تصبح نظريات العاطفة نظريات عن المشاعر الشبكية مما يستدعي الاعتراف بالمكانة المركزية المتزايدة للعاطفة في الثقافات الرقمية .

الاتجاه الرابع : الدراسات التي تبحث نشاط المستخدمين في المواقع وفي شبكات التواصل .

و هذه النوعية من الدراسات تهتم بفحص أنشطة المستخدم user activity أو سلوكياته مثل البحث عن موضوعات معينة باستخدام محركات البحث أو كيفية تفاعله مع الموقع بفحص سجلات التصفح وغيرها و " عادة ما تستخدم الدراسات التي تبحث في أنشطة المستخدم قواعد بيانات تم تجميعها وتنظيمها من قِبل أطراف أخرى. وتمثل ملفات تسجيل زيارات المواقع مثلاً لقاعدة البيانات هذه التي تسجل تفاصيل حول المستخدمين والمحتوى الذي شاهدوه. ويقتصر الوصول إلى قواعد البيانات هذه عادة على مشغلي الموقع ، وإن كان بعضها يسمح بالإطلاع على بعض الأنشطة في مواقعها المختلفة. إن Google Trends مثلاً تسمح للباحثين بدراسة ومقارنة طلبات البحث في Google أو الاستفسارات عن موضوعات معينة في مناطق مختلفة". (Baram-Tsabari, Segev, & Sharon, 2017)

أحد الأمثلة على هذه النوعية من الدراسات دراسة تجريبية (Arapakis, et.al,) 2014 أجريت في إطار سعى بوابات الأخبار للتوصل إلى طرق جديدة للاحتفاظ بالمستخدمين وتشجيع تفاعلهم مع المحتوى الإخباري، قام فريق البحث بإجراء تحليل على ١٣٣١٩ مقالة إخبارية مأخوذة من البوابة الإخبارية الشهيرة بالولايات المتحدة



الأمريكية Yahoo News US واستخدم عدة مقاييس للكشف عن البعد العاطفي لتفاعل المستخدمين مع المحتوى وقياس المشاعر المعبر عنها ضمن المحتوى الإخباري وكيفية تغيرها اعتماداً على متغيرات مثل نوع المقالة ووقت نشرها ومصدرها واستخدمت النتائج في إعادة ترتيب قائمة التعليقات على القصص الإخبارية . واتضح أن مشاعر المستخدمين تعد أحد العوامل التي تحدد تفاعلهم مع المقالات الإخبارية وحاولت الدراسة فحص التأثيرات الفعلية للمشاعر على تجربة قراءة الأخبار وأظهرت النتائج أن المحتوى الإخباري أظهر تبايناً كبيراً من حيث العاطفة وقطبية المحتوى بناء على متغيرات مثل مزود خدمة الأخبار ومتغير النوع واتضح أن ترتيب التعليقات المختلفة أثر على القراء ومدى تفاعلهم مع القصص الإخبارية .

الاتجاه الخامس: الدراسات التي تجمع بين تنقيب البيانات والأبحاث التقليدية.

إن معظم هذه الدراسات تنطلق من فكرة المقارنة بين نتائج تنقيب البيانات ونتائج الأبحاث التقليدية وهي تحاول أن تدرس الظواهر من أكثر من زاوية وتطرح تساؤلات من قبيل إمكانية استبدال الأبحاث التقليدية بعمليات تنقيب البيانات. ونذكر هنا نماذج من هذه الدراسات.

في دراسة استقصائية حاولت التوصل إلى فهم عميق لتفاعل الجمهور مع حدث ثقافي يتمثل في معرض ملبورن للكتاب عام ٢٠١٣ م من خلال التعبير عن مشاعرهم على وسائل التواصل الاجتماعي وبالاعتماد على مجموعتين من البيانات الأولى هي الهاشتاج الرسمي لذلك الحدث والثانية استطلاع رأي للجماهير المشاركة أجرى عبر الأنترنت من جانب منظمي المهرجان وكشفت الدراسة_ من خلال تحليل ٢٠١٨٩ تغريدة و ٩٢١ مفردة ممن استجابوا للاستطلاع _ عن كيفية تفاعل الجمهور مع المهرجان من خلال قراءة التغريدات الأكثر إثارة للمشاعر واتضح أن ٣٨% من



ردود الجمهور تعبر عن مشاعر إيجابية في حين عبر ٢٠% عن مشاعر سلبية وكانت المشاعر المعبر عنها عموماً تميل إلى الاعتدال وهكذا كشفت الدراسة عن ثروة من البيانات المعبرة عن خبرات ثقافية للجمهور وتشكل تجمعات ثقافية عبر المشاركة في الحدث والتعبير عن المشاعر من خلال التغريدات وأن تويتر أعاد تشكيل هذه التجمعات الثقافية بتوفير خطوط اتصال مباشرة بين منظّمى المهرجان وكل من الحاضرين والجمهور وأن هذه الخبرات الثقافية وتلك التفاعلات مع الحدث الثقافي لم يكن بالإمكان الكشف عنها من خلال وسائل الإعلام والاتصالات المباشرة وجهاً لوجه. (Driscoll, 2015)

حاولت أنديرا سيرون وزملاؤها (Ceron, Curini, Iacus, & Porro, 2013) في دراستها حول استخدام وسائل التواصل الاجتماعي كوسيلة للتنبؤ بنتائج الانتخابات وتقييم شعبية السياسيين، مقارنة الأفضليات السياسية التي يعبر عنها المواطنين عبر الأنترنت مع تلك التي تشير إليها نتائج استطلاعات الرأي التقليدية وركزت الدراسة على تقييم مدى شعبية القادة السياسيين الإيطاليين على الشبكة خلال عام ٢٠١١ م وكذلك نية التصويت لمستخدمي الأنترنت الفرنسيين في الانتخابات الرئاسية عام ٢٠١٢ م والانتخابات التشريعية اللاحقة ومن خلال تحليل المشاعر لجميع التغريدات المتعلقة بكل قائد سياسي في الشهر السابق لإجراء المسح الشامل (أكثر من ٠٧٠٠٠ تغريدة) بالإضافة إلى متابعة تفضيلات مستخدمي الأنترنت يومياً من خلال تغريداتهم (٢٤٤٠٠٠ تغريدة في الفترة من ٢٧ أبريل حتى ٥ مايو) تم رصد وتحليل تفاعل المستخدمين مع الأجندة السياسية والحملة الانتخابية الجارية واتضح من النتائج أن تحليل المشاعر في وسائل التواصل الاجتماعي (تويتر) يوفر تنبؤات أكثر دقة عند التركيز على القادة الأكثر شعبية وفيما يتعلق بشعبية القادة السياسيين كانت استطلاعات الرأي في المتوسط أعلى في تقديرها لشعبية القادة السياسيين من وسائل التواصل



الإجتماعى كما كشفت نتائج التغريدات في الدراسة عن توقع انتصار للمرشح الاشتراكي ، هولاند في جولة الإعادة ، بنسبة ٥٤.٩ ٪ من الأصوات. واتفق ذلك مع التنبؤات التي أجرتها شركات المسح التي توقعت نسبة تتراوح بين ٥٢.٥ و ٥٣.٥ ٪ في آخر استطلاعاتها المنشورة. و كنتيجة عامة تظهر الدراسة قدرة ملحوظة للإعلام الاجتماعى على التنبؤ بالنتائج الانتخابية و وجود علاقة جديرة بالملاحظة بين وسائل التواصل الاجتماعي ونتائج الدراسات الاستقصائية التقليدية ويكشف ذلك عن القدرة التنبؤية لتحليلات هذه الوسائل ، التي تتعزز مع زيادة عدد المواطنين الذين يعبرون عن آرائهم على الأنترنت شريطة أن يتصرفوا باستمرار بناء على هذه الآراء.

فى دراسة (Oliveira, Bermejo, & Santos, 2016) عنوانها هل يمكن لوسائل الإعلام الاجتماعية أن تكشف عن تفضيلات الناخبين؟ مقارنة بين تحليل المشاعر واستطلاعات الرأي التقليدية قام الفريق البحثى بجمع ٩٢٤٤١ تغريدة متعلقة بمرشحي الرئاسة البرازيلية خلال الجولة الثانية من الانتخابات في عام ٢٠١٤ م . وتمت مقارنة نتائج تحليل المشاعر بنتائج التحليل المستخلصة من ستة استطلاعات للرأي رافقت التصويت و اجراها معهد الأبحاث المسمى Datafolha وبشكل عام ، كشفت النتائج أن الاثني كانا متشابهين للغاية. على الرغم من أن تحليل المشاعر لم يصل إلى نفس مستوى الدقة مثل استطلاعات الرأي التقليدية وكان الاختلاف بين تحليل المشاعر ونتائج الاستطلاعات فقط في مستوى دقة النتائج من ١% الى ٨% مما يعنى أن للتقنية نتائج مرضية توضح آراء الناخبين. وبالتالي ، يمكن استخدام تدفق المعلومات من تويتر لتحديد تفضيلات الناخبين السياسية .

فى دراسة (Neuman, Guggenheim, Jang, & Bae, 2014) حاولت اختيار فروض نظرية ترتيب الاولويات في بيئة وسائل التواصل الاجتماعي والبيانات



الكبيرة للإجابة على سؤال من يحدد جدول الاعمال للقضايا العامة في كل يوم وعمّا إذا كان هناك دلائل على ارتباط تسلسل زمني للقضايا من الوسائل التقليدية الى وسائل التواصل والعكس ، واعتمدت على تجميع بيانات على مدار اليوم ولمدة عام كامل ٢٠١٢ م من المقالات الإخبارية على الإنترنت وتغريدات تويتر ومشاركات المدونات ومنديات النقاش وفي مقارنة بين الوسائل التقليدية ووسائل التواصل الاجتماعي وقدرتهم على ترتيب الأولويات فيما يختص ب ٢٩ قضية مختلفة بالمجتمع الأمريكي شملت قضايا اجتماعية كالأجهاض وقضايا بيئية مثل الاحترار العالمي وقضايا اقتصادية مثل البطالة وسياسية مثل العراق وغيرها وفي محاولة لبحث دينامية ترتيب القضايا سواء في الوسائل التقليدية أم في وسائل التواصل _ خلال أحد الأيام فقط تم المقارنة بين ٥٠٠٠ مقال مطبوع و٤٥٣٧ قصة إخبارية تقليدية في التليفزيون والراديو و١٠٠٠٠ تغريدة على تويتر و٢٠٠٠ مشاركة في منديات النقاش و٦٠٠ مشاركة في مدونة و١٣٣٦٢ تعليق على وسائل التواصل الاجتماعي المختلفة_ اتضح للباحثين أن العلاقة بين النقاش الدائر في الوسائل التقليدية و وسائل التواصل فيما يختص بترتيب أولويات القضايا تتميز بالتفاعل والرنين المتبادل وأن الأمر أبعد من أن يكون علاقة سببية آلية لأن كل منهما يستجيب بطريقته الخاصة للأحداث اليومية وأن الديناميات تحدث لدقائق محدودة ولا تستمر لأيام والنتيجة النهائية أنه لا يزال الوقت مبكراً لاختبار النظريات الاجتماعية في بيئة البيانات الكبيرة خاصة وأن الادوات والطرق البحثية المعنية بذلك لا تزال تتطور .

الاتجاه السادس : الدراسات ذات الطابع النقدي (الدراسات النقدية).

و يتركز هدف هذه الدراسات في طرح عدة قضايا نقدية تثيرها أبحاث تنقيب البيانات وتحليل المشاعر ، يتعلق بعضها بنقد الأسس المنهجية لدراسات التنقيب مع تناول



إشكالياتها النظرية والمنهجية المرتبطة بذلك في حين يركز البعض الآخر على السياق الاجتماعي الأوسع

في مقالة مبكرة لهيلين كينيدي (Kennedy, 2012) تدور حول تحليل المشاعر حاولت فيها تقديم نظرة شاملة على المنظورات المهيمنة في دراسات تحليل المشاعر ورصد وجهات نظر الجماعات الفاعلة فيها مثل القائمين بعمليات تحليل المشاعر ووجهات نظر كل من مستخدمى وسائل التواصل و الباحثين ذوى التوجه الماركسي النقدي، توصلت الى أن التحليل الدقيق للمشاعر ليس سهلاً بسبب الطرق المعقدة التي يعبر بها الناس عن مشاعرهم التي لا تخلوا من سخرية أو فكاهة أو اختصارات حيث عبر محللوا المشاعر عن مخاوفهم بشأن "نظافة" البيانات وتحدثوا عن تنظيف المشاعر بمجرد تحديدها واعتبروا أن نسبة دقة ٧٠% تعتبر كافية في ذلك المجال كما دافعوا عن الانتقادات الموجهة اليهم والى شركاتهم بأنها تطفلية وتنتهك خصوصيات المستخدمين وتمثلت وجهة نظرهم بأن الهدف هو قياس الشعور العام وليس مشاعر فردية وأنهم لا يولون اهتماماً كبيراً بالبيانات الفردية مع قدرتهم على تحديد الأشخاص إذا ما أرادوا وأن هدفهم توليد معرفة مفيدة اجتماعياً . في حين اعتبر الباحثون النقديون أن المواقع والشركات التي تعمل تحت غطاء حماية المستهلكين وخدمة مصالحهم _ وتوفر مساحة للمستخدمين للتعبير عن آرائهم ومشاعرهم _ تستخدم هذه البيانات في تنمية حقوق العلامات التجارية وتحولها إلى مورد قابل للإستغلال في إطار ما يسمى الإقتصادات العاطفية ، بل أن شركات تحليل المشاعر التي تستخدم كلمة الاستماع Listening لوصف الخدمات التي توفرها هدفها هو المراقبة والإشراف وليس الإصغاء فعلاً وأن ممارسات تنقيب بيانات وسائل التواصل الاجتماعي هي استخلاص للقيمة من الشعور وهي آلية رأسمالية أخرى



لاستخراج القيمة من المشاعر وأن ناتج ما يستخرج من تحليل المشاعر يمثل قيمة فقط لأولئك الذين يقومون بتطوير التقنيات لمراقبة المشاعر وليس من يعبرون عن مشاعرهم. وفي النهاية لا تظهر وجهات نظر مستخدمى وسائل التواصل بشأن تحليل مشاعرهم وربما لا يدركون ذلك ومعظمهم ليسوا على علم بأن مواقع الويب مسموح لها بمشاركة معلوماتهم مع الشركات التابعة لها . وهناك حاجة الى مزيد من الابحاث حول هذا الموضوع.

و في دراسة أخرى (Kennedy, 2016) للباحثة ذاتها نشرتها في كتاب بعنوان ما الذى يجب أن يهنا حول تنقيب بيانات وسائل التواصل الإجتماعى عرضت فيها للنقاشات الرئيسية الدائرة حول الموضوع وانتقدت الأسس المعرفية لمجال دراسات تحليل المشاعر والمزاعم الإيستمولوجية القائلة بأن البيانات الكبيرة بما تنتجه من فرص لدراسة السلوك البشرى والرأى العام على نطاق واسع وفي الوقت الحقيقى _ تعد أكثر فاعلية في فهم المجتمع دون حاجة الى التنظير والتفسير، و ركزت على نقد تلك الضجة التى يثيرها المتحمسين لأبحاث البيانات الكبيرة التى لم تقتصر على العلوم التقنية فحسب بل امتدت الى العلوم الاجتماعية . ومن أمثلة ذلك انتقاد ما عبر عنه تشيز اندرسون Chis Anderson رئيس تحرير مجلة وايرد Wired بقوله " إن كل نظريات السلوك البشرى من علم اللغة الى علم الاجتماع الى علم النفس يمكن استبعادها بسهولة فى عصر بيتابايت Peta byte age حيث تتحدث البيانات عن نفسها" ولقد كشفت عن كثير من الانتقادات والمخاوف المنهجية حول الطرق التى يتم بها استخراج البيانات والآثار المترتبة على ذلك وبخاصة مخاطر ظهور فجوات رقمية جديدة نتيجة قلة أعداد من يستطيعون الحصول على البيانات الكبيرة وقلة أعداد من يستطيعون معالجتها بالفعل . وعرضت كذلك بالتفصيل للقضايا المتعلقة بانتهاك



الخصوصية وزيادة أشكال المراقبة المعاصرة وحاولت تقديم إجابات حول ما الذى ينبغي للأفراد والشركات الصغيرة القيام به تجاه تنقيب بيانات وسائل التواصل الاجتماعى والسؤال الرئيسى لعملها المتمثل فى ما إذا كان يمكن القيام بتنقيب بيانات وسائل التواصل الاجتماعى بطرق تعتبر مقبولة أخلاقيا وبشكل يساهم إيجابيا فى الحياة الاجتماعية.

قام سميث (Smith, 2012) فى دراسة مبكرة باستعراض وجهات النظر التي تطالب باستبدال الأبحاث المسحية واستطلاعات الرأي التقليدية بأبحاث تنقيب البيانات وقام بتفنيد المزاعم القائلة بأن استطلاعات الرأي أصبحت عتيقة وقديمة وأنه يمكن من خلال تنقيب البيانات للإنترنت ووسائل التواصل الاجتماعى خاصة_ والدمج بين مصادر معلومات متنوعة_ الحصول على نتائج عن الرأي العام والسلوكيات الاستهلاكية وغيرها من قضايا اجتماعية وسياسية بشكل أفضل وأقل تكلفة وأكثر سرعة وسهولة من المسوح والاستطلاعات التقليدية وأشار سميث الى مجموعة من النقاط التي توضح أن هذه البدائل لم تثبت بعد تفوقها على الاستطلاعات التقليدية من الناحية العملية أو التجريبية ومن ذلك القيود المتعلقة بقضية التغطية وعدم تمثيل العينات للمجتمع الأسمى فكثير من الناس لا يستخدمون وسائل التواصل أو لا يستخدمونها بانتظام وأن المستخدمين لا يمثلون جميع السكان كما ان هذه الوسائل تسمح بالتلاعب وإخفاء الهوية وملئة بالهويات الزائفة وتتغير بياناتها باستمرار وأن تنقيب البيانات وتحليلها يقتصر فقط على تلك التي يمكن الحصول عليها فى ضوء قيود الخصوصية وحتى إذا ما تم الدمج بين السجلات الإدارية للمواطنين وقواعد البيانات الحكومية المختلفة_ وقد تم ذلك في عدد من الدول الأوروبية_ لخلق صورة أكثر اكتمالا للأفراد فقد واجه ذلك عوائق قانونية وأخلاقية كثيرة وعلى الرغم من أوجه القصور



هذه وغيرها فإنه يمكن دراسة وسائل التواصل بعمق لكونها شكلاً مهماً من أشكال التفاعل الاجتماعي ولكنها لن تكون بديلاً لدراسات الرأي العام ويمكن أن توفر هذه البدائل معلومات عن الرأي العام لكنها أكثر محدودية من حيث تغطية النطاق الكامل ومن حيث التعقيد ولا تستطيع الإجابة عن جميع أسئلة البحث . ويخلص سميث الى نتيجة عامة وهي أن وسائل التواصل الاجتماعي وإن كانت تمثل مصدراً ثرياً للمعلومات يزيد من فهمنا للرأي العام فإنها لن تحل محل الاستطلاعات في المستقبل القريب .

خامساً : تنقيب بيانات وسائل التواصل الاجتماعي وتحليل المشاعر رؤية نقدية.

الفرص البحثية

شهدت العلوم الاجتماعية سيلاً من البيانات من خلال منصات التواصل الاجتماعي ومئات الملايين من المستخدمين النشطين يومياً الأمر الذي مكن العلماء الاجتماعيين من دراسة مجموعات كبيرة ومتنوعة من السلوكيات وفي الوقت الفعلي وبطريقة غاية في الدقة (Baram-Tsabari, Segev, & Sharon, 2017,p:94) إن الزيادة الأخيرة في البيانات المتاحة رقمياً والأدوات القادرة على معالجة هذه البيانات والتطبيقات الجديدة في مجالات تحليل النصوص والشبكات عملت على تعزيز الأساليب المحوسبة لدراسة الاتصال ويوفر ذلك فرصاً بحثية لتحليل السلوك الاجتماعي ودراسة الاتصال بطرق لم تكن ممكنة من قبل ويمكن ان يغير ذلك تخصصنا بطريقة جذرية وبعده طرق أهمها الإنتقال من التقارير الذاتية عن السلوك الى دراسة السلوك الفعلي، ومن ذلك مثلاً يمكن باستخدام تطبيقات الهاتف المحمول ذات الأغراض الخاصة وملحقات برنامج التصفح تتبع استهلاك الأخبار في الوقت الفعلي ودمجها مع



بيانات الاستقصاء للحصول على مقياس أكثر تعقيداً لاستهلاك الأخبار وتأثيراتها والأكثر أهمية هو الانتقال من البحوث الفردية الى البحوث التعاونية إذ من الصعب للغاية على باحث بمفرده امتلاك المهارات والموارد اللازمة للقيام بجميع مراحل وخطوات البحث المحوسب بنفسه ولا شك أن هذا يغير من طبيعة إجراء الأبحاث بتعزيز التعاون متعدد التخصصات اللازم للتعامل مع مجموعات البيانات الضخمة وتقنيات الحوسبة الأكثر تعقيداً كما أنه يدعم في الوقت ذاته الترابط بين البحث الكمي والكيفي ، بحيث تسهم الأبحاث الكيفية في تعزيز فهمنا للبيانات وبناء النظرية ثم تأتي المقاييس الكمية للبحث الكمي لاختبار الفرضيات الناتجة. (Atteveldt & Peng, 2018,p:81-85

بقدر ما يحمله ذلك التحول فى طبيعة أبحاث الاتصال_ فى بيئة البيانات الكبيرة _ من فرص بحثية فإنه ينطوى على مجموعة من التحديات أو الصعوبات التقنية والمنهجية أثارها الدراسات النقدية. تتطلب رؤية نقدية واعية لما تطرحه من قضايا علمية ومجتمعية نعرضها هنا بمزيد من التفصيل .

الصعوبات التقنية والانتقادات المنهجية.

صعوبات تقنية ومنهجية:

هناك قدر كبير من الصعوبات التقنية لأبحاث تنقيب البيانات يتعلق بعضها بالقدرة على تقليل نسبة الخطأ ومشكلة نقل النطاق أو نقل المجال والصعوبات المختصة بحجم البيانات الضخم . و يرى سميث (Smith, 2012,p:221) أن الكثير من تقنيات التنقيب غير مجربة نسبياً ولم يتم تحسينها ويشمل ذلك معاجم البحث وخوارزميات القطبية وتحليل المشاعر وتقنيات الترشيح لعزل مجموعات البيانات المستهدفة والتي تشكل أهمية بالنسبة للبحث وكذلك ما يتعلق بعملية التجميع ومشكلة انخفاض الثقة.



ويضيف تريلنج Trilling الى تلك المشكلات ثلاث نقاط تتمثل الأولى فى نقص سعة التخزين فمجموعة البيانات الضخمة قد تتجاوز حجم محركات الأقراص الصلبة المتوفرة حالياً _ كمثال هنا اصدرت ياهو نيوز مجموعة بيانات تمثل أربعة اشهر فقط من البيانات بحجم ١٣,٥ تيرا بايت _ والثانية نقص ذاكرة الحاسب لأن مجموعة البيانات لبعض الحسابات كبيرة جداً ولا يمكن تحميلها على أجهزة الحاسب الشخصية ، والثالثة : الافتقار الى القدرة الحاسوبية بمعنى قوة المعالجة مما يجعل الحاسب يستغرق مزيداً من الوقت لإجراء التحليلات. (Trilling & Jonkman, 2018,p:161)

وفى الواقع و كما رأينا فى محور الدراسات النقدية السابق هناك نقاش علمى محتدم حول النواحي المنهجية فى أبحاث تنقيب بيانات وسائل التواصل الاجتماعى ويتركز النقاش على عدة مستويات:

على مستوى طبيعة البيانات ذاتها توصف هذه البيانات بأنها بيانات كثيفة data rush وكبيرة ومن مصادر متنوعة وملينة بالشوائب ، تتطلب إجراءات كثيرة من أجل تنقيتها وهى أيضاً متدفقة باستمرار ويصف العباسى ذلك بقوله " إن أحد التحديات الكبيرة التى تواجه أبحاث تنقيب بيانات وسائل التواصل الاجتماعى هو مفارقة البيانات الكبيرة واختيار عينات ممثلة ومعالجة أو إزالة الشوائب ومشكلة التقييم الاجتماعى (Zafarani, &Abbasi, 2014,p:3,4)

على مستوى إتاحة البيانات للبحث : هناك كميات هائلة من البيانات التى تم إنشاؤها من جانب المستخدمين وهذا يفرض قيوداً كبيرة إذ لم يتم فهرسة أى من هذه المواد فى وسيلة ذات مغزى ولا يمكن النظر إليها نظرة شاملة فهناك بالفعل مادة بحثية (



بيانات) تتناسب العديد من الأهداف البحثية ولكن التساؤل حول كيفية الوصول إلى هذه البيانات واختيارها أو سحب العينات منها لا يمكن الإجابة عليه بسهولة . (Mahrt & Scharkow, 2013,p:21) يضاف الى ذلك" أن مجموعات البيانات التي يتم استخلاصها من شبكات التواصل الاجتماعي _و التي قد تكون عشوائية أو غير عشوائية _ هي عينات لم يخترها الباحث وعلى سبيل المثال عندما يستخدم الباحثون مؤشرات جوجل فإنهم يقومون بدراسة العينات التي توفرها لهم الشركة في حين يتطلب الحصول على بيانات كبيرة موارد ضخمة مثل تلك المتوافرة في الوكالات الحكومية والشركات الخاصة وغالباً ما تكون مجموعة البيانات هذه متاحة لمجموعة متميزة من الباحثين. (Baram-Tsabari, Segev, & Sharon, 2017,p:96)

و تشير هنا الى بعض الجهود التي بذلت من أجل إتاحة مجموعات من بيانات وسائل التواصل الاجتماعي للأغراض البحثية لتكون بمثابة أرشيف دائم للعلوم الاجتماعية أو قاعدة بيانات قياسية لتعزيز قابلية تبادل البيانات الخاصة بوسائل التواصل الاجتماعي لأغراض تكرار الأبحاث و المقارنة بينها ، و من تلك المحاولات أرشيف بيانات جيسيس للعلوم الاجتماعية The GESIS Data Archive for the Social Sciences الممول من القطاع العام بهدف الأرشفة الآمنة ، و طويلة الأجل لبيانات العلوم الاجتماعية و تم أرشفة و توثيق نصف مليار من التغريدات المدموغة بالعلامات الجغرافية (تمثل عدة أشهر فقط من التغريدات في الولايات المتحدة في الفترة من ٢٠١٤ م الى ٢٠١٥ م) مع مراعاة الاعتبارات الأخلاقية و القانونية و شروط خدمة منصة تويتر (Kinder-Kurlanda, et.al, 2017,p:1) و لا يخفى ما يعترض تلك الجهود من تحديات كثيرة، وأنه لا يوجد ما يضمن أن تكون البيانات



المؤرشفة ممثلة للعينات التى يريد الباحثون دراستها مستقبلاً أو مناسبة لأهدافهم البحثية.

على مستوى معالجة البيانات واستخلاص النتائج: إن الدراسات التقليدية تقوم بحفظ بياناتها فى ملف مسطح بسيط أما دراسات تنقيب البيانات فتحتاج إلى تخزين مجموعات بيانات كبيرة فى قواعد بيانات بتصميمات قابلة للتطوير وذات كفاءة للحصول على البيانات فى الوقت المناسب وفى مقابل أن البحث الكمى التقليدى ليس مكثفاً من الناحية الحاسوبية ويمكن إجراء معالجة بياناته على جهاز حاسب شخصى وبواسطة مستخدم لديه فهم للإحصاء الوصفى والاستدلالي ومعرفة حزم البرمجيات الاحصائية مثل SPSS و R studio وغيرها تأتى أبحاث تنقيب البيانات على النقيض من ذلك تتطلب العمل على قواعد البيانات الكبيرة ومعرفة كيفية الوصول الى مجموعات البيانات من خلال واجهة برمجة التطبيقات . (Baram-Tsabari, Segev, & Sharon, 2017,p:98)

هناك أيضاً نقاش علمى أو جدل دائر حول مسألة التراكم المعرفى و مدى مناسبة نتائج تنقيب البيانات للبناء عليها ودرجة الثقة فيها وإمكانية تكرار الأبحاث واستخدام ذات المجموعات من البيانات لإعادة التجارب والتحقق من هذه النتائج. " ويزيد الأمر صعوبة أن الباحث فى مجال تنقيب البيانات قد يستخدم برنامجاً معيناً أو خدمة مدفوعة لاستخلاص بعض المحتوى من الويب ثم برنامجاً آخر لمعالجة البيانات الناتجة وإجراء التحليلات الاحصائية وبرنامجاً ثالثاً لعرض المرئيات وهذا يجعل من الصعب بناء تدفقات او نتائج قابلة للتكرار وأكثر صعوبة فى البناء على نتائج الدراسات السابقة ((Trilling & Jonkman, 2018,p:158

بصرف النظر عن أن الباحث قد يلجأ لخدمة مدفوعة تساعده في إنجاز إحدى مهمات أو مراحل التنقيب _ ومعظم الشركات تقدم خدماتها كصندوق أسود دون الكشف عن آليات المعالجة أو طبيعة عمل ونوعية الخوارزميات التي استخدمت فضلاً عن التكلفة المرتفعة لهذه الخدمات _ فإن هناك قضية أخرى متعلقة بتفسير النتائج " رغم ما تقدمه من رؤى جديدة حول خيارات المستخدمين وسلوكياتهم هناك تساؤل عما إذا كانت تحليلات البيانات الكبيرة وعلى نطاق واسع تسمح في الواقع للباحثين والممارسين بفهم ما تعنيه هذه النتائج ومن ثم هناك اقتراح بأن يستخدم الباحثين نهجاً يجمع بين الاساليب القائمة على البيانات واساليب أكثر إثنوجرافية" (Mahrt & Scharkow, 2013,p:23,24).

2013,p:23,24)

على مستوى الأيديولوجية المتضمنة في أبحاث البيانات الكبيرة: اقترنت حركة البيانات الكبيرة بأيديولوجية أو أسطورة تؤكد على قيمتها التحويلية ، وكان ثمة غلواً فيما يسمى " ثورة البيانات الكبيرة " التي جلبت معها كثيراً من الادعاءات ، أكثرها تطرفاً هو أن البيانات الضخمة أو الكبيرة ستجعل العلوم ذاتها بالية أو على الأقل لن تكون هناك حاجة إلى النظرية أو النماذج التفسيرية و أنه مع وجود ما يكفى من البيانات فإن الأرقام تتحدث عن نفسها و أن الارتباطات بين البيانات سيكون كافياً و لن يكون هناك حاجة الى اختبار الفرضيات أو التحليلات السببية لدفع التطور العلمى . و قد يكون الرأى الأكثر واقعية هو الاعتراف بقيمة البيانات الكبيرة و الاستفادة من البيانات واسعة النطاق ، و إعطاء مزيد من الاهتمام للأسئلة ذات الأهمية النظرية و الاجتماعية (Parks, 2014,p:356-358)

يمكن للبيانات الكبيرة من خلال أدواتها المحوسبة تعزيز قدرة الباحثين على جمع بياناتهم من قواعد البيانات المتعددة ، في أوقات مختلفة وإجراء أبحاث على



مستوى أعلى من التعقيد و على نطاق واسع أو غير مسبوق غير أن هذا يثير أيضاً مزيداً من التساؤلات حول صحة القياس و ضمان جودة هذه البيانات و مدى تمثيلها حيث تعتمد الأبحاث على البيانات المتاحة بدلاً من البيانات المطلوبة ، و هناك حاجة إلى تقديم أدلة أقوى على صحة أنظمة الترميز الآلى و خوارزميات تعلم الآلة و الأدوات الجديدة المستخدمة في تحليل المشاعر . (Parks, 2014,p:358,359)

التحديات الأخلاقية.

أثارت أبحاث تحليل المشاعر مجموعة من التحديات أخلاقية تشير فى مجملها إلى قضية أساسية وهى أنه لا يمكن فصل تنقيب محتوى وسائل التواصل الاجتماعى عن المناقشات الأوسع حول أخلاقيات ممارسة الأبحاث فى علم البيانات وفيما يلي توضيح لأهم التحديات .

مخاوف من انتهاك خصوصية المستخدمين:

" تتمثل القضية الرئيسية في بحوث وسائل الاعلام الجديدة هذه في كيفية الحصول على بيانات مفيدة ومتاحة ودقيقة وفي الوقت ذاته احترام الخصوصية الفردية " (Nevius , 2017,p:1040) ، ففي عام ٢٠١٠ ، أعلن مارك زوكربيرج ، المؤسس والرئيس التنفيذي لـ Facebook ، أنه في عصر وسائل الإعلام الاجتماعية ، لم تعد الخصوصية "معياراً اجتماعياً" وبالطبع فإن وجهة النظر هذه تتفق مع مصالح شركاته التى تجنى أموالاً طائلة عن طريق بيع البيانات التى يشاركها المستخدمون على منصات التواصل الاجتماعى (Kennedy, 2016,p: 44) هذه القضية دفعت الباحثين الى فهم ما يسمونه "تناقض الخصوصية" ، أو حقيقة أن ممارسات مشاركة مستخدمى الوسائط الاجتماعية لبياناتهم تبدو متناقضة مع مخاوف الخصوصية التى أعربوا عنها.

وقد دافع بعضهم عن وجهة نظر زوكربيرج بالتميز بين مفهومين للخصوصية : الخصوصية الاجتماعية (وهي تشير الى رغبة المستخدمين في التحكم فى بياناتهم وتحديد أى من الاشخاص الذين يمكنهم الحصول على بياناتهم الشخصية) والخصوصية المؤسسية (و هي المختصة بتتقيب المعلومات الشخصية بواسطة منصات وسائل الإعلام الاجتماعية والشركات التجارية الأخرى) وأن مخاوف مستخدمي وسائل الإعلام الاجتماعية حول السيطرة على معلوماتهم الشخصية تتعلق بالأول ، وليس الأخير(Kennedy, 2016,p: 45)

وفى مقابل وجهة النظر هذه نجد من يؤكد أن الخصوصية لا تزال ، فى الواقع ، هي القاعدة الاجتماعية حيث تحدث الفيلسوف هيلين نيسينباوم Helen Nissenbaum عن مفهوم آخر للخصوصية أسماه "خصوصية فى السياق" 'privacy in context' أو التكامل السياقي "contextual integrity". وفى رأيه أنه عندما يكشف الناس عن معلومات شخصية فى سياق معين ، فإن لديهم توقعات حول ما سيحدث لبياناتهم فى هذا السياق. و يُعد تجاهل هذه التوقعات انتهاكاً لحقوق الأشخاص. إن ما نحتاجه إذن ، هو احترام توقعات الناس حول ما يمكن أن يحدث على نحو معقول لبياناتهم فى سياق معين . (Kennedy, 2016,p: 44)

أشكال جديدة من التمييز والإقصاء : بعيدا عن المخاوف المنهجية حول الطرق التي يتم بها جمع البيانات وتحليلها ، والآثار المترتبة على ذلك على المعرفة، هناك قضية أوسع تتعلق بمسألة من يمكنه القيام بعمليات التنقيب و يحمل ذلك إمكانية وجود أشكال جديدة من التمييز والإقصاء " بينما نقوم جميعاً بإنشاء البيانات ، لا يستطيع إلا عدد قليل منا الوصول إليها ، وعدد أقل يمكنه معالجتها ، وهذا يؤدي إلى خطر وجود فجوات رقمية جديدة قائمة على البيانات" وتذكر مانوفيتش Manovich ثلاث قضايا



ترتبط بذلك فهناك "أولئك الذين ينشئون البيانات (سواء بوعي أو بترك بصمات رقمية) ، وأولئك الذين يملكون وسائل جمعها ، وأولئك الذين لديهم الخبرة اللازمة لتحليلها". (Kennedy, 2016,p:53)

إن معظم البيانات ليست متاحة للجمهور ، وتحفظ بها شركات خاصة مثل شركات الاتصالات. فالشركات ، مثل جوجل ، والوكالات الحكومية ، تجمع يومياً كمية غير معروفة من البيانات المقدرة بنحو ٢.٥ مليار جيجابايت ، تغطي هذه البيانات كل شيء في الاستخدام الفردي للإنترنت ، مثل النشرات وسجلات البحث والوقت الذى يقضيه الأشخاص مع الوسائط وسجلات المواقع. وبما أن هذه البيانات غالباً ما تكون مصدراً للميزات التجارية أو الحكومية ، فهي نادراً ما تكون متاحة للباحثين. (Nevius, 2017,p:1039)

ولقد أشار بويد وكروفورد Boyd & Crawford الى قضية التمايز بين الباحثين والعلماء نتيجة لذلك بقوله " هل من السخرية أو السذاجة أن تسأل عن الحصول على مجموعات البيانات البحثية في عصر البيانات الكبيرة؟ ...إن الوصول المميز إلى البيانات الضخمة من قبل مجموعة صغيرة من الباحثين _ يرتبطون أو يعملون لحساب شركات مثل جوجل أو أمازون أو فيس بوك_ سيجعل هؤلاء الباحثين الذين يتمتعون بقدر غير عادل من الاهتمام يتميزون على نظرائهم الموهوبين بنفس القدر من دون هذه الروابط"(Atteveldt & Peng, 2018)

المراقبة والتحكم وتهديد الحريات العامة: تتمثل القضية هنا في ما اذا كان تنقيب البيانات أداة للتحرر أم للهيمنة ويتعلق ذلك بأهداف عمليات تنقيب البيانات ونتائجها التى يمكن أن تعنى زيادة فى الرقابة والهيمنة و التحكم . إن التركيز على مسألة

الخصوصية وحق المستخدم في الحفاظ على بياناته يصرف النظر عن القضية البنائية الأكثر أهمية والمتمثلة في زيادة المراقبة والهيمنة من جانب الحكومات ولا يتعلق الأمر هنا بالتمييز والإقصاء بل بإثارة أسئلة أكثر جدية فيما يتعلق بالحريات وحقوق الإنسان.

في وقت مبكر من عام ٢٠٠٤ كتب Mark Andrejevic موضحاً كيف أن التقنيات الرقمية الجديدة بما في ذلك منصات وسائل التواصل الاجتماعي مدمجة ومتداخلة في أشكال المراقبة المعاصرة على نطاق واسع ، وتعمل على تمديد ممارسات المراقبة ، وجادل بأن التكنولوجيات التفاعلية تسمح للأنشطة التي لم تخضع للرصد من قبل أن تخضع للمراقبة ، وبالتالي تصبح ممارسات الاتصال الرقمي الخاصة بنا جزءاً من مجتمع المراقبة. في الآونة الأخيرة. (Kennedy, 2016,p: 46)

وفي الوقت الحالي تتنوع الجهات الفاعلة التي يمكن أن تتقب البيانات الاجتماعية. وتتضمن تلك العملية حساب إبداءات الإعجاب والمشاركات ، أو تحليلات أكثر تقدماً لقوة الآراء والمشاعر التي تتضمنها ، وامتدادها ، وخصائصها الكمية الأخرى مثل (لفت الإنتباه لمحتوى معين mentions ، وعلامات التصنيف (hashtags). كما يتم على نطاق واسع تنقيب البيانات الوصفية الموجودة خلف محتوى وسائل التواصل الاجتماعي _ وهي أكثر قيمة من المحتوى نفسه _ وهذه البيانات الوصفية تتضمن معلومات عن المستخدمين مثل: من يتحدث ومن يشارك ، وموقعه الجغرافي، وبمن يرتبط ، ومتى يكون نشطاً وكيف يكون لنشاطه تأثير ، وكيف تبدو أنماط أنشطة المستخدمين السابقة وما يشير إليه ذلك من تفضيلات محتملة وأنشطة مستقبلية. (Kennedy & Moss, 2015,p: 1)



مثلما يمكن لتقنيات تنقيب البيانات أن تسهم في خدمة المصالح المجتمعية إنسانية و اجتماعية فإنها يمكن أيضا أن تشكل خطراً وتهديداً محتملاً وذلك بحسب الجهة التي تقوم بعملية التنقيب وبحسب أهدافها وهناك بعض الأمثلة التي يمكن أن تثير القلق حقاً بهذا الصدد، وتعنى أن تنقيب البيانات الكبيرة لا يحمل الخير دائماً بل يمكن استخدامها لخدمة أهداف سلطوية من خلال ما يسمى "بمراقبة البيانات (المراقبة من خلال سجلات البيانات) ، وزحف السيطرة (البيانات التي نتجت عن المستخدم والتي تجمع لغرض معين يمكن أن تستخدم لغرض آخر)" (Kitchin, 2016,p:31) ولتوضيح ذلك نعطي مثالا افتراضياً:

حينما ينتشر هاشتاج في تويتر يختص بحقوق الانسان في بلد ما ، يمكن لمنظمات حقوق الإنسان أن تكون على علم بالإنتهاكات من خلال تنقيب بيانات المواقع وشبكات التواصل التي تحفل بشهادات واقعية وصور وفيديوهات لا تجد لها مكانا فى وسائل الإعلام التقليدية وفى الوقت ذاته يمكن لنظام قمعى أو ديكتاتورى _ ينكر وجود مثل تلك الإنتهاكات_ أن يتعقب الناشطين وباستخدام التقنيات ذاتها يتم تحديد الأشخاص المؤثرين ومن ثم يقبض عليهم بدعاوى نشر الشائعات وتهديد استقرار الدولة.

إن زيادة المشاعر السلبية تجاه نظام سياسى معين أو مسئوليهِ بسبب قضايا الفساد أو زيادة فرض الضرائب أو إنفلات الأسعار وتدنى مستويات المعيشة يعطى مؤشرا على الرأى العام تجاه هذه المسائل لنظام ديموقراطى فيعمل على اتخاذ إجراءات إصلاحية تجنباً للهبات الشعبية أو الثورات فى حين أن ذات التقنيات يمكن ان تستخدم فى نظام على الطرف النقيض لإنشاء شبكات لمراقبة الرأى العام وتوظيف ما

يسمى اللجان الإلكترونية لإخراص المعارضين ومحاصرتهم بردود نمطية موجهة أو حتى بتهديدهم ليمتنعوا عن نشر آرائهم.

آلية رأسمالية أخرى لاستخراج القيمة: ثمة انتقاد لأبحاث تنقيب البيانات من الباحثين ذوى التوجه الماركسي ومنهم مثلاً Turow الذى اعتبر أن "صناعات تنقيب البيانات " تؤدي إلى التمييز الاجتماعي. يحدث ذلك ، من خلال عمليات تحليل البيانات ، التى تقوم بتحويل ملفات التعريف "profiles" الخاصة بالمستخدمين إلى "تقييمات فردية" وهذه البيانات تحدد هويتنا وقيمتنا ويتم إنتاج حسابات القيمة التسويقية لكل شخص ، استناداً إلى سلوكياته على شبكات التواصل ومن خلال أشكال أخرى لتتبع البيانات ، يتم تصنيف كل فرد ، كى يكون هدفاً للإعلانات أو يصنف بوصفه لايمثل أية قيمة. ولقد عبر هيرن عن ذلك بقوله إن قدرة استخراج البيانات على تحديد المشاعر القيمة تمثل آلية رأسمالية أخرى لاستخراج القيمة. (Kennedy, 2016,p: 48)

سادساً: تحليل المشاعر وبحوث الإعلام نظرة مستقبلية.

شكلت المفاهيم العلمية وطرق القياس وأدواته المنضبطة منهجياً دوماً الأسلوب المتعارف عليه أكاديمياً للتوصل الى المعرفة وبناء النظريات و القوانين العلمية لذا نعرض هنا لأهم ما يمكن أن يسهم به مدخل تحليل المشاعر فى دراسات الإعلام على المستويات السابقة المشار إليها.

مستوى المفاهيم : ليس من الغريب أن يحدث اقتباس بعض المفاهيم من مجالات معرفية أخرى والأدبيات العلمية لدراسات الإعلام وغيره من تخصصات العلوم الاجتماعية تكشف عن ذلك بوضوح ومثلما اقتبست دراسات الإعلام فى بدايتها مفاهيم من علم النفس والسياسة وغيرها ومؤخراً فى أحدث نظريات الإعلام الجديد مثل نظرية

التشكل العضوى التى نجد فيها مصطلحات مثل الإحلال والتعايش والتطور المشترك.. الخ (نلاحظ أن المفاهيم تحمل دلالات توحى بمجال علم البيولوجيا). وبمفهوم يتسق مع بحثنا الحالى كشفت القراءة المتعمقة للأدبيات عن وجود " سحابة " من المفاهيم فيما يختص بتقيب بيانات وسائل التواصل الاجتماعى وتحليل المشاعر. هذه المفاهيم تكشف أولاً عن مدى التنوع واتساع المصطلحات المستخدمة وتكشف ثانياً عن أن كل مجال بحثى يستخدم مصطلحاته الخاصة ولذلك ينبغى على باحثى الاتصال محاولة الاستفادة من المفاهيم الناشئة أيا كان مجالها وبمعنى " تهيئة " هذه المفاهيم كى تتوافق ومجالنا البحثى أو تخصصنا مع عدم إغفال ضرورة ضبط المفاهيم وتحديدتها بدقة حتى لا يشكل الاختلاط فى دلالاتها عائقاً للأبحاث المستقبلية . وتسهم هذه الخطوة فى إثراء دراسات الإعلام وإعطائها أبعاداً تحليلية وتفسيرية جديدة .

مستوى طرق القياس والأدوات البحثية: ينبغى ضرورة الاستفادة من تقنيات تتقيب البيانات خاصة وأن التقنيات التحليلية التقليدية التى كانت تناسب تحليل محتوى وسائل الإعلام " القديمة " تبدو غير كافية لمعالجة البيانات الضخمة المستخلصة من بيئات وسائل التواصل الاجتماعى .و يلاحظ هنا أن تقنيات تتقيب البيانات لا تزال تُستخدم لتمييز المشاعر والآراء فى النصوص فقط فى حين أن محتوى الويب ووسائل التواصل يشتمل على وسائط متعددة تتنوع بين النصوص والصور والملفات الصوتية والفيديو ولا ينوفاً لدى الباحث معلومات عما إذا كان هناك برامج أو تقنيات للتتقيب فى مثل هذه الوسائط.

مستوى النظريات العلمية فى تخصص الاتصال: هناك حاجة ملحة إلى اختبار نظريات الاتصال التقليدية فى بيئة وسائل التواصل والبيانات الكبيرة ، ويتعلق الأمر هنا بإعادة اختبار الفرضيات النظرية التى نشأت نتيجة تجارب معملية أو دراسات استقصائية



تقليدية لقياس الفروض والعلاقات السببية لعينات " صغيرة " منضبطة منهجياً والانطلاق الى إعادة إختبار هذه العلاقات في بيئاتها الطبيعية أو على نطاقات واسعة وعينات ممتدة زمنياً ومتنوعة مكانياً وثقافياً وكمثال على ذلك يمكن لباحثي الاتصال إعادة اختبار نظرية لولب الصمت *The theory of the spiral of silence* في إطار مناقشات المنتديات أو تعليقات المستخدمين على القضايا السياسية أو الاجتماعية كما يمكن إعادة اختبار فرضيات تدفق الأخبار على المستوى الدولي أيضاً.

مستوى الموضوعات البحثية: لا بد لباحثي الاتصال من أن يدركوا ويتعلموا كيفية الاستفادة من هذا الكم الضخم من البيانات المتوافرة في وسائل التواصل وأن يدركوا الفروق بين مختلف المنصات الاجتماعية وأن يشكلوا فرق بحثية مشتركة تضم تخصصات مختلفة أهمها معالجة اللغات الطبيعية أو اللغويات الحاسوبية والذكاء الاصطناعي وتعلم الآلة حتى لا نتفاجأ وبحسب قول العباسي (Zafarani, &Abbasi, 2014,p:1) "إننا نواجه مشكلة متفاقمة، تتمثل في الغرق في البيانات الضخمة في حين أننا نتعطش الى المعرفة "

من بين الموضوعات البحثية الجديدة والتي تحتاج الى دراسات مثلاً تحليل المشاعر المقارن عبر منصات مختلفة من وسائل التواصل الاجتماعي بخلاف تويتر لأن معظم الدراسات تستخلص بيانات من تويتر بوصفه أسهل المنصات للحصول على البيانات . وهناك حاجة أيضاً لإجراء دراسات حول تحليل مشاعر البيانات المتدفقة ومحاولة الاستفادة من تقنيات بناء السلاسل الزمنية ودراسة ديناميات الآراء والمشاعر بحسب اختلاف الفترات الزمنية في قضايا وموضوعات محددة ، ومثال ذلك دراسة تأثير أحداث أو أزمات معينة_ مثل الاحتجاجات الجماهيرية_ على المشاعر العامة



عبر فترات زمنية قبل الأحداث وأثناء حدوثها وبعدها ومحاولة بناء تصورات أو تقديرات لحالة المشاعر ودرجاتها من حيث الشدة وتركزاتها المكانية أو الجغرافية.

اتضح من خلال الأدبيات المنشورة في أبحاث تحليل المشاعر تركيز معظمها على تحليل النصوص باللغة الانجليزية_ مع بعض الاستثناءات فيما يختص بلغات اخرى _ غير أن أقلها هو تحليل المشاعر للنصوص التي ينتجها المستخدمون باللغة العربية ومن ثم تعد دراسات تحليل المشاعر باللغة العربية مجالاً بحثياً واعداً، آخذاً في النمو ومطلوباً في دراسات الاتصال في الوقت ذاته ، ويمتد ذلك بالطبع الى تحليل المشاعر بلهجات مختلفة تسود في البلاد العربية مثل اللهجة المصرية أو الشامية واللهجات الخليجية أو لهجات دول المغرب العربي وهذه جميعاً تشكل تحديات لأبحاث ومشروعات تحليل المشاعر وتتطلب أن يكون التعاون البحثي ليس فقط عبر التخصصات بل عبر الدول للتوصل الى رؤية شاملة من خلال قياسات واقعية للثقافات المحلية

و بشكل أكثر تحديدا فإن دراسات الإعلام ينبغي أن تتطرق الى منطقة بحثية جديدة لدراسة قضاياها وموضوعاتها واختبار فروضها أو قضاياها النظرية والمنهجية في بيئة وسائل التواصل الاجتماعي والبيانات الكبيرة ومن بين الموضوعات التي يمكن أن تمثل محورا لدراسات مستقبلية وكمجرد أمثلة نذكر مثلا :

- دراسات عابرة للثقافات حول المشاعر تجاه أخبار الصراعات والنزاعات في المنطقة العربية ، وهذا يصب في مجال دراسات الإعلام الدولي من ناحية كما يمكن أن يزود صانعي القرار في بلداننا من ناحية أخرى بتوجهات الشعوب تجاه قضايا محددة ومن ثم يمكن إدماج نتائج مثل هذه الدراسات مع



جهود الدبلوماسية الرسمية أو الشعبية للتأثير فى القرارات الدولية تجاه القضايا العربية.

- دراسات فى مجال الصورة الذهنية والصور النمطية السائدة فى فضاء شبكات التواصل الاجتماعى تجاه جماعات نوعية معينة من خلال تحليل آراء مستخدمى هذه الوسائل ومشاعرهم، ومثال ذلك المشاعر تجاه الأقليات الدينية أو العرقية أو المهاجرين غير الشرعيين وهذه النوعية من الدراسات تكشف عن التحيزات الأيديولوجية بل والنزعات العنصرية ودرجة الشوفينية الثقافية (أو الإنغلاق الثقافى) للمجتمعات والشعوب الغربية والعربية على حد سواء.



أهم النتائج والتوصيات

١- تشكل دراسات تنقيب البيانات وتحليل المشاعر مجالاً بحثياً جديداً متعدد التخصصات عميق الجذور فى علوم الحاسب ينتمى الى ما يسمى العلوم الاجتماعية المحوسبة أو الرقمية أو الموجهة بالبيانات ذا طبيعة استكشافية يمكن الباحثين من استخراج الآراء ألياً من كميات ضخمة من البيانات ويحمل إمكانية التوصل الى معلومات وبناء أنماط لم تكن معروفة من قبل.

٢- ترجع البدايات الأولى لذلك المجال الناشئ الى عام ٢٠٠٣ وقد غلب عليه التوجه التجارى ولم يكن هناك اتفاق على المصطلحات حيث استخدمت مصطلحات كثيرة مثل مراقبة العلامات التجارية والمعالجة الحاسوبية للرأى والمشاعر وتحليل الذاتية فى النص واستمرت جهود تنقيب البيانات فى التطور واتسعت استخداماتها وتطبيقاتها وأصبح هناك اتفاق على بعض المفاهيم التى تستخدم على نطاق واسع وعبر تخصصات بينية وانتشرت مفهومات مثل تحليل المشاعر أو استخراج الرأى والبيانات الكبيرة كمفهومات انتقلت من علوم الحاسبات إلى العلوم الاجتماعية . ولا يزال هذا المجال البحثى يُنتج مزيداً من المفاهيم منها تحليل المشاعر المقارن وتنقيب الويب الدلالي وتحليل البيانات المتدفقة إضافة الى مجموعة واسعة من التعريفات المستخدمة فى سياق دراساته وخلال مراحل التنقيب المختلفة مثل مفهوم المجال والكيان وقطبية الرأى والميزة أو الجانب والخوارزميات إضافة الى مفاهيم أخرى تستخدم عند عرض النتائج مثل سحابة الكلمات والخرائط الحرارية للمشاعر وهى فى مجموعها يمكن أن تثرى الدراسات الإعلامية وتعطيها أبعاداً تفسيرية جديدة.



٣- تتعدد مستويات التحليل في أبحاث تحليل المشاعر بين مستوى المستند أو الجملة أو الكلمة أو الجانب كما تتنوع مراحل وتقنياته أو برامجه .و تتمثل أهم مراحل هذه الأبحاث في جمع البيانات من منصات التواصل ويلى ذلك مرحلة تأهيل البيانات أو المعالجة المسبقة باستخدام تقنيات معالجة اللغة الطبيعية وفي مرحلة ثالثة يتم استخدام الخوارزميات لإجراء التحليلات واستخراج الآراء أو المشاعر المتضمنة في النصوص الذاتية وتحديد قطبيتها ،ثم تأتى المرحلة الختامية لعرض النتائج باستخدام الأشكال والرسوم البيانية.

٤- تتطلب دراسات تنقيب البيانات ، تعزيز التعاون متعدد التخصصات وتشكيل فرق بحثية تضم متخصصين من مجالات علمية مختلفة أهمها معالجة اللغة الطبيعية والمتخصصين في تصميم الخوارزميات والمبرمجين فى علوم الحاسب ومعهم متخصصى العلوم الإنسانية هذا من منطلق أنه يصعب على باحث بمفرده إتقان جميع المهارات اللازمة لإجراء دراسة بداية من جمع البيانات من وسائل التواصل أو المواقع الالكترونية مروراً بمرحلة تنقية البيانات وهيئتها واختيار التقنية المناسبة لأهداف البحث أوبناء المعاجم ومجموعات الكلمات اللازمة لتصنيف الآراء وتحديد قطبيتها .

٥- تنوعت الاتجاهات البحثية فى موضوع تنقيب البيانات وتحليل المشاعر ويمكن رصد ستة اتجاهات رئيسية أولها يأتى من تخصصات بعيدة عن العلوم الإنسانية وتهتم غالباً بتطوير تقنيات التعرف على المعلومات الذاتية واستخراجها من النصوص والعمل على تحسين مستوى دقتها ، ويتمثل الاتجاه الثانى فى الدراسات المهمة بتنقيب بيانات وسائل الاتصال السائدة وتتميز هذه الدراسات بزيادة حجم العينات الخاضعة للتحليل ، ويركز الاتجاه الثالث فى التنقيب فى المحتوى الذى يُنشئه المستخدمون وهو يتميز



بإمكانية الكشف عن الأشكال البنائية للاتصال ، ويعنى الاتجاه الرابع بسلوكيات أو أنشطة المستخدمين فى المواقع وشبكات التواصل ، أما الاتجاه الخامس فيجمع بين تنقيب البيانات والأبحاث التقليدية بهدف دراسة ظواهر الاتصال من أكثر من زاوية وأخيراً يأتى الاتجاه السادس ليتمثل فى الدراسات ذات الطابع النقدي التى تناقش الأسس المنهجية والجوانب الأخلاقية لأبحاث تنقيب البيانات وتحليل المشاعر .

٦- تنطوى طريقة تحليل المشاعر على كثير من الفرص البحثية التى يمكن أن تشرى دراسات الاتصال حيث تسمح تقنيات التنقيب_ وبرامجها الحاسوبية التى تتعامل مع مجموعات ضخمة من البيانات _ بدراسة السلوك الاتصالي الفعلى فى الوقت الحقيقى ، والانتقال من الأبحاث الفردية الى الأبحاث التعاونية متعددة التخصصات .

٧ - يوجد كثير من الإنتقادات المنهجية لأبحاث تنقيب البيانات، سواء طبيعة بياناتها الكثيفة وغير المفهومة أم على مستوى إتاحة هذه البيانات لإجراء الأبحاث أم على مستوى معالجتها وتعذر تكرار الأبحاث على ذات المجموعات من البيانات ، وانعكاس ذلك على دقة النتائج والثقة فيها من أجل التراكم المعرفى و البناء عليها للأبحاث التالية، كما تفرض دراسات تحليل المشاعر عدة تحديات أخلاقية تعنى أنه لا يمكن الفصل بين تنقيب محتوى وسائل التواصل والنقاش الأوسع حول أخلاقيات ممارسة الأبحاث فى علم البيانات ، وينطوى ذلك على قضايا أهمها ضرورة احترام الخصوصية الفردية وإمكانية حدوث أشكال جديدة من التمييز والإقصاء وبروز فجوات رقمية جديدة بين باحثى الاتصال أنفسهم فضلاً عن القضايا البنائية الأوسع المتعلقة بزيادة الرقابة والهيمنة وتهديد الحريات العامة .

التوصيات:

إن الحلول الجاهزة والشركات التي تقدم خدمات تحليل المشاعر التجارية مكلفة ولا يستطيع باحث بمفرده تحمل نفقاتها ، لذا من المقترح أن تنتبى الكليات المتخصصة عقد دورات تدريبية للباحثين على البرامج مفتوحة المصدر المخصصة للأغراض الأكاديمية من أجل توظيفها في أبحاثهم على أن توفر الجامعة التقنيات اللازمة لذلك حينها سيصبح الباحثين الإعلاميين أكثر إنتاجية في تعاملهم مع البيانات الكبيرة مع دعم الفرق البحثية في هذا المجال لتعزيز التعاون متعدد التخصصات .

لا ينبغي الإسراف في المبالغة تجاه توظيف تنقيب البيانات في دراسات الإعلام خاصة وأن هناك محتوى أيديولوجي أو ترويجي مصدره شركات بيع البيانات والتحليلات الجاهزة للمشاعر لأغراض تجارية لذا ينبغي أن يكون استخدام مجموعات البيانات الكبيرة وتوظيفها في أبحاث الاتصال مترافقاً مع التفكير في حدودها المنهجية وقضاياها الأخلاقية . ، كما لا ينبغي الاقتصار في التحليلات على التحليلات الكمية في حالة البيانات الضخمة إذ أننا في النهاية علوم إنسانية تبحث في المعاني ، إن تصورات الناس واتجاهاتهم وقيمهم هي أمور لا يمكن الاكتفاء معها بالدراسات الكمية ، فالبشر وتفاعلاتهم ومشاعرهم ليسوا مجرد أرقام ومؤشرات بل تفاعلات إنسانية بين كائنات تعي وتفهم وتُدرك وتُبدع .



المراجع والهوامش

- 1- Abdullah, M., Almasawa, M., Makki, I., Alsolmi, M., & Mahrous, S. (2018). Emotions extraction from Arabic tweets. *International Journal of Computers and Applications*, 1-15. doi:10.1080/1206212x.2018.1482395
- 2- Aldayel, H. K., & Azmi, A. M. (2016). Arabic tweets sentiment analysis – a hybrid scheme. *Journal of Information Science*, 42(6), 782-797. doi:10.1177/0165551515610513
- 3- Al-Kabi, M., Al-Ayyoub, M., & Alsmadi, I. (2016). A Prototype for a Standard Arabic Sentiment Analysis Corpus. *The International Arab Journal of Information Technology*, 13(1A).
- 4- Al-Kabi, M., Alsmadi, I., Khasawneh, .., & Wahsheh, H. (2018). Evaluating Social Context in Arabic Opinion Mining. *The International Arab Journal of Information Technology*, 15(6), 974-982.
- 5- Al-Rawi, A. (2017). Assessing public sentiments and news preferences on Al Jazeera and Al Arabiya. *International Communication Gazette*, 79(1), 26-44. doi:10.1177/1748048516642732
- 6- Arapakis, I., Lalmas, M., Cambazoglu, B. B., Marcos, M., & Jose, J. M. (2014). User engagement in online News: Under the scope of sentiment, interest, affect, and gaze. *Journal of the Association for Information Science and Technology*, 65(10), 1988-2005. doi:10.1002/asi.23096
- 7- Aston, N., Liddle, J., & Hu, W. (2014). Twitter Sentiment in Data Streams with Perceptron. *Journal of Computer and Communications*, 02(03), 11-16. doi:10.4236/jcc.2014.23002
- 8- Atteveldt, W. V., & Peng, T. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, 12(2-3), 81-92. doi:10.1080/19312458.2018.1458084
- 9- Baram-Tsabari, A., Segev, E., & Sharon, A. J. (2017). What's New? The Applications of Data Mining and Big Data in the Social Sciences. *The SAGE Handbook of Online Research Methods*, 92-106. doi:10.4135/9781473957992.n6

- 10- Barbier, G., & Liu, H. (2011). Data Mining in Social Media. *Social Network Data Analytics*, 327-352. doi:10.1007/978-1-4419-8462-3_12
- 11- Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2013). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, 16(2), 340-358. doi:10.1177/1461444813480466
- 12- Driscoll, B. (2015). Sentiment analysis and the literary festival audience. *Media & Cultural Studies*, 29(6), 861-873. doi:10.1080/10304312.2015.1040729
- 13- Giaxoglou, K., Döveling, K., & Pitsillides, S. (2017). Networked Emotions: Interdisciplinary Perspectives on Sharing Loss Online. *Journal of Broadcasting & Electronic Media*, 61(1), 1-10. doi:10.1080/08838151.2016.1273927
- 14- Hadi, W. (2015). Classification of Arabic Social Media Data. *Advances in Computational Sciences and Technology*, 8(1), 29-34. Retrieved from <http://www.ripublication.com/>
- 15- Harvey, K. (2014). Encyclopedia of Social Media and Politics. 349-350. doi:10.4135/9781452244723
- 16- Kemp, S. (2019, January 30). Digital 2019: *Global Internet Use Accelerates*. Retrieved February 8, 2019, from <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>
- 17- Kennedy, H. (2012). Perspectives on Sentiment Analysis. *Journal of Broadcasting & Electronic Media*, 56(4), 435-450. doi:10.1080/08838151.2012.732141
- 18- Kennedy, H. (2016). What Should Concern Us About Social Media Data Mining? *Key Debates. Post, Mine, Repeat*, 41-66. Palgrave Macmillan, London doi:10.1057/978-1-137-35398-6_3
- 19- Kennedy, H., & Moss, G. (2015). Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society*, 2(2), 1-11, 205395171561114. doi:10.1177/2053951715611145



- 20- Kim, Y., & Jeong, S. R. (2015). Opinion-Mining Methodology for Social Media Analytics. *KSH Transactions on Internet and Information Systems*,9(1). doi:10.3837/tiis.2015.01.024
- 21- Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*,4(2), 205395171773633. doi:10.1177/2053951717736336
- 22- Kitchin, R. (2016). Big Data – Hype or Revolution? *The SAGE Handbook of Social Media Research Methods*,27-38. doi:10.4135/9781473983847.n3
- 23- Library, W. E. (٢٠١٧). Data mining. Retrieved November 7,2018 from [http://community.ebooklibrary.org/Results.aspx?PageIndex=1&SearchAuthor=World Heritage Encyclopedia&AuthorType=1](http://community.ebooklibrary.org/Results.aspx?PageIndex=1&SearchAuthor=World+Heritage+Encyclopedia&AuthorType=1)
- 24- Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. *Mining Text Data*,415-463. doi:10.1007/978-1-4614-3223-4_13
- 25- M. (2009, October 23). Weka برنامج للتعامل مع خوارزميات الذكاء الصناعي Retrieved January 27, 2019, from <https://linuxac.org/forum/-/برامج-وتطبيقات-30934/العامه-التطبيقات-البرامج-و-المكتب-سطح-weka--الذكاء-مع-خوارزميات-الذكاء-الصنعي>
- 26- Mahrt, M., & Scharkow, M. (2013). The Value of Big Data in Digital Media Research. *Journal of Broadcasting & Electronic Media*, 57(1), 20-33. doi:10.1080/08838151.2012.761700
- 27- Mostafa, M. M., & Nebot, N. R. (2017). Sentiment analysis of Arabic language influence on Spanish vocabulary: An El País newspaper and Twitter case study. *Journal of Information Technology Case and Application Research*,19(3), 145-157. doi:10.1080/15228053.2017.1394143
- 28- Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. *Journal of Communication*,64(2), 193-214. doi:10.1111/jcom.12088
- 29- Nevius, F. (2017). New Media Analysis. *The SAGE Encyclopedia of Communication Research Methods*,1091-1093. doi:10.4135/9781483381411.n375

- 30- Oliveira, D. J., Bermejo, P. H., & Santos, P. A. (2016). Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls. *Journal of Information Technology & Politics*, 14(1), 34-45. doi:10.1080/19331681.2016.1214094
- 31- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis . *Foundations and Trends® in Information Retrieval*, 2(1-2), 1-135. doi:10.1561/15000000011
- 32- Pendergast, T., & Pendergast, S. (2000). *St. James encyclopedia of popular culture* (Vol. 2). Saint James Press.
- 33- Ramírez-Tinoco, F. J., Alor-Hernández, G., Sánchez-Cervantes, J. L., Olivares-Zepahua, B. A., & Rodríguez-Mazahua, L. (2018). A Brief Review on the Use of Sentiment Analysis Approaches in Social Networks. *Advances in Intelligent Systems and Computing Trends and Applications in Software Engineering*, 263-273. doi:10.1007/978-3-319-69341-5_24
- 34- Salem, F. (2017). The Arab World Online 2017: Digital Transformations and Societal Trends in the Age of the 4th Industrial Revolution. *SSRN Electronic Journal*. doi:10.2139/ssrn.3059445
- 35- Shoukry, A., & Rafea, A. (2015). A Hybrid Approach for Sentiment Classification of Egyptian Dialect Tweets. *2015 First International Conference on Arabic Computational Linguistics (ACLing)*, 78-85. doi:10.1109/acling.2015.18
- 36- Smith, T. W. (2012). Survey-Research Paradigms Old and New. *International Journal of Public Opinion Research*, 25 (2), 218-229. doi:10.1093/ijpor/eds040
- 37- The State of the Internet in Q4 2018. (2018, October 26). Retrieved from <https://wearesocial.com/blog/2018/10/the-state-of-the-internet-in-q4-2018>
- 38- Thelwall, M. (2017). Sentiment Analysis for Small and Big Data. *The SAGE Handbook of Online Research Methods*, 344-360. doi:10.4135/9781473957992.n20
- 39- Thelwall, M., Buckley, K., & Paltoglou, G. (2010). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2), 406-418. doi:10.1002/asi.21462



- 40- Trilling, D., & Jonkman, J. G. (2018). Scaling up Content Analysis. *Communication Methods and Measures*, 12(2-3), 158-174. doi:10.1080/19312458.2018.1447655
- 41- Varathan, K. D., Giachanou, A., & Crestani, F. (2016). Comparative opinion mining: A review. *Journal of the Association for Information Science and Technology*, 68(4), 811-829. doi:10.1002/asi.23716
- 42- Yusof, N. F., Lin, C., & He, Y. (2018). Sentiment Analysis in Social Media. *Encyclopedia of Social Network Analysis and Mining*, 2386-2399. doi:10.1007/978-1-4939-7131-2_120
- 43- Zafarani, R. & Abbasi, M. A.,. (2014). *Social Media Mining: An Introduction*. CAMBRIDGE UNIVERSITY PRESS.
- 44- Zhu, Y., Kim, M.C. & Chen, C. (2017). An investigation of the intellectual structure of opinion mining research *Information Research*, 22(1), paper 739. Retrieved from <http://InformationR.net/ir/22-1/paper739.html>. Archived by WebCite at <http://www.webcitation.org/6oGbzTjxJ>
- ٤٥ - تحليل البيانات العربية - ٢٠١٧). Retrieved January 27, 2019, from <https://ethraarab.wordpress.com/category/أداة-weka/>

(*) من المتوقع أن يصل حجم بيانات الشبكات الرقمية والاجتماعية على الإنترنت إلى ٤ زيتابايت بحلول نهاية عام ٢٠٢٠ م (Mostafa & Nebot, 2016)

(*) الخوارزميات **Algorithms**: ضرورة للطريقة التي تعمل بها أجهزة الحاسبات. وتحتوي العديد من برامج الحاسبات على الخوارزميات التي تتضمن تفاصيل التعليمات المحددة التي ينبغي ان يقوم بها الحاسب (في امر محدد) للقيام بمهمة محددة، ويمكن اعتبار الخوارزمية اي سلسلة من العمليات التي يمكن محاكاتها. وتهدف معظم الخوارزميات الى تنفيذها باعتبارها برامج حاسوبية ، وفي النظم الحاسوبية، تعتبر الخوارزمية أساسا مثلا للمنطق المكتوب في البرمجيات بواسطة مطوري البرمجيات

ICT in Our Lives International Conference. (n.d.). Retrieved January 8, 2019, from https://www.facebook.com/pg/ICTinOurLives/posts/?ref=page_internal

(*) **زاحف الويب Web crawler** : عبارة عن برنامج آلي يقوم بالزحف الى المواقع وتنزيل صفحات الويب والتتقيب عن مسار الصفحات والروابط التشعبية للعثور على معلومات مفيدة

ملحق رقم (١)

نموذج تطبيقي (كيفية تحليل المشاعر)

لإعطاء فكرة مبسطة عن الكيفية التي يتم بها تتقيب البيانات من أجل استخراج الرأى أو تحليل المشاعر لابد من أن نشير بداية إلى بعض المفاهيم التي تعد بمثابة تعريفات إجرائية تستخدم فى هذا المجال. (Varathan, Giachanou, & Crestani, 2016)

- استخراج الرأى opinion Mining : مصطلح يشير إلى استخدام معالجة اللغة الطبيعية وتحليل النص واللغويات الحاسوبية لتحديد واستخلاص المعلومات الذاتية فى المواد النصية وهو يعرف أيضا بتحليل المشاعر.
- تحليل المشاعر sentiment analysis : هو أحد المجالات الناشئة التي تركز على تطوير أساليب يمكنها تلقائياً الكشف عن المعلومات التي يتم رصدها والمتعلقة بهدف محدد وتحديد قطبية الرأى تجاه ذلك الهدف.
- هدف الرأى The opinion target : هو عادة كيان مسمى مثل منظمة أو فرد أو حدث.



• الرأى Opinion هو تقرير ذاتى أو وجهة نظر أو موقف أو عاطفة أو تقييم حول كيان أو جانب من جوانب كيان ، يصدر من صاحب رأى .

• الكيان Entity: هو كائن ملموس أو مجرد مثل منتج أو شخص أو حدث أو منظمة والذي يمكن تمثيله كتسلسل هرمى للمكونات أو المكونات الفرعية لسماته. و على سبيل المثال فى جملة مثل " إن جودة الصورة الخاصة ب كاميرا نيكون الجديدة ممتازة" فإن الكيان الذى تم التعبير عنه هو كاميرا نيكون ، وجانب الكيان هو جودة الصور، وقطبية الرأى إيجابية .

ولمعرفة كيف تتم عملية تنقيب البيانات واستخراج الرأى أو تحليل المشاعر فان هناك مجموعة مهام يتم إنجازها كما يلى (Liu & Zhang, 2012)

المهمة ١: استخراج الكيانات من وثائق الرأى وتجميعها واستخراج جميع تعبيرات الكيان

المهمة ٢ : استخراج الجوانب الخاصة بالكيان وتجميعها .

المهمة ٣ : تحديد صاحب الرأى واستخراج الوقت من البيانات أو من النص.

المهمة ٤ : تصنيف معنويات الجانب بمعنى تحديد ما إذا كان الرأى إيجابياً أو سلبياً أو محايداً

المهمة ٥ : استخراج الرأى

فى إحدى المراجعات الخاصة بشراء جهاز Iphone يمكن أن نوضح الخطوات والمهام السابقة: (Liu & Zhang, 2012)

١- اشترت جهاز iPhone قبل بضعة أيام ٢- كان الهاتف جميلاً



٣- كانت الشاشة رائعة حقاً

٤- كانت جودة الصوت واضحة

٥- ومع ذلك غضبت والدتي منى جداً لأنى لم أخبرها قبل أن أشتريه

٦ - واعتقدت أيضاً أن الهاتف باهظ الثمن وطلبت منى إعادته للمحل .

ما الذى نريد أن نستخلصه من هذا الاستعراض؟ أول شيء نلاحظه أن هناك العديد من الآراء في هذا الشأن، وبإعادة النظر في الجمل ٢ و ٣ و ٤ نجدها تعبر عن آراء إيجابية في حين تعبر الجمل ٥ و ٦ عن آراء أو عواطف سلبية، ويلاحظ أن جميع الآراء لديها بعض الأهداف فالهدف من الجملة ٢ مثلاً هو جهاز ال iPhone ككل وأهداف الآراء في الجمل ٣ و ٤ هي الشاشة التي تعمل باللمس وجودة الصوت للآيفون على التوالي، الهدف من الرأي في جملة ٦ هو سعر الآيفون ولكن الهدف من الرأي أو العاطفة في الجملة ٥ هو أنا وليس الآيفون ويلاحظ أن صاحب الآراء في الجمل ٢ و ٣ و ٤ هو المؤلف للمراجعة (أنا) ولكن في الجمل ٥ و ٦ هو والدتي. وهذا تمثيل بسيط للطريقة التي تعمل بها برامج وخوارزميات تحليل المشاعر .