توظيف تقنيات معالجة اللغة الطبيعية (NLP) في تحليل المشاعر لآراء القراء باستخدام Orange Data Mining: دراسة تطبيقية لمراجعات كتب علوم المكتبات والمعلومات في LibraryThing

Employing Natural Language Processing (NLP) Techniques in Sentiment Analysis of Readers' Opinions Using Orange Data Mining: An Applied Study of Library and Information Science Book Reviews in LibraryThing.

د. أحمد محمد علي عبد المختار

مدرس علوم المكتبات والمعلومات كلية الآداب - جامعة المنيا

تاريخ الإرسال	تاريخ القبول	تاريخ النشر
٢٠٢٤/٧/١١	٢٠٢٤/٧/٢٣	٢٠٢٤/١٠/١

المستخلص:

تهدف الدراسة إلى بحث سبل تعزيز الإفادة من تقنيات التعلم الآلي، ومعالجة اللغة الطبيعية، وأدوات تحليل المشاعر، في فهم تفضيلات القراء، وتحليل سلوكهم. وآرائهم تجاه مصادر وخدمات المعلومات. وذلك من خلال بناء وتصميم سير عمل (Workflow) قائم على تقنيات التعلم الآلي باستخدام Orange Data Mining لتحليل مراجعات القراء للكتب. واعتمدت الدراسة في تحقيق ذلك على المنهج الوصفي التحليلي. وأُجْرِيَت الدراسة التطبيقية على مراجعات القراء حول كتب علوم المكتبات والمعلومات في LibraryThing والتي بلغت (١٦٥٤) مراجعة. وتم استخدام برنامج Orange Data Mining والذي يتكامل مع لغة البرمجة الشهيرة Python، ويدعم العديد من مكتباتها؛ NumPy، NLTK. بالإضافة لبعض الخوارزميات الأخرى مثل؛ Vader، (TF-IDF)، (LDA). وتوصلت الدراسة إلى العديد من النتائج أبرزها؛ أن كتاب Information Representation and Retrieval in the Digital Age جاء في المرتبة الأولى من حيث المشاعر الإيجابية حيث سجل مؤشر(0.6114) compound. وإن المشاعر الإيجابية هي السائدة في مراجعات القراء للكتب، حيث بلغت (١٢٨٣) مراجعة إيجابية بما يمثل (٧٧,٦٪)، في حين بلغ عدد المراجعات السلبية نحو (٢٠٢) مراجعة سلبية، بنسبة (١٢٪) من إجمالي مراجعات القراء. كما أثبتت الدراسة أن نموذج Plutchik يُعد هو الأكثر دقة في تصنيف مراجعات وتعليقات القراء وفقا للمشاعر والعواطف التي تعبر عنها، مقارنة بنموذج Ekman. وأوصت الدراسة المكتبات ومؤسسات المعلومات بضرورة التوسع في استثمار تقنيات معالجة اللغة الطبيعية، في معالجة العديد من القضايا بما يضمن تلبية احتياجات وتوقعات المستفيدين منها. كما أوصت الباحثين بإجراء المزيد من الدراسات المستقبلية البينية بالاشتراك مع المتخصصين في علوم الذكاء الاصطناعي لإعداد خوارزميات مناسبة، وبناء نماذج التعلم الآلي التي تسهم في تطوير وتحسين تجربة المستخدم في المكتبات ومؤسسات المعلومات.

الكلمات المفتاحية: معالجة اللغة الطبيعية (NLP) - تحليل المشاعر – مراجعات الكتب - **LibraryThing – Orange Data Mining.**

Abstract:

The study aims to investigate ways to enhance the utility of artificial intelligence, natural language processing, and sentiment analysis tools in understanding readers' preferences and analyzing their behavior and opinions towards information sources and services. This is achieved through constructing and designing a workflow based

on artificial intelligence techniques using Orange Data Mining to analyze reader reviews of books. The study adopted a descriptive-analytical approach and conducted an empirical study on reader reviews of library and information science books on LibraryThing, totaling 1654 reviews. Orange Data Mining, integrated with the popular Python programming language and supporting libraries such as NumPy and NLTK, was used along with algorithms including Vader, TF-IDF, and LDA. The study revealed several findings, notably that the book "Information Representation and Retrieval in the Digital Age" ranked first in positive sentiment with a compound score of 0.6114. Positive sentiments predominated in reader reviews, totaling 1283 positive reviews compared to approximately 202 negative reviews. Additionally, the study demonstrated that the Plutchik model is more accurate in classifying reader reviews based on emotions compared to the Ekman model. The study recommended that libraries and information institutions expand their investment in natural language processing technologies to address various issues and meet the needs and expectations of their users. Furthermore, researchers recommended conducting more collaborative future studies with artificial intelligence experts to develop suitable algorithms and build machine learning models that contribute to enhancing user experience in libraries and information institutions.

Keywords: Natural Language Processing (NLP), Sentiment Analysis, Book Reviews, Orange Data Mining, LibraryThing

١. مشكلة الدراسة:

في ظل التطور التقني المستمر، والتحول الرقمي الذي يشهده العالم، أصبحت منصات مراجعات الكتب عبر الإنترنت مثل LibraryThing بمثابة أدوات فعالة للتعبير عن آراء وتقييمات القراء تجاه الكتب. حيث تتيح هذه المنصات للقراء فرصة مشاركة آرائهم وتوجهاتهم تجاه محتوى الكتب المختلفة، مما يوفر ثروة من البيانات النصية، والتي تحمل في مَكمنِها العديد من المؤشرات الهامه. إذ تُعد مورداً قيّماً يمكن من خلاله فهم وتقييم توجهات ومشاعر القراء. والذي بدوره يوفر رؤى قيمة تُسهم في تطوير الخدمات المعلوماتية والمعرفية بالمكتبات ومؤسسات المعلومات من ناحية، وتلبية احتياجات القراء من جانب المؤلفين والناشرين من ناحية أخرى.

وبالرغم من ذلك، إلا أن تحليل هذا الكم الهائل من المراجعات النصية يدوياً يُعد أمراً معقداً، ويستغرق وقتاً طويلاً. لذا تسعى الدراسة الحالية إلى بحث سبل تعزيز الإفادة من تقنيات التعلم الآلي، ومعالجة اللغة الطبيعية، في تحليل المشاعر لنصوص مراجعات القراء حول الكتب، بغرض فهم تفضيلات القراء، وتحليل سلوكهم، وآرائهم تجاه مصادر وخدمات المعلومات. وذلك من خلال الدراسة التطبيقية لتحليل مراجعات القراء لكتب علوم المكتبات والمعلومات في LibraryThing.

٢. أهداف الدراسة:

يتمثل الهدف الرئيس للدراسة في بحث سبل تعزيز الإفادة من تقنيات التعلم الآلي، ومعالجة اللغة الطبيعية، وتقنيات تحليل المشاعر، في فهم تفضيلات القراء، وتحليل سلوكهم. وآرائهم تجاه مصادر وخدمات المعلومات. ويتفرع من هذا الهدف مجموعة من الأهداف الفرعية على النحو التالي:

١- إعداد مجموعة بيانات لمراجعات كتب علوم المكتبات والمعلومات بموقع LibraryThing.

٢- عرض الخصائص العددية، والزمنية، واللغوية، إلخ، لكتب علوم المكتبات والمعلومات التي قُيِّمَت من خلال قراء LibraryThing.

٣- تحليل المشاعر لمراجعات وآراء القراء حول كتب علوم المكتبات والمعلومات، وتصنيفها من حيث كونها؛ إيجابية، سلبية، أو محايدة، باستخدام خوارزمية Vader. وتحديد أبرز الكتب التي تلقت أكبر قدر من المراجعات الإيجابية، والسلبية.

٤- عرض نتائج تحليل المشاعر لمراجعات الكتب من خلال خرائط Heat Map.

٥- تحليل مراجعات القراء وفقاً لمقياس الحالة المزاجية (POMS).

٦- مقارنة دقة نتائج تحليل العواطف لمراجعات القراء بين نموذجي Ekman، Plutchik.

٧- تحديد الموضوعات الرئيسية في مراجعات الكتب من خلال نمذجة الموضوعات Topic Modelling باستخدام خوارزمية (LDA)، وتحديد أنسب تقنيات تحليل وتصور نتائج نمذجة الموضوعات.

٨- رصد الكلمات الرئيسية الأكثر تكرارًا في مراجعات القراء لكتب علوم المكتبات والمعلومات من خلال خوارزمية (TF-IDF).

٩- بناء وتصميم سير عمل (Workflow) قائم على تقنيات التعلم الآلي باستخدام Orange Data Mining لتحليل مراجعات القراء للكتب.

٣. تساؤلات الدراسة:

تسعى الدراسة للإجابة عن التساؤلات التالية:

١- ما الخطوات اللازمة لإعداد مجموعة بيانات قابلة للتحليل، والمعالجة النصية، لمراجعات كتب علوم المكتبات والمعلومات؟

٢- ما الخصائص العددية، والزمنية، واللغوية لكتب علوم المكتبات والمعلومات التي قُيِّمَت من قبّل قراء LibraryThing؟

٣- كيف يمكن إجراء تحليل المشاعر لمراجعات وآراء القراء حول كتب علوم المكتبات والمعلومات، وتصنيفها من حيث كونها؛ إيجابية، سلبية، أو محايدة، باستخدام خوارزمية Vader؟ وما أبرز الكتب التي تلقت أكبر قدر من المراجعات الإيجابية، والسلبية؟

٤- كيف يمكن تصور نتائج تحليل المشاعر لمراجعات الكتب من خلال خرائط Heat Map؟

٥- ما نتائج تحليل مراجعات القراء وفقاً لمقياس الحالة المزاجية (POMS)؟

٦- ما مدى دقة نتائج تحليل العواطف لمراجعات القراء باستخدام نموذج Ekman، مقارنة بنموذج Plutchik ؟

٧- ما الموضوعات الرئيسية في مراجعات الكتب التي تم اكتشافها من خلال نمذجة الموضوعات Topic Modelling، وما أنسب تقنيات تحليل وتصور نتائج نمذجة الموضوعات؟

٨- ما الكلمات الرئيسية الأكثر تكرارًا في مراجعات القراء لكتب علوم المكتبات والمعلومات من خلال خوارزمية (TF-IDF)؟

٩- كيف يمكن بناء وتصميم سير عمل (Workflow)، قائم على تقنيات التعلم الآلي باستخدام Orange Data Mining لتحليل مراجعات القراء للكتب؟

٤. أهمية الدراسة:

تتمثل أهمية الدراسة في النقاط التالية:

١. إلقاء الضوء على أهمية الدراسات البينية؛ بالربط بين مجال المكتبات والمعلومات، ومجال الذكاء الاصطناعي وتعلم الآلة. والذي يمكن أن يؤدي إلى تطوير أدوات وخوارزميات حديثة تسهم في تقديم حلول، مبتكرة لتطوير وتحسين كفاءة الخدمات المعلوماتية المقدمة.

٢. يُعد من أهم مخرجات الدراسة إعداد مجموعة بيانات لمراجعات كتب علوم المكتبات والمعلومات في LibraryThing. والتي أتاحها الباحث للإفادة منها في دراسات تطوير وتحسين الخوارزميات المستخدمة في تحليل النصوص، وتدريب نماذج تعلم الآلة من جانب المطورين.

٣. كما تفتح الدراسة مجالًا أمام الباحثين لاستخدام نفس (Workflow) الذي قام بتصميمه الباحث باستخدام Orange Data Mining في تحليل مراجعات مجموعات مختلفة من الكتب من خلال المنصات الأخرى.

٤. إتاحة الفرصة للمكتبات ومؤسسات المعلومات في استكشاف قدرات تقنيات التعلم الآلي، ومعالجة اللغة الطبيعية في فهم تفضيلات القراء، واتجاهاتهم الموضوعية، وبالتالي اتخاذ قرارات صائبة نحو اقتناء مصادر المعلومات التي تلبي احتياجات المستفيدين منها.

٥. كما تُبرز الدراسة أهمية تحليل المشاعر لمراجعات القراء للكتب سواء للناشرين، أو المؤلفين فيما يتعلق بتحديد الموضوعات التي تقع ضمن اهتمامات القراء في فترات معينة، وتقييم المحتوى ووجهات النظر المعروضة، فضلًا عن تحديث خطط النشر المستقبلية، بما يلبي توقعات القراء واهتماماتهم.

٥. منهج الدراسة:

نظرًا لطبيعة الدراسة الحالية، اعتمد الباحث على المنهج الوصفي التحليلي، لوصف ومناقشة سبل الإفادة من تقنيات التعلم الآلي، وتقنيات معالجة اللغة الطبيعية، في تحليل المشاعر لمراجعات القراء لكتب علوم المكتبات والمعلومات بمنصة LibraryThing.

١/٥ مجتمع وعينة الدراسة:

يتمثل مجتمع الدراسة في كُلِّ المراجعات التي كُتبت من قَبْل القراء تعْليقًا على كتب علوم المكتبات والمعلومات، حيث اتبعت الدراسة أسلوب الحصر الشامل، وتم رصد (١٦٥٤) مراجعة قُدِّمَت من قَبْل قراء LibraryThing كتقييم لمحتوى كتب علوم المكتبات والمعلومات.

ويرجع السبب وراء اعتماد الباحث على موقع LibraryThing للمبررات التالية:

١- لم يرصد الباحث أية دراسات سابقة لتحليل مراجعات الكتب من خلال موقع LibraryThing بشكل عام، وفي مجال المكتبات والمعلومات بصفة خاصة.

٢- يُعد LibraryThing في رأي الباحث مجتمعًا نشطًا من القراء والباحثين المهتمين بالكتب العلمية الأكاديمية، مقارنة بالعديد من منصات الكتب الأخرى، والتي يغلب عليها الروايات والقصص، ودواوين الشعر، إلخ.

٣- استعرض الباحث مراجعات كتب علوم المكتبات والمعلومات في بعض منصات ومواقع الكتب على سبيل المثال لا الحصر؛ موقع Amazon، حيث تبين أن مستوى التفاعل والمشاركة والتقييم منخفض، مقارنة بموقع LibraryThing، والذي يتمتع بحجم أكبر من المشاركة

والتفاعل، ويتضمن مراجعات ذات جودة عالية، ومناقشات وتحليلات عميقة حول كتب علوم المكتبات والمعلومات، وهو ما دعا الباحث للاعتماد عليه.

٢/٥ أدوات الدراسة:

اعتمدت الدراسة على الأدوات التالية:

١. برنامج MS Excel: لتخزين وترتيب بيانات مراجعات القراء في LibraryThing.

٢. Orange Data Mining: أداة برمجية مفتوحة المصدر تستخدم للتنقيب عن البيانات والتعلم الآلي. طُوِّرَت في جامعة Ljubljana في سلوفينيا، وتتكامل هذه الأداة مع لغة البرمجة الشهيرة python، والمكتبات الخاصة بها. ويمكن من خلالها؛ تحليل البيانات وتصورها، وبناء نماذج التعلم الآلي وتدريبها واختبارها ونشرها (Salem et al,2021).

٣. بعض مكتبات لغة البرمجة بايثون؛ NLTK،NumPy.

٤. خوارزمية Vader بغرض تصنيف وتحليل المشاعر للنصوص.

٥. نموذجي Ekman، Plutchik لتحليل العواطف والمشاعر للقراء.

٦. نمذجة الموضوعات (Topic Modelling) بالاعتماد على خوارزمية، Latent Dirichlet Allocation (LDA).

٧. خوارزمية (TF-IDF) لتحديد الكلمات الرئيسية في النصوص.

٦. حدود الدراسة:

الحدود الموضوعية: تناقش الدراسة سبل الإفادة من تقنيات التعلم الآلي، وتقنيات معالجة اللغة الطبيعية، في تحليل المشاعر لمراجعات القراء حول كتب علوم المكتبات والمعلومات بموقع LibraryThing.

الحدود الزمنية: تم جمع وإعداد مجموعة البيانات الخاصة بمراجعات كتب علوم المكتبات والمعلومات من خلال موقع LibraryThing في الفترة من ٢٥ نوفمبر ٢٠٢٣م حتى ٩ فبراير ٢٠٢٤م.

الحدود اللغوية: جميع مراجعات القراء التي جُمِعَت من خلال موقع LibraryThing باللغة الإنجليزية.

الحدود النوعية: اقتصرت الدراسة على تحليل المراجعات المنشورة حول كتب علوم المكتبات والمعلومات فقط.

٧. الإطار النظري للدراسة:

١/٧ المراجعة العلمية للإنتاج العلمي:

اعتمد الباحث في مراجعة الإنتاج الفكري المتصل بموضوع الدراسة، على العديد من أدوات بحث الإنتاج الفكري ومنها؛ قواعد البيانات المتاحة من خلال بنك المعرفة المصري؛ قواعد بيانات دار المنظومة، والكشاف العربي للاستشهادات المرجعية - Arabic Citation Index - Web of Science – Scopus - ProQuest - science direct، بالإضافة لمحركات البحث العلمية Semantic Scholar – core - google scholar. فضلا عن فهرس اتحاد المكتبات الجامعية المصرية، والدوريات والمؤتمرات المتخصصة في مجال المكتبات والمعلومات).

وقد أسفرت نتائج البحث عن العديد من الدراسات العربية والأجنبية ذات الصلة بموضوع التنقيب عن البيانات، ومعالجة اللغة الطبيعية، وتقنيات تحليل المشاعر، ونظراً لكثرة تلك الدراسات وارتباطها بكافة فروع المعرفة البشرية، وبالعديد من القضايا المجتمعية، تم الاقتصار فقط على عرض الدراسات ذات الصلة بمجال المكتبات والمعلومات، والتي تتناول مجالات الإفادة من تقنيات معالجة اللغة الطبيعية، وتحليل المشاعر في المكتبات ومؤسسات المعلومات، وفيما يلي عرض لأبرز تلك الدراسات.

فعلى المستوى العربي؛ رُصِدت العديد من الدراسات التي تطرقت لموضوع معالجة اللغة الطبيعية، والإفادة من تقنيات تحليل المشاعر والآراء في المجالات المختلفة، منها على سبيل المثال لا الحصر في مجال الإعلام؛ دراسات تحليل مشاعر الرأي العام تجاه الانتخابات الرئاسية الأمريكية (الشريف،٢٠٢٢). ودراسة تحليل مشاعر الجمهور العربي نحو العلاقات العامة عبر تويتر (عطية،٢٠٢٣)، وتحليل المشاعر في وسائل التواصل الاجتماعي (الخليفي،٢٠١٩)، وتحليل اتجاهات مستخدمي مواقع التواصل الاجتماعي نحو جائحة كورونا (سامي،٢٠٢١)، وفي مجال التسويق ودراسة مشاعر العملاء (رية وآخرون، ٢٠٢٢)، وفي مجال السياحة ودراسات تحليل مشاعر السائحين تجاه المزارات السياحية المختلفة (مصطفاوي، ٢٠٢٢). وبالرغم من ذلك، إلا أن موضوع معالجة اللغة الطبيعية وتحليل المشاعر، لم يحظ بالتغطية الكافية حتى الآن في مجال المكتبات والمعلومات على الصعيد العربي.

وتعد دراسة (سميع،٢٠٢٢) أولى الدراسات العربية التي ناقشت أسلوب تحليل المشاعر، كأحد تطبيقات الذكاء الاصطناعي، وذلك بهدف الكشف عن فرص الاستفادة منه كمؤشر لتقييم خدمات المعلومات من خلال قياس رضاء المستفيدين عن طريق التحليل الآلي لما ينشر من تعليقات عبر وسائل التواصل الاجتماعي للصفحات الخاصة بمؤسسات المعلومات، واعتمدت

الدراسة على المنهج الاستكشافي من خلال إجراء دراسة على عينة بلغت (٦٦٥) تغريدة، أو تعليقا للمستفيدين عبر وسائل التواصل الاجتماعي (تويتر وفيس بوك)، واعتمدت الدراسة على منصة Google Colab. وأثبتت الدراسة أهمية الاستفادة من البيانات الناتجة مثل: تحديد قطبية آراء المستفيدين (إيجابية – محايدة – سلبية) نحو ما تنشره المكتبة من منشورات، في قياس رضاء المستفيدين بما يسهم في دعم عمليات اتخاذ القرار لتحسين مستوى تقديم الخدمات، وأوصت الدراسة بضرورة تبني دراسات بينية يتعاون فيها باحثون من تخصصات علوم المكتبات والمعلومات، والحاسبات، واللغويات لتصميم وبناء أدوات برمجية جاهزة وموارد معجمية تدعم إمكانات اللغة العربية، يمكن أن تستخدمها المكتبات دون الحاجة إلى التعقيدات البرمجية لعمليات تحليل الآراء بصورة مفصلة.

أما على الصعيد العالمي؛ تم رصد العديد من الدراسات التي تناولت موضوع تحليل المشاعر في المكتبات ومؤسسات المعلومات بمختلف أنواعها على النحو الأتي.

فعلى مستوى المكتبات العامة؛ جاءت دراسة (Collins & Karami, 2018) تهدف لاستكشاف محتوى التغريدات التي نشرتها تسع مكتبات عامة؛ في شمال شرق الولايات المتحدة الأمريكية وذلك في ديسمبر ٢٠١٧م، وجمعت الدراسة أكثر من ١٩,٠٠٠ ألف تغريدة من تويتر من خلال الحسابات الخاصة بالمكتبات محل الدراسة؛ سبع مكتبات حكومية، ومكتبتين عامتين حضريتين. وطُبِّقَت الأساليب الحسابية لجمع التغريدات واكتشاف موضوعات تشملها ذات معنى. وأشارت النتائج إلى أن معظم التغريدات كانت تتعلق بخدمات المكتبة، والموارد، فضلاً عن الفعاليات، والبرامج، الخ. كما أثبتت الدراسة أن المكتبات العامة تستخدم حساباتها على تويتر بشكل رئيسي للتواصل مع روادها والمجتمع. وهدفت دراسة (Çakmak & Eroğlu,2021) إلى تحليل المنشورات التي تشاركها المكتبات العامة في تركيا على صفحاتها الرسمية على فيسبوك. وذلك بغرض التعرف على مستوى تفاعل المستخدمين مع المنشورات التي تشاركها المكتبات (استنادًا إلى الإعجاب، التعليق والمشاركة)؟ ورصد فئات المحتوى التي تشاركها المكتبات، وتناولت الدراسة البيانات المتعلقة بـ٢٣,٠٢٧ منشورًا، والتي شاركتها ١١٦ مكتبة عامة من خلال التحليلات الوصفية وتحليل المحتوى. واعتمدت الدراسة على تطبيقات تحليل الكلمات التركية، وتقنية نمذجة المواضيع LDA، وتوصلت الدراسة إلى عدة نتائج منها؛ أن تفاعل المستخدمين مع المكتبات يحدث بشكل عام في شكل إعجاب، وأن عددًا كبيرًا من المكتبات العامة في تركيا استهدفت المستخدمين من الأطفال في مرحلة ما قبل المدرسة إلى الشباب، وأوصت الدراسة

بإجراء دراسات مستقبلية لبحث التحليلات بناءً على معالجة الصور واكتشاف أنماط أكثر تعقيدًا، يمكن الإفادة منها في مؤسسات المعلومات.

وجاءت دراسة كل من (Khan & Loan,2022) تهدف إلى تحليل مراجعات وآراء المستفيدين حول المكتبات العامة في مدينة دلهي في الهند، من خلال منصة خرائط جوجل. واُخْتِيرَت المكتبات التي تملك أكثر من (٥٠) مراجعة. وتم إجراء تحليل المشاعر للمراجعات وآراء المستفيدين بغرض تسليط الضوء على الجوانب الإيجابية والسلبية والمحايدة للمكتبات المختارة. وتوصلت الدراسة إلى عدة نتائج منها؛ أن عدد المراجعات التي تمت تحليلها بلغت (٦٨٥) مراجعة، منها (٧٠,٥١٪) تعبر عن مشاعر إيجابية، و (١٥,٣٢٪) سلبية، و (١٤,١٪) محايدة. وأن الموضوعات السلبية تتعلق بفئة الموظفين، تليها المرافق. في حين أن التكنولوجيا، ومناطق الأطفال حصلت على ردود فعل إيجابية. وأثبتت الدراسة أنه يمكن استخدام خرائط جوجل، كأداة تفاعلية جديدة، يمكن لإدارة المكتبة استخدامها للتواصل مع المستخدمين، عِوَضا عن صناديق الاقتراحات التقليدية والنماذج عبر الانترنت، وأوصت الدراسة بتبني أسلوب تحليل المراجعات في مكتبات دلهي العامة من خلال منصة خرائط جوجل حتى تتمكن من وضع استراتيجيات فعالة لتحسين جودة خدماتها. ومؤخرًا جاءت دراسة (Kim & Noh,2022) تهدف إلى فهم التصورات العامة من قبل أفراد المجتمع تجاه المكتبات العامة، واعتمدت الدراسة على مجموعة بيانات جُمِعَت من خلال المدونات ومواقع الإنترنت، في الفترة من ١ يناير ٢٠١٦، وحتى ٣٠ نوفمبر ٢٠٢٠م. ومن نتائج الدراسة؛ أن أكثر الكلمات الرئيسية تكرارًا وفقًا لخوارزمية TF-IDF؛ كانت "القراءة" (٦,٦٦٨ مرة)، تليها كلمة "كتاب" (٤,٨٥٥ مرة).

وعلى مستوى المكتبات الأكاديمية؛ جاءت دراسة كل من (Al-Daihani & Abrahams,2016) تهدف إلى التنقيب عن النصوص لتغريدات عشر مكتبات أكاديمية من خلال منصة تويتر. وشملت مجموعة البيانات ٢٣,٧٠٧ تغريدة، تتضمن ١٧,٨٤٨ إشارة، و٧,٦٢٥ وسمًا، و٥,٩٧٤ إعادة تغريد. وأظهرت النتائج تباينًا بين المكتبات الأكاديمية في توزيع التغريدات على مدار الوقت. وأن الكلمة الأكثر تكرارًا كانت "open "، والتي استخدمتها المكتبات في سياقات متنوعة. وأكثر كلمتين تكرارًا في التغريدات المجمعة هو مصطلح المجموعات الخاصة "special collections "، وكانت الفئة الأكثر شيوعًا للتغريدات هي الموارد "Resources " بين جميع المكتبات الأكاديمية المختارة. وأوصت الدراسة المكتبات الأكاديمية بأهمية استخدام تقنيات تحليل البيانات والنصوص لفهم البيانات الاجتماعية، مما يساعد في اتخاذ القرارات، والتخطيط الاستراتيجي للتواصل مع المستفيدين، وتسويق خدمات المعلومات. كما هدفت دراسة (Parabhoi, L,2019)

إلى تحليل المشاعر لتعليقات تويتر الخاصة بعشر مكتبات جامعية. واعتمدت الدارسة على تصنيف التايمز للجامعات العالمية في تحديد هذه الجامعات. وتم جمع (١٥٨٥٠) تغريدة في الفترة من يناير ٢٠١٣ وحتى سبتمبر ٢٠١٩ من خلال واجهة برمجة تطبيقات تويتر (API). وتوصلت الدراسة إلى العديد من النتائج منها؛ أن عدد تغريدات مكتبات Bodleian بجامعة أكسفورد بلغ (٢٧٦٠) وبلغ عدد المتابعين لديها (٧٦١٨٠) متابع. وحازت تغريدات مكتبات Bodleian على النصيب الأكبر من المشاعر الإيجابية برصيد (١,٧٧٢٨). وأن أكثر الكلمات تكرارًا؛ كلمة " exhibition" حيث استخدمت (٤٩٩) مرة، وكلمة "archive" (٤٠١) مرة في جميع التغريدات. وأوصت الدراسة بأن المكتبات يمكنها استخدام مواقع الشبكات الاجتماعية وفحص تعليقات المستخدمين، والمراجعات التي يقدمها المستخدمون على المنشورات. في تحسين جودة العمل، والتغلب على المشكلات، واتخاذ قرارات أفضل للمستقبل.

أما فيما يتعلق بالدراسات التي تناولت تحليل مراجعات الكتب؛ فجاءت دراسة (Almjawel,2019) تهدف إلى تحليل المشاعر لعينة من مراجعات الكتب بموقع أمازون، واعتمدت الدراسة على مجموعة بيانات تضم حوالي ١٠٠٠ مراجعة، وتم استخدام برنامج Tableauولغة R، في تحليل المراجعات، بغرض بحث إمكانية اكتشاف وتحديد المراجعات الأكثر تأثيرًا على كتاب معين، سواء كانت إيجابية أو سلبية أو محايدة. وكذلك تحديد أي الكتب حصلت على أكبر عدد من المراجعات. وتوصلت الدراسة إلى عدة نتائج منها؛ إن كتاب " Water for Elephants" حصل على أكبر عدد من المراجعات. وأن نسبة المراجعات الإيجابية هي الأعلى بالمراجعات السلبية والمحايدة. وجاءت دراسة (Marinho et al.,2022) تهدف إلى تطبيق تقنية تحليل المشاعر على المراجعات المنشورة على منصة Skoob من أجل اقتراح معيار تقييم جديد لمساعدة المستخدمين في اتخاذ قراراتهم بشأن قراءة الكتاب من عدمه. واعتمدت الدراسة على المنهج الاستكشافي، حيث استخدمت تقنية تحليل المشاعر، لتحديد درجة القطبية للآراء، والتي يمكن أن تكون إيجابية، سلبية أو محايدة. واعتمدت الدراسة على مجموعة بيانات تضم (٤٥,١١٤) مراجعة تتعلق بأكثر ٢٠ كتابًا قرأها مستخدمو منصة Skoob، وأثبتت الدراسة أن هذا الأسلوب ساعد مستخدمي منصة Skoob في اتخاذ قراراتهم بشأن الكتب التي يبدؤون بقراءتها أو الكتب التي يضعونها في قائمة قراءاتهم المقبلة. كما أثبتت الدراسة أن مراجعات الكتب يمكنها التأثير على تفضيلات القراءة للمستخدمين، وكذلك تقديم الخصائص الإيجابية والسلبية لكتاب معين. مما يجعل من الممكن قياس سلوك المستخدمين تجاه الكتب التي قَرَءوهَا، وكذلك استخدامها كمعيار بديل لتقييم الكتب.

وفيما يتعلق باستخدام برنامج Orange Data Mining:

رصد الباحث العديد من الدراسات التي اعتمدت على برنامج Orang Data Mining في التنقيب عن البيانات، ومعالجة النصوص، وتحليل المشاعر، وبناء النماذج القائمة على تقنيات التعلم الآلي في المجالات المختلفة ومنها؛ دراسة (Peker et al.,2018)، ودراسة (Ferhi & Djillali,2023) والتي هدفت كل منها إلى بناء نموذج قائم على تقنيات التعلم الآلي لتشخيص مرضى السكري. ودراسة (Umer,2020) التي قدمت نموذَجًا قادرًا على التنبؤ بأمراض القلب، ودراسة (Yahyia & Elsanoussi,2023) التي هدفت إلى توظيف أدوات تنقيب البيانات في التنبؤ بمرض فقر الدم. وكذلك دراسة (Koleva & Filipov,2023) التي هدفت إلى بناء نموذج لتصنيف صور الأشعة السينية. وقدمت دراسة (Thange,2021) تحليلاً لبيانات كوفيد-١٩، واعتمدت دراسة (Abdulqader,2023) على البرنامج أيضاً في التنبؤ بانبعاثات ثاني أكسيد الكربون. كما هدفت دراسة (Tiwari & Gunjan,2022) إلى تحليل تأثير البيئة المحيطة على أداء الطلاب، وكذلك تحليل مبيعات السوق لمنتجات معينة من خلال دراسة (Martinez et al.,2021).

ودراسة (Sajwan & Ranjan,2029) التي قدمت نموذجاً لتصنيف صور الزهور وفقا لتقنيات التعلم الآلي. ونظراً لكثرة تلك الدراسات التي اعتمدت على برنامج Orange Data Mining، وتنوع أهدافها، ومجالاتها، يعرض الباحث الدراسات ذات الصلة بموضوع معالجة النصوص، وتحليل المشاعر فقط، على النحو الاتي:

حيث جاءت دراسة (Marcu & Danubianu,2020) تهدف إلى تحليل المشاعر لآراء الطلاب وذلك من خلال الإجابة على مجموعة من الأسئلة التي تتعلق بجوانب مختلفة من العملية التعليمية، باستخدام نموذج Google docs ، وتم الحصول على ١٩٩ إجابه من إحدى عشرة مدرسة ثانوية في سوتشافا، واعتمدت الدراسة في معالجة البيانات وتحليل المشاعر على Orange Data Minin. وعرضت الدراسة نتائج تحليل المشاعر باستخدام نموذجي Ekman، Plutchik وفقاً لفئات تحليل المشاعر المختلفة؛ الغضب، الاشمئزاز، الخوف، الفرح، الحزن، والدهشة. ودراسة (Nahar,K et al.,2020) التي هدفت إلى تحليل المشاعر وتصنيف تعليقات الفيسبوك العربية الأردنية تجاه مجموعة من شركات الاتصالات الأردنية وهي؛ (Zain, Orange and Umniah) وذلك باستخدام المنهج القائم على المعجم وتقنيات التعلم الآلي. بالاعتماد على برنامجي KNIME وorange وكانت النتائج التي تم التوصل إليها فيما يتعلق بمعجم المشاعر العربية الذي تم تقييمه واعدة. حيث جاءت المشاعر الايجابية بنسبة (٦٠٪)، في حين بلغت نسبة المشاعر السلبية (٤٠٪). وبلغت دقة المعجم العام (٩٨٪).

وجاءت دراسة (Al Assyam & Hasan,2023) تهدف إلى تقييم ردود الفعل للمجتمع الإندونيسي تجاه نقل العاصمة الوطنية إلى مدينة IKN Nusantara، واعتمدت الدراسة في جمع بياناتها على تويتر، حيث تم جمع (١٠٠) مراجعة. واستخدمت الدراسة برنامج Orang Data Mining، في عملية معالجة وتنظيف البيانات، وتحليل المشاعر. وتوصلت الدراسة إلى عدة نتائج؛ حيث أظهرت سحابة الكلمات أن أكثر الكلمات تكراراً؛ "Nusantara،IKN Nusantara"، وأن نتائج تحليل المشاعر تمثلت في ست فئات على النحو الاتي؛ الدهشة (Surprise) بنسبة ٨٠٪، والفرح (Joy) بنسبة ٥٠٪، والحزن (Sadness) بنسبة ١٥٪، والخوف (Fear) بنسبة ١٠٪، والاشمئزاز (Disgust) بنسبة ٠٪، والغضب (Anger) بنسبة ٠٪. في حين ركزت دراسة (باتي،٢٠٢٣) على بناء نموذج قائم على تقنيات التعلم الآلي لتصنيف التمور السعودية (العجوه، المجدول). وذلك من خلال إنشاء مجموعة بيانات متوازنة الصنفين تتضمن خصائص تلك التمور، وصور كل نوع. واعتمدت الدراسة في ذلك على برنامج Orange Data Mining، وموقع Teachable Machine. وتم تدريب واختبار النموذج، وتوصلت الدراسة إلى تحقيق نتائج عالية الدقة في تصنيف التمور بشكل صحيح.

١/١/٧ التعليق على الدراسات السابقة:

اتضح من خلال مراجعة الانتاج الفكري في موضوع الدراسة الاتي:

١. لم تحظ دراسات التنقيب عن البيانات، وتحليل المشاعر في مجال المكتبات والمعلومات، بالتغطية الكافية على الصعيد العربي. حيث لم يتم رصد أية دراسات عربية متخصصة سوى دراسة (سميع،٢٠٢٢). كما أن غالبية الكتابات العربية في هذا الصدد يُغلب عليها الطابع النظري، والمنهج الاستشرافي، وفي رأي الباحث أننا بحاجة ماسة للانتقال من المرحلة الاستشرافية، والتي خلفت مئات الدراسات، إلى مرحلة الدراسات، التجريبية، والتطبيقية، لتعظيم الإفادة من التقنيات الناشئة في مجال المكتبات والمعلومات.

٢. لم يرصد الباحث أية دراسات على المستوى العربي تناولت تحليل المشاعر لمراجعات الكتب. وبالرغم من رصد بعض الدراسات التي تناولت مراجعات الكتب على المستوى الأجنبي. إلا أنها لم تطرق لمراجعات كتب علوم المكتبات والمعلومات، أو تحليل مراجعات الكتب من خلال LibraryThing. وبذلك تُعد الدراسة الحالية أولى الدراسات التي تطرقت لتحليل المشاعر لمراجعات القراء حول كتب علوم المكتبات والمعلومات.

٣. بالرغم أن بيئة Orange Data Mining تدعم التعامل مع النصوص العربية، إلا أنه لم يتم رصد أي دراسات عربية اعتمدت على البرنامج سوى دراستين؛ دراسة (باتي،٢٠٢٣)، والتي

حاولت بناء نموذج قائم على تقنيات التعلم الآلي لتصنيف التمور السعودية. ودراسة (Yahyia & Elsanoussi,2023) التي هدفت لبناء نموذج للتنبؤ بحالات مرض فقر الدم، بالاعتماد على مجموعة بيانات لعينة من زوار أحد المختبرات بليبيا. – وإن كانت هاتان الدراستان لم تَطّرقا لتحليل المشاعر والآراء - وعلى الجانب الآخر من الصعيد العالمي، رُصِدَت عشرات الدراسات الأجنبية التي استخدمت البرنامج في مراحل محتفلة؛ كمعالجة النصوص، والتنقيب عن البيانات، وتحليل المشاعر، ونمذجة الموضوعات، وبناء نماذج التعلم الآلي، في مختلف المجالات الموضوعية، وهو ما استدعى الاستعانة به في الدراسة الحالية.

٤. اتفقت الدراسات السابقة التي أُشير إليها على أهمية الاعتماد على تقنيات معالجة اللغة الطبيعة، وتقنيات التعلم الآلي في تحليل المشاعر والآراء، ومدى دقتها في تصنيف مشاعر الأفراد وتوجهاتهم حيال القضايا المختلفة. وكذلك أهمية هذا الأمر لمتخذي القرار، وصانعي السياسات، سواء على مستوى المؤسسات، أو الدول. وهو الأمر الذي يُعزز توجه البحث الحالي، وأسلوب التحليل المتبع فيه، في ضوء أهمية مراجعات كتب علوم المكتبات والمعلومات، وأهمية تحليلها سواء للمكتبات، المؤلفين، والقراء.

٢/٧ تحليل المشاعر Sentiment Analysis.

يرجع ظهور مصطلح تحليل المشاعر إلى أوائل العقد الأول من القرن الحادي والعشرين، ومن أوائل الدراسات التي تناولت مصطلح تحليل المشاعر؛ (Pang & Vaithyanathan,2002) (Nasukawa &Yi.,2003) (Dave et al.,2003). وهناك العديد من المصطلحات الأخرى التي تُتَدَاوَل للتعبير عن موضوع تحليل المشاعر sentiment analysis والتي تتمثل في المصطلحات التالية؛ التنقيب عن الرأي opinion mining، ، التنقيب عن المشاعر sentiment mining، استخلاص الرأي opinion extraction، تحليل التأثير affect analysis، تحليل العاطفة emotion analysis، التحليل الذاتيsubjectivity analysis، التنقيب عن المراجعات review mining، إلخ.

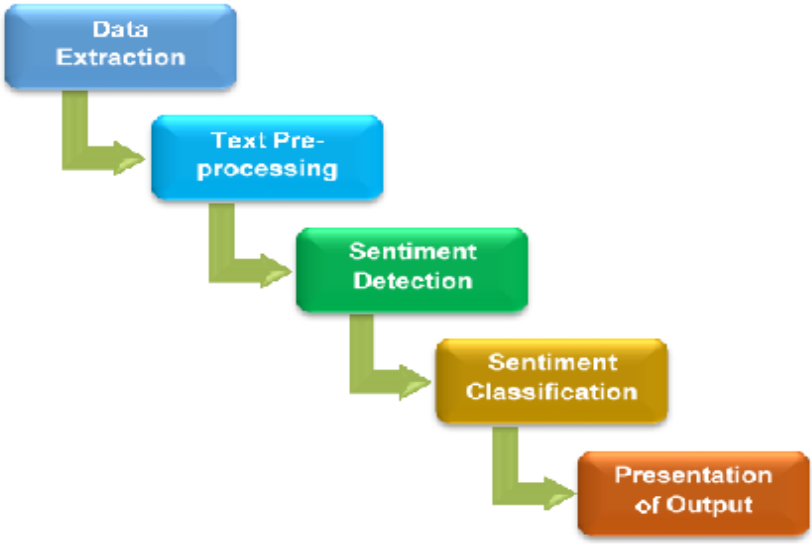
ويهتم مجال تحليل المشاعر بدراسة وتحليل آراء الأشخاص، ومشاعرهم، وتقييماتهم، واتجاهاتهم، تجاه الكيانات المختلفة مثل؛ الخدمات، والمنتجات، والسلع، والهيئات، والأفراد والقضايا، والأحداث الجارية، إلخ. (Liu,2022). وتُعد تقنية تحليل المشاعر وسيلة فعالة لتحليل المشاعر الشخصية للكاتب عن موضوع معين، حيث يمكن من خلالها تصنيف اتجاهات ومشاعر الشخص من حيث كونها؛ إيجابية أو سلبية. كما يمكن لتقنيات تحليل المشاعر التعامل مع النصوص من خلال الكلمات الدالة على المشاعر المرتبطة بالسياق، بحيث يمكن

تصنيف الآراء، والتعليقات حول أحداث معينة من حيث كونها إيجابية، سلبية، أو محايدة بشكل موضوعي ودقيق. وبذلك تتعدد استخداماتها في مختلف القطاعات من خلال تحليل ردود الفعل للجمهور، بغرض حل المشكلات، وتحسين الخدمات المقدمة (Noh, Kim,2022). كما يستخدم محللو البيانات في الشركات الكبيرة تقنيات معالجة اللغة الطبيعية في تحليل المشاعر لقياس الرأي العام، وتتبع صور العلامة التجارية والمنتجات، وتقييم تجارب المستهلكين، وإجراء أبحاث السوق، بهدف تحسين خدمة العملاء، وإدارة الأزمات، وبالتالي زيادة الإيرادات من المبيعات (De et al.,2022).

كما تُعرف تقنيات تحليل المشاعر بأنها إحدى المهام الأكثر نجاحًا في معالجة اللغة الطبيعية (NLP) حيث تستخدم في تحديد وتحليل الموضوعات الدقيقة في النصوص التي يكتبها المستخدم على هيئة آراء، أو تعليقات، أو مراجعات. كما أن لها قيمة معترف بها في العديد من التطبيقات والأعمال التجارية مثل مراجعات الأفلام التي يمكن أن تعزز أو تضر بإيرادات الفيلم، ومراجعات المنتجات التي يمكن أن تعزز جودة المنتج، أو تضر بمبيعاته (Marreddy & Mamidi,2023).

٣/٧ خطوات تحليل المشاعر.

تشمل عملية تحليل المشاعر للنصوص خمس مراحل أساسية، على النحو الأتي (Zishumba,2019):



شكل (١) خطوات تحليل المشاعر

المصدر: (Ahmed et al., 2020,p.438)

١/٣/٧ استخراج البيانات Data Extraction : تُعد المرحلة الأساسية في عملية تحليل المشاعر، وتتضمن جمع البيانات المراد تحليلها من المصادر المختلفة مثل شبكات التواصل الاجتماعي وغيرها.

وهناك العديد من الأدوات التي يمكن الاعتماد عليها في جمع البيانات، ويتوقف اختيار تلك الأدوات على حسب طبيعة مصدر البيانات. فيمكن استخدام Web Scraping لاستخراج البيانات من مواقع الويب. أو استخدام واجهات برمجة التطبيقات (APIs) لجمع البيانات من الشبكات الاجتماعية مثل استخدامAPI Twitter لجمع تغريدات موضوع معين. أو الاعتماد على مجموعات البيانات المتاحة من خلال الانترنت مثل Kaggle.

٢/٣/٧ معالجة النصوص Text Preprocessing :

وتضمن هذه المرحلة تنظيف البيانات قبل إجراء عملية التحليل، وذلك من خلال حذف البيانات غير الضرورية مثل؛ الروابط، والعلامات الخاصة، والرموز التعبيرية، إلخ.

٣/٣/٧ اكتشاف المشاعر Sentiment Detection

تتضمن هذه المرحلة استخلاص الآراء من المراجعات باستخدام تقنيات معالجة النصوص. حيث تُقَسَّم البيانات النصية إلى جمل، ثم تحليل كل جملة للتحقق من مدى موضوعيتها. مع استبعاد الجمل التي ليس لها طابع شخصي. وبذلك تصبح النصوص جاهزة لاكتشاف المشاعر، ويمكن إجراء ذلك سواء على مستوى الكلمة الواحدة، أو العبارة، أو الجملة، أو كامل الوثيقة.

٤/٣/٧ تصنيف المشاعر Sentiment Classification

يتم في هذه المرحلة تحديد القطبية، بحيث يتم تصنيف الجمل، والكلمات إلى ثلاث مستويات؛ إيجابية، سلبية، أو محايدة.

٥/٣/٧ عرض النتائج Presentation of Output

يُعد الهدف الأساسي من عملية تحليل المشاعر هو تحويل تلك النصوص غير المفهومة إلى مؤشرات ورؤي واضحة. ولذلك لا بد من عرض نتائج التحليل من خلال تقنيات التمثيل المرئي للبيانات باستخدام الخرائط والاشكال الرسومية، حتي يسهل قراءتها وفهمها من جانب متخذ القرار.

٨. الإطار العملي للدراسة:

١/٨ إجراءات الدراسة:

مرت الدراسة بعدة مراحل على النحو الاتي:

١/١/٨ مرحلة جمع البيانات:

جمع الباحث بيانات الدراسة، والتي تتعلق بمراجعات كتب علوم المكتبات والمعلومات، وذلك من خلال الولوج لموقع LibraryThing. واستعراض كتب علوم المكتبات والمعلومات، وفحص المراجعات التي كُتب من جانب القراء كتقييم لمحتوى الكتب. وبلغ إجمالي مراجعات القراء التي جمعها (١٦٥٤) مراجعة، كُتبت من جانب قراء LibraryThing على (199) كتاباً في مجال علوم المكتبات والمعلومات. وتضمنت مجموعة البيانات العناصر التالية؛

Original language - text - date of publication - Author Name - (The title of the book reviews - rate)

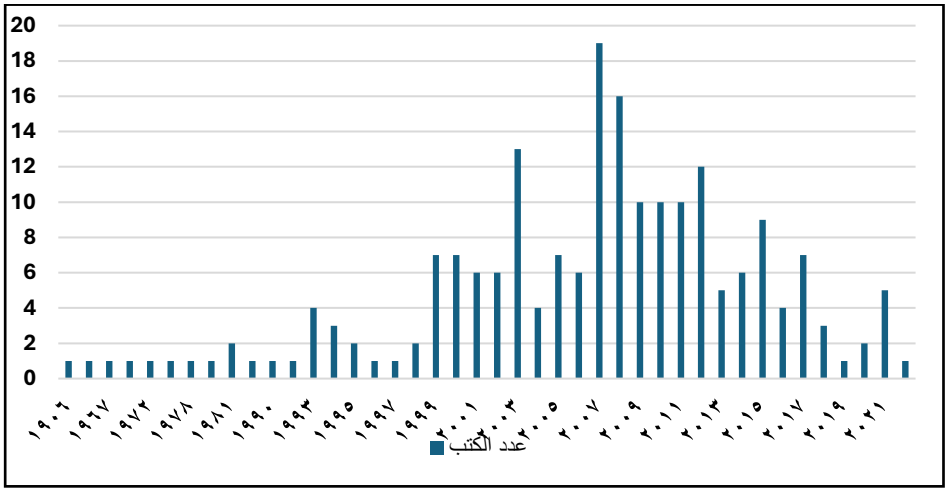
وقام الباحث بتنظيم البيانات وترتيبها في شكل ملف Excel. وإتاحتها من خلال موقع Kaggle [١] وذلك للإفادة منها في الأغراض التعليمية والبحثية من جانب الباحثين والمهتمين.

	A	B	C	D	E	F	G
1	title	Author	date	Original language	url	rate	textreviews
2	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	5	I just finished a course on Information Organization with one of the authors, Professor Joud
3	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	0	great information for graduate students
4	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	0	Comprehensive textbook, with just the right mix for information classification, including hi
5	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	2	Has a lot of technical data and sometimes requires reading and re-reading passages.
6	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	3	An interesting overview of the various methods used to organize information.
7	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	4	Clearly explains technical information and includes humorous sentences.
8	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	3	This is good book for gaining an entry level understanding of the organization of informatio
9	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	3	Heavy, tedious reading but it has everything I need in very clear language. The content is jus
10	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	3	Informative, but harder to read than the Chan, which covered much of the same material. Pe
11	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	0	Well organized text book. Revised updated second edition of this popular textbook, provide
12	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	2	It was difficult to get through but it did its job. There is a newer edition of this book availab
13	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	2	Similar to Taylor's Introduction to Cataloguing and Classification. This one takes a broader
14	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	3	This is a required textbook for a class. It is NOT an exciting read. Sorry. It's not helped by th
15	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	3	Good Text, clear explanations. A good reference.
16	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	2.5	A book I read for cataloging class. Textbooks have no reason to be so readable and easy to
17	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	3	While I certainly don't expect page-turning excitement in a book introducing the concepts o
18	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	4.5	not exactly light reading, but it does the job it set out to do. Easy to read than intro to class
19	The Organization of Information	Joudrey, Daniel N.& Arlene G. Taylor	2009	English	https://www.librarything.com/	0	In the process of reading and enjoying the larger font size.
20	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	Classic textbook.
21	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	This is a book that I will reach for again and again as I advance in my studies, and in my pr
22	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	nice book, very benefical, important key points
23	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	library data
24	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	4	For students
25	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	for information
26	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	for library science students
27	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	3	This is a nice introductory textbook. In the history chapter I learned about a librarian of hig
28	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	It covers the practice of librarianship,the place of librarianship in information structure,an
29	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	5	This was very helpful for an overview of the profession. It came in handy for the e-Porfolio s
30	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	here are no stars...what's to like (or not) about a standard text?This is a standard SLIS/MLIS
31	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	3	A bit dry, but what do you expect from a textbook! I found some of the material outdated, bu
32	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	3.5	First of all, yes, this is a textbook, so it already loses points for excitement and story-telling
33	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	3	This book gave an excellent overview of library and information science. It should provide a
34	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	2	hahah reading this ALL SEMESTER LONG I mean, it's a textbook. It was fine, but nothing extra
35	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	this book is poised to equipping students with the fundamentals of library and information
36	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	A good reading for the MLIS professional. Rubin was easy to follow in studying the issues fa
37	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	2	I had to read this for class. There were a few interesting chapters, but I had a hard time pay
38	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	0	A good overall introduction to the library sciences. It does not explain any one area in-dept
39	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	5	I was required to purchase this book for my first class in my LIS graduate studies. The book
40	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	3	boring, wordy and long. However, it does give you the basics on library history and practice
41	Foundations of Library and Information Science	Richard Rubin	1978	English	https://www.librarything.com/	3	Kind of wordy and unnecessarily redundant. Could've used some more editing. Very informa

شكل (٢) مقطع عرضي لمجموعة بيانات مراجعات كتب علوم المكتبات والمعلومات التي تم تجميعها من LibraryThing

١/١/١/٨ الخصائص العددية والزمنية واللغوية لكتب علوم المكتبات والمعلومات بموقع LibraryThing

يتراوح تاريخ نشر كتب علوم المكتبات والمعلومات التي تم مراجعتها من قبل قراء موقع LibraryThing من عام ١٩٠٦م، وهو تاريخ نشر أو طبعة من كتاب The Chicago Manual of Style، وحتى عام ٢٠٢٢م، تاريخ نشر اخر كتاب تم رصده من خلال مراجعات القراء في LibraryThing. وجاء عام ٢٠٠٧م في المرتبة الأولى برصيد (١٩) كتاب، يليه عام ٢٠٠٨ برصيد (١٦) كتاب من إجمالي كتب علوم المكتبات والمعلومات التي تم مراجعتها بـ LibraryThing.



شكل (٣) التوزيع الزمني لكتب علوم المكتبات والمعلومات التي تم مراجعتها من قبل القراء بموقع LibraryThing

وجاء كتاب The Library Book في المرتبة الأولى من حيث عدد مراجعات القراء برصيد (178) مراجعة. ويعرض الجدول التالي أعلى عشرة كتب تلقت أكبر عدد من مراجعات القراء.

جدول (١) أعلى عشرة كتب تلقت أعلى عدد من مراجعات القراء بموقع LibraryThing

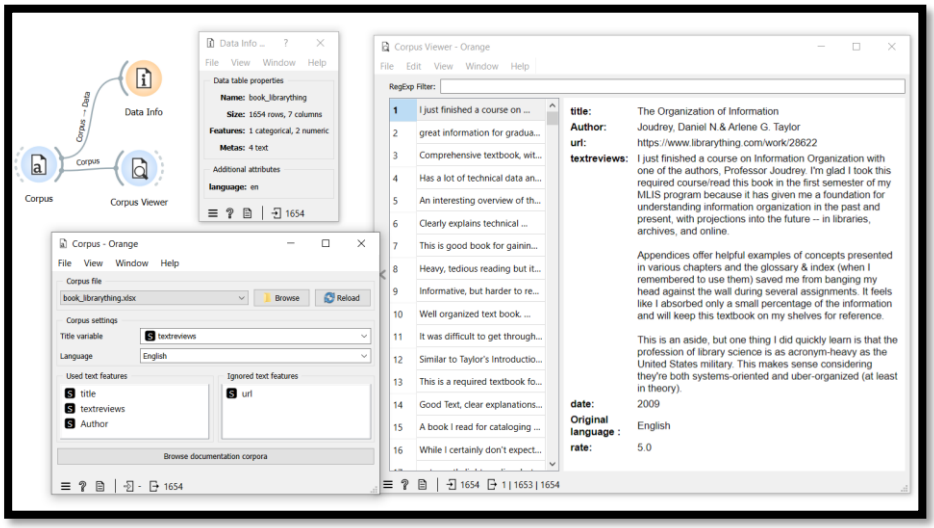
id	Book Title	Author Name	Number	% of 1654
1.	The Library Book	Susan Orlean	178	10.8%
2.	This Book is Overdue!: How Librarians and Cybrarians Can Save Us All	Marilyn Johnson & Huber, Hillary	110	6.7%
3.	The Information: A History, a Theory, a Flood	James Gleick	102	6.2%
4.	The World's Strongest Librarian: A Memoir of Tourette's, Faith, Strength, and the Power of Family	Josh Hanagarne	88	5.3%

5.	The Bad-Ass Librarians of Timbuktu: And Their Race to Save the World's Most Precious Manuscripts	Joshua Hammer	66	4.0%
6.	The Library at Night	Alberto Manguel	58	3.5%
7.	How to Read a Book, Revised and Updated Edition	Mortimer J. Adler & Charles Van Doren &, Charles Van Doran	57	3.4%
8.	Everything Is Miscellaneous: The Power of the New Digital Disorder	David Weinberger	54	3.3%
9.	A History of Reading	Alberto Manguel	47	2.8%
10.	Library: An Unquiet History	Matthew Battles	41	2.5%

وفيما يتعلق بالخصائص اللغوية، فنجد اللغة السائدة لكتب علوم المكتبات والمعلومات التي تم مراجعتها من جانب قراء موقع LibraryThing؛ هي الإنجليزية برصيد (١٦٥١) كتاب، في حين تم رصد ثلاث كتب بلغات مختلفة؛ الألمانية، والدنماركية، والسويدية.

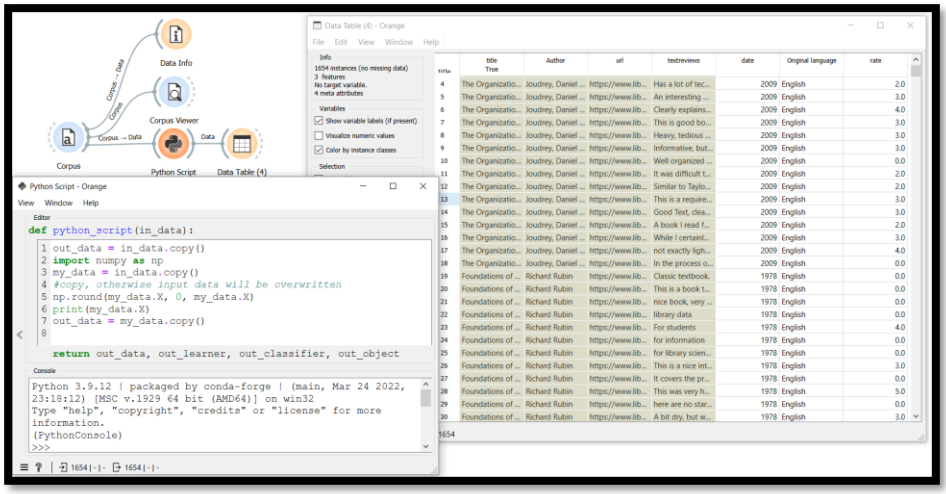
٢/١/٨ مرحلة تهيئة وقراءة البيانات من خلال Orange data mining

تم تحميل مجموعة البيانات التي قام الباحث بإعدادها بالامتداد (csv.)، لبرنامج Orange data mining، من خلال الأداة corpus، تمهيداً لإجراء معالجة البيانات وتحليلها، وتوضح الأداة Data Info جميع المعلومات التي تتعلق بمجموعة البيانات. ومن خلال الأداة Corpus Viewer يمكن عرض وفحص كل تسجيله داخل مجموعة البيانات. كما موضح بالشكل (٤).



شكل (٤) تحميل وعرض المعلومات الخاصة بمجموعة البيانات من خلال Orang data mining

ولضمان الدقة في إخراج البيانات وعرضها، تم كتابة كود بايثون من خلال الأداة python script، لعرض البيانات بشكل مناسب من خلال الأداة Data Table. كما موضح بالشكل التالي.



شكل (٥) عرض مجموعة البيانات باستخدام python script & Data Table

٣/١/٨ مرحلة المعالجة النصية للبيانات

تم إجراء عملية المعالجة لنصوص مراجعات القراء التي تم جمعها من خلال موقع LibraryThing، بالاعتماد على أداة Preprocess Text، وذلك بغرض تنظيف البيانات، من أجل الحصول على أفضل دقة أثناء تحليلها. وتضمنت عملية المعالجة الاستراتيجيات والأساليب الاتية:

١/٣/١/٨ التحويل Tokenization: يتم تحويل النص من الحروف الكبيرة إلى الحروف الصغيرة، وإزالة عناوين URL من النص، وإزالة جميع علامات التشكيل، واللكنات في النص.

٢/٣/١/٨ الترميز Tokenization: وفي هذه العملية يتم تقسيم النص إلى أجزاء أصغر لتسهيل معالجتها. وفق عدة أساليب منها؛ Word & Punctuation حيث يتم تقسيم النص إلى كلمات مع الاحتفاظ بعلامات الترقيم، كذلك وفقًا للمسافات البيضاء Whitespace، والجملة Sentence حيث يتم تقسيم النص وفقا للنقطة بعد نهاية كل جملة مكتملة (Irawan et al.,2019).

٣/٣/١/٨ عملية التطبيع Normalization: أحد أساليب معالجة اللغة الطبيعية، بهدف تحويل الكلمات والنصوص إلى شكل أكثر بساطة واتساقًا، حتي يسهل معالجتها وتحليلها. وتشمل هذه العملية بعض التقنيات مثل:

- **التقطيع (Stemming):** حيث يقوم على أساس إرجاع الكلمات إلى جذورها وذلك من خلال إزالة اللواحق (suffixes) والنهايات (endings) في الكلمات والجمل (Manasa & Prasad,2024).

ومن أشهر الخوارزميات المستخدمة في عملية التقطيع تشمل:

- **Porter Stemmer**: وهي تعد من أشهر خوارزميات معالجة اللغة الطبيعية (NLP)، تم تطويرها من قبل Martin Porter عام ١٩٧٩م (Porter,2006).
- **Snowball Stemmer:** تعد هذه الخوارزمية نسخة محسنة من Porter Stemmer، وتدعم العديد من اللغات.
- **WordNet Lemmatizer:** واحدة من أهم أدوات معالجة اللغة الطبيعية (NLP)، تستخدم قاعدة بيانات معجمية كبيرة للغة الإنجليزية. تتضمن الأسماء والأفعال والصفات والأحوال في مجموعات من المرادفات المعرفية بحيث تستخدم لفهم السياق والمعني الفعلي للكلمات (Miller,1995).
- **UDPipe، و Lemmagen:** عبارة عن أدوات متكاملة لمعالجة اللغة الطبيعية، وتعتمد على نماذج مدربة، لتطبيع البيانات اللغوية، وتدعم العديد من اللغات.

٤/٣/١/٨ التصفية Filtering

تعد عملية التصفية أخر مرحلة في عملية المعالجة النصية، حيث يتم بمقتضاها أي تصفية الكلمات والرموز التي لا حاجة لها في عملية (تحليل المشاعر). وتشمل الاتي:

- **كلمات التوقف Stopwords:**

تم إزالة كلمات التوقف من النص مثل ('and', 'or', 'in, et al...) كما تم تحميل ملف Txt يضم قائمة من كلمات التوقف في اللغة الإنجليزية. حتي يتم فلترة المراجعات التي تم جمعها من موقع LibraryThing بدقة عالية. ومن أمثلة كلمات التوقف التي تم استبعادها ما يلي:

I, me, my, myself, we, our, ours, ourselves, you, you're, you've, you'll, you'd, your, yours, yourself, yourselves, he, him, his, himself, she, she's, her, hers, herself, it, it's, its, itself, they, them, their, theirs, themselves, what, which, who, whom, dll.

- **Regexp** هي اختصار Regular Expression وتم تمكين هذا الخيار بغرض تنقية نصوص مراجعات القراء قيد الدراسة من الرموز وعلامات الترقيم التالية؛

\.|,|:|;|!|\?|\(|\)|||\+|'|"| _|'|—|‖|'|\'|...|\-|–|—|\$|&|*|#|@|%|_|>|<|\/|\[|\]|1|2|3|4|5|6| 7|8|9|10

٢/٨ تحليل المشاعر لمراجعات القراء:

يوفر Orange data mining العديد من خوارزميات تحليل النصوص، والتي تسمح بتحليل المشاعر للنصوص، وقام الباحث بتوظيف تلك الخوارزميات في تحليل مجموعة البيانات التي تم إعدادها من خلال مراجعات القراء التي تم جمعها من LibraryThing حول كتب علوم المكتبات والمعلومات.

١/٢/٨ خوارزمية Vader:

اعتمد الباحث على خوارزمية Vader بغرض تصنيف وتحليل المشاعر لآراء القراء حول كتب المكتبات والمعلومات من حيث كونها مشاعر؛ إيجابية، وسلبية، أو محايده.

وتعد Valence Aware Dictionary and Sentiment Reasoner (VADER) من أدوات تحليل المشاعر القائمة على القواعد، والتي تم تصميمها خصيصًا للمشاعر التي يتم التعبير عنها على مواقع الويب، ووسائل التواصل الاجتماعي، وتستند علي مجموعة غنية من المفردات (Deo et al.,2020). ويتم حساب المؤشر المركب **(Compound Score)** من خلال مجموع كل تقييمات القاموس، ثم تطبيعها بين -١ (سلبي جدًا) و +١ (إيجابي جدًا) من خلال المعادلة التالية (Molenaar et al.,2024):

$$\text{Compound} = \frac{x}{\sqrt{x2 + \alpha}}$$

حيث x = مجموع نقاط الكلمات.

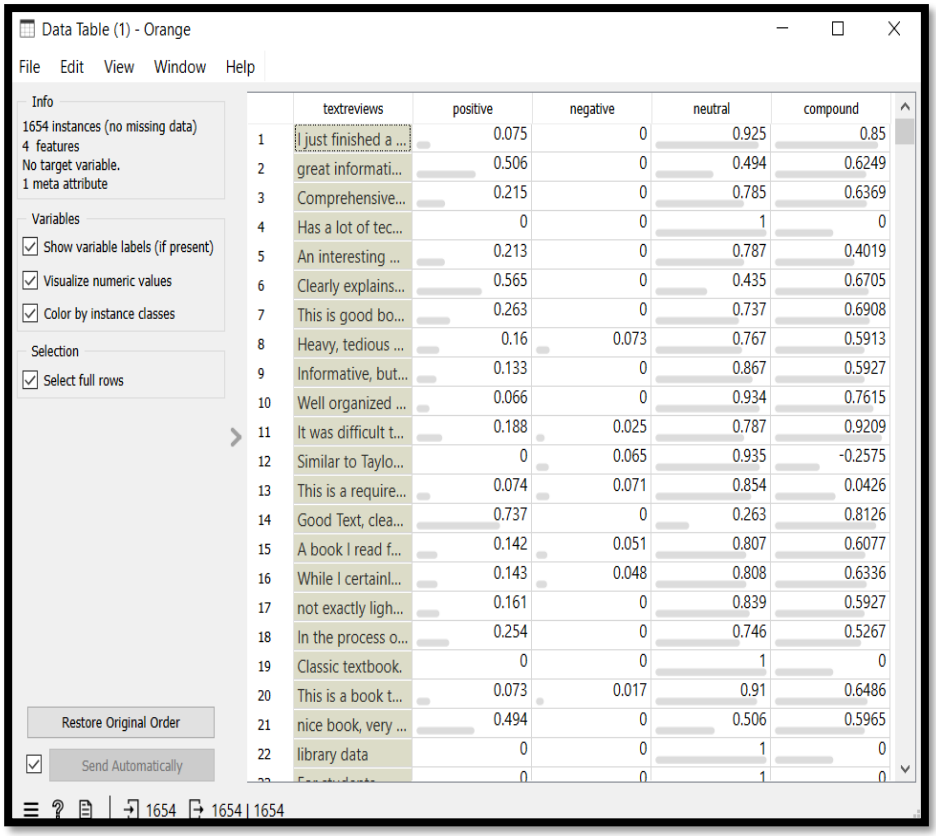
و α = معامل التطبيع، وقيمته ١٥.

ويعني ذلك أن المؤشر المركب Compound Score يعبر عن النتيجة النهائية لمستوي المشاعر في النص، فكلما اقتربت القيم من +١ تشير إلى مشاعر إيجابية جدا، في حين القيم القريبة من (-١) تشير إلى مشاعر سلبية جدا، أما القيم القريبة من (٠) تعبر عن مشاعر محايدة. وتم استخدام معامل التطبيع (α) بقيمة ثابتة (١٥) لتحقيق توازن مناسب في القياسات (Noviana,2021).

وبالرغم من توافق Orange data mining مع خوارزميات أخري مثل Liu Hu ، SentiArt، LiLaH sentiment، إلا أن Vader تعد الأنسب لمجموعة البيانات محل الدراسة. وذلك لعدة أسباب منها؛ أنه تم تصميمه خصيصًا لتحليل النصوص القصيرة مثل التغريدات والتعليقات على منصات ومواقع التواصل الاجتماعي، فضلًا عن تكامله مع مكتبة NLTK في بايثون، كما يتسم Vader بالدقة والسرعة في معالجة اللغة الطبيعية، كما أنه نموذج مفتوح المصدر ومن

ثم متاح للمطورين والمستخدمين لتحسينة ورفع كفاءته باستمرا. فضلًا عن إمكانية تحليل الرموز التعبيرية التي يتم استخدامها من قبل المستخدمين في كتابة تعليقاتهم وآرائهم.

وللحصول على نتائج تحليل المشاعر لمراجعات الكتب بأعلى دقة، قام الباحث بتحديد الجزء الخاص بـ Textreviews من مجموعة البيانات التي تم إنشاءها، حيث يشمل نصوص مراجعات وآراء القراء حول كتب علوم المكتبات والمعلومات في LibraryThing، باستخدام الأداة Select Columns، وتم تطبيق عمليات المعالجة النصية كما ذكرنا سابقاً، وبالاعتماد على الـ Widget الخاص بـ Sentiment Analysis، ثم إجراء تحليل المشاعر لمراجعات القراء وفق خوارزمية Vader والتي تعتمد علي مكتبة NLTK في بايثون، والتي توضح تصنيف مشاعر القراء من خلال تعليقاتهم على كتب المكتبات والمعلومات بـ Library Thing إلى إيجابية، وسالبية، ومحايدة، والتي تم عرضها بشكل مفصل من خلال الأدة Data Table كما موضح بالشكل التالي(٦).



	textreviews	positive	negative	neutral	compound
1	I just finished a ...	0.075	0	0.925	0.85
2	great informati...	0.506	0	0.494	0.6249
3	Comprehensive...	0.215	0	0.785	0.6369
4	Has a lot of tec...	0	0	1	0
5	An interesting ...	0.213	0	0.787	0.4019
6	Clearly explains...	0.565	0	0.435	0.6705
7	This is good bo...	0.263	0	0.737	0.6908
8	Heavy, tedious ...	0.16	0.073	0.767	0.5913
9	Informative, but...	0.133	0	0.867	0.5927
10	Well organized ...	0.066	0	0.934	0.7615
11	It was difficult t...	0.188	0.025	0.787	0.9209
12	Similar to Taylo...	0	0.065	0.935	-0.2575
13	This is a require...	0.074	0.071	0.854	0.0426
14	Good Text, clea...	0.737	0	0.263	0.8126
15	A book I read f...	0.142	0.051	0.807	0.6077
16	While I certainl...	0.143	0.048	0.808	0.6336
17	not exactly ligh...	0.161	0	0.839	0.5927
18	In the process o...	0.254	0	0.746	0.5267
19	Classic textbook.	0	0	1	0
20	This is a book t...	0.073	0.017	0.91	0.6486
21	nice book, very ...	0.494	0	0.506	0.5965
22	library data	0	0	1	0

شكل (٦) مقطع عرضي لنتائج تحليل المشاعر لمراجعات القراء حول كتب علوم المكتبات والمعلومات.

ومن خلال الأداة feature Statistics تم عرض الإحصاءات الخاصة بنتائج تحليل المشاعر من حيث عرض اسم الخاصية، والقيم الحسابية لكل من الدالة، والمنوال، والمتوسط، والوسط، والتشتت، وأصغر قيمة وأكبرها، ونسبة البيانات المفقودة. كما موضح بالشكل التالي.

شكل (٧) البيانات الإحصائية لنتائج تحليل المشاعر من خلال الأداة feature Statistics

٢/٢/٨ المراجعات الإيجابية للقراء حول كتب علوم المكتبات والمعلومات.

يوضح الباحث في هذا الجزء نتائج تحليل المشاعر لمراجعات القراء حول كتب علوم المكتبات والمعلومات بـ LibraryThing وفقًا للمشاعر الأكثر إيجابية. ويوضح الجدول (٢) أعلى عشرين مراجعة إيجابية.

جدول (٢) أعلى عشرين مراجعة إيجابية للقراء حول كتب علوم المكتبات والمعلومات بـ LibraryThing

id	code (2)	textreviews	positive	negative	neutral	compound
١.	377	Excellent!	1	0	0	0.6114
٢.	496	Breathtaking.	1	0	0	0.4588
٣.	541	Brilliant	1	0	0	0.5859
٤.	1061	Fun	1	0	0	0.5106
٥.	1107	Excellent	1	0	0	0.5719
٦.	630	Great resource.	0.804	0	0.196	0.6249
٧.	1197	great book	0.804	0	0.196	0.6249
٨.	391	Excellent book!	0.8	0	0.2	0.6114
٩.	1035	Excellent book.	0.787	0	0.213	0.5719
١٠.	370	good reader	0.744	0	0.256	0.4404

١١.	1653	Good starter	0.744	0	0.256	0.4404
١٢.	47	helpful. good practical examples	0.74	0	0.26	0.6908
١٣.	974	A true love letter to libraries. Great writing!	0.74	0	0.26	0.908
١٤.	14	Good Text, clear explanations.	0.737	0	0.263	0.8126
١٥.	574	Actually helpful.	0.737	0	0.263	0.4215
١٦.	589	Just wonderful and serendipitous. A joy!	0.722	0	0.278	0.8313
١٧.	1582	Great book easy to read!	0.709	0	0.291	0.807
١٨.	258	Excellent book. Lots of inspiration.	0.703	0	0.297	0.7964
١٩.	351	Cool book. Libraries are neat. Recommended.	0.703	0	0.297	0.7269
٢٠.	1605	Definite;y a niche interest. Still, fascinating.	0.684	0	0.316	0.7579

ومن خلال قراءة بيانات الجدول السابق يمكننا الخروج بالمؤشرات التالية:

- أن المراجعات العشرين الأعلى وفقاً للمشاعر الإيجابية، حصلت على درجة (٠) في الفئة negative، بالإضافة أنها سجلت درجات تقترب من (١) في المعامل compound، وذلك وفق المعادلة المذكورة سابقًا.
- ويتبين من خلال الجدول أنه يوجد خمس مراجعات فقط، حصلت على درجة (+١) أعلى درجة إيجابية. وجاءت الكلمات المعبرة عنها؛ (Fun، Brilliant، Breathtaking،Excellent) وبالرجوع لمجموعة البيانات محل الدراسة من خلال الكود المسجل به كل مراجعة، تبين أن تلك الآراء الإيجابية تخص ثلاث كتب يوضحها الجدول التالي:

جدول (٣) كتب علوم المكتبات والمعلومات التي سجلت أعلى درجة للمشاعر الإيجابية لدى القراء بموقع LibraryThing

code	book title	Text reviews	Rate	Result	
				Positive	compound
377	Information Representation and Retrieval in the Digital Age	Excellent	٥	+1	0.6114
496	The Information: A History, a Theory, a Flood	Breathtaking.	4	+1	0.4588
541	The Information: A History, a Theory, a Flood	Brilliant	٤	+1	0.5859
1061	The Library Book	Fun	3	+1	0.5106
1107	The Library Book	Excellent	5	+1	0.5719

٣/٢/٨ المراجعات السلبية للقراء حول كتب علوم المكتبات والمعلومات.

يوضح الجدول (٤) نتائج تحليل المشاعر لمراجعات القراء حول كتب علوم المكتبات والمعلومات بـ LibraryThing وفقًا للمشاعر الأكثر سلبية. وكما يتضح من الجدول أن مؤشر compound للمراجعات بالدرجة السالبة.

جدول (٤) أعلى عشرين مراجعة سلبية للقراء حول كتب علوم المكتبات والمعلومات بـ LibraryThing

id	code	text reviews	positive	negative	neutral	compound
1.	344	boring	0	1	0	-0.3182
2.	732	BORING	0	1	0	-0.3182
3.	1267	Boring.	0	1	0	-0.3182
4.	392	horrible index	0	0.778	0.222	-0.5423
5.	1372	not as thrilling as one would suspect	0	0.487	0.513	-0.5795
6.	273	categorization, classification, order, chaos, Web	0	0.481	0.519	-0.5719
7.	1453	Profoundly moving, often depressing.	0	0.464	0.536	-0.3818
8.	1178	Didn't like the book..	0	0.413	0.587	-0.2755
9.	1051	Sorry. Against the grain, I didn't enjoy this one. It's possibly Orlean's harsh, dry reading.	0	0.383	0.617	-0.703
10.	368	Duller than dirt, but useful.	0.41	0.378	0.213	0.3182
11.	321	history; sort of, of libraries and library science, many sad tales of books lost	0	0.31	0.69	-0.6597
12.	1002	nonfiction, history, Los Angeles, true crime, ebook, library	0.228	0.285	0.488	-0.1779
13.	1004	nonfiction, history, Los Angeles, true crime, ebook, library	0.228	0.285	0.488	-0.1779
14.	1363	Snobby but useful	0.575	0.276	0.149	0.4588
15.	916	ReadingIsResistance to ignoring the rest of the world's issues THE BAD-ASS LIBRARIANS OF TIMBUKTU yabscspl Ignorance imposed from above is a global problem. Resist with all your might! Read read read!	0	0.273	0.727	-0.8309

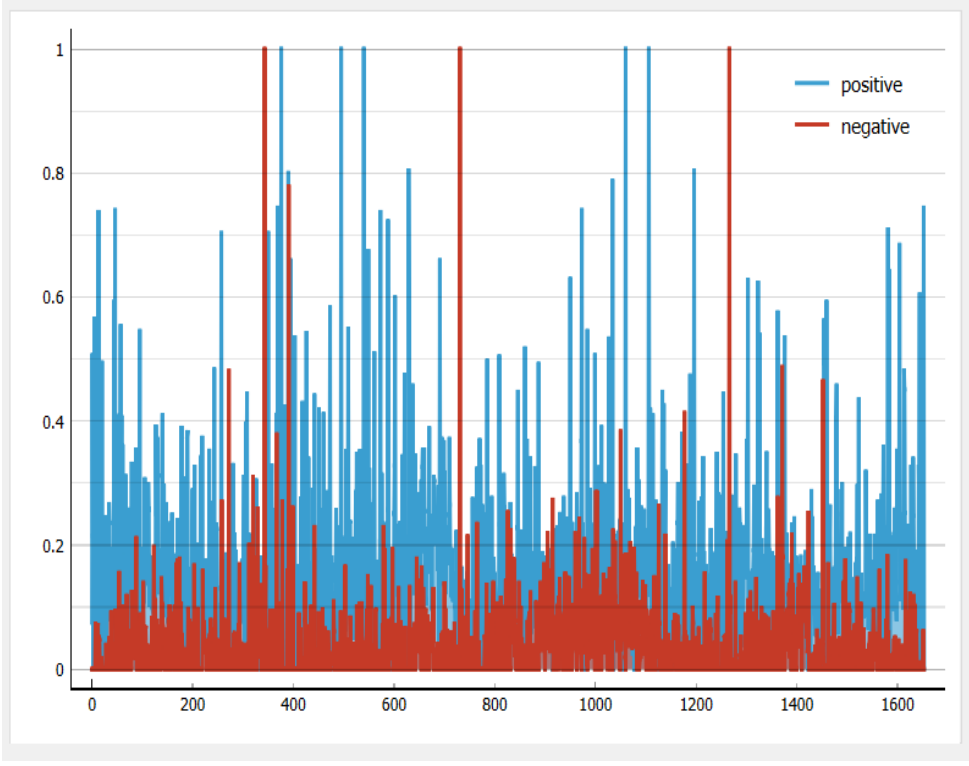
16.	379	This book is conceptually well-written, but suffers from a frequent, jarring misuse of English syntax and idiom, resulting in, at worst, confusion and, at best, an interruption in the flow of reading. I hold the editor/publisher responsible for this, not the author.	0.149	0.27	0.58	-0.7964
17.	259	Kind of mind blowing. I loved when he attacked the alphabet, but don't agree with him that Oscar's messy tags won't necessarily affect Felix.	0.083	0.27	0.647	-0.6668
18.	1127	The Los Angeles Public Library suffered a horrible fire on the same day as the Chernobyl disaster, which is one of the reasons that I had never heard about this event. More than 1,000,000 volumes were damaged or destroyed. The author speculates about the cause of the fire, discusses the history of this library and the ways in which libraries have evolved over the centuries. Interesting and informative.	0.033	0.263	0.704	-0.9584
19.	400	Great book on a difficult subject	0.427	0.26	0.312	0.3818
20.	330	Got bored with it. Never finished it.	0	0.259	0.741	-0.2732

ويتبين من خلال الجدول (٥) أنه يوجد ثلاثة آراء حصلت على أعلى درجة سلبية (1-)، حيث تضمنت تلك المراجعات كلمة boring بالشكل التالي؛ (boring – BORING - Boring)، حيث تعد وفقاً لمعجم التصنيف في خوارزمية Vader من أشد الكلمات التي تعكس المشاعر السلبية. وبالرجوع لمجموعة البيانات تبين أن هذه المراجعات السلبية الثلاث تخص ثلاثة كتب مختلفة يبينها الجدول التالي:

جدول (٥) كتب علوم المكتبات والمعلومات التي سجلت أعلى درجة للمشاعر السلبية لدى القراء بموقع LibraryThing

Id	book title	Textreviews	Rate	Result	
				Negative	compound
377	Library: An Unquiet History	boring	١	-1	- 0,3182
541	The Social Life of Information	BORING	٢	-1	- 0,3182
1107	Indexing and Abstracting in Theory and Practice	Boring.	١	-1	- 0,3182

ويوضح الشكل التالي نسبة المشاعر الإيجابية إلى المشاعر السلبية حيث يعكس اللون الأزرق مراجعات وتعليقات القراء الايجابية في LibraryThing حول كتب علوم المكتبات والمعلومات. في حين يعكس اللون الأحمر مراجعات القراء السلبية والتي تضح من خلال الشكل أنها أقل من المشاعر الإيجابية.



شكل (٨) نسبة مراجعات القراء الإيجابية، والسلبية حول كتب المكتبات والمعلومات في LibraryThing.

٣/٢/٨ المراجعات المحايدة للقراء حول كتب علوم المكتبات والمعلومات.

وهناك بعض المراجعات والآراء للقراء جاءت محايدة كما يوضح جدول (٦) أعلى (٢٠) مراجعة للقراء محايدة، حيث حصلت على درجة (١) في الفئة neutral، كما نلاحظ أن هذه الآراء لا تنتمي للمشاعر الايجابية أو السلبية، لذا نلاحظ الدرجة (٠)، وبالتالي أصبح المؤشر المركب للنص (٠) أيضًا. وتتمثل تلك الآراء والمراجعات في الكلمات والعبارات الآتية؛ (Classic textbook. library data، For students، for information، for library science students) وبالفعل لا يمكن حسم مشاعر تلك التعليقات كونها إيجابية، أو سلبية. لذا تم تصنيفها وفق معجم Vader لمشاعر محايدة.

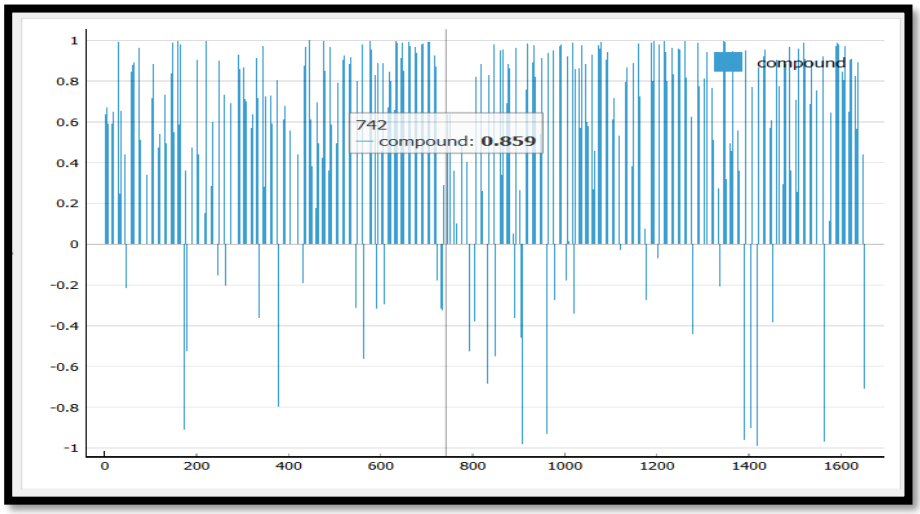
جدول (٦) أعلى عشرين مراجعة محايدة للقراء حول كتب علوم المكتبات والمعلومات بـ LibraryThing

	CODE	text reviews	positive	negative	neutral	compound
1.	4	Has a lot of technical data and sometimes requires reading and re-reading passages.	0	0	1	0
2.	19	Classic textbook.	0	0	1	0
3.	22	library data	0	0	1	0
4.	23	For students	0	0	1	0
5.	24	for information	0	0	1	0
6.	25	for library science students	0	0	1	0
7.	27	It covers the practice of librarianship,the place of librarianship in information structure,and more.	0	0	1	0
8.	30	A bit dry, but what do you expect from a textbook! I found some of the material outdated, but it provided a decent foundation and prompted a lot of class discussion.	0	0	1	0
9.	40	Kind of wordy and unnecessarily redundant. Could've used some more editing. Very informative though.	0	0	1	0
10	41	I remembered that I have borrowed this book from CU library and I had read the whole book.	0	0	1	0
11	51	A handy guide for cataloging.	0	0	1	0

12	54	I don't know - we'll see at the end of the semester exam.	0	0	1	0
13	90	A must-read for every librarian!	0	0	1	0
14	164	For librarians and their fans, a must read!	0	0	1	0
15	168	"I was under the librarians' protection. Civil servants and servants of civility, they had my back." pg.252	0	0	1	0
16	170	for course related to learn something new about the field	0	0	1	0
17	173	Standards management concepts applied to libraries and information centers	0	0	1	0
18	187	I cannot say enough about how well-written and informative Mann's text is. I don't know how I got through high school, much less college and a required senior thesis, without this book (except that it hadn't been published yet).	0	0	1	0
19	188	Written more as a guide for students needing to learn how to use a library rather than for librarians. However, it needs to be updated to be of use.	0	0	1	0
20	189	Practical book gives an overview of how to do research in a library. It can make a reference librarian out of anyone.	0	0	1	0

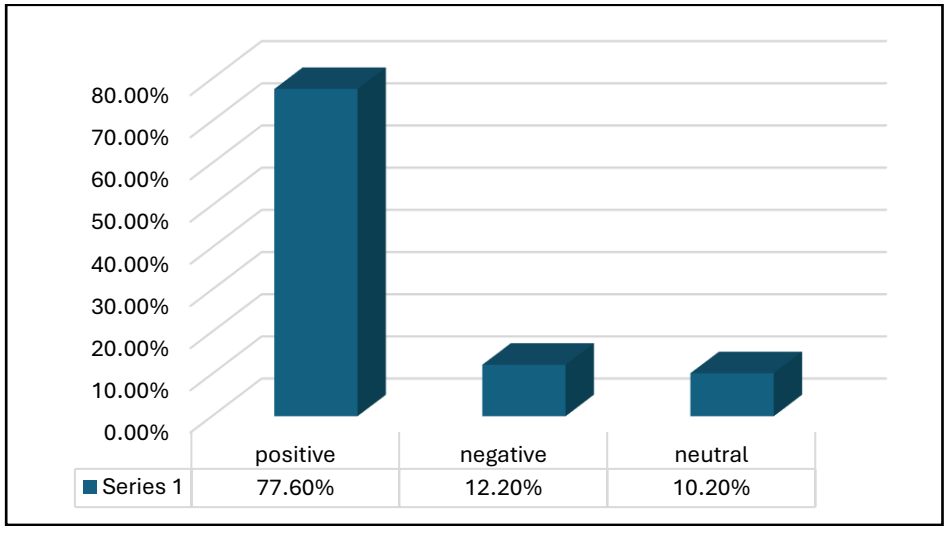
٤/٢/٨ نتائج تحليل المشاعر لآراء ومراجعات القراء وفقًا لـ Compound Score

تم إجراء ترتيب نتائج تحليل المشاعر لآراء ومراجعات القراء حول كتب علوم المكتبات والمعلومات وفقًا للمؤشر المركب Compound Score والذي يعبر عن النتيجة النهائية لمستوي المشاعر السائدة في النص وفقاً لطريقة الحساب السابق ذكرها.



شكل (٩) نتائج تحليل المشاعر لآراء ومراجعات القراء حول كتب علوم المكتبات والمعلومات وفقًا للمؤشر المركب Compound Score

ويتضح من خلال الشكل (١٠) أن المشاعر الإيجابية هي السائدة في مراجعات القراء للكتب، حيث تتراوح غالبيتها ما بين (٠ ، +١)، والتي بلغت (١٢٨٣) مراجعة إيجابية بما يمثل (٧٧,٦٪)، في حين بلغ عدد المراجعات السلبية والتي تتراوح ما بين (٠ ، -١) نحو (٢٠٢) مراجعة سلبية بنسبة (١٢٪)، وبلغ عدد المراجعات التي تم تصنيفها بأنها محايدة (١٦٩) مراجعة محايدة بنسبة (١٠٪) من إجمالي مراجعات القراء.

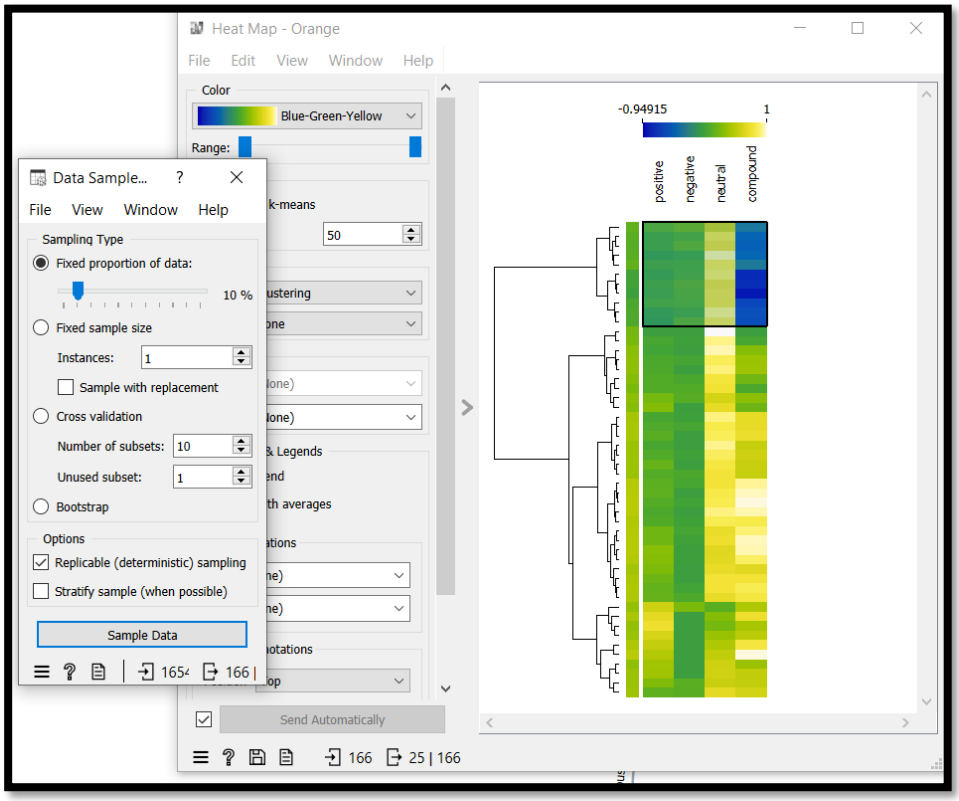


شكل (١٠) نسب تصنيف المشاعر لمراجعات القراء لكتب علوم المكتبات والمعلومات

٥/٢/٨ عرض نتائج تحليل المشاعر باستخدام الخرائط الحرارية Heatmap:

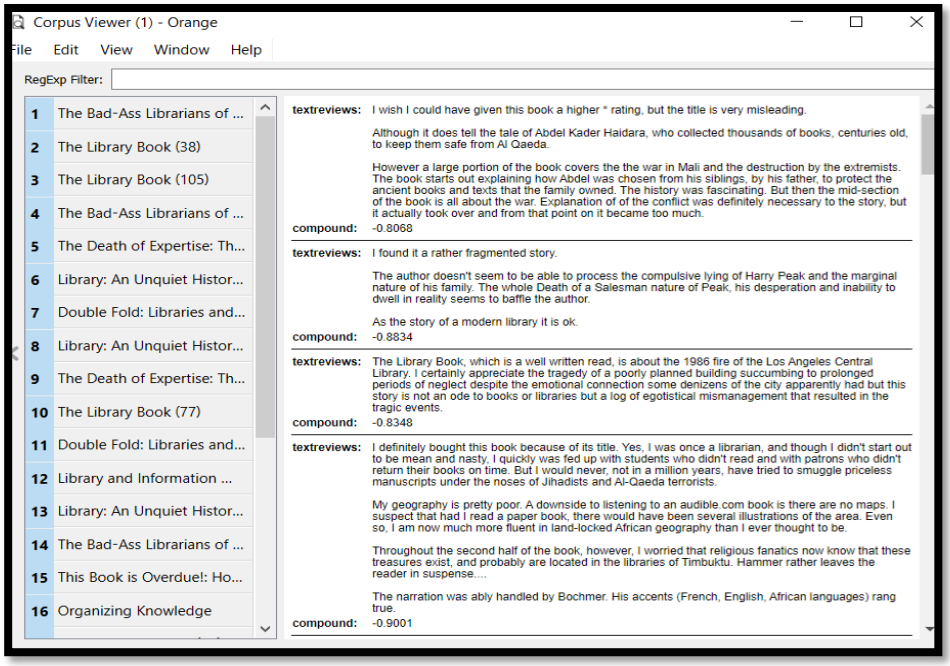
تُعد Heatmap أداة فعالة في تمثيل البيانات التي تتضمن متغيرات رقمية بالاعتماد على الألوان، حيث تظهر القيم في مصفوفة بألوان مختلفة. ويعتمد لون كل نقطة على قيمة البيانات التي تمثلها. وتستخدم مقياس لتدرج اللون من الألوان الباردة إلى الألوان الساخنة كالأحمر (Guo et al,2020).

قام الباحث بتحديد عينة (١٠٪) من إجمالي مراجعات الكتب من خلال الأداة Data Sampler، لتمثيلها مرئياً باستخدام خرائط Heatmap كما موضح بالشكل (١١)



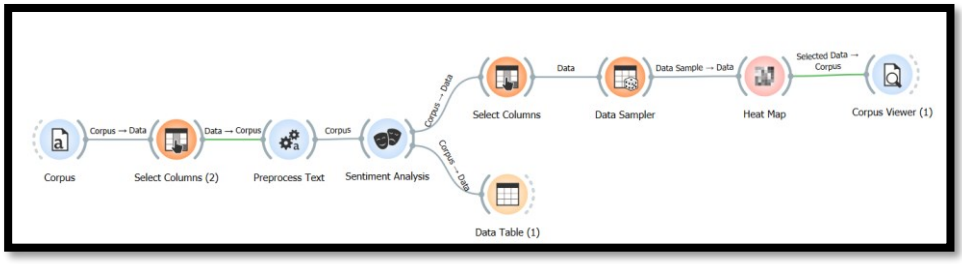
يتبين من خلال الشكل (١١) أن مخرجات تحليل المشاعر تتمثل في أربعة أعمدة، وهي درجة المشاعر الإيجابية والسلبية والمحايدة، ودرجة مركبة تجمع الدرجات السابقة في رقم واحد، وتعبر عن المشاعر السائدة. وتمثل القيم المركبة الموجبة (اللون الابيض) مراجعات القراء الإيجابية لكتب علوم المكتبات والمعلومات، بينما تمثل القيم السالبة (اللون الأزرق) مراجعات القراء السلبية لكتب علوم المكتبات والمعلومات. بينما تمثل القيم المحايدة (اللون الأخضر) مراجعات القراء المحايدة حول كتب علوم المكتبات والمعلومات.

وعند تحديد الجزء الخاص بالقيم المركبة الممثلة باللون الأزرق على سبيل المثال، يمكن عرض مراجعات وآراء القراء السلبية لكتب علوم المكتبات والمعلومات بموقع LibraryThing. من خلال الأداة Corpus Viewer، ومن ثم نلاحظ قيمة Compound Score السالبة، كما بالشكل التالي.



شكل (١٢) عرض مراجعات وآراء القراء السلبية لكتب علوم المكتبات والمعلومات من خلال الأداة Corpus Viewer،

ويوضح الشكل (١٣) سير عمل Workflow الذي قام بإعداده الباحث من خلال برنامج Orange Data Mining، باستخدام الأدوات والأساليب السابقة في تحليل المشاعر لمراجعات وآراء القراء لكتب علوم المكتبات والمعلومات بموقع LibraryThing.



شكل (١٣) سير عمل Workflow المستخدم في تحليل المشاعر لمراجعات القراء لكتب علوم المكتبات والمعلومات.

(من إعداد الباحث باستخدام Orange Data Mining)

٣/٨ تحليل مراجعات القراء وفقا لمقياس الحالة المزاجية Profile of Mood States

Profile of Mood States

يتيح Orange Data Mining أداة Tweet Profiler، والتي تعتمد على مقياس الحالة المزاجية Profile of Mood States (POMS) الذي يعد بمثابة أداة تقييم نفسية تم تطويرها كل من (McNair, Droppleman, and Lorr) في عام ١٩٧١م. بغرض تحليل المشاعر والمزاج لدى الفرد من خلال مجموعة من الصفات التي تعكس شعور الفرد بكل عاطفة. (Duong,2023).

جدير بالذكر أن Tweet Profiler تدعم ثلاث نماذج مختلفة من تحليل العاطفة. وهي؛ نموذج بول إكمان للعواطف الأساسية إكمان (Ekman)، ونموذج روبرت بلوتشيك (Plutchik)، ونمو ج (POMS) لتقييم العواطف بشكل عام.

وتتم عملية تحليل العواطف والشعور بناء على نمطين (Al Assyam & Hasan,2023)؛

- تصنيف متعدد الفئات (Multi-class classification) حيث يسجل شعورًا واحدًا محتملاً لكل تغريدة أو مستند
- تصنيف متعدد العلامات (Multi-tag classification) يسجل قيما متعددة بناء على العديد من المشاعر المختلفة مع إعداد عمود يوضح نسبة كل شعور.

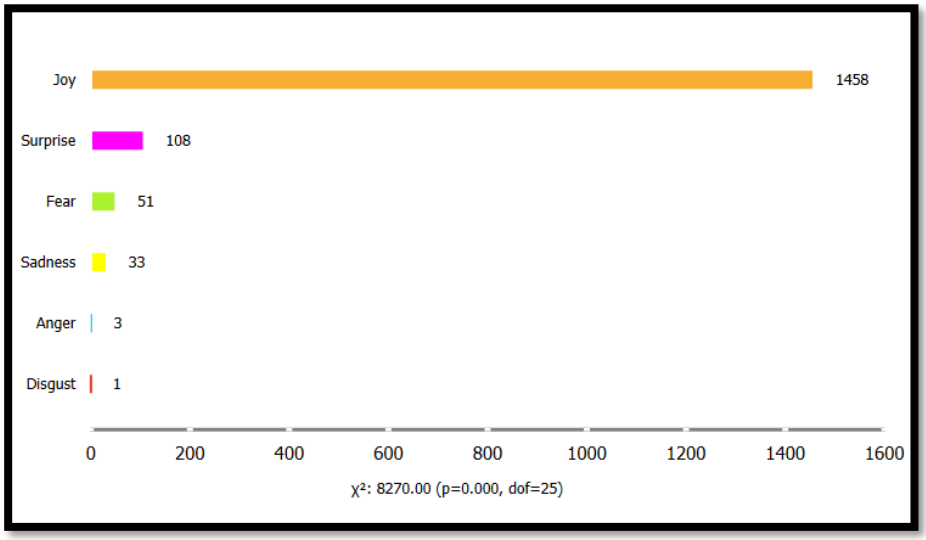
١/٣/٨ تطبيق تحليل العاطفة لمراجعات القراء لكتب علوم المكتبات والمعلومات

قام الباحث بإعداد Workflow من خلال برنامج Orange Data Mining يتضمن الأدوات المناسبة لإجراء تحليل العاطفة والشعور لمراجعات القراء حول كتب المكتبات والمعلومات بموقع LibraryThing وذلك من خلال أداة Tweet Profiler باستخدام نموذج كل من؛ Ekman، Plutchikوبالاعتماد على نمط (Multi-class) بالإضافة لأدوات تحليل البيانات Box Plot، Distributions.

١/١/٣/٨ تحليل العاطفة لمراجعات القراء وفقًا لنموذج Ekman's

وضع بول إيكمان هذا النموذج بهدف تحديد المشاعر والعواطف الأساسية للفرد، ويشمل النموذج ست فئات منفصلة لا تختلط مع بعضها البعض وهي؛ الفرح Joy، الدهشة Surprise، الخوف Fear، الحزن Sadnes، الاشمئزاز Disgust، الغضب Anger. وهذه العواطف يتم تصنيفها إما إيجابية أو سلبية (Sajwan et al., 2023).

ويبين الشكل التالي نتائج تحليل مراجعات القراء حول كتب علوم المكتبات والمعلومات في LibraryThing، وفقًا لخوارزمية Ekman's والذي يتبين من خلاله أن غالبية المراجعات تم تصنيفها في الفئة Joy برصيد (١٤٥٨) مما يدل أن العاطفة السائدة في مشاعر القراء المتعة والفرح. في حين بلغ عدد المراجعات التي تم تصنيفها في الفئة الحزن (٣٣)، وفي فئة الغضب (٣)، وفي فئة الاشمئزاز (١) من إجمالي المراجعات في مجموعة البيانات محل الدراسة.



شكل (١٤) نتائج تحليل العاطفة لمراجعات القراء لكتب علوم المكتبات والمعلومات في LibraryThing، وفقًا لخوارزمية Ekman's

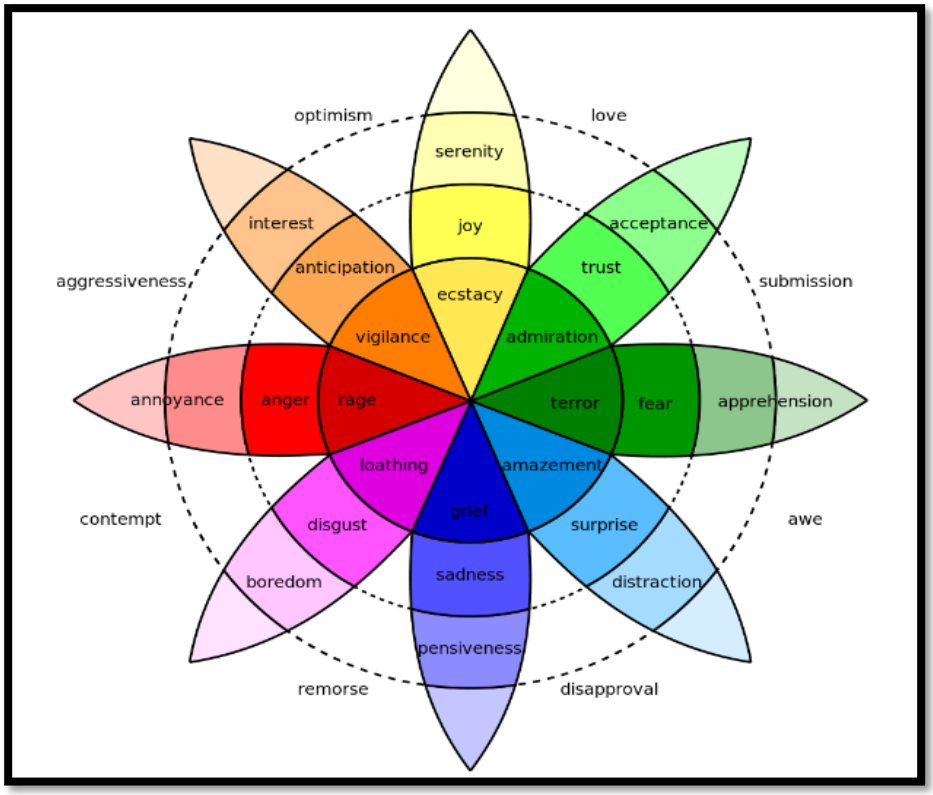
٢/١/٣/٨ تحليل العاطفة لمراجعات القراء وفقًا لنموذج Plutchik's

تعتمد نظرية Plutchik النفسية التطورية للعواطف على مجموعة من ثمانية عواطف ثنائية القطب، بدلاً من الفئات الست للعواطف التي حددها Ekman's وهذه العواطف الثمانية منظمة في أربع مجموعات ثنائية القطب: الفرح مقابل الحزن، الغضب مقابل الخوف، الثقة مقابل الاشمئزاز، المفاجأة مقابل الترقب (Suttles et al.,2013).

حيث شمل نموذج Plutchik الفئات الست التي أدرجها Ekman بالإضافة إلى فئتين آخرتين؛ الثقة Trust والترقب Anticipation. ويطلق عليه نموذج Wheel of Emotions (عجلة العواطف).

وتقوم فكرة عجلة العواطف في نموذج Plutchik على أساس كونها رسم دائري يضم ثماني فئات تمثل العواطف الأساسية في نموذج Plutchik، وتم ترتيبها بحيث تكون كل عاطفة تقابل عكسها في الاتجاه المقابل من الدائرة. علي سبيل المثال (الفرح مقابل الحزن)، (الثقة مقابل الاشمئزاز)

كما يمكن لهذه العواطف الاساسية أن تتحد مع بعضها البعض لتعبر عن مشاعر أكثر تعقيدًا وتفاوتًا في شِدَّتهَا. على سبيل المثال: الفرح + الثقة = الحب، والخوف + المفاجأة = الدهشة إلخ، ويعتمد نموذج Plutchik في التعبير عن شدة العاطفة ودرجتها بتطبيق آلية اللون والترتيب. بحيث تصبح العواطف أكثر شدة كلما اقتربت من مركز العجلة (Abbasi & Beltiukov,2019).

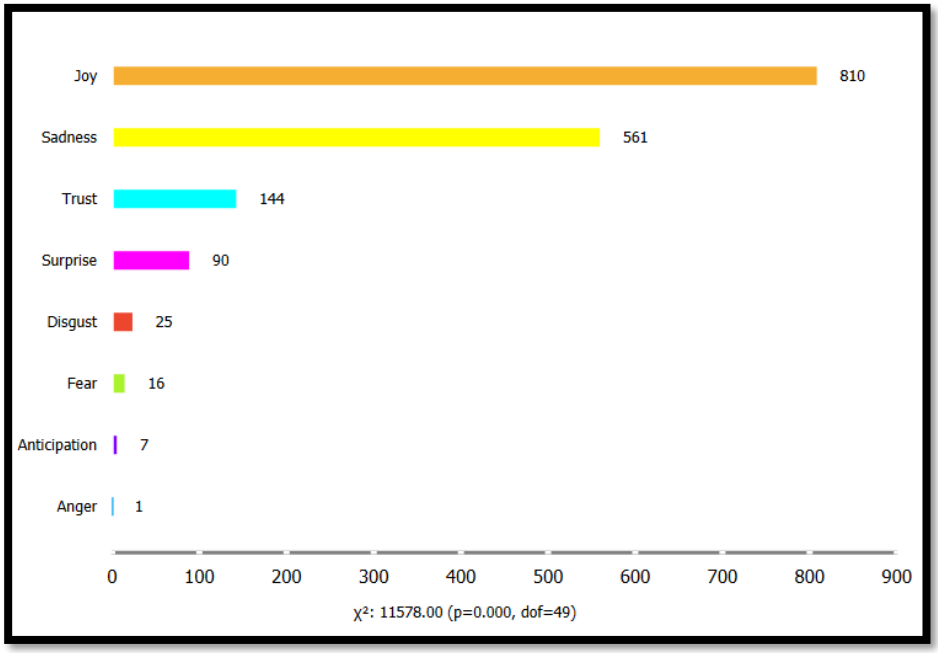


شكل (١٥) عجلة العواطف (Wheel of Emotions) في نموذج Plutchik's

المصدر: (Marcu & Danubianu,2020,p.205)

ويبين الشكل التالي نتائج تحليل مراجعات القراء حول كتب علوم المكتبات والمعلومات في LibraryThing، وفقًا لنموذج Plutchik's والذي يتبين من خلاله أن غالبية المراجعات تم تصنيفها في الفئة Joy أيضا كما الحال في نموذج Ekman's ولكن هنا بنسبة أقل، وتحديداً (٨١٠) من إجمالي مراجعات القراء في مجموعة البيانات. مما يدل أن العاطفة السائدة في مشاعر القراء المتعة والفرح. تليها في المرتبة الثانية عاطفة الحزن Sadness والتي استحوذت على (٥٦١) من مراجعات القراء. تليها في المرتبة الثالثة عاطفة الثقة Trust، برصيد (١٤٤) وفي فئة

الترقب Anticipation (٧) مراجعات فقط، وفي فئة الغضب تم تصنيف تعليق واحد فقط من إجمالي المراجعات في مجموعة البيانات محل الدراسة.



شكل (١٦) نتائج تحليل العاطفة لمراجعات القراء لكتب علوم المكتبات والمعلومات في LibraryThing، وفقًا لخوارزمية Plutchik's

٣/١/٣/٨ مقارنة نتائج تحليل العواطف بين نموذج Ekman، Plutchik

قام الباحث بإجراء تقييم لنتائج تحليل العواطف في كل من نموذج Ekman، Plutchik وذلك من خلال مراجعة عينة من تعليقات ومراجعات القراء قيد الدراسة، ومن ثم تقييم نتائج التحليل بين النموذجين، حتي يتضح أي النموذجين أدق في تصنيف العاطفة.

وذلك من خلال تحديد كل فئة على حدة من فئات تحليل العواطف من خلال الأداة Box Plot، ثم استعراض تلك المراجعات والتعليقات التي تمثل العاطفة من خلال الأداة Data Tabel.

ويوضح الجدول (٧) عينة من التعليقات والمراجعات وتصنيفها وفقًا لكلا النموذجين، وكذلك التقييم الأدق من وجهة نظر الباحث.

جدول (٧) مقارنة نتائج تحليل العواطف بين نموذج Ekman، Plutchik

م	code	Text Reviews	compound	Ekman's	Plutchik's	الأكثر دقة
1.	1648	A good introductory book on web search	0.4404	Surprise	Trust	Plutchik's
2.	51	A handy guide for cataloging.	0	Joy	Trust	Plutchik's
3.	409	Boolean	0	Sadness	Trust	Plutchik's
4.	324	Boring. Not the book I expected it to be. Could not get through the first chapter.	-0.3182	Surprise	Surprise	=
5.	541	Brilliant	0.5859	Joy	Trust	Ekman's
6.	386	Duller than dirt, but useful.	0.3182	Sadness	Trust	Plutchik's
7.	391	Excellent book!	0.6114	Joy	Trust	Plutchik's
8.	786	Fascinating little book. I wish it were longer/more detailed.	0.7115	Sadness	Anticipation	Plutchik's
9.	1538	Good resource to help librarians understand mashups and to use it in their libraries	0.6808	Fear	Joy	Plutchik's
10	89	Good, kind of inspiring, kind of kooky, kind of sad	0.3197	Fear	Fear	=
11	330	Got bored with it. Never finished it.	-0.2732	Joy	Sadness	Plutchik's
12	631	Great for ref services!	0.6588	Joy	Trust	Plutchik's
13	686	How could I not read this---it's about how libraries saved this guys life!	0.4753	Surprise	Disgust	Ekman's
14	474	i liked this book.	0.4215	Surprise	Trust	Plutchik's
15	96	I really enjoyed this book.	0.5563	Surprise	Joy	Plutchik's
16	831	Indispensable	0	Fear	Joy	Plutchik's
17	631	Loved it enough to read a second time!	0.6588	Joy	Trust	Ekman's
18	1612	Loved the pictures and the history of not just the library of congress but all card catalogs and in many ways libraries themselves.	0.3506	Sadness	Joy	Plutchik's
19	84	Made me want to become a librarian!	0.1511	Surprise	Surprise	=

20	173	Standards management concepts applied to libraries and information centers	0	Anger	Trust	Plutchik's
21	1456	Super subject for book focus...library	0.5994	Fear	Surprise	Plutchik's
22	1198	This is great....I want a copy of my own!	0.1511	Sadness	Joy	Plutchik's
23	1524	Too long, but has useful advice.	0.5927	Sadness	Trust	Plutchik's
24	191	Very helpful for basic database design.	0.4754	Joy	Joy	=
25	385	Very technical book, but easy to understand	0.5927	Fear	Trust	Plutchik's

اتضح من خلال الجدول السابق أن غالبية مراجعات القراء التي تم تصنيفها من قبل نموذج Plutchik في كل من فئة الثقة Trust والترقب Anticipation. تم التعبير عنها في نموذج Ekman كمشاعر تصف الفرح غالبًا. وتتفق هذه النتيجة مع دراسة (Marcu & Danubianu,2020). ومنها على سبيل المثال لا الحصر؛ المراجعات التي تحمل كود (٥١)، (٥٤١)، (٣٩١)، (٦٣١) تم تصنيفها في نموذج Plutchik تحت فئة Trust، في حين تم تصنيفها من قبل نموذج Ekman في الفئة Joy. وفي رأي الباحث أن التصنيف الأدق للمراجعات (٥١)، (٣٩١)، هو تصنيف نموذج Plutchik حيث تدل في مفهومها على مشاعر الثقة بالفعل. في حين يعد التصنيف الأقرب للمراجعات (٥٤١)، (٦٣١) هو تصنيف نموذج Ekman حيث تعبر عن مشاعر الفرح أكثر من مشاعر الثقة.

جدير بالذكر أنه تم رصد بعض المشاعر التي تم تصنيفها بشكل متناقض من قبل النموذجين ومنها على سبيل المثال؛ المراجعات كود (٣٨٦، ٤٠٩) حيث تم تصنيفها في نموذج Ekman بأنها تدل على مشاعر الحزن Sadness، في حين تم تصنيفها في نموذج Plutchik بأنها تدل على الثقة Trust، ويعد الأخير هو الأقرب من وجهة نظر الباحث.

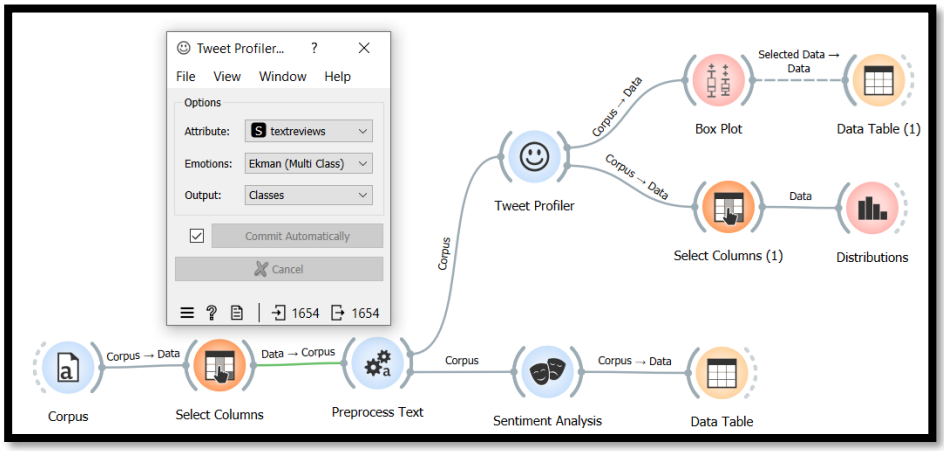
أيضًا المراجعة (٨٣١) والتي تحمل نص (Indispensable) تم تصنيفها من قبل نموذج Ekman بأنها تدل على مشاعر الخوف Sadness، في حين تم تصنيفها في نموذج Plutchik بأنها تدل على الفرح Joy، وبمراجعة سياق هذه المراجعة تبين أنها كُتبت من قبّل أحد القراء كتقييم لمحتوى كتاب The Chicago Manual of Style وهو دليل إرشادي لكتابة الاستشهادات بنمط Chicago، ومن ثم فلا غني عنه بالفعل سواء لأي مكتبة أو لأي باحث، مما يعكس أهمية اقتناء الكتاب. وبالتالي فيعد تصنيف هذا التعليق من قبل نموذج Plutchik هو الأكثر دقة.

كما نلاحظ التناقض أيضا في تصنيف المراجعة (١١٩٨) حيث تم تصنيفها من قبل نموذج Ekman بأنها تدل على مشاعر الخوف Sadness، في حين تم تصنيفها في نموذج Plutchik بأنها تدل على الفرح Joy، ويعد التصنيف وفق نموذج Plutchik هو الأكثر دقة.

وهناك الكثير من أوجه الاتفاق في درجات تصنيف مراجعات القراء من قبل كل من نموذج Ekman، Plutchik، كما المثال في المراجعات (٨٤)، (٨٩)، (١٩١)، (٣٢٤)، (٣٨٥)، (٦٨٦).

ومن خلال العرض السابق يتبين أن نموذج Plutchik يُعد هو الأكثر دقة في تصنيف مراجعات وتعليقات القراء وفقا للمشاعر والعواطف التي تعبر عنها، وربما يعود ذلك لزيادة فئات التصنيف في النموذج مقارنة بنموذج Ekman، من خلال إضافة فئات مختلفة؛ الثقة Trust والترقب Anticipation، مما أتاح إمكانية تصنيف بعض الآراء بشكل أكثر دقة.

ويوضح الشكل (١٧) سير عمل Workflow الذي قام بإعداده الباحث من خلال برنامج Orange Data Mining، لإجراء تحليل آراء ومراجعات القراء لكتب علوم المكتبات والمعلومات بموقع LibraryThing، وفقا لمقياس الحالة المزاجية (POMS)، باستخدام نموذج Ekman، Plutchik.



شكل (١٧) سير عمل Workflow المستخدم في تحليل مراجعات القراء وفقا لمقياس الحالة المزاجية (POMS)

(من إعداد الباحث باستخدام Orange Data Mining)

٤/8 نمذجة الموضوعات لنصوص مراجعات القراء لكتب علوم المكتبات والمعلومات:

نمذجة الموضوعات (Topic Modelling) هي تقنية من تقنيات التعلم الآلي (Machine Learning) ومعالجة اللغة الطبيعية (NLP)، والتي يتم الاعتماد عليها في اكتشاف الموضوعات المخفية داخل مجموعة من النصوص أو مجموعة مستندات نصية، وذلك بغية تحديد الكلمات الرئيسية التي تشملها نصوص معينة (Tong & Zhang,2016).

ويعرض الباحث في هذا الجزء لنمذجة الموضوعات، واكتشاف الموضوعات الرئيسية التي تتضمنها مراجعات القراء حول كتب علوم المكتبات والمعلومات بـ LibraryThing.

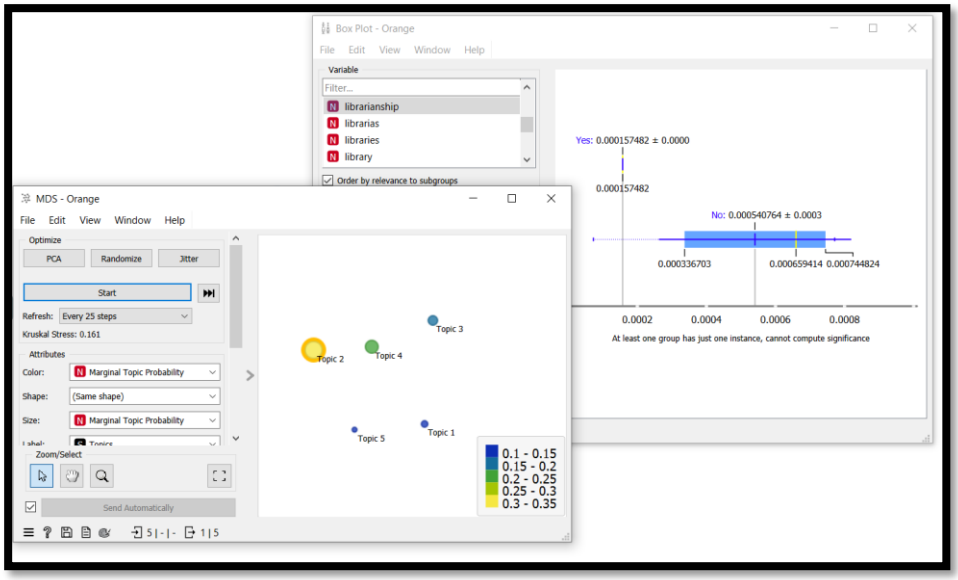
جدير بالذكر أن برنامج Orange Data Mining يوفر ثلاث خوارزميات مختلفة لإجراء نمذجة الموضوعات من خلال الأداة Topic Modelling؛ (LSI) Latent Semantic Indexing، Latent Dirichlet Allocation (LDA)، Hierarchical Dirichlet Process (HDP).

واعتمد الباحث في هذه الدراسة على خوارزمية، (LDA) Latent Dirichlet Allocation، والتي تم إنشاؤها وتطويرها بواسطة كل من ديفيد بلي (David Blei)، أندرو نغ (Andrew Ng)، ومايكل جوردان (Michael Jordan) في عام ٢٠٠٣. وقدم هؤلاء الباحثون ورقة بحثية تضمنت آليات نموذج LDA (Blei et al.,2003). كما تُعد الأكثر شهرة في نمذجة الموضوعات، ويتم الاعتماد عليها في العديد من أدوات ونماذج تعلم الآلة. حيث تقوم خوارزمية LDA على افتراض أن النصوص والمستندات التي تتناول موضوعات متشابهة، تستخدم مصطلحات وكلمات متشابهة. فمن خلال البحث عن مجموعات الكلمات التي تتكرر كثيرًا في النصوص والمستندات داخل مجموعة نصوص معينة، يمكن العثور على موضوعات ضمنية (Egger,2022).

ومن خلال الأداة Topic Modelling تم تحديد أكثر خمس موضوعات تتناولها مراجعات القراء قيد الدراسة، وفقًا لخوارزمية LDA. تتمثل في؛ قراءة الكتب وأسلوب المؤلفين – المكتبات والمعلومات – الكتب وأمناء المكتبات – المكتبات والتاريخ والمعلومات – المكتبات والتاريخ والقراءة. وقام الباحث بربط نتائج نمذجة الموضوعات بتقنيات LDAvis، MDS التي تستخدم لتحليل وتصور نتائج نمذجة الموضوعات، وفهم توزيع وتشابه الموضوعات في المجموعات النصية.

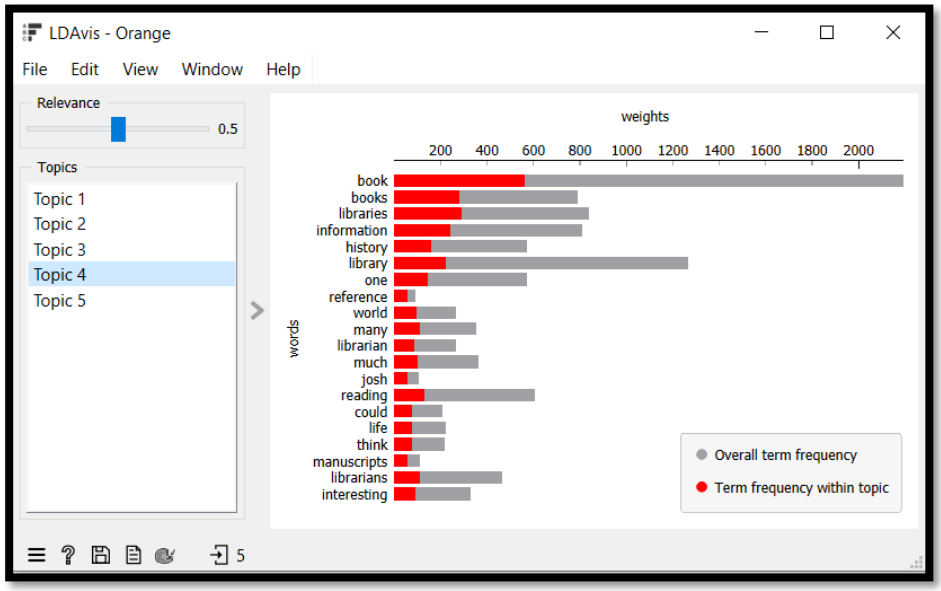
وتستخدم تقنية Multidimensional scaling (MDS) لتصور التشابه بين الموضوعات. في النص الذي تتم معالجته، وفهم العلاقات بين الموضوعات المتشابه. حيث يتم ضبط لون وحجم النقاط في MDS بناءً على الاحتمالية الهامشية لكل موضوع. وبالتالي يعكس مدى انتشار وتكرار كل موضوع في مجموعة النصوص (Saeed et al.,2018). ويتضح من خلال الشكل (18) أن

Topic 2 هو الأكثر انتشاراً، ولاستعراض أبرز الموضوعات الرئيسية في هذه النصوص، تم تحديد Topic 2 ثم عرضه من خلال الأداة Box plot، فتظهر قائمة بالكلمات والموضوعات الرئيسية التي تشملها مجموعة النصوص المتشابه في Topic2. والأوزان والإحصاءات التي حصلت عليها في عملية التحليل. كما موضح بالشكل التالي.



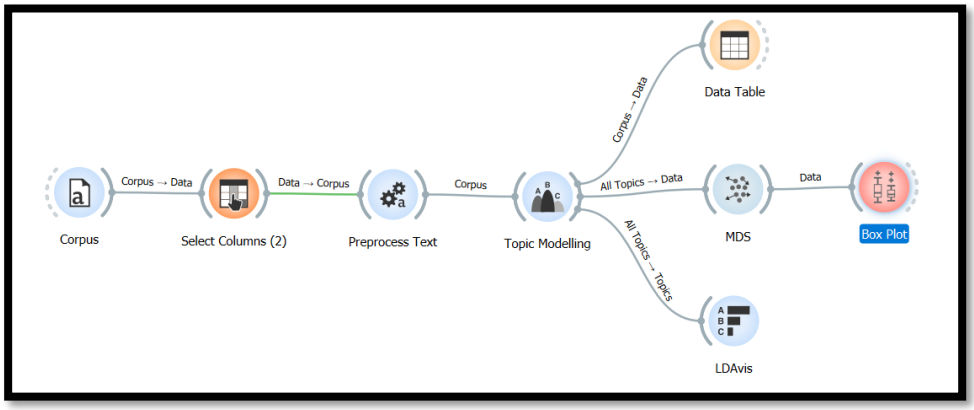
شكل (١٨) نمذجة الموضوعات لمراجعات القراء وفقًا بتقنية بتقنيات (MDS)

كما تعرض أداة LDAvis قائمة بالموضوعات التي تشملها مجموعة نصوص معينة، فعند تحديد Topic 4 على سبيل المثال، تظهر على الجهة اليمني قائمة بأعلى عشرين كلمة حسب الأهمية كما في الشكل (١٩)، ويمثل اللون الأحمر تكرار الكلمة او الموضوع في النص المحدد، بينما يمثل اللون الرمادي مدى تكرار الكلمة او الموضوع في كافة النصوص التي تتم معالجتها.



شكل (١٩) نمذجة الموضوعات لمراجعات القراء وفقًا لتقنية (LDAvis)

ويوضح الشكل (٢٠) سير عمل Workflow الذي قام بإعداده الباحث من خلال برنامج Orange Data Mining، لنمذجة الموضوعات لنصوص مراجعات القراء لكتب المكتبات والمعلومات بموقع LibraryThing، باستخدام الأداة Topic Modelling، وتقنيات التصور MDS، LDAvis.



شكل (٢٠) سير عمل Workflow المستخدم في نمذجة الموضوعات لنصوص مراجعات القراء لكتب المكتبات والمعلومات بـ LibraryThing.

(من إعداد الباحث باستخدام Orange Data Mining)

٥/٨ الكلمات الرئيسية في مراجعات القراء لكتب علوم المكتبات والمعلومات

في هذه الجزء نستعرض أبرز الكلمات الرئيسية التي تضمنتها مراجعات القراء في موقع LibraryThing حول كتب علوم المكتبات والمعلومات، ولإتمام هذه العملية استخدم الباحث بعض الأدوات المناسبة التي يوفرها برنامج Orange Data Mining ومنها؛

Bag of Words: تُعد إحدى الطرق البسيطة التي يتم الاعتماد عليها في معالجة اللغة الطبيعية (NLP) لتحويل النصوص أو المستندات إلى شكل رقمي قبل إدخالها في نماذج التعلم الآلي حتي يمكن استخدامها في عمليات التحليل والنمذجة (Topper,2023). ولذلك تم استخدام هذه الأداة لتهيئة النصوص الخاصة بمراجعات القراء في مجموعة البيانات محل الدراسة تمهيدا لاستخراج الكلمات الرئيسية.

Extract Keywords : حيث تستخدم هذه الأداة في استخراج الكلمات الرئيسية من النصوص في برنامج Data Mining Orange. وتتيح الأداة Extract Keywords عدة خوارزميات يمكن من خلالها استخراج الكلمات الرئيسية وهي؛ (TF-IDF, YAKE, Rake) واعتمد الباحث في هذه الدراسة على خوارزمية TF-IDF باعتبار أنها الأكثر شيوعًا، فضلًا عن دقتها في حساب تكرار الكلمات الرئيسية في النص.

١/٥/٨ خوارزمية (TF-IDF) :

تعد (TF-IDF) Term Frequency-Inverse Document Frequency إحدى تقنيات معالجة اللغة الطبيعية (NLP) التي تستخدم لتقييم أهمية الكلمات المختلفة في الجملة. ويُعْتَمَد عليها في تصنيف النصوص، ومساعدة نماذج التعلم الآلي على قراءة الكلمات، وتحليلها (Hamdaoui,2024). كما تُعد هذه الخوارزمية بمثابة أسلوب إحصائي يقيس مدى أهمية المصطلح داخل المستند أو النص بالنسبة إلى مجموعة من المستندات (أي بالنسبة إلى مجموعة الوثائق). (Kuncoro,Iswanto,2015). كما تجمع خوارزمية (TF-IDF) بين اثنين من مفاهيم حساب الوزن؛ تكرار ظهور الكلمة في وثيقة معينة، والتكرار العكسي للمستندات التي تحتوي على الكلمة. (Trihanto et al.,2017)

وتُحْسَب خوارزمية (TF-IDF) من خلال المعادلة الآتية: (Karabiber,2024).

تكرار المصطلح (TF) Term Frequency: يقصد به عدد المرات التي يظهر فيها المصطلح في المستند مقارنة بإجمالي عدد الكلمات في المستند.

$$TF = = \frac{\textbf{Number of times term appears in the document}}{\textbf{Total number of terms in the document}}$$

تكرار المستند العكسي: Inverse Document Frequency (IDF): ويقيس هذا المؤشر أهمية الكلمة في جميع المستندات أو النصوص. من خلال قسمة إجمالي عدد المستندات أو النصوص الكلية / عدد المستندات أو النصوص التي ظهرت بها المصطلح.

$$\text{IDF} = \textbf{Log}\left(\frac{\textbf{Total number of } \textit{the} \textbf{ documents in the corpus}}{\textbf{number of } \textit{the} \textbf{ documents in the corpus contain the term}}\right)$$

يتم حساب TF-IDF للمصطلح عن طريق ضرب ناتج كل من TF و IDF.

$$\textbf{TF-IDF} = \textbf{TF} \times \textbf{IDF}$$

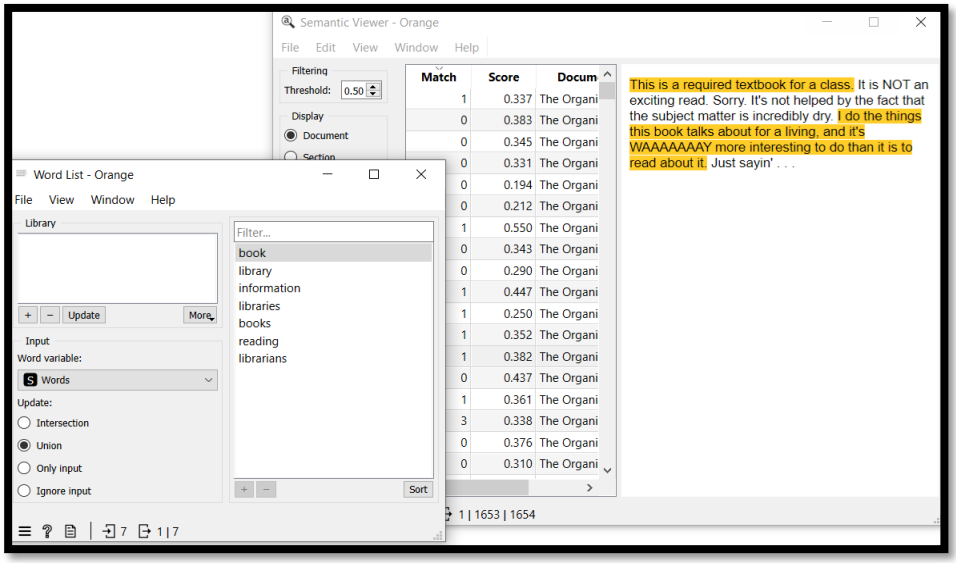
وبلغ عدد الكلمات الرئيسة التي أُسْتُخْرِجَت من نصوص مراجعات القراء حول كتب علوم المكتبات والمعلومات بموقع LibraryThing (١٦٠٠٣) كلمة من إجمالي (١٦٥٤) مراجعة. وذلك وفقًا لخوارزمية TF-IDF، وتم ترتيبها وفقا للمتوسط Mean. ويوضح الجدول (٨) أعلى عشرين كلمة رئيسية.

جدول (٨) الكلمات الرئيسية الأكثر تكرارًا في مراجعات كتب علوم المكتبات والمعلومات وفقًا لخوارزمية TF-IDF.

id	Word	TF-IDF
1.	library	0.024
2.	book	0.023
3.	information	0.022
4.	libraries	0.021
5.	read	0.019
6.	books	0.017
7.	history	0.017
8.	reading	0.016
9.	librarians	0.015
10.	good	0.014
11.	interesting	0.014
12.	one	0.013
13.	great	0.013
14.	fire	0.011

15.	well	0.011
16.	like	0.011
17.	really	0.011
18.	much	0.011
19.	would	0.01
20.	librarian	0.01

كما أضاف الباحث بعض الأدوات التي يمكن من خلالها عرض موضع الكلمات الرئيسة والسياق الخاص بها في نصوص مراجعات القراء، وذلك من خلال تحديد بعض الكلمات المراد عرضها من خلال الأداة Extract Keywords، ومن ثم يتم عرض تلك الكلمات في أداة Word list، للتعديل عليها بالإضافة أو الحذف، ومن خلال الأداة Semantic Viewer يتم استعراض موضع كل كلمة تم تحديدها في النصوص التي تتم معالجتها وتحليلها كما يتضح في الشكل التالي (٢١).



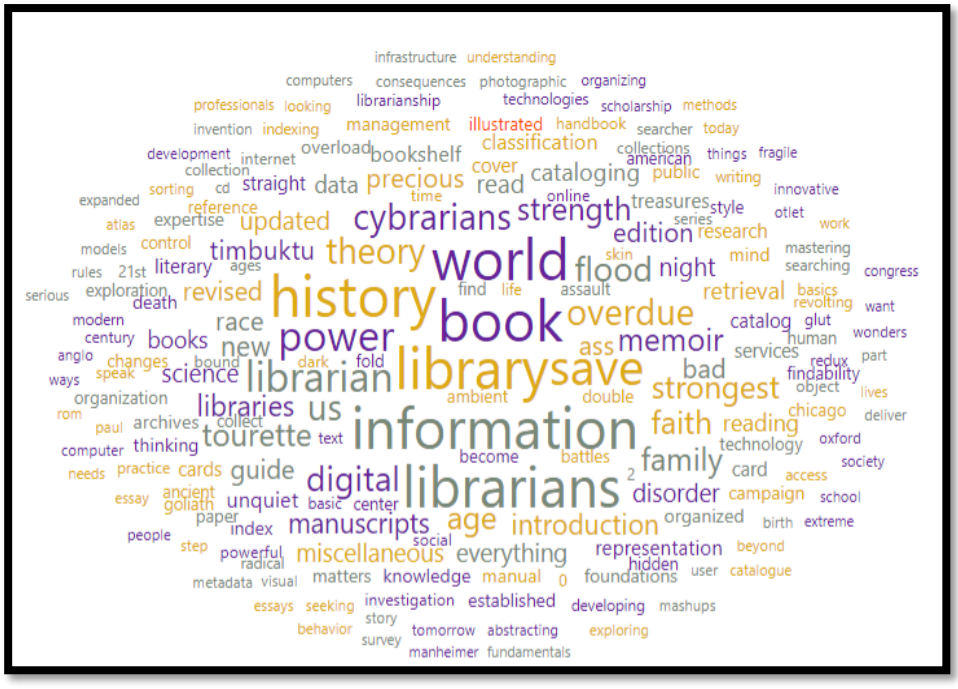
شكل (٢١) عرض الكلمات الرئيسية في نصوص مراجعات الكتب باستخدام أداة Word list، Semantic Viewer

(من إعداد الباحث باستخدام Orange Data Mining)

٦/8 سحابة الكلمات Word Cloud

أُنْشِئَت سحابة الكلمات Word Cloud والتي تعطي رؤى مرئية، وسريعة لأبرز الكلمات والعبارات الشائعة في النص، وتوضح سحابة الكلمات في شكل (٢٢) مدى تكرار الكلمات الرئيسية التي تناولها القراء من خلال مراجعاتهم لكتب علوم المكتبات والمعلومات من خلال LibraryThing، حيث تعطي هذه الأداة وزناً لكل كلمة أو مصطلح بناء على مرات تكرارها في النصوص، فكلما ظهرت الكلمة أكثر في مجموعة النصوص، كلما زاد وزنها وحجمها في السحابة. وأُسْتُخْدِم الكود التالي لتمثيل البيانات في سحابة الكلمات بشكل أكثر دقة.

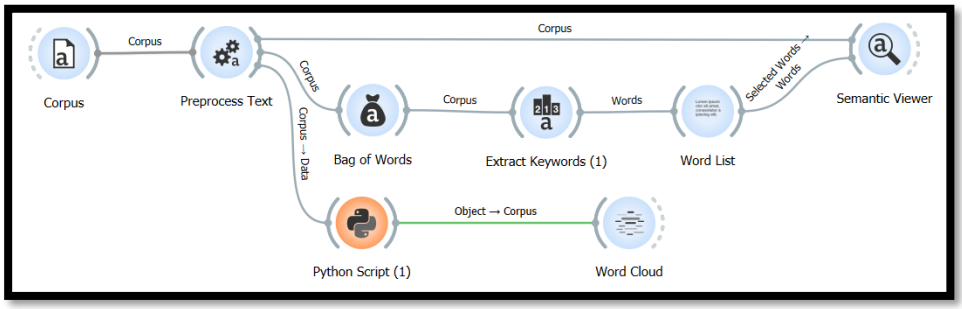
```
print('Running Preprocessing ...')
tokens = [doc.split(' ') for doc in in_data.documents]
print('Tokens:', tokens)
out_object = in_data
out_object.store_tokens(tokens)
```



شكل (٢٢) سحابة الكلمات Word Cloud لنصوص مراجعات القراء حول كتب علوم المكتبات والمعلومات.

ومن خلال الشكل السابق يتضح أن أكثر الكلمات تكرارا في نصوص مراجعات القراء حول كتب علوم المكتبات والمعلومات؛ Library (٤٩٤)، Book (٤١٦)، Information (٣٥٦). ومن خلال عرض سحابة الكلمات، تبين أن غالبية الكلمات السائدة في مراجعات الكتب قيد الدراسة، تعبر عن مشاعر إيجابية

ويوضح الشكل (٢٣) سير عمل Workflow الذي قام بإعداده الباحث من خلال برنامج Orange Data Mining، لتحليل واستخراج الكلمات الرئيسية من نصوص مراجعات القراء لكتب المكتبات والمعلومات بموقع LibraryThing باستخدام الأدوات؛ Bag of Words، Extract Keywords، Word list، Semantic Viewer.



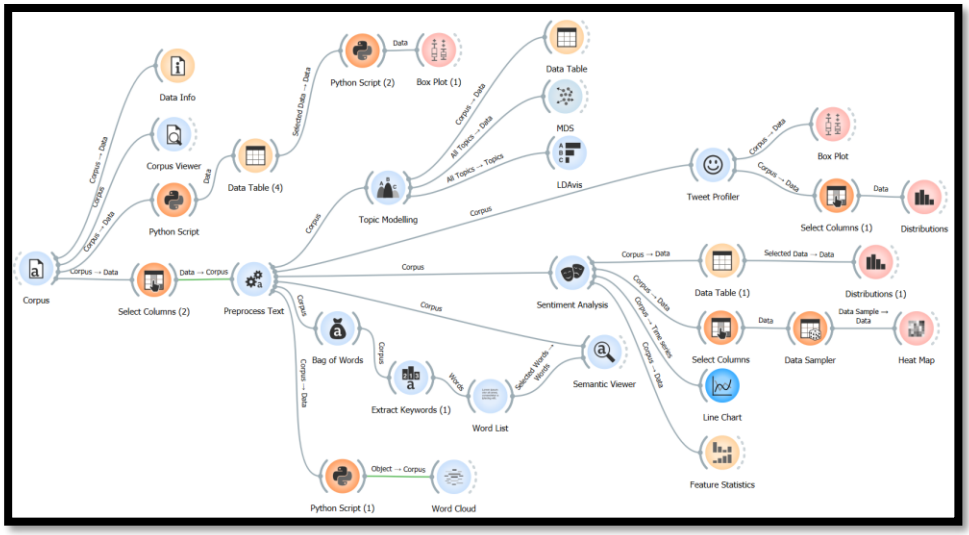
شكل (٢٣) سير عمل Workflow المستخدم في تحليل واستخراج الكلمات الرئيسية من نصوص مراجعات القراء لكتب المكتبات والمعلومات.

(من إعداد الباحث باستخدام Orange Data Mining)

٧/٨ التصميم النهائي لسير عمل (Workflow) المستخدم في الدراسة لتحليل المشاعر لمراجعات القراء.

يعرض الشكل (٢٤) النموذج النهائي لـ (**Workflow**) والذي أعده الباحث من خلال برنامج Orange Data Mining. حيث يشمل النموذج جميع خطوات ومراحل تحليل البيانات التي تمت خلال الدراسة، بما يعكس مسار تدفق البيانات، ومعالجتها، وتنظيفها، واستخراج البيانات الإحصائية، واستكشاف المشاعر، والعواطف التي تحملها نصوص مراجعات الكتب. وتقنيات تصور وعرض البيانات وتمثيلها مرئياً، فضلًا عن تقنيات نمذجة الموضوعات الرئيسية، والكلمات الأكثر تكراراً في نصوص مراجعات الكتب. وغيرها من مراحل العمل التي تمت على مجموعة بيانات مراجعات كتب علوم المكتبات والمعلومات بموقع LibraryThing.

ويتيح هذا النموذج إمكانية الاعتماد عليه سواء من قبل المكتبات ومؤسسات المعلومات، أو من خلال الباحثين المهتمين بغرض إجراء تحليل المشاعر لمراجعات القراء حول مجموعة كتب مختلفة.



شكل (٢٤) التصميم النهائي لـ (Workflow) المستخدم في الدراسة لتحليل المشاعر لمراجعات القراء من خلال Orange Data Mining

٩. نتائج الدراسة:

توصلت الدراسة إلى النتائج الأتية:

١. بلغ إجمالي مراجعات القراء التي جمعها (١٦٥٤) مراجعة كُتبت من جانب قراء LibraryThing على (199) كتاباً في مجال علوم المكتبات والمعلومات

٢. يتراوح تاريخ نشر كتب علوم المكتبات والمعلومات التي رُوجِعَت من قبل قراء موقع LibraryThing من عام ١٩٠٦م، وهو تاريخ نشر أو طبعة من كتاب The Chicago Manual of Style، وحتى عام ٢٠٢٢م.

٣. جاء عام ٢٠٠٧م في المرتبة الأولى برصيد (١٩) كتاباً، يليه عام ٢٠٠٨ برصيد (١٦) كتاب من إجمالي كتب علوم المكتبات والمعلومات التي رُوجِعَت بـ LibraryThing.

٤. جاء كتاب The Library Book في المرتبة الأولى من حيث عدد مراجعات القراء برصيد (178) مراجعة.

٥. إن اللغة السائدة لكتب علوم المكتبات والمعلومات التي تم مراجعتها من جانب قراء موقع LibraryThing؛ هي الإنجليزية وذلك برصيد (١٦٥١) كتاب، في حين تم رصد ثلاثة كتب بلغات مختلفة؛ الألمانية، والدنماركية، والسويدية.

٦. جاء كتاب Information Representation and Retrieval in the Digital Age في المرتبة الأولى من حيث المشاعر الإيجابية حيث سجل مؤشر compound (٠,٦١١٤). يليه كتاب The Information: A History, a Theory, a Flood بمعامل compound (٠,٥٨٥٩). ثم كتاب The Library Book بمعامل (٠,٥١٠٦).

٧. يعد كل من كتاب؛ Library: An Unquiet History، The Social Life of Information، Indexing and Abstracting in Theory and Practice أكثر الكتب التي تلقت مشاعر سلبية من خلال مراجعات القراء، حيث سجل مؤشر compound لكل منها (- ٠,٣١٨٢).

٨. إن المشاعر الإيجابية هي السائدة في مراجعات القراء للكتب، حيث بلغت (١٢٨٣) مراجعة إيجابية، بما يمثل (٧٧,٦٪). في حين بلغ عدد المراجعات السلبية نحو (٢٠٢) مراجعة سلبية بنسبة (١٢٪)، وبلغ عدد المراجعات التي صُنِّفَت بأنها محايدة (١٦٩) مراجعة محايدة بنسبة (١٠٪) من إجمالي مراجعات القراء.

٩. إن العاطفة السائدة في مشاعر القراء وفقاً لنموذج Ekman's هي المتعة والفرح حيث تم بلغ عدد المراجعات التي تم تصنيفها في الفئة Joy (١٤٥٨) مراجعة. في حين بلغ عدد المراجعات التي تم صُنِّفَت في كل من؛ الحزن (٣٣) مراجعة، وفي فئة الغضب (٣) مراجعات، وفي فئة الاشمئزاز مراجعة واحدة فقط من إجمالي المراجعات في مجموعة البيانات محل الدراسة.

١٠. إن غالبية مراجعات القراء التي تم تصنيفها من قبل نموذج Plutchik في كل من فئة الثقة Trustوالترقب Anticipation. تم التعبير عنها في نموذج Ekman كمشاعر تصف الفرح غالبًا.

١١. إن نموذج Plutchik يُعد هو الأكثر دقة في تصنيف مراجعات وتعليقات القراء وفقا للمشاعر والعواطف التي تعبر عنها، وربما يعود ذلك لزيادة فئات التصنيف في النموذج مقارنة بنموذج Ekman، من خلال إضافة فئات مختلفة؛ الثقة Trust والترقب Anticipation، مما أتاح إمكانية تصنيف بعض الآراء بشكل أكثر دقة.

١٢. نتج عن نمذجة الموضوعات تحديد أكثر خمس موضوعات تتناولها مراجعات القراء قيد الدراسة، وفقًا لخوارزمية LDA. تتمثل في؛ قراءة الكتب وأسلوب المؤلفين – المكتبات والمعلومات – الكتب وأمناء المكتبات – المكتبات والتاريخ والمعلومات – المكتبات والتاريخ والقراءة.

١٣. تُعد تقنيات LDAvis، MDS أفضل التقنيات التي يمكن من خلالها تحليل وتصور نتائج نمذجة الموضوعات، وفهم توزيع وتشابه الموضوعات في المجموعات النصية.

١٤. بلغ عدد الكلمات الرئيسة التي أُسْتُخْرِجَت من نصوص مراجعات القراء حول كتب علوم المكتبات والمعلومات بموقع LibraryThing (١٦٠٠٣) كلمة من إجمالي (١٦٥٤) مراجعة. وذلك وفقًا لخوارزمية TF-IDF.

١٥. تبين من خلال سحابة الكلمات أن أكثر الكلمات تكرارا في نصوص مراجعات القراء حول كتب علوم المكتبات والمعلومات؛ Library (٤٩٤)، Book (٤١٦)، Information (٣٥٦). وأن غالبية الكلمات السائدة في مراجعات الكتب قيد الدراسة، تعبر عن مشاعر إيجابية.

١٠.توصيات الدراسة:

من خلال الدراسة خرج الباحث بعدد من التوصيات المقترحة، التي تسهم في تعزيز سبل الإفادة من تقنيات التعلم الآلي، ومعالجة اللغة الطبيعية في تطوير الخدمات المعرفية والمعلوماتية بالمكتبات ومؤسسات المعلومات ومنها؛ في تطوير وتحسين تجربة المستخدم في المكتبات ومؤسسات المعلومات، ومنها ما يلي:

أولًا: توصيات موجهة للقائمين على للمكتبات ومؤسسات المعلومات العربية:

١- ضرورة التوسع من قبل المكتبات ومؤسسات المعلومات في الاعتماد على تقنيات معالجة اللغة الطبيعية، وأدوات تحليل المشاعر حتى يمكنها تعظيم الإفادة من تلك التقنيات في العديد من القضايا على النحو الأتي:

- تحديد الفئات المستهدفة من المكتبات ومؤسسات المعلومات بشكل أكثر دقة، وبالتالي تحسين استراتيجيات التسويق لجذب تلك الفئات.

- تنظيم ورش العمل، الندوات، ونوادي القراءة التي تتماشى مع اهتمامات وتوجهات المستفيدين من المكتبة، والتي رُصِدَت من خلال تحليل المشاعر.

- ضبط وتحديث سياسات الاقتناء لدى المكتبات ومؤسسات المعلومات من خلال تحديد مصادر المعلومات المختلفة التي تحظى بتقييمات إيجابية.

- كما أن الكشف عن مشاعر القراء تجاه المجموعات الحالية، يعزز بدوره اتخاذ القرارات الصائبة بشأن الاحتفاظ بالمواد أو التخلص منها.

- الاعتماد على نتائج تحليل مشاعر رواد المكتبة، يسهم في توجيه سياسات المكتبة وبرامجها، بما يضمن تلبية توقعات واحتياجات المستفيدين منها.

٢- إعداد برامج وورش تدريبية للمكتبيين حول كيفية استخدام تقنيات معالجة اللغة الطبيعية (NLP) وأدوات تحليل المشاعر لفهم وتحليل توجهات القراء.

ثانيًا: توصيات موجهة للباحثين العرب والمهتمين بالمجال.

٣- ضرورة إجراء المزيد من الدراسات المستقبلية البينية بالاشتراك مع المتخصصين في علوم الذكاء الاصطناعي لإعداد خوارزميات مناسبة، وبناء نماذج التعلم الآلي التي تسهم في تطوير وتحسين تجربة المستخدم في المكتبات ومؤسسات المعلومات. ويقترح الباحث مجموعة من النقاط البحثية تتمثل في الأتي:

- دراسة تحليل المشاعر لمراجعات وآراء القراء حول الكتب في التخصصات الموضوعية الأخرى، بالاعتماد على أدوات أخري بخلاف Orange Data Mining. ومقارنة نتائجها.

- إجراء دراسات مقارنة لتحليل مشاعر القراء تجاه نفس الكتب في عدة مكتبات مختلفة، لقياس مدى تأثير السياسات والخدمات المختلفة بالمكتبات على مشاعر وتقييمات القراء.

- دراسة مشاعر وتوجهات المستفيدين تجاه خدمات المكتبات ومؤسسات المعلومات، وبخاصة الخدمات الرقمية، بغية رصد المشكلات القائمة، وبحث سبل حلها لتلبية احتياجات المستفيدين.

- دراسة العلاقة بين مشاعر القراء، ومعدلات استعارة الكتب في المكتبات، بغرض تطوير إدارة المجموعات بالمكتبة.

١١.المراجع:

أولاً: المراجع العربية:

باتي، غسان بن فاروق. (٢٠٢٣). عجوة أو مجدول: مجموعة بيانات متوازنة الصنفين لتدريس تعلم الآلة. مجلة دراسات المعلومات والتكنولوجيا، 6(٢)،١ - ١٣.

الخليفى، طارق. (٢٠١٩). تنقيب بيانات وسائل التواصل الاجتماعي واستخداماته في البحوث الاعلامية: تحليل المشاعر نموذجا. مجلة البحوث والدراسات الإعلامية، (٨)، ٢٧٩ - ٣٥١.

رية، آية الشربيني حسين، عبدالمجيد، طلعت أسعد، و يسن، هشام محمد محمد. (٢٠٢٢). توسيط مشاعر العميل في العلاقة بين استعادة ثقة العميل ونية إعادة الشراء بالتطبيق على عملاء مطاعم الوجبات السريعة بمحافظة الدقهلية. المجلة المصرية للدراسات التجارية، ٤٦ (١)، ١ - ٢٥.

سامي، ريهام (٢٠٢١) اتجاهات مستخدمي مواقع التواصل الاجتماعي نحو جائحة كورونا (كوفيد-١٩): تحليل من المستوى الثاني لدراسات مدخل معالجة اللغة الطبيعية. مجلة البحوث الإعلامية، ٥٦ (٣)، ١٢٧٧-١٣١٨.

سميع، ليلى سيد. (٢٠٢٢). تحليل المشاعر القائم على الذكاء الاصطناعي كأداة لتقييم فعالية خدمات المعلومات. المجلة العلمية للمكتبات والوثائق والمعلومات، ٤ (١١)، ٧ - ٣٢.

الشريف، سلوى أحمد محمد أبو العلا. (٢٠٢٢). تحليل مشاعر تغريدات توتير أثناء الانتخابات الرئاسية الأمريكية ٢٠٢٠ باستخدام إطار البيانات الضخمة. المجلة المصرية لبحوث الرأي العام، ٢١ (٢)، ١ - ٤٩.

عطية، عطية محمد (٢٠٢٣). تحليل مشاعر الجمهور العربي نحو العلاقات العامة وممارسيها عبر تويتر: دراسة باستخدام تحليلات البيانات الضخمة .Big Data Analytics المجلة العلمية لبحوث العلاقات العامة والاعلان. ٢٠٢٣ (٢٦)، ٥٧ – ٩٧.

مصطفاوي، الطيب. (٢٠٢٢). تحليل المشاعر لمراجعات الشبكات الاجتماعية للوجهات السياحية باستخدام التقنيات الذكية. مجلة العلوم الاقتصادية والتسيير والعلوم التجارية، ١٥ (٢)، ١ - ١٦.

ثانياً: المراجع الأجنبية:

Abdulqader, A. W. (2023). Predicting carbon dioxide emissions with the orange application: an empirical analysis. Mesopotamian Journal of Computer Science, 2023, 53-63.

Ahmad, I. S., Bakar, A. A., Yaakub, M. R., & Darwich, M. (2020). Beyond sentiment classification: A novel approach for utilizing social media data for business intelligence. International Journal of Advanced Computer Science and Applications, 11(3).

Al Assyam, H. D., & Hasan, F. N. (2023). Analisis sentimen Twitter terhadap perpindahan ibu kota negara ke IKN nusantara menggunakan orange data mining. KLIK: Kajian Ilmiah Informatika Dan Komputer, 4(1), 341-349.

Al-Daihani, S. M., & Abrahams, A. (2016). A text mining analysis of academic libraries' tweets. The journal of academic librarianship, 42(2), 135-143.

Almjawel, A., Bayoumi, S., Alshehri, D., Alzahrani, S., & Alotaibi, M. (2019, May). Sentiment analysis and visualization of amazon books' reviews. In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Çakmak, T., & Eroğlu, Ş. (2021). Public libraries on Facebook: content analysis of Turkish public libraries' posts. Library Hi Tech, 39(2), 602-624.

Collins, M., & Karami, A. (2018). Social media analysis for organizations: Us northeastern public and state libraries case study. arXiv preprint arXiv:1803.09133.

Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web (pp. 519-528).

De, S., Dey, S., Bhatia, S., & Bhattacharyya, S. (2022). An introduction to data mining in social networks. In Advanced data mining tools and methods for social computing (pp. 1-25). Academic Press.

Deo, G. S., Mishra, A., Jalaluddin, Z. M., & Mahamuni, C. V. (2020, September). Predictive analysis of resource usage data in academic libraries using the vader sentiment algorithm. In 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 221-228). IEEE.

Duong, T. T. L. (2023). *Exploring psychological topics and sentiments in how people tweet about chronic pain: A text-mining approach* (Master's thesis, University of Twente).

Egger, R. (2022). Topic modelling: modelling hidden semantic structures in textual data. In Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications (pp. 375-403). Cham: Springer International Publishing.

Ferhi, W., Hadjila, M., & Djillali, D. M. (2023, November). Machine Learning-based Classification of Diabetes Disease: A Case Study with Orange Data Mining. In 2023 International Conference on Electrical Engineering and Advanced Technology (ICEEAT) (Vol. 1, pp. 1-6). IEEE.

Hamdaoui,Yassine (2024) A Guide to TF-IDF. Retrieved at 24 June 2024,from: https://builtin.com/articles/tf-idf

Irawan, H., Akmalia, G., & Masrury, R. A. (2019, September). Mining tourist's perception toward Indonesia tourism destination using sentiment analysis and topic modelling. In Proceedings of the 2019 4th International Conference on Cloud Computing and Internet of Things (pp. 7-12).

Karabiber, Fatih (2024).TF-IDF — Term Frequency-Inverse Document Frequency . Retrieved at 24 June 2024,from: https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/

Khan, A. M., & Loan, F. A. (2022). Exploring the reviews of Google Maps to assess the user opinions about public libraries. Library Management, 43(8-9), 601-615.

Koleva, L. S., & Filipov, G. A. (2023). Classification of chest X-Ray images using Orange Data Mining Tool. Electrotechnica & Electronica (E+ E), 58(2).

Kuncoro, B. A., & Iswanto, B. H. (2015, November). TF-IDF method in ranking keywords of Instagram users' image captions. In 2015 International Conference on Information Technology Systems and Innovation (ICITSI) (pp. 1-5). IEEE.

Liu, B. (2022). Sentiment analysis and opinion mining. Springer Nature.p7.

Manasa, P., & Prasad, U. D. (2024). Journey Through Opinions: Unravelling Heritage Hotel Experience And Customer Sentiment Through Topic Modelling In Online Reviews–A Case Study Of Taj Udaipur. *Educational Administration: Theory and Practice*, *30*(5), 1674-1682.

MARCU, D., & DANUBIANU, M. (2020, May). Sentiment analysis from students' feedback: a Romanian high school case study. In 2020 International Conference on Development and Application Systems (DAS) (pp. 204-209). IEEE.

MARCU, D., & DANUBIANU, M. (2020, May). Sentiment analysis from students' feedback: a Romanian high school case study. In 2020 International Conference on Development and Application Systems (DAS) (pp. 204-209). IEEE.

Marinho, R. S., Pereira, C. M., & Segundo, J. E. S. (2022). Análise de sentimento como apoio à seleção de livros: um estudo aplicado à plataforma Skoob. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, 27, 1-20.

Marreddy, M., & Mamidi, R. (2023). Learning sentiment analysis with word embeddings. In Computational Intelligence Applications for Text and Sentiment Data Analysis (pp. 141-161). Academic Press.

Martinez, M., Escobar, B., Maria-Elena, G. D., & Pinto-Roa, D. P. (2021). Market basket analysis with association rules in the retail sector using Orange. Case Study: Appliances Sales Company. CLEI Electronic Journal, 24(2), 12-1.

Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), 39-41.

Mohsin, M. A., & Beltiukov, A. (2019, May). Summarizing emotions from text using Plutchik's wheel of emotions. In 7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019) (pp. 291-294). Atlantis Press.

Molenaar, A., Lukose, D., Brennan, L., Jenkins, E. L., & McCaffrey, T. A. (2024). Using Natural Language Processing to Explore Social Media Opinions on Food Security: Sentiment Analysis and Topic Modeling Study. Journal of Medical Internet Research, 26, e47826.

Nahar, K. M., Jaradat, A., Atoum, M. S., & Ibrahim, F. (2020). Sentiment analysis and classification of arab jordanian facebook comments for jordanian telecom companies using lexicon-based approach and machine learning. Jordanian Journal of Computers and Information Technology, 6(3).

Nasukawa, T., & Yi, J. (2003, October). Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture (pp. 70-77).

Noh, Y., & Kim, D. (2022). A study on social perceptions of public libraries utilizing the sentiment analysis. International Journal of Knowledge Content Development & Technology, 12(4).

Noviana, R. (2021). Perancangan Web Sistem Analisis Sentimen Media Sosial Twitter Dengan Metode Valence Aware Dictionary And Sentimen Reasoner (Vader) Menggunakan PHP & MysSQL pada Pemerintah Kota Bekasi: Array. *Jurnal Ilmiah Komputasi*, *20*(1), 1-14.

Palomino, M., Allen, R., Aider, F., Tirotto, F. A., Giorgi, I., Alexander, H., & Masala, G. (2022, September). The mood of the silver economy: a data science analysis of the mood states of older adults and the implications on their wellbeing. PTI.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. ArXiv, cs.CL/0205070.

Parabhoi, L. (2019). Twitter sentiment analysis: A case study of ten university libraries. INFLIBNET Centre, Gandhinagar.

Peker, M., Özkaraca, O., & Şaşar, A. (2018). Use Of Orange Data Mining Toolbox for Data Analysis In Clinical Decision Making: The Diagnosis Of Diabetes Disease. In Expert System Techniques In Biomedical Science Practice (Pp. 143-167). Igi Global.

Porter, Martin (2006) The Porter Stemming Algorithm. Retrieved at ١٠ May2024,from: https://tartarus.org/martin/PorterStemmer/

Saeed, N., Nam, H., Haq, M. I. U., & Muhammad Saqib, D. B. (2018). A survey on multidimensional scaling. ACM Computing Surveys (CSUR), 51(3), 1-25.

Sajwan, V., & Ranjan, R. (2019). Classifying flowers images by using different classifiers in orange. International Journal of Engineering and Advanced Technology, 8(6), 1057-1061.

Sajwan, V., Awasthi, M., Goel, A., & Sharma, P. (2023). Sentiment analysis of Twitter data regarding the agnipath scheme of the defense forces. Indonesian Journal of Electrical Engineering and Computer Science, 30(3), 1643-1650.

Salem, I. E., & Abdulqader, A. W. (2021). Predicting Carbon Dioxide Emissions with the Orange Application: An Empirical Analysis.

Suttles, J., & Ide, N. (2013, March). Distant supervision for emotion classification with discrete binary values. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 121-136). Berlin, Heidelberg: Springer Berlin Heidelberg.

Thange, U., Shukla, V. K., Punhani, R., & Grobbelaar, W. (2021, January). Analyzing COVID-19 dataset through data mining tool "orange". In 2021 2nd international conference on computation, automation and knowledge management (ICCAKM) (pp. 198-203). IEEE.

Tiwari, R., Kumar, G., & Gunjan, V. K. (2022, December). Effect of Environment on Students Performance Through Orange Tool of Data Mining. In International Conference on Data Science, Machine Learning and Applications (pp. 283-292). Singapore: Springer Nature Singapore.

Tong, Z., & Zhang, H. (2016). A text mining research based on LDA topic modelling. In International conference on computer science, engineering and information technology (pp. 201-210).

Topper, Noah (2023) Bag of Words Model in NLP Explained. Retrieved at 25 June 2024,from: https://builtin.com/machine-learning/bag-of-words

Trihanto, W. B., Arifudin, R., & Muslim, M. A. (2017). Information Retrieval System for Determining The Title of Journal Trends in Indonesian Language Using TF-IDF and Na? ve Bayes Classifier. Scientific Journal of Informatics, 4(2), 179-190.

Umer, D. (2020). Analysis Of Heart Patients Disease Using Data Mining Tool Orange. Vol, 9, 1146-1150.

Zishumba, K. (2019). Sentiment Analysis Based on Social Media Data (Doctoral dissertation).

ثالثا: الهوامش:

[1] Ahmed Mohamed Ali (٢٠٢٤) books reviews at LibraryThing. At: https://www.kaggle.com/datasets/ahmedalielmokhtar/books-reviews-at-librarything

[2] (Code): يشير رقم الكود في جداول تحليل البيانات الواردة في الدراسة إلى نصوص مراجعات القراء بمجموعة البيانات، حيث أعطى الباحث لكل مراجعة كود مختلف بشكل متسلسل، بحيث يسهل الرجوع للبيانات من جانب القارئ.

Copyright of Egyptian Journal of Information Sciences is the property of Beni Suef University and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.