تحليل الاتجاهات الموضوعية للأخبار السياسية العراقية باستخدام نموذج (*BERTopic*)

المدرس الدكتور ايناس جاسم هادي
كلية الآداب / الجامعة المستنصرية

المستخلص

يهدف هذا البحث إلى تحليل الاتجاهات الموضوعية في الأخبار السياسية العراقية خلال الفترة من ٢٠١٥ إلى ٢٠٢٣، وذلك من خلال توظيف تقنيات حديثة في معالجة اللغة الطبيعية وتحليل البيانات، مع التركيز على نمذجة الموضوعات باستخدام نموذج (BERTopic) المستند إلى تقنيات تعلم الآلة العميق. تم جمع بيانات ضخمة من الأخبار السياسية وتنقيحها ومعالجتها لغويًا باستخدام أدوات عربية مخصصة وتمثيل النصوص عدديًا، وأخيرا تطبيق خوارزميات الذكاء الاصطناعي لاكتشاف الموضوعات المتكررة وتحليلها. وعلى وفق ذلك تم استخراج أبرز عشر موضوعات إخبارية، حيث اجري تحليل دلالي وزمني لرصد تطورها ومكوناتها اللغوية عبر السنوات، مع إيلاء اهتمام خاص لتحليل المعلومات السياقية وتحول الخطاب السياسي داخل الإعلام العراقي. كما تم استخدام أدوات بصرية مثل سحابة الكلمات والتحليل السنوي للكشف عن التحولات في الخطاب الموضوعي للإعلام. أثبتت النتائج فعالية نمذجة الموضوعات في تحويل النصوص السياسية إلى بنى معلوماتية قابلة للتحليل، مما يعزز من دور علم المعلومات في فهم وتحليل الخطاب السياسي، ويوفر آفاقًا جديدة لاستثمار البيانات النصية في دعم صناع القرار وتوجيه البحوث الإعلامية والاجتماعية.

الكلمات المفتاحية: **تحليل الاتجاهات الموضوعية، تحليل الاخبار، نمذجة الموضوعات، نموذج BERTopic، علم المعلومات.**

تاريخ الاستلام: ٢٠٢٥/٠٤/٢٣ **تاريخ القبول: ١١ /٠٥/٢٠٢٥**

Analyzing Thematic Trends in Iraqi Political News Using the BERTopic Model

Lecturer Dr. Enas Jasim Hadi
College of Arts, Al-Mustansiriyah University

Abstract

This study aims to analyze thematic trends in Iraqi political news from 2015 to 2023 by employing advanced techniques in natural language processing and data analysis, with a particular focus on topic modeling using the BERTopic model based on deep machine learning methods. A large corpus of political news articles was collected, cleaned, and linguistically processed using specialized Arabic language tools. The texts were then transformed into numerical representations, after which artificial intelligence algorithms were applied to identify and analyze recurring topics.

As a result, the study extracted the ten most prominent news topics. These topics were subjected to semantic and temporal analysis to trace their development and linguistic features over time, with particular attention given to contextual information and shifts in political discourse within Iraqi media. Visual analytical tools, such as word clouds and annual trend analysis, were also employed to reveal transformations in thematic discourse across the studied period.

The findings demonstrate the effectiveness of topic modeling in converting political texts into structured, analyzable information, thereby reinforcing the role of information science in understanding and analyzing political discourse. Moreover, the study opens new avenues for leveraging textual data to support decision-makers and guide media and social research.

Keywords: thematic trend analysis, news analysis, topic modeling, BERTopic model, information science.

Received: 23/04/2025 **Accepted: 11/05/2025**

المقدمة

بشكل عام ومنذ عقود، تشكل الاخبار أحد اهم الوسائل الرئيسة التي يعتمد عليها الجمهور في الحصول على المعلومة وتشكيل رؤية محددة حول مختلف القضايا المحلية والدولية، الامر الذي يجعل من الاخبار من الأمور الضرورية التي يسعى المتلقي لتحريها ومواكبتها (Park, 2022)، ولفهم الاتجاهات العامة للأخبار والتفاعل معها، لا بد من توفر تقنيات متطورة لتحليل البيانات النصية للأخبار (Rouhana, 2023). وفي ظل التطور السريع لتقنيات الذكاء الاصطناعي وتعلم الالة، ظهرت عدد من الأدوات والتقنيات المتطورة والتي تهدف الى تمكين الباحثين من تحليل كميات هائلة من بيانات النصية واستخلاص المعلومات والمعرفة منها بشكل فعال وكفوء (Amara et al., 2021). ومن جانب اخر، اثبتت الدراسات السابقة في هذا السياق ان توظيف تقنيات لذكاء الاصطناعي وتعلم الالة يتيح إمكانات لاستخلاص الأنماط الخفية والكامنة في البيانات النصية الإعلامية وتحويلها الى معلومات عميقة تعطي اتجاهات وتفسيرات قد يصعب اكتشافها باستخدام الأساليب التقليدية لتحليل البيانات. ومن امثلة هذه التقنيات هي تقنيات نمذجة الموضوعات (Topic Modeling) التي تعمل على كشف الأنماط الكامنة وتوجهات الخطاب الإعلامي في النصوص ومن بينها النصوص العربية (Barberá et al., 2021; Timmerman & Bronselaer, 2022). حيث تساهم هذه النماذج والأدوات في اجراء التصنيفات التلقائية للنصوص بشكل دقيق مما يدعم عمليات تحليل الخطاب النصي وتقديم رؤى مختلفة للخطاب الإعلامي. وبالنظر لأهمية اللغة العربية وكثرة عدد مستخدميها حول العالم واعتبارها أحد اهم اللغات الحيوية حول العالم (Alshammary et al., 2024; Ibrahim et al., 2021)، الا ان توظيف هذه التقنيات ما يزال يعاني من عدد من التعقيدات المتعلقة بطبيعة النص العربي من جهة مثل التنوع اللغوي والتشكيل، وصعوبة تنفيذ إجراءات معالجة اللغات الطبيعية من جهة أخرى.. وبالأخص إذا ما تم مقرنة النص العربي من نصوص اللغات الأخرى كالإنجليزية مثلا، التي لا تتسم بالتعقيد الموجود في اللغة العربية. حيث تعتمد اللغة الإنجليزية على الاحرف فقط ولا يوجد تشكيل او حركات اعرابية تتحكم بالمعنى بشكل جذري (Khelil et al., 2024). وفي ظل ذلك ظهرت عدة نماذج مستندة الى الذكاء الاصطناعي تختص بالتعامل مع اللغة العربية ونمذجة موضوعاتها مثل (BERTopic)، الذي يختص بتنفيذ نمذجة الموضوعات ومعالجتها في اللغة العربية (Jannani et al., 2025).

تتمثل أهمية بحثنا الحالي في تقديم إطار عملي متكامل يساهم في تحليل التوجهات العامة للخطاب الإعلامي عن طريق اجراء تحليل للموضوعات الخاصة بالإعلام السياسي العراقي، الامر الذي يساهم في تطوير اليات وتقنيات تمكن من اكتشاف الأنماط الكامنة في البيانات النصية الكبيرة.

١. مشكلة البحث

- هل يمكن استخدام تقنيات الذكاء الاصطناعي وتعلم الالة في حليل المحتوى النصي للأخبار

السياسية العراقية؟

- ماهي أبرز الأنماط والتوجهات التي يمكن استخلاصها من المحتوى النصي العربي.
- هل يمكن تقديم تحليل موضوعي للخطاب الإعلامي والكشف عن اهم الانماط الكامنة للخطاب الإعلامي.

٢. أهداف البحث

- تقديم نموذج فعال مستند الى تقنيات الذكاء الاصطناعي وتعلم الالة لتحليل البيانات النصية العربية للأخبار السياسية العراقية.
- استخراج أنماط وتوجهات للخطاب الإعلامي العراقي من خلال تنفيذ تقنيات نمذجة الموضوعات.
- تقديم تحليل موضوعي للخطاب الإعلامي والكشف عن الأنماط الكامنة.

٣. فرضيات البحث

في ضوء أهداف هذا البحث، تم اعتماد فرضيات بحثية تساعد في توجيه العمل نحو اختبار قدرة الأدوات الحديثة على تحليل الموضوعات السياسية داخل المحتوى الإخباري العراقي. وقد بُنيت هذه الفرضيات استنادًا إلى الدراسات السابقة ومنهجية نمذجة الموضوعات المعتمدة على تقنيات تعلم الآلة:

- الفرضية الأولى: تُساهم نمذجة الموضوعات باستخدام BERTopic في تصنيف المحتوى السياسي العراقي إلى موضوعات متمايزة تعكس الاتجاهات العامة للخطاب الإعلامي.

- الفرضية الثانية: يُمكن من خلال تحليل تطور الموضوعات عبر الزمن الكشف عن التحولات الدلالية في تغطية القضايا السياسية، مما يعزز فهم بنية المعلومات وتغيرها في السياق الإعلامي.

٤. أهمية البحث

تبرز أهمية هذا البحث في توظيف نمذجة الموضوعات كأداة معرفية فعالة لتحليل البيانات النصية غير المنظمة، التي تُعد من التحديات الجوهرية في ميدان علم المعلومات. ففي ظلّ الكم الهائل من المحتوى السياسي المنتج يوميًا، تصبح القدرة على تحويل النصوص إلى هياكل موضوعية قابلة للتحليل أمرًا ضروريًا لفهم التوجهات العامة والخطاب الإعلامي. ويُضيف هذا البحث قيمة علمية مضاعفة من خلال اعتماده على نماذج لغوية متقدمة، لمعالجة النصوص العربية التي تتميز بتعقيدها التركيبي وثرائها الدلالي، وهو ما يشكّل تحديًا أمام العديد من نماذج الذكاء الاصطناعي المصممة أساسًا للغة الإنجليزية. كما يُساهم البحث في دعم تخصص علم المعلومات والمعرفة من خلال تقديم نموذج تطبيقي لتحليل البيانات النصية سياسيًا وزمنيًا، بما يعكس القدرة على استخلاص المعرفة من النصوص وتحويلها إلى مؤشرات ذات قيمة تفسيرية يمكن استثمارها في الإعلام، وصناعة القرار، والدراسات الاجتماعية.

٥. **حدود البحث** :بغية ضمان دقة النتائج ووضوح معالمها، يشمل البحث الحدود الزمانية والموضوعية التالية:

- الحدود الزمانية: تتمثل الحدود الزمانية للبحث في البيانات النصية الخاصة بالأخبار السياسية العراقية التي تم جمعها من الوكالة الوطنية العراقية للأنباء والتي تمتد من العام ٢٠١٥ ولغاية لعام ٢٠٢٣. وتم تحديد هذه الفقرة بناءا على ما شملته من احداث سياسية مختلفة على المستوى العراقي والإقليمي والعالمي. الامر الذي يجعلها فترة غنية بالأحداث المهمة التي تثري نتائج البحث.

- الحدود الموضوعية: من الناحية الموضوعية، يتحدد البحث بتحليل الاخبار السياسية المنشورة في الوكالة الوطنية العراقية للأخبار دون التطرق الى الاخبار ذات الأنواع الأخرى غير السياسية. حيث يقتصر التحليل على النصوص المكتوبة باللغة العربية فقط بغية التطرق الى تحديات وصعوبات هذه اللغة في عمليات التحليل المستند الى تقنيات الذكاء الاصطناعي وتعلم الالة. ويعتمد البحث أدوات نمذجة الموضوعات (Topic Modeling) كمنهج أساسي من بين مناهج تعلم الالة لاستخلاص الأنماط الكامنة وتحليلها.

٦. **الدراسات السابقة**

في هذا القسم، نقوم بعرض اهم الدراسات التي تناولت الموضوعات المرتبطة ببحثنا الحالية وهي نمذجة الموضوعات وتحليل النصوص العربية. ويجري عرض لهذه الدراسات والتقنيات المستخدمة بغية الوقوف على أحدث التقنيات والبحوث التي تم إنجازها بهذا الصدد. وكذلك لإعطاء نظرية تحليلية عن هذه الدراسات وكف تطورت عبر الزمن. في دراسة قدمها (Barberá et al., 2021) تم تقديم نموذج لتصنيف نصوص الاخبار اعتمادا على خوارزميات الذكاء الاصطناعي وتعلم الالة. تم من خلال الدراسة القيام بتحليل الجمل والفقرات النصية الإخبارية عن طريق استخدام خوارزمية (Support Vector Machine (SVM)) وبينت النتائج قدرة النماذج الذكية على تحسين قدرة التصنيف وزيادة دقتها مقارنة بالنماذج التقليدية. وفي دراسة أخرى (Amara et al., 2021) تم استخدام نمذجة الموضوعات متعددة اللغات باستخدام خوارزمية ((Latent Dirichlet Allocation (LDA) لاستخراج الموضوعات الكامنة داخل نصوص تعليقات لمنصة (Facebook) بغية تحليل اتجاهات المواضيع في ظل جائحة كوفيد -١٩. ولقد نجحت الدراسة في تحقيق نتائج مميزة في عملية تتبع التوجهات وردود الفعل الاجتماعية حول القضايا المتعلقة بالجائحة. كما وفرت الدراسة رؤية واضحة حول إمكانية استخدام بيانات متعددة اللغات لفهم الاحداث المعينة والآراء حولها. اما في الدراسة المقدمة من قبل (Alswedani et al., 2022) فتم فيها استخدام بيانات نصية من منصة تويتر (المعروفة حاليا بمنصة X). أظهرت النتائج انه يمكن تقديم تحديدا دقيقا لموضوعات تتعلق بالحوكمة والتعليم عن بعد في المملكة العربية السعودية. الامر الذي يساهم بشكل كبير في دعم القرارات وتسهيل الإجراءات خلال فترة جائحة كوفيد-١٩.

في دراسة (Almuzaini & Azmi, 2022) تم تطبيق نمذجة الموضوعات باستخدام الخوارزميات الجينية (Genetic Algorithms). هدفت الدراسة الى تحليل النصوص العربية المكتوبة باليد وتصنيفها. بينت النتائج ان هنالك تحسنا كبيرا في دقة التصنيف واستخلاص عدد جيد من الموضوعات بعد تحليل المحتوى النصي. وفي دارسة أخرى تم تقديمها في (Rouhana, 2023) تم اعتماد على الدمج فيما بين الموضوعات الخاصة بتحليل الاخبار. حيث تم تحليل الخطاب النقدي لقناة الجزيرة الإخبارية لنصوص مكتوبة باللغة العربية. وقدمت الدراسة تحليلا مفصلا عن الأطر الخطابية وماهي الخطابات السائدة والمهيمنة على الاحداث. وبينت نتائج الدراسة ان هنالك فعالية كبيرة للنماذج المستخدمة في استكشاف الأنماط الخطابية في الوسائل الإعلامية.

وقامت دراسة أخرى تم تقديمها في (Tahat et al., 2024) بتوظيف نمذجة الموضوعات لتحليل الخطاب الإعلامي للصحافة في دولة الامارات العربية المتحدة. وركزت الدراسة على استراتيجيات القوى الناعمة ودورها في التوجهات الإعلامية. أظهرت النتائج قدرة النماذج المستخرجة في الكشف عن الاستراتيجيات الإعلامية وبوضوح وكفاءة، الامر الذي يتيح اليات فهم وتحليل الخطاب الإعلامي والموضوعات التي تسيطر عليه وتؤثر فيه. واستخدم الباحثون في دراسة تقم تقديمها في (Jannani et al., 2025) تقنيات التعلم التقليدية إضافة الى تقنيات تعلم الالة لأجراء تحليل معمق للعناوين الإخبارية في دولة المغرب. وتم أيضا اعتماد تقنيات نمذجة الموضوعات، وحققت الدراسة نتائج مهمة في دقة التصنيف وتحديد الموضوعات فضلا عن تحليل المشاعر. اما في الدراسة المقدمة من قبل (Dogra et al., 2022)، فتم تطبيق تقنيات الذكاء الاصطناعي لتنفيذ نظام تصنيف شامل للموضوعات عن طريق خوارزميات الشبكات العصبية. وقدمت الدراسة أيضا نتائج جيدة وناجحة في التصنيف واستخراج الموضوعات. وتم في دراسة (Alshanik et al., 2025) تحليل المحتوى النصي واستخراج الموضوعات عن طريق تقنيات التعلم العميق (Deep Learning) للأخبار العربية وأيضا تمكنت الدراسة من استخراج وتحليل الموضوعات بمستوى جيد من الدقة والكفاءة لفهم توجهات الجمهور وتحليل اتجاهات الخطاب الإعلامي. ان الدراسات السابقة التي تم عرضها تظهر بوضوح مستوى الكفاءة والفاعلية الذي يمكن ان تقدمه تقنيات الذكاء الاصطناعي وتعلم الالة في تحليل البيانات النصية وخاصة في سياق الخطاب الإعلامي والاخبار. كما تبرهن الدراسات التي تناولت مختلف المواضيع قدرة التقنيات الذكية على انشاء نماذج تحليل وتصنيف الاخبار العربية بمستوى عال من الدقة والموضوعية. وعلى الرغم من النجاح الذي حققته النتائج في الدراسات السابقة، ما يزال هنالك ضرورة لتطوير نماذج أكثر تطورا وتتناول مختلف التوجهات الإخبارية في اللغة العربية. وهو ما يسعى له بحثنا من خلال تقديم نموذج تحليل الموضوعات لبيانات تم جمعها من الوكالة الوطنية العراقية للأنباء. الامر الذي يساهم في رفد الدراسات السابقة وتقديم نماذج تمتار بتحليل النصوص بيان مختلف الموضوعات والرؤى فيما يخص تحليل الخطاب النصي.

٧. الإطار النظري للبحث (Theoretical Framework)

يرتكز هذا القسم على مجموعة من المفاهيم التي تدعم عملية تحليل البيانات النصية، وبالأخص نمذجة الموضوعات (Topic Modeling) كإحدى أبرز تقنيات معالجة اللغة الطبيعية. وتُعد هذه التقنية جزءًا من توجه علمي أوسع في علم المعلومات يهدف إلى تنظيم النصوص غير المنظمة، واستخلاص الأنماط الدلالية والمعرفية منها.

١-٨ مفهوم تحليل البيانات (Data Analysis)

ان عملية تحليل البيانات هي عملية فحص وتدقيق إجراءات تحويل البيانات من شكلها الخام الى معلومات مفيدة وقابلة للاستخدام في مختلف المجالات. وتشتمل هذه العملية على توظيف مختلف التقنيات الممنهجة والاحصائية والرياضية بهدف إيجاد الأنماط المعرفية والكشف عن العلاقة فيما بين المتغيرات المختلفة. الامر الذي ينتج فهم عميق للبيانات الخام التي تم معالجتها (Albright & Winston, 2020). وبشكل عام تعد عملية تحليل البيانات من أكثر المراحل أهمية في نطاق اتخاذ القرارات في المجالات المختلفة كالتجارة والصحة والتعليم والاعلام وغيرها. حيث يتم الانتقال من مرحلة المعلومات العشوائية الى معرفة واضحة ومنظمة يمكن الاعتماد عليها في اتخاذ القرارات ووضع السياسات وكشف المواطن الخفية في البيانات (Provost & Lange, 2021). وفي ظل التطور التكنولوجي الذي يشهده العالم في وقتنا الحالي، تتزايد أهمية تحليل البيانات بسبب التوسع الهائل في البيانات التي تنتج يوميا وهو ما يصطلح عليه بالبيانات الضخمة (Big Data) (Gandomi & Haider, 2015; Oussous et al., 2018). حيث تشير الإحصاءات العالمية الى ان حجم البيانات ازداد بشكل ملحوظ وكبير في السنوات الأخيرة، الامر الذي يتوجب بدوره مواكبة هذا التوسع بتقديم الطرق والأدوات المناسبة لتحليل هذا الكم الهائل من البيانات والتعامل معها، وبالتالي تحقيق فائدة أكبر من استخدامها (Brown, 2024). وفي ادناه نوضح ماهي الأنواع الرئيسية في سياق تحليل البيانات

أ. التحليل الوصفي (Descriptive Analysis): في هذا النوع من التحليل، يتم توظيف المقاييس الإحصائية البسيطة لغرض القيام بتحليل البيانات، على سبيل المثال المتوسطات الحسابية والانحرافات المعيارية والجداول البيانية. ان الهدف من هكذا أنواع من تحليل البيانات هو لإعطاء نظرة من منطلق شامل واضفاء صورة واضحة عن ماهية البيانات دونما التطرق الى تفسيرات عميقة. ويعد هذا النوع من ابسط أنواع تحليل البيانات وأكثرها شيوعا، ويستخدم بكثرة كمرحلة أولية ضمن سلسلة إجراءات تحليل البيانات. وله فوائد عديدة من ضمنها فهم البنية الأساسية للبيانات وكيفية توزيعها الامر الذي يساهم في اختيار أنماط أخرى من التحليل اعتمادا على النظرة الأولية للبيانات (Thompson, 2009).

ب. التحليل الاستكشافي (Exploratory Analysis): يتمحور هذا النوع من التحليل على استكشاف الأنماط والعلاقات الخفية وغير الظاهرة فيما بين متغيرات البيانات. ويعد أكثر تعقيدا من التحليل الوصفي. ويتم فذا التحليل استخدام نماذج تخدم المبدأ الاستدلالي كأدوات تعلم الالة او التحليل الاحصائي

الاستدلالي، كالمخططات التفاعلية، التحليلات العنقودية، وتحليلات التباين (Abelairas-Etxebarria & Astorkiza, 2020).

ج. التحليل الاستدلالي (Inferential Analysis): ويعد أحد اهم أنواع التحليلات الخاصة بالبيانات، حيث يتجاوز موضوع وصف البيانات الى مواضيع تتعلق باستنتاج العلاقات بين المتغيرات ودراسة مدى إمكانية تعميم النتائج التي تم الحصول عليها على بيانات ونماذج أخرى. ومن بين التقنيات والأدوات المستخدمة في التحليل الاستدلالي هو اختبار الفرضيات، تحليل الانحدار التي تهدف الى تقديم أساس علمي دقيق لاتخاذ القرارات والتوجهات المستقبلية (Fossum & Montoya, 2023). ان الدور الذي تلعبه عملية تحلل البيانات هو دور محوري وخصوصا في عملية بناء القرارات والاستنتاجات المختلفة وذلك عن طريق توفير الأسس العلمية والمنهجية يمكن تبنيها في حل المشكلات المختلفة واختيار البدائل المتوفرة. حيث من خلا الاستخدام السليم لأدوات تحليل البينات يمكن الاستجابة والتفاعل مع مشكلات علمية معينة وتقديم حلول أكثر جدوى وفاعلية. وعليه في بحثنا هذا يتمحور عملنا على جراء تحليل معمق للبيانات النصية الضحمة وتقديم تفسيرا والكشف عن الأنماط الكامنة في هذا الكم الهائل من البيانات عن طريق استخدام أدوات الذكاء الاصطناعي وتعلم الالة الذي يعد من اهم أدوات التحليل للبيانات الضخمة، مما يساهم في تقديم تفسيرا ورؤى مختلفة مرتبطة بالخطاب السياسي لمنصات الاخيار العراقي.

٢-٨ تعلم الالة (Machine Learning)

وهي فرع من فروع الذكاء الاصطناعي يهدف الى بناء وتطوير أنظمة وأدوات لها القابلية على التعلم التلقائي وتحسين الأداء عن طريق تراكم الخبرة والتعلم النتائج من معالجة البيانات الخام (Mahesh, 2020). وتستند تقنيات تعلم الالة على توظيف خوارزميات ونماذج راضية لتحليل البيانات الامر الذي يسهم في الكشف عن الأنماط المخفية وبالتالي استخدام هذه الأنماط في بناء القرارات والتنبؤ بالاتجاهات المستقبلية. وتمتاز اليات تعلم الالة بالقدرة على التعامل مع الكميات الضخمة للبيانات والمعقدة، الامر الذي يصعب بشكل كبير التعامل معه بالأساليب التقليدية، فضلا عن تقليل الوقت والجهد اللازمين لتحليل البيانات الضخمة من خلال اتمتة الإجراءات التحليلية المعقدة. وبشكل عام هنالك ثلاث أنواع رئيسية لتعلم الالة (Roscher et al., 2020):

أ. التعلم المشرف عليه (Supervised Learning): ان هذا النوع من تعلم الالة يشير الى اجراء عملية تدريب النماذج على بيانات لها تسميات معينة مسبقا (Labeled Data). اين ان البيانات الموجودة يتوفر فيها أنماط المدخلات والمخرجات. مثلا تتضمن البيانات الخبر ونوعه. ويتم هنا تدريب النماذج على هذه البيانات ومن ثم اعطاءه بيانات جديدة لم يرها مسبقا لغرض تقييم النتائج ومعرفة مدى دقة التدريب. في الكثير من الأحيان يشار الى هذا النوع من التعلم بتصنيف البيانات. وهو عملية تقسيم البيانات الى تصنيفات محددة مثلا تصنيف الاخبار، التنبؤ بالأسعار والتشخيص الطبي. وهنالك عدد من الخوارزميات التي تستخدم لتنفيذ هذا النوع من تعلم الالة.

ب. التعلم غير المشرف عليه (Unsupervised Learning): على العكس من النوع السابق في التعلم، هذا النوع من التعلم تكون بيانات غير موسومة مسبقا (Unlabeled Data). حيث تقوم النماذج الذكية بمحاولة اكتشاف أنماط وهياكل معينة لتجميع البيانات الى مجاميع تشترك في صفات معينة دونما معرفة مسبقة بالتصنيف الخاص بالبيانات. ومن أشهر تطبيقات هذا النوع من تعلم الالة هو التجميع (Clustering)، ونمذجة الموضوعات، حيث يتم انشاء عناقيد معينة ووضع البينات المتشابه في نفس المجموعة.

ج. التعلم شبه المشرف (Semi-supervised Learning): هو نوع من تعلم الالة يمثل حالة وسطية بين النوعين المذكورين أعلاه. حيث تكون البيانات موسومة جزئيا. وفي الغالي تكون البيانات الموسومة اقل من البيانات غير الموسومة بشكل كبير. وتستخدم هذه النماذج من التعلم عندما تكون البيانات المتوفرة ليست موسومة بشكل كامل وهو ما قد يتم مواجهته في البيانات النصية. وتتنوع خوارزميات تعلم الالة بتنوع الغرض من استخداماتها حيث من أشهر الأمثلة على هذه الخوارزميات هي شجرات القرار (Decision Trees)، الة المتجهات الداعمة (Support Vector Machine) وغيرها من الخوارزميات. وبغية التحقق من أداء الخوارزميات المختلفة هنالك عدد من المعايير يتم استخدامها لتقييم أداء النماذج وتقييم النتائج وكما يلي:

أ. الدقة (Accuracy): يتم في هذا المعيار قياس نسبة التنبؤات الصحيحة من اجمالي التنبؤات التي اجراها النموذج، أي مدى صحة توقعات النموذج بشكل عام، ويتم ذلك على وفق المعادلة التالية:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

ب. الدقة التنبؤية (Precision): وهو المعيار الذي يقيس قدرة النموذج على تحديد النتائج الصحيحة من بين النتائج التي توقعها إيجابيا. أي هنا يتم التركيز على مسالة التوقع الإيجابي، أي جودة توقع الحالات الإيجابية، وكما في المعادلة التالية:

$$Precision = \frac{TP}{TP + FP}$$

ج. الدقة التذكيرية (Recall): هنا يتم تقييم قدرة النموذج على تحديد جميع الحالات الإيجابية الفعلية بشكل صحيح، أي قدرة النموذج على اكتشاف الحالات الإيجابية، ويتم ذلك على وفق المعادلة التالية:

$$Recall = \frac{TP}{TP + FN}$$

د. معامل (F1): هنا يتم الدمج بين الدقة التنبؤية والدقة التذكيرية في نفس المعيار، أي متوسط توافقي بين الدقة التنبؤية والتذكيرية. وهذه الحالة مفيدة جدا في حالة وجود تباين واضح في توزيع البيانات وكما في المعادلة التالية:

$$F1 = \frac{Precision * Recall}{Precision + Recall} * 2$$

حيث:

- TP: تشير التوقع الإيجابي الحقيقي (True Positive)، وتعني الحالة التي يتوقعها النموذج بشكل صحيح إيجابي وهي بالفعل إيجابية. على سبيل المثال ان يكون الخبر المراد تصنيفه هو خبر سياسي ويتم التنبؤ به على انه خبر سياسي بالفعل. أي
- TN : التوقع السلبي الحقيقي (True Negative)، وفي هذه الحالة يتوقع النموذج حالة معينة انها سلبيه وتكون بالفعل هي سلبية، مثلا يكون الخبر غير سياسي وبالفعل يتم التوقع على انه خبر غير سياسي كأن يكون خبر رياضي مثلا.
- FP: التوقع الإيجابي الخاطئ (False Positive)، هي الحالية التي يتوقعها النموذج على انها إيجابية ولكنها في الواقع سلبية، على سبيل المثال ان يتوقع النموذج خبرا رياضيا على انه سياسي.
- FN: التوقع السلبي الخاطئ (False Negative)، وهي ان يتوقع النموذج حالة معينة على انها سلبية ولكنها في الواقع إيجابية، مثلا ان يكون الخبر سياسي ولكن النموذج يتوقعه رياضي مثلا.

٣-٨ معالجة اللغات الطبيعية (Natural Language Processing - NLP)

وهي فرع من فروع الذكاء الاصطناعي يهدف الى تمكن الالة (الحاسوب) من فهم ومعالجة وتفسير اللغات بطريقة تحاكي قدرة البشر على عمل ذلك (Khurana et al., 2023). تستخدم هذه التقنيات في تحويل النصوص المكتوبة الى صيغ يمكن للآلة التعامل معها وتحليلها من قبل الخوارزميات الذكية. وتبرز أهمية تقنيات معالجة اللغات الطبيعية من خلال اتساع رقعة تطبيقاتها كتصنيف النصوص وأنظمة الإجابة الالية وتحليل المحتوى النصي وغيرها. وتشمل عملية معالجة اللغات الطبيعة عدد من المراحل الأساسية التي تهدف الى تجهيز النص بشكل ملائم للآلة ليتم تحليله لاحقا وكما يلي (Chanda & Pal, 2023):

أ. تقطيع النصوص (Tokenization): وهي المرحلة التي يجري خلالها تحويل الجمل والفقرات النصية الى وحدات صغيرة تسمى (Tokens). والتي قد تكون كلمات او جمل قصيرة. ان هذه المرحلة تساهم في تهيأة النصوص للمراحل اللاحقة من المعالجة على سبيل يتم تقطيع الجملة التالية: "اجتماع جامعة الدول

العربية في بغداد"، حيث تعمل الية تقطيع النصوص الى تحويل الجملة أعلاه الى مصفوفة كلمات بالشكل التالي: ["اجتماع"، "جامعة"، "الدول"، "العربية"، "في"، "بغداد"].

ب. إزالة كلمات التوقف (Stop Words Removal): وهي المرحلة التي يتم من خلال التخلص من الكلمات الشائعة في النصوص. حيث ان هذه الكلمات لا تحمل قيمة معنوية ودلالية في النص ووجودها لا يخدم التحليل كونها متواجدة في الغالبية العظمى في النصوص ولا تميز نصا عن اخر. وتساهم هذه العملية في تقليل حجم البيانات المراد معالجها. وتشمل كلمات التوقف الضمائر وحروف الجر وبعض العلامات الاملائية مثلا علامة التعجب او علامة السؤال والنقاط والفارزة وغيرها. ان وجود هذه النصوص لا يشكل فرقا معنويا في النص ولذلك فان حذفها يساهم في تقليل البيانات من جهة وتحسين سرعة المعالجة من جهة أخرى.

ج. التجذير (Stemming): في الكثير من الأحيان يتم اللجوء الى التجذير في معالجة النصوص، وتشير هذه لعملية الى تحويل الكلمات المراد معالجتها الى جذورها الاصلية. ان هذه العملية مفيدة جدا في تقليل التباين في الكلمات مما يساهم في زيادة دقة تحليل النماذج النصية. على سبيل المثال الكلمات ["يستخدم"، "مستخدم"، "يستخدمون"، "مستخدمين"، "استخدم"] يمكن ارجاعها الى جذر واحد وعلى سبي المثال يتم اعادتها الى كلمة "استخدام". حيث في هذه الحالة يتم التخلص من الاشكال المتنوعة للكلمات ذات الدلالة الواحدة. ولكن في بعض الأحيان قد يكون الجذر غير دقيق لغويا في بعض اللغات ومن بينها اللغة العربية. لذلك يجب التعامل بدقة من أدوات التجذير للنصوص (Jaafar et al., 2017).

د. تمثيل النصوص رقميًا (Vectorization and Embeddings): بغية تنفيذ اليات وخوارزميات تحليل النصوص من خلال أدوات تعلم الالة، يجب ان يتم تحويل النصوص الى صيغ تفهمها هذه الخوارزميات. وبطبيعة الحالة ان جميع خوارزميات تعلم الالة لا تتعامل مع النصوص بشكلها الأصلي ولكن يجب ان يتم تحويلها الى تمثيلات رقمية معينة. ومن أشهر هذه التقنيات حقيبة الكلمات (Bag of Words)، التي تستخدم لتوليد اوزان للكلمات على وفق أسس معينة وبالتالي انشاء متجه لخزن هذه الاوزان. وكذلك أيضا توجد تقنية أخرى مشهورة جدة تسمى تضمين الكلمات (Word Embeddings)، حيث يتم تمثيل الكلمات في فضاء واسع من المتجهات المستمرة الامر الذي يعكس دلالات الكلمات حسب وجودها في سياقاتها النصية. وبحسب الدراسات السابقة، يتم استخدام تقنيات معالجة اللغات الطبيعية في تحليل النصوص الخاصة بالخطاب الإعلامي، حيث يستخدمها الباحثين في بناء نماذج ذكية لتحليل المحتوى الاخباري. وفي نفس السياق تعد نمذجة الموضوعات التي هي موضوع بحثنا الحالي، أحد اهم التقنيات المستخدمة في تحليل النصوص الإخبارية والخروج

بنتائج ومعالم كامنة من شأنها إعطاء مختلف التفسيرات والمعلومات حول اتجاهات الخطاب الإعلامي لا سيماء السياسي العراقي. شكل رقم (١) يوضح اهم مراحل معالجة اللغات الطبيعية.

شكل (1): مراحل معالجة اللغات الطبيعية

٨-٣-١ تحديات معالجة اللغة العربية:ان اللغة العربية تعد من اللغات التي لها خصائص فريدة ومعقدة مما يشكل تحديا كبيرا امام الباحثين وخصوصا في معالجتها آليا. ان تعقيد اللغة العربية يبرز من طبيعتها الصرفية والنحوية المعقدة والتي تختلف عن الكثير من اللغات الحية الأخرى. حيث تمتاز اللغة العربية بكونها اشتقاقية تسمح بتكوين عدد غير قليل من الكلمات من نفس الجذر اللغوي الامر الذي يزيد من صعوبة تحديد معاني الكلمات اليا بدقة (Guellil et al., 2021). وكذلك من اهم التحديات التي تواجه معالجة اللغة العربية هي صعوبة التجهيز الاولي للنصوص العربية. حيث تتنوع اشكال الحروف في اللغة العربية حسب موقعها من الكلمة، أي ان اغلب الحروف العربية لها ثلاثة اشكال كتابية، حيث يختلف شكل الحرف عند وضعه في بداية، وسط، او نهاية الكلمة. إضافة لذلك اختلاف طرق كتابة بعض الحروف مثل الهمزة ولهاء والتاء المربوطة وغيرها. كل هذه الأمور تجعل من اللغة العربية ليست سهلة فيما يخص معالجتها اليا. وبالرغم ن تعقيدات اللغة العربية ونتيجة لأهمية اللغة وسعة رقعة انتشارها على الانترنت كأحد اهم اللغات الحية عالميا، ظهرت في السنوات الأخيرة عددا من التقنيات الحديثة والذكية والمخصصة لمعالجة اللغة العربية. ومن أبرز هذه النماذج وأكثرها انتشارا هي نموذج (BERTopic). حيث يعتمد هذا النموذج على خوارزميات التعلم العميق والمحولات (Transformers) لمبنية على قواعد بيانات نصية كبيرة جدا من اللغة العربية الفصحى (El-Alami et al., 2022). ويستخدم هذا النموذج لأداء العديد من المهام المتعلقة بمعالجة اللغة العربية كتحليل البيانات وتصنيفها. وهذه النماذج تقدم حلولا مبتكرة لمواجهة تحديثات اللغة العربية واظهرت تفوقا واضحا في الأبحاث ذات الصلة. وعليه يتم في بحثنا هذا اعتماد هذه النماذج لغرض تنفيذ الجانب العملي للبحث.

٨-٤ نمذجة الموضوعات (Topic Modeling)

وهي أحد تقنيات معالجة اللغات الطبيعية الهدف منها هو الكشف عن المواضيع الخفية والكامنة او الأنماط الموضوعية داخل مجموعة كبيرة من البيانات النصية. وفي هذه التقنيات يتم تصنيف الكلمات ضمن مجموعات من الموضوعات المتصلة. وتعتمد هذه التقنية على فرضية ان النصوص تتكون بطبيعتها من خليط من الموضوعات وكل موضوع بدوره يحوي كلمات ذات صلة متقاربة بشكل كبير (Weston et al., 2023). وفي سياق النصوص الإعلامية، فيمكن الاستفادة بشكل كبير من نمذجة الموضوعات في الكشف عن الموضوعات الأكثر تداولا وكيف تتغير هذه الموضوعات بمرور الزمن، أي ان هذه التقنيات تمتلك القدرة على تلخيص الاتجاهات الإعلامية بشكل جيد. مما يساعد على توفير فهم أعمق للمحتوى الإعلامي من خلال تحليل عميق للنصوص الإعلامية الضخمة. ومن بين الخوارزميات التي تعمل على نمذجة الموضوع، نموذج (BERTopic)، التي تستخدم لتحليل المحتوى النصي العربي بكفاءة ودقة، وتم استخدامها في العديد من البحوث ذات الصلة. وتعمل هذه الخوارزمية على انشاء تضمينات عصبية (Embeddings) وتتميز بقدرتها على التعامل مع النصوص القصيرة والمعقدة. ويتم تقييم نماذج الموضوعات باعتماد عدد من المقاييس مثلا درجة التماسك (Coherence Score) التي تقيس مدى الترابط المعنوي بين الكلمات داخل الموضوع الواحد وكلما زادت هذه القيمة زاد الترابط بين الكلمات. كذلك هنالك مقياس يعرف بمقياس الحيرة (Perplexity) الذي يقيس مدى قدرة النموذج على التنبؤ بكلمات جديدة في النص، وكلما قلت قيمة الحيرة كانت النتائج أفضل كون هذا الموضوع يدل على نسبة تأكد اعلى في التنبؤ (Chehal et al., 2021).

٨. الإطار العملي (Practical Framework)

يرتكز هذا القسم على عرض جميع الإجراءات والمفاهيم الخاصة بالجانب العملي للبحث وكل ما يتعلق به وكما يلي:

٩-١ الأدوات والتقنيات المستخدمة في البحث

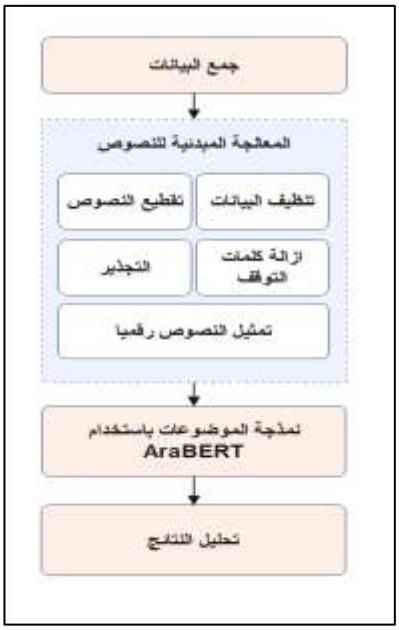
أ. لغة البرمجة: لتحقيق الأهداف المتوخاة من البحث، تم الاعتماد على لغة برمجة بايثون (Python) بالنظر لكفاءتها العالية وانشار استخدامها في تحليل البيانات ومعالجتها وخصوصا عند التعامل مع أدوات وتقنيات الذكاء الاصطناعي ومعالجة اللغات الطبيعية.

ب. المكتبات الخاصة بمعالجة البيانات: تم الاستعانة بعد من الأدوات والمكتبات التي ترتبط بالجانب العملي والبرمجي للبحث. حيث تم الاستعادة بمكتبات (Pandas) التي تعد من المكتبات الأساسية التي تستخدم مع لغة برمجة بايثون حيث توفر قدرات عالية في التعامل مع مجاميع البيانات والمصفوفات. وكذلك تم الاستعادة بمكتبات (Numpy) وهي متخصصة بتنفيذ العمليات الرياضية والاحصائية المعقدة على البيانات وتجهيز البيانات للمهام التحليلية.

ج. مكتبات معالجة اللغات الطبيعية: تم الاستعانة بهذه المكتبات لغرض تنفيذ اعمال محددة لمعالجة النصوص ومن بين المكتبات مكتبة (NLTK) المتخصصة بتقطيع النصوص واجراء عمليات التجذير

والمعالجة المبدئية للنصوص واستخراج الخصائص اللغوية. وكذلك مكتبة (SpaCy)التي تعمل على استخراج الكيانات والسياقات النصية وتحليلها على نحو كفوء

د. مكتبات نمذجة الموضوعات: تم استخدام مكتبة (BERTopic) المتخصصة في نمذجة الموضوعات الخاصة باللغة العربية. حيث تعمل هذه المكتبة على استخدام تضمينات النصوص ولها قدرة عالية على استخلاص الموضوعات المتماسكة. وهي تحول الكلمات إلى تمثيلات رقمية مستمرة تعبر عن العلاقات الدلالية بين الكلمات بشكل فعّال ودقيق، وتساهم بشكل واضح في تحسين جودة النماذج المستخدمة.



شكل (٢): منهجية البحث

٢-٩ منهجية البحث (Methodology)

في هذا القسم يتم توضيح المنهجية المتبعة لتحقيق أهداف البحث والإجابة على أسئلته، وذلك عبر استخدام تقنيات تحليل البيانات ومعالجة اللغة الطبيعية (NLP)، وتطبيق نمذجة الموضوعات باستخدام نموذج BERTopic. ستشمل هذه المنهجية المراحل التالية والتي يوضحها شكل رقم (٢):

١-٢-٩ جمع البيانات (Data Collection)

تم جمع بيانات البحث من خلال مجموعة البيانات الخاصة من وكالة أنباء عراقية تُعرف باسم الوكالة الوطنية العراقية للأنباء (NINA, 2023). تُعد هذه الوكالة واحدة من أبرز وكالات الأنباء في العراق. في تشرين الأول من العام ٢٠٠٦، تم إطلاق أخبار الوكالة باللغتين الإنجليزية والعربية. وانضمت الوكالة إلى

منظمة وكالات أنباء آسيا والمحيط الهادئ (OANA) في تشرين الثاني من العام ٢٠١٠. ومنذ ذلك الحين، غطت الوكالة مجموعة واسعة من الأخبار وقدمتها للجمهور. التم جمع مجموعة البيانات باستخدام تقنيات الكشط (Scarping)، وهي تقنيات تسمح باستخدام الأدوات البرمجية لجمع عدد من البيانات التي تكون متاحة على الانترنت بسرعة وسهولة. تضمنت المقالات الإخبارية التي تم جمعها الأخبار المنشورة في الفترة (١ كانون الثاني ٢٠١٦) حتى (٣١ كانون الاول ٢٠٢٣). تجاوز مجموع البيانات التي تم جمعها (١٠٠) ألف خبر من الأخبار المكتوبة باللغة العربية وموزعة على أربع فئات إخبارية، وهي (السياسة، الأمن، الاقتصاد، والرياضة). يختلف عدد الأخبار بحسب كل فئة، حيث تعد الأخبار السياسية الجزء الأكثر أهمية في مجموعة البيانات، تليها أخبار الأمن والاقتصاد والرياضة.

يعود هيمنة الأخبار السياسية والأمنية إلى الأوضاع التي شهدتها البلاد في العقود الأخيرة. فقد مر العراق بتقلبات سياسية وأمنية جعلت هذه الأخبار تتصدر قوائم الأخبار. ونظرا لهيمنة الاخبار السياسية على المشهد السياسي في العراق، فقط تم اختيار هذه الفئة الإخبارية لغرض اجراء التحليل لها وبيانات كيف تغيرت مع مرور الزمن. وبلغ عدد الاخبار السياسية (٤١٢٣٧) خبرا سيتم اجراء التحليل عليها جميعها. وتشمل مجموعة البيانات على المواصفات التالية:

- حقل العنوان: وهو العنوان الرئيسي للخبر
- حقل المحتوى النصي، ويشمل النص الفعلي للخبر.
- حقل التاريخ، ويشمل التاريخ الذي تم نشر الخبر به، ويشمل اسم اليوم والتاريخ والوقت.

٩-٢-٢ المعالجة المبدئية للبيانات النصية (Preprocessing)

تعد هذه المرحلة من اهم المراحل في أي نظام يعتمد على معالجة النصوص وتحليلها. وتهدف هذه المرحلة على اجراء سلسلة من المعالجات على البيانات النصية الخام بغية تحويلها الى صيغ يمكن للخوارزميات التعامل معها. ولهذه المرحلة أهمية خاصة أيضا فميا يخص معالجة النصوص العربية نظرا للتركيب اللغوي والنحوي المعقد للغة العربية. وفي بحثنا الحالي تم تنفيذ عدد من عمليات المعالجة المبدئية للبيانات وكما يلي:

- تنظيف النصوص: وفي هذه المرحلة تم التخلص من النصوص غير الضرورية والتي لا تشكل فرقا معنويا في المعنى كالروابط الالكترونية والأرقام والحروف غير العربية وعلامات الترقيم الزائدة. وبعدها تم أيضا إزالة العلامات الاعرابية ان وجدت.
- ازالة كلمات التوقف، حيث تم ازالة كلمات التوقف من جميع البيانات النصية، حيث ان هذه الكلمات لا تحمل دلالات معينة تميز نصا عن الاخر بل هي موجودة في كل النصوص مثل أسماء الإشارة والاسماء الموصولة، وتوفر المكتبات البرمجية قوائم مخصصة لكلمات التوقف يمكن ازالتها من النصوص بكل سهولة.

- التجذير، حث يتم إعادة الكلمات الى جذورها بغية توحيد الكلمات متعددة الشكل والتي تكون متشابهة في المعنى والدلالة استنادا لقواعد اللغة العربية، واضا توفر المكتبات البرمجية عدد كبير من الأدوات التي تسمح بإنجاز هذه المهمة بسهولة.
- تمثيل النصوص رقميا، حيث بعد تنفيذ عمليات التصفية على البيانات النصية والبقاء على الكلمات المراد تحليلها، يجري استخدام مكتبات برمجية تقوم بتحويل الكلمات النصية الى تضمينات رقمية يمكن بسهولة ارسالها الى الخوارزميات بغية تحليلها بشكل أعمق.

ان الإجراءات أعلاه والمتعلقة بالمعالجة المبدئية للبيانات النصية تساهم في تقليل البيانات من حيث استبعاد البيانات غير الضرورية والإبقاء على البيانات ذات المعنوية العالية والتي تؤثر الى حد كبير في استخراج الموضوعات والانماط الكامنة، حيث تؤثر المعالجة المبدئية بشكل مباشر في نتائج التحليل والتصنيف التي يتم تطبيقها بعد هذه المرحلة.

٩-٢-٣ نمذجة الموضوعات (Topic Modeling)

تمثّل نمذجة الموضوعات أحد أكثر الأساليب فعالية في استخراج المعرفة من البيانات النصية غير المنظمة، حيث تُستخدم لتحديد البنية الكامنة للمواضيع التي تتكرر ضمن مجموعة نصوص كبيرة. وفي هذا البحث، تم اعتماد نموذج BERTopic كإطار أساسي لتنفيذ هذه المرحلة، لما يتمتع به من دقة ومرونة في التعامل مع النصوص باللغة العربية، لاسيما عند دمجه مع نماذج التعلم العميق المدربة مسبقًا.

تبدأ عملية نمذجة الموضوعات بمعالجة مبدئية للنصوص الإخبارية، شملت تنظيف البيانات، إزالة كلمات التوقف، والتجذير الجزئي، بالإضافة لاستخراج تمثيلات عددية (Embeddings) تعكس المعاني العميقة للنصوص. بعد ذلك، تم استخدام ادوات لتقليل أبعاد التمثيلات العددية، مما يسمح بتجميع النصوص المتقاربة في فضاء دلالي مشترك. ثم تم تجميع النصوص ضمن موضوعات مترابطة، بناءً على الكثافة الدلالية. ان ميزة نموذج (BERTopic) تكمن في اعتماده على المعالجة السياقية للنصوص، حيث لا يتم تحليل تكرار الكلمات فقط، بل أيضًا السياق الذي ترد فيه، مما أتاح اكتشاف موضوعات أكثر دقة وواقعية. بعد ذلك، تم استخراج الكلمات المفتاحية لكل موضوع، ثم تحليلها زمنيًا ودلاليًا لرصد التحولات في الخطاب السياسي والإعلامي العراقي. ساهمت هذه المنهجية في تحويل مئات الآلاف من الكلمات إلى هيكل موضوعي قابل للتحليل، مما وفّر للبحث إطارًا معرفيًا واضحًا لفهم محتوى الأخبار السياسية من منظور علمي دقيق يعزز من قيمة التحليل المعلوماتي ضمن حقل علم المعلومات والمعرفة.

٩. تحليل الاخبار السياسية :في هذا القسم نقوم بعرض النتائج التي تم الحصول عليها بعد تنفيذ نمذجة الموضوعات وكما يلي:

١٠-١ عرض اهم الموضوعات :تم تنفيذ عملية استخراج الموضوعات باستخدام نموذج BERTopic، الذي يُعد من أحدث النماذج في تحليل البيانات النصية. ويتميّز هذا النموذج بدمجه بين التمثيلات العددية

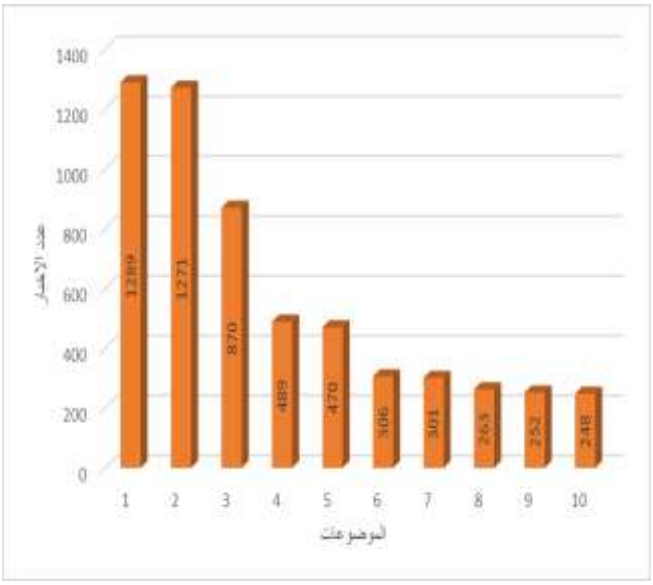
السياقية للنصوص وتقنيات غير مشرفة لتجميع النصوص وتحليلها. بدأت العملية باستخراج التمثيلات العددية من النصوص السياسية، وذلك بهدف التقاط المعاني السياقية الدقيقة للكلمات ضمن السياق العربي. حيث تم تحويل كل فقرة نصية إلى تمثيل عددي. تم تجميع النصوص في مجموعات موضوعية بناءً على كثافة التوزيع للكلمات المتشابهة في الدلالة، حيث تم استخراج الكلمات الأكثر تكرارًا وتميزا داخل كل مجموعة، مكونا بذلك كلمات مفتاحية تعبر عن جوهر كل موضوع. عليه، فإن كل موضوع مكتشف يمثل مجموعة من النصوص السياسية التي تشترك في مفاهيم دلالية وسياقية مشتركة، وقد تم تصنيف كل نص بناءً على انتمائه لأحد هذه الموضوعات. بعد تنفيذ نمذجة الموضوعات على البيانات التي تم معالجتها مبدئيا، وذلك باستخدام نموذج BERTopic لنمذجة الموضوعات، حيث تم استخراج عدد كبير من الموضوعات السياسية التي مثلت توجهات الخطاب الإعلامي للوكالة الإخبارية المستخدمة في البحث ولجميع الاخبار السيارية الممتدة للفترة من العام ٢٠١٦ ولغاية ٢٠٢٣. حيث تم تجميع النصوص المتشابهة في تمثيلاتها الرقمية الى مجموعات بناءا على التشابه في التمثيل العددي لكل منها. وبينت النتائج الأولية لنمذجة الموضوعات ان توزيع الموضوعات كان غير متوازنا مما يعكس الطبيعة الديناميكية والمتغير للخطاب الإعلامي في المنطقة. ومن بين الموضوعات التي تم استخراجها، تم اختيار عشرة موضوعات مثلت الموضوعات الأعلى تكرارا في البيانات النصية وكما هو موضح في جدول رقم (١)، الذي يوضح الموضوعات المستخرجة وعدد الأخبار لكل موضوع فضلا عن الكلمات المفتاحية للموضوعات.

جدول (1): توزيع اعلى عشر موضوعات مستخرجة من الاخبار السياسية

الموضوع	عدد الاخبار	الكلمات المفتاحية	التسمية المقترحة
1	1289	النجف، الصدر، الحكومة، الصدري، المحافظة، التيار، السيد	التيار الصدري
2	1271	التركي، اسطنبول، رجب، طيب، اردوغان، أنقرة، أوغلو	تركيا والسياسة الخارجية
3	870	روسيا، بوتين، موسكو، أوكرانيا، فلاديمير، لافروف، سوريا، الرئيس	روسيا والتحركات الدولية
4	489	صالح، برهم، الجمهورية، رئاسي، رئيس، ضرورة، أهمية، المنطقة،	رئاسة الجمهورية
5	470	الموازنة، المالية، اللجنة، العامة، دينار، قانون، رواتب،	البرلمان

		سعر، العجز	والتشريعات
6	306	المدنيين، الهجوم، استهدف، استشهاد، التفجير، واصابة، الضحايا، داعش	الإدارة المحلية
7	301	ساحة، المتظاهرين، وسط، التحرير، الخضراء، المحتجين، التظاهرة	تظاهرات
8	263	المهدي، المرشحين، لمنصب، مرشح، عبد، عادل، الوزارية، الكتل	انتخابات
9	252	الاصلاح، تحالف، السياسية، الكتلة، الدولة، العبادي، تشكيل، الفساد	السياسات والتنقلات
10	248	النووي، طهران، الذرية، إيران، فيينا، المفاوضات، اليورانيوم	المفاوضات

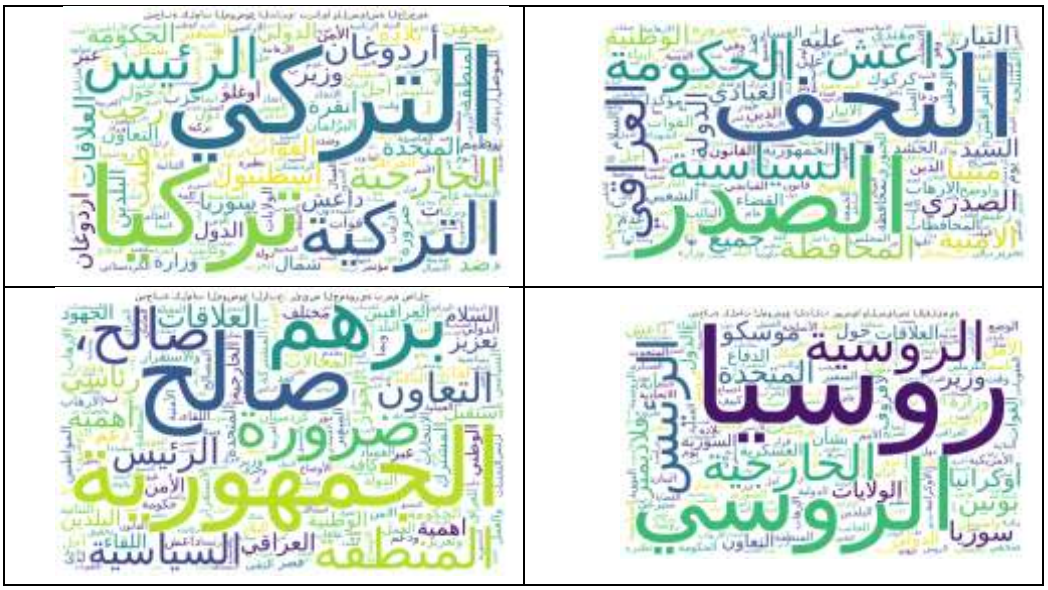
ويتبين من الجدول أعلاه ان الموضوع الأول والذي شمل الكلمات المفتاحية (النجف، الصدر، الحكومة، الصدري، المحافظة، التيار، السيد) تصدر المشهد الإخباري وشكل اعلى تكرارات للمقالات الإخبارية خلال فترة بيانات البحث. اما المواضيع التي تخص (النووي، طهران، الذرية، إيران، فيينا، المفاوضات، اليورانيوم)، فقد تذيلت المشهد الاخباري للسنوات ٢٠١٦-٢٠٢٣. الشكل رقم (٣) يوضح توزيع الاخبار بحسب الموضوعات المستخرجة.



شكل (2): توزيع الاخبار حسب الموضوعات المستخرجة

ولتوضيح المواضيع والكلمات التابعة لها، يمكن الاستعانة بالأدوات البرمجية لرسم سحابة الكلمات (Word Cloud)، وهي من الأدوات البصرية الفعالة في نمذجة الموضوعات وخصوصا مع البيانات الضحمة. وتكمن أهمية هذه التقنية في قدرتها على تقديم تمثيل بصري سريع للكلمات الأكثر تكرارا في مجموعة نصوص معينة. ويتم إعطاء حجم نص للكلمة اعتمادا على عدد تكراراتها.. فالكلمة ذات التكرارات الكثيرة تظهر أكبر من غيرها. مما يعطي فهما بصريا لتوزيع الكلمات داخل الموضوعات. والشكل رقم (٤) يمثل رسم سحابة الكلمات للمواضيع الأربع الأولى.

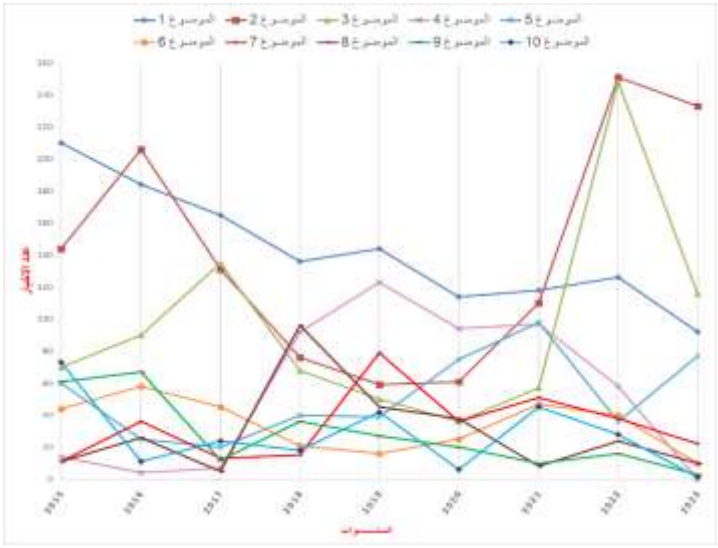
شكل (3): سحابة الكلمات للمواضيع الأربعة الاولى



٢-١٠ الاتجاه الزمني لأهم الموضوعات

تم اجراء تحليل زمني لتطور الموضوعات العشر التي تم استخراجها، واظهر التحليل الزمني للموضوعات السياسية في الخطاب الإعلامي العراق ان المواضيع شهدت تباينا ملحوظا للأعوام ٢٠١٥-٢٠٢٣، وكما هو موضح في الشكل رقم (٥). الذي يبين كيف تتباين اعداد الاخبار سنويا. وهذا يدل على ان الخطاب الإعلامي يركز في مراحل معينة وفترات محددة على احداث مهمة قد تكون محلية او دولية تعكس تطورات ميدانية معينة. فعلى سبي المثال الموضوع الأول والمرتبط بالمصطلحات (النجف، الصدر، الحكومة، الصدري، المحافظة، التيار، السيد) نلاحظ انه يرتبط بشكل وثيق بمدينة النجف والتيار الصدري بشكل عام ويتبين ذلك من خلال كلمات الموضوع المفتاحية، وهذا يشير الى ان هنالك تغطية إعلامية للحراك السياسي والديني في المدينة، ويظهر التحليل انه خلال الأعوام ٢٠٢٢،٢٠١٩،٢٠١٦ شهدت ارتفاعا ملحوظا في عدد الاخبار حث ارتبط هذه السنوات بالمظاهرات والانتخابات وهي محطات سياسية حاسمة في المشهد العراقي. نرى ان هنالك تغطية إعلامية متفاوتة الوتيرة مع استمراراه في جميع الأعوام كالأمور المتعلقة بالموازنة والرواتب وكما هو في الموضوع الخامس (الموازنة، المالية، اللجنة، العامة، دينار، قانون، رواتب، سعر، العجز). حيث ان

اغلب المواضيع المحلية كانت تغطيها متواجدة عبر السنوات مع تفاوت نسبي. شكل (4): تباين المواضيع السياسية العراقية زمنيا



شكل (5): تباين المواضيع السياسية العراقية زمنيا

اما فيما يخص الاحداث العالمية، فنرى هنالك طفرات في اعداد الاخبار وحسب سنوات معينة فلو نلاحظ ان الموضوع الثاني والذي يرتبط المصطلحات (التركي، اسطنبول، رجب، طيب، اردوغان، أنقرة، أوغلو) والتي تتمحور حول تركيا. نلاحظ ان نتائج التحليل توضح ان هذا الموضوع مرتبط بتركيا وزعامتها، الامر الذي يشير الى ظهور هذه المصطلحات في الخطاب الإعلامي على مدى السنوات ٢٠١٥-٢٠٢٣. حيث نرى ان هنالك، ونرى انه هنالك طفرات في الشكل رقم (٥) للسنوات ٢٠١٧، ٢٠٢٠، و٢٠٢٢، وهي سنوات شهدت احداث محورية على صعيد الجانب السياسي والتطورات على الحدود العراقية التركية. وكذلك السنوات التي تجري بها الانتخابات التركية. وكذاك الحال فيما يخص الموضوع الرابع (روسيا، بوتين، موسكو، أوكرانيا، فلاديمير، لافروف، سوريا، الرئيس) والذي يعنى بالشأن الروسي، حيث نلاحظ طفرات في أوقات معينة تعكس تطور الاحداث في هذا الموضوع وهو ما يشير الى الاحداث السياسية بين أوكرانيا وروسيا والحرب الدائرة بين البلدين. والشكل رقم (٦) يبين هذا التباين في الحركة الديناميكية للأحداث المحلية مقارنة بالأحداث لعالمية. ونلاحظ ان نمذجة الموضوعات اتاحت إمكانية التعرف على توزيعات المواضيع سنويا، وتتبع تطور الاحداث ديناميكيا، الامر الذي يوفر اطارا معرفيا لفهم اهم محاور الاخبار السياسية العراقية لفترة معينة.

شكل (6): اتجاهات الاحداث المحلية والعالمية

٣-١٠ تحليل التحوّل الدلالي الموضوعي للمحتوى الإعلامي

يمثل التحليل الدلالي أحد الأساليب المتقدمة في تحليل المحتوى النصي، ويُستخدم لفهم كيفية تغير المعنى والسياق المرتبط بمحتوى معين على امتداد فترة زمنية. ويكتسب هذا النوع من التحليل أهمية خاصة في الدراسات الإعلامية والسياسية، حيث تُظهر اللغة المستخدمة في التغطيات الإخبارية تطورات الخطاب السياسي، وتحوّلات الرأي العام، وتغيّر أولويات الفاعلين السياسيين أو الإعلاميين. في هذا البحث، تم تطبيق التحليل الدلالي على المواضيع المستخرجة عبر من البيانات التي تم جمعها والتي تغطي الفترة بين ٢٠١٥ و٢٠٢٣. من خلال هذا التحليل، سعينا إلى تتبع تطور الخطاب السياسي والإعلامي في المواضيع الأكثر تناولًا، وفهم التحول في المفردات والدلالات المرتبطة بكل موضوع بمرور الوقت، مما يساهم في بناء صورة أكثر عمقًا للمشهد السياسي وتفاعلاته الإعلامية. وفي ادناه عرض لتحليل التحول الدلالي للموضوعات الخمسة الأولى.

أ. الموضوع الأول (التيار الصدري)

يظهر هذا الموضوع تحولا دلاليا في الخطاب السياسي والإعلامي للأمور المرتبطة بالموضوع، حيث من خلال الكلمات المفتاحية يتبين ان الموضوع يغلب عليه الطابع الاخباري فيما يخص التيار لصدري وزعامته. والجدول رقم (٢) يمثل تحليلا لتغير الخطاب السياسي.

جدول (2): تحليل التحول الدلالي للموضوع الأول (التيار الصدري)

الفترة الزمنية	الكلمات البارزة	التحليل	وصف الخطاب الدلالي
2015 – 2016	المرجعية، الجمعة، النجف، السيد،	اتسم الخطاب خلال هذه المرحلة بصبغة دينية واضحة، حيث طغت مفردات تدل على الأنشطة الدينية والخطب والصلاة، وتمحورت التغطيات حول المرجعية الدينية في النجف وأدوارها الروحية والمجتمعية.	خطاب ديني ومحلي
2017 – 2018	التيار الصدري، المتظاهرين، الفساد، خدمات	بدأت النبرة تتحول نحو التركيز على القضايا الخدمية والاجتماعية، وأصبح الخطاب الصدري يحمل دلالات سياسية إصلاحية، مع تصاعد لهجة نقد الفساد والدعوة إلى التغيير،	الخطاب الاجتماعي والسياسي
2019 – 2020	تظاهرات، ساحة، المحتجين، إصلاح، بغداد، الأحزاب	في هذه الفترة، يتغير الخطاب جذريًا نحو الحراك الشعبي، وتصبح مفردات الثورة والاعتصام والاحتجاج محورية	تصاعد الحراك الشعبي
2021 – 2022	مرشحين، الكتلة، الانتخابات، انسحاب، الأغلبية	تحول الموضوع إلى خطاب انتخابي، وبرزت مفردات تدل على الاستعداد لخوض الانتخابات أو الانسحاب منها، مما يعكس موقف التيار الصدري داخل البرلمان والساحة السياسية	الخطاب الانتخابي والسياسي
2023	اعتكاف، انسحاب، الصدر، السيد، التيار، أنصار	تتزايد مفردات الإصلاح والاحتجاج مع خطاب يوحي بالانسحاب والانكفاء، ما يعكس ربما أزمة داخلية أو إعادة تموضع سياسي بعد التغيرات الكبرى في التوازنات العراقية.	خطاب الأزمة والانكفاء

ويظهر الجدول أعلاه ان التحليل الدلالي لأخبار الخاصة بموضوع التيار الصدري يمر بخمسة مراحل دلالية، كما هي موضحة في شكل (٧)، وهو وهذا يشير الى ديناميكية عالية في الخطاب الإعلامي وتصدره المشهد العراقي لعدة سنوات، وهو ما يعكس تفاعل الشارع مع هذا الخطاب واهميته العالية في الوضع السياسي العراقي.

شكل (7): المراحل الدلالية للموضوع الأول (التيار الصدري)

ب. الموضوع الثاني (تركيا والسياسة الخارجية)

يتناول هذا الموضوع التغطية الإعلامية المرتبطة بتركيا، ولا سيما سياساتها الإقليمية في الشأن العراقي، بما في ذلك الأمن، العلاقات الثنائية، والتدخلات العسكرية والدبلوماسية. ويُظهر التحليل تحولات واضحة في الخطاب الإعلامي من زاوية الأمن والتدخل العسكري إلى زاوية العلاقات السياسية والانتخابية التركية. جدول رقم (٣).

جدول (3): تحليل التحول الدلالي للموضوع الثاني (تركيا والسياسة الخارجية)

الفترة الزمنية	الكلمات البارزة	التحليل	وصف الخطاب الدلالي
2015 – 2016	تركيا، الجيش التركي، بعشيقة، التوغل، الحدود	ركزت التغطية في هذه الفترة على الوجود العسكري التركي داخل الأراضي العراقية، خصوصًا في شمال البلاد. كانت النبرة تميل إلى الاتهام والدعوة للانسحاب، وتُظهر استخدامًا كثيفًا لمفردات مثل "انتهاك السيادة" و"العدوان"	التوغّل التركي
2017 – 2018	أردوغان، زيارة، أنقرة، اجتماع، العلاقات، الموصل	رغم استمرار التوترات الحدودية، بدأ الخطاب الإعلامي يشمل جوانب دبلوماسية واقتصادية، مثل الزيارات المتبادلة والاجتماعات بين المسؤولين. وهنا يظهر تغير في اللغة باتجاه تهدئة نسبية وتغطية أكثر توازنًا.	خطاب دبلوماسي

2019 – 2020	زيارة، ملف المياه، التجارة، النفط، لقاء.	توسعت التغطية لتشمل ملفات إقليمية متشابكة مثل ملف المياه والنفط والتبادل التجاري. وتحول الحديث من الصراع الحدودي إلى التفاوض والتأثير السياسي، مما يعكس تغيرًا في النظرة الإعلامية لدور تركيا كفاعل إقليمي لا كخصم أمني	التداخل الإقليمي
2021 – ٢٠٢٣	المعارضة، أردوغان، الحزب، إسطنبول، الانتخابات	شهدت التغطية تحولًا ملحوظًا نحو الداخل التركي، حيث ركزت الأخبار على الانتخابات التركية، والمعارضة، والمواقف من أردوغان. أصبح الخطاب أكثر تحليلًا للتوازنات السياسية داخل تركيا، بدلًا من التركيز على سياساتها تجاه العراق فقط.	تصاعد التداخل السياسي

وعلى وفق ذلك نجد ان الخطاب الإعلامي فيما يخص تركيا يمر بأربع تحولات دلالية وكما يل، يوضحها الشكل (٨). أيضا يتميز هذا الخطاب بتنوع الدلالات مما يعكس كثر الاحداث السياسية المرتبطة بهذا الموضوع.

شكل (8): المراحل الدلالية للموضوع الثاني (تركيا والسياسة الخارجية)

ج. الموضوع الثالث (روسيا والتحركات الدولية)

يرتبط هذا الموضوع بالتغطيات الإعلامية المتعلقة بروسيا في الشأن الإقليمي والدولي، وخاصة ما يتعلق بعلاقاتها مع العراق، وسوريا، والولايات المتحدة، والحرب في أوكرانيا. ويظهر في التحليل الدلالي أن التناول الإعلامي للملف الروسي قد تأثر بشدة بالأحداث الدولية الكبرى، مما جعله يمر بثلاث مراحل دلالية رئيسية. جدول رقم (٤).

جدول (4): تحليل التحول الدلالي للموضوع الثالث (روسيا والتحركات الدولية)

الفترة الزمنية	الكلمات البارزة	التحليل	وصف الخطاب الدلالي
2015– ٢٠١٧	بوتين، سوريا، لافروف،	اتسم الخطاب الإعلامي برؤية إيجابية لروسيا، باعتبارها حليفًا إقليميًا في الحرب	روسيا كقوة إقليمية داعمة

	موسكو، دعم، الإرهاب، العمليات	على الإرهاب. وظهر ذلك في سياق دعم النظام السوري ومواقفها في مجلس الأمن، إضافة إلى مشاركات رمزية مع العراق في التنسيق الأمني	
٢٠١٨– ٢٠٢١	تدخل، سلاح، أمريكا، حلف الناتو، أوكرانيا، العقوبات	بدأت صورة روسيا تتغير في الخطاب الإعلامي العراقي تدريجيًا، خاصة مع تصاعد التوتر بينها وبين الغرب. أصبحت التغطية أكثر تركيزًا على صراعات النفوذ، وظهرت مفردات مثل "العقوبات" و"الاصطفاف"، ما يعكس تموضعًا جديدًا لروسيا كقوة مضادة للغرب	التوتر الجيوسياسي
2022– 2023	بوتين، الحرب، أوكرانيا، الرئيس، العقوبات، روسيا	تغيّر الخطاب بشكل جذري مع اندلاع الحرب الروسية – الأوكرانية، حيث أصبحت روسيا المحور الرئيسي في الإعلام الدولي والمحلي. التغطية ركزت على الحرب، ردود الأفعال الدولية، التأثيرات الاقتصادية، والموقف العراقي من الأزمة	حرب أوكرانيا

وعلى وفق ذلك نجد ان الخطاب الإعلامي فيما يخص روسيا يمر بثلاث تحولات دلالية، وكما موضح في الشكل (٩)، الذي يوضح تحولات جذرية في الخطاب الإعلامي لما يخص الوضع الروسي

شكل (9): المراحل الدلالية للموضوع الثالث (روسيا والتحركات الدولية)

د. الموضوع الرابع (رئاسة الجمهورية)

يرتبط هذا الموضوع بالتغطيات الإعلامية التي تناولت شخصية رئيس الجمهورية السابق برهم صالح خلال فترة ولايته الممتدة من ٢٠١٨ وحتى ٢٠٢٢، وهي فترة شهدت تحولات سياسية حادة في العراق، بدءًا من الاحتجاجات، مرورًا باستقالات حكومية، وصولًا إلى انتخابات مبكرة وأزمات دستورية. وقد عكست التغطية الإعلامية المرتبطة ببرهم صالح هذه الديناميكية في ثلاث مراحل دلالية أساسية. وكما في جدول رقم (٥).

جدول (5): تحليل التحول الدلالي للموضوع الرابع (رئاسة الجمهورية)

الفترة الزمنية	الكلمات البارزة	التحليل	وصف الخطاب الدلالي
٢٠١٨–٢٠١٩	الجمهورية، برهم، الرئيس، الحوار، الدستور، الإصلاح	ظهرت تغطيات تميل إلى تقديم برهم صالح كشخصية وطنية توافقية، تمركزت التغطية حول تصريحاته الدبلوماسية، وزياراته الرسمية، ومحاولاته لإظهار رئاسة الجمهورية كجسر تواصل بين الأطراف السياسية.	رئاسة توافقية ومواقف وطنية
٢٠١٩–٢٠٢١	التظاهرات، الاستقالة، الشعب، المتظاهرين، الضغط	مع تصاعد الاحتجاجات في ٢٠١٩، تغيّر الخطاب الإعلامي تجاه رئيس الجمهورية، وأصبح يُركّز على مواقفه من الأزمات، ظهرت تعبيرات مثل "صالح يهدد بالاستقالة"، و"موقف الرئيس من التظاهرات"، مما يُشير إلى تحوّله من رئيس شكلي إلى لاعب له استقلال نسبي في ظرف سياسي معقّد.	الأزمات ومواقف حاسمة
٢٠٢٢–٢٠٢٣	منصب، الترشيح، المنافسة، كردستان، عبد اللطيف رشيد	خفت الحضور الإعلامي لرئيس الجمهورية، وتركزت التغطيات على الجدل حول ترشحه لولاية ثانية، وعلى الصراع داخل القوى الكردية. الخطاب أصبح أقرب إلى "التقييم" أو "المقارنة"، بدلًا من كونه تقريرًا عن نشاط سياسي مباشر.	دور محدود ونهاية ولاية

وعلى وفق ذلك نجد ان الخطاب الإعلامي فيما يخص رئاسة الجمهورية يمر بثلاث تحولات دلالية، شكل (١٠).

شكل (10): المراحل الدلالية للموضوع الرابع (رئاسة الجمهورية)

هـ. الموضوع الخامس (البرلمان والتشريعات)

يرتبط هذا الموضوع بالتغطيات الإخبارية التي تناولت أداء مجلس النواب العراقي، من حيث جلساته، قوانينه، الخلافات التشريعية، والمواقف المرتبطة بالتصويت على قرارات مصيرية. وقد تراوح الخطاب الإعلامي لهذا الموضوع بين التغطية الروتينية للجلسات، إلى التركيز على النزاعات السياسية حول مشاريع القوانين الكبرى، إلى تسليط الضوء على الأزمات البرلمانية المتكررة. جدول رقم (٦).

جدول (6): تحليل التحول الدلالي للموضوع الخامس (البرلمان والتشريعات)

الفترة الزمنية	الكلمات البارزة	التحليل	وصف الخطاب الدلالي
٢٠١٥– ٢٠١٧	البرلمان، جلسة، قانون، القراءة، اللجنة، المقترح	هيمنت التغطيات ذات الطابع الإجرائي والتشريعي، حيث تناول الإعلام مجريات الجلسات، التصويتات، والقراءات للقوانين. كان الخطاب يعكس وظائف البرلمان.	الأداء التشريعي التقليدي
٢٠١٨– ٢٠٢٠	تأجيل، خلاف، تصويت، مكونات، الأغلبية، الكتلة	مع تصاعد الأزمات السياسية بعد الانتخابات، بدأت التغطية الإعلامية تتحول نحو التركيز على الاصطفاف داخل البرلمان. أصبحت المفردات مثل "خلاف بين الكتل"، "عدم التوافق"، و"تعطيل الجلسة" سائدة، مما يعكس تحوّل البرلمان من مؤسسة تشريعية إلى ساحة نزاع بين الكتل والأحزاب.	الخلافات والانسدادات
٢٠٢١– ٢٠٢٣	المحكمة الاتحادية، الطعن، الانسحاب، الرئاسة، الأغلبية	في هذه الفترة، أصبح الخطاب الإعلامي منصبًا على الأزمة الدستورية بعد الانتخابات، خصوصًا ملف رئاسة البرلمان، وانسحاب كتل رئيسية مثل التيار الصدري. وتكررت مفردات الطعن، المحكمة الاتحادية، الأغلبية السياسية، مما يعكس انسدادًا حادًا في البنية السياسية التشريعية	الأزمة الدستورية والانقسام

وعلى وفق ذلك نجد ان الخطاب الإعلامي فيما يخص البرلمان والتشريعات يمر بثلاث تحولات دلالية، شكل (١١).

شكل (11): المراحل الدلالية للموضوع الخامس (البرلمان والتشريعات)

من خلال أدوات نمذجة الموضوعات وتحليل الترددات والكلمات المفتاحية، تم تقسيم المحتوى السياسي إلى مجموعة من الموضوعات المحورية، ثم تطبيق منهجية التحليل الدلالي الموضوعي لتتبع تطور كل موضوع عبر

الزمن، بناءً على الكلمات والمفردات الأكثر تداولًا سنويًا. لم يكن التركيز على المواقف أو المشاعر، بل على تحولات المعلومات والسياقات المصاحبة لها، مما يمكن من فهم كيف يعاد إنتاج الخطاب السياسي داخل المنظومة الإعلامية، وكيف تنتقل التغطية من قضايا محلية إلى إقليمية، ومن مؤسسات إلى شخصيات، ومن الأداء السياسي إلى الأزمات. تُبرز هذه الدراسة قدرة تحليل المعلومات النصية على كشف أنماط خفية ومعاني عميقة في التغطيات الإعلامية، وتسلط الضوء على أهمية اعتماد إطار علمي متكامل يجمع بين الأدوات الإحصائية والخوارزميات اللغوية لتحويل النصوص إلى مؤشرات قابلة للتفسير والتحليل. وهذا يعزز دور تحليل البيانات النصية في دعم دراسات الإعلام وصناعة القرار، من خلال قراءة التطور الموضوعي للمعلومات لاستخلاص توجهات الرأي العام والتحولات السياسية الكبرى.

١٠. الاستنتاجات

- فعالية نمذجة الموضوعات في تحليل المعلومات: أثبت نموذج BERTopic كفاءته العالية في استخراج الموضوعات السياسية الرئيسة من الكم الهائل من الأخبار، وقد تم تحديد موضوعات كبرى عكست التوجهات السياسية والإعلامية في العراق خلال الفترة الزمنية المدروسة.
- التحليل الزمني والدلالي يعكس بنية الخطاب الإعلامي: أظهر تتبع تطور الموضوعات عبر السنوات قدرة على رصد تحولات جوهرية في مضمون الخطاب، مثل الانتقال من الخطاب الديني إلى الاحتجاجي أو من الأمني إلى الانتخابي، مما يعكس مرونة عالية في تحليل المعلومات وتفسير السياقات.
- على الرغم من التحديات الكبيرة في تحليل اللغة العربية، إلا أن توظيف نماذج حديثة وذكية قد ساعد في معالجة هذه التحديات بكفاءة.
- من خلال هذا البحث، تبيّن أن تحليل المعلومات لا يقتصر على معالجة البيانات، بل يمتد لفهم السياقات، كشف الأنماط الخفية، واستنباط مؤشرات تحليلية لها قيمة عالية في مجالات متعددة مثل الصحافة، السياسة، والأمن.

١١. التوصيات

- بتوسيع قواعد البيانات العربية ودعم أدوات المعالجة اللغوية المخصصة للنصوص العربية لتعزيز الدقة التحليلية.
- على المؤسسات الإعلامية أن تستفيد من نتائج تحليل الموضوعات لتحسين تغطيتها الإخبارية وفهم تأثيرها على الجمهور.
- تبني نماذج تحليل المعلومات في الدراسات السياسية والإعلامية لرصد التغيرات الفكرية والدلالية.
- الاستفادة من مخرجات هذا النوع من الأبحاث لتطوير منصات ذكية تتيح تتبع الخطاب السياسي وتحليل محتوى الأخبار بشكل دوري آلي.

- يُقترح توسيع نطاق البحث مستقبلًا ليشمل تحليل الأخبار المصورة والفيديوهات، لدمج التحليل النصي مع أنماط بيانات متعددة.

المصادر

1. Abelairas-Etxebarria, P., & Astorkiza, I. (2020). From exploratory data analysis to exploratory spatial data analysis. Mathematics and Statistics, 8(2). https://doi.org/10.13189/ms.2020.080202
2. Albright, S. Christian., & Winston, W. (2020). Business analytics : data analysis and decision making. 882. https://thuvienso.hoasen.edu.vn/handle/123456789/11614
3. Almuzaini, H. A., & Azmi, A. M. (2022). An unsupervised annotation of Arabic texts using multi-label topic modeling and genetic algorithm. Expert Systems with Applications, 203. https://doi.org/10.1016/j.eswa.2022.117384
4. Alshammary, H., Ibrahim, M. F., & Hussein, H. A. (2024). Evaluating The Impact of Feature Extraction Techniques on Arabic Reviews Classification. InfoTech Spectrum: Iraqi Journal of Data Science, 42–54. https://doi.org/10.51173/ijds.v1i1.10
5. Alshanik, F., Khasawneh, R., Dalky, A., & Qawasmeh, E. (2025). Unveiling Topics and Emotions in Arabic Tweets Surrounding the COVID-19 Pandemic: Topic Modeling and Sentiment Analysis Approach. JMIR Infodemiology, 5(1), e53434. https://doi.org/10.2196/53434
6. Alswedani, S., Katib, I., Abozinadah, E., & Mehmood, R. (2022). Discovering Urban Governance Parameters for Online Learning in Saudi Arabia During COVID-19 Using Topic Modeling of Twitter Data. Frontiers in Sustainable Cities, 4. https://doi.org/10.3389/frsc.2022.751681
7. Amara, A., Hadj Taieb, M. A., & Ben Aouicha, M. (2021). Multilingual topic modeling for tracking COVID-19 trends based on Facebook data analysis. Applied Intelligence, 51(5). https://doi.org/10.1007/s10489-020-02033-3
8. Barberá, P., Boydstun, A. E., Linn, S., McMahon, R., & Nagler, J. (2021). Automated Text Classification of News Articles: A Practical Guide. In Political Analysis (Vol. 29, Issue 1). https://doi.org/10.1017/pan.2020.8
9. Brown, Iain. (2024). Mastering marketing data science : a comprehensive guide for today's marketers. https://books.google.com/books/about/Mastering_Marketing_Data_Science.html?id=MAYFEQAAQBAJ

10. Chanda, S., & Pal, S. (2023). The Effect of Stopword Removal on Information Retrieval for Code-Mixed Data Obtained Via Social Media. SN Computer Science, 4(5). https://doi.org/10.1007/s42979-023-01942-7
11. Chehal, D., Gupta, P., & Gulati, P. (2021). Implementation and comparison of topic modeling techniques based on user reviews in e-commerce recommendations. In Journal of Ambient Intelligence and Humanized Computing (Vol. 12, Issue 5). https://doi.org/10.1007/s12652-020-01956-6
12. Dogra, V., Verma, S., Kavita, Chatterjee, P., Shafi, J., Choi, J., & Ijaz, M. F. (2022). A Complete Process of Text Classification System Using State-of-the-Art NLP Models. In Computational Intelligence and Neuroscience (Vol. 2022). https://doi.org/10.1155/2022/1883698
13. El-Alami, F. zahra, Ouatik El Alaoui, S., & En Nahnahi, N. (2022). Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization. Journal of King Saud University - Computer and Information Sciences, 34(10). https://doi.org/10.1016/j.jksuci.2021.02.005
14. Fossum, J. L., & Montoya, A. K. (2023). When to Use Different Inferential Methods for Power Analysis and Data Analysis for Between-Subjects Mediation. Advances in Methods and Practices in Psychological Science, 6(2). https://doi.org/10.1177/25152459231156606
15. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2). https://doi.org/10.1016/j.ijinfomgt.2014.10.007
16. Guellil, I., Saâdane, H., Azouaou, F., Gueni, B., & Nouvel, D. (2021). Arabic natural language processing: An overview. In Journal of King Saud University - Computer and Information Sciences (Vol. 33, Issue 5). https://doi.org/10.1016/j.jksuci.2019.02.006
17. Ibrahim, M. F., Alhakeem, M. A., & Fadhil, N. A. (2021). Evaluation of Naïve Bayes Classification in Arabic Short Text Classification. Al-Mustansiriyah Journal of Science, 32(4), 42–50. https://doi.org/10.23851/mjs.v32i4.994
18. Jaafar, Y., Namly, D., Bouzoubaa, K., & Yousfi, A. (2017). Enhancing Arabic stemming process using resources and benchmarking tools. Journal of King Saud University - Computer and Information Sciences, 29(2). https://doi.org/10.1016/j.jksuci.2016.11.010
19. Jannani, A., Bouhsissin, S., Sael, N., & Benabbou, F. (2025). Topic Modeling and Sentiment Analysis of Arabic News Headlines for a Societal Well-Being Scoring and Monitoring System: Moroccan Use Case. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3538888
20. Khelil, H. F., Ibrahim, M. F., & Hussein, H. A. (2024). Classifications of Arabic Customer Reviews Using Stemming and Deep Learning. Indonesian Journal of Electrical Engineering and Informatics, 12(3), 489–502. https://doi.org/10.52549/ijeei.v12i3.5452

21. Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia Tools and Applications, 82(3). https://doi.org/10.1007/s11042-022-13428-4
22. Mahesh, B. (2020). Machine Learning Algorithms - A Review. International Journal of Science and Research (IJSR), 9(1). https://doi.org/10.21275/art20203995
23. NINA. (2023). National Iraqi News Agency. https://www.ninanews.com
24. Oussous, A., Benjelloun, F. Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. In Journal of King Saud University - Computer and Information Sciences (Vol. 30, Issue 4). https://doi.org/10.1016/j.jksuci.2017.06.001
25. Park, C. S. (2022). Reading a Snippet on a News Aggregator vs. Clicking through the Full Story: Roles of Perceived News Importance, News Efficacy, and News-Finds-Me Perception. Journalism Studies, 23(11). https://doi.org/10.1080/1461670X.2022.2086160
26. Provost, Foster., & Lange, Benjamin. (2021). Data Science for Business. Analytic Thinking.
27. Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access, 8. https://doi.org/10.1109/ACCESS.2020.2976199
28. Rouhana, T. (2023). Critical discourse analysis guided topic modeling: the case of Al-Jazeera Arabic. Information Communication and Society, 26(5). https://doi.org/10.1080/1369118X.2023.2166364
29. Tahat, K., Mansoori, A., Tahat, D. N., Habes, M., & Salloum, S. (2024). Leveraging Soft Power: A Study of Emirati Online Journalism Through Arabic Topic Modeling. Lecture Notes in Networks and Systems, 923 LNNS. https://doi.org/10.1007/978-3-031-55911-2_2
30. Thompson, C. B. (2009). Descriptive Data Analysis. Air Medical Journal, 28(2). https://doi.org/10.1016/j.amj.2008.12.001
31. Timmerman, Y., & Bronselaer, A. (2022). Automated monitoring of online news accuracy with change classification models. Information Processing and Management, 59(6). https://doi.org/10.1016/j.ipm.2022.103105
32. Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the Number and Labels of Topics in Topic Modeling: A Tutorial. Advances in Methods and Practices in Psychological Science, 6(2). https://doi.org/10.1177/25152459231160105