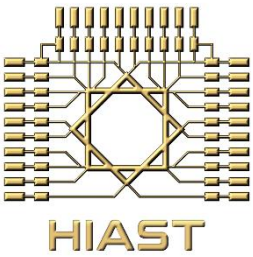
الجمهورية العربية السورية
المعهد العالي للعلوم التطبيقية والتكنولوجيا
قسم المعلوميات



أطروحة دكتوراه

كشف الشذوذ والتوجه في النصوص القصيرة المكتوبة باللغة العربية في شبكات التّواصل الاجتماعي

إعداد

م. مدين عبد الحميد

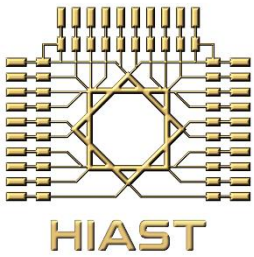
إشراف

د. آصف جعفر **د. ياسر رحال**

أعدت هذه الأطروحة لنيل درجة الدكتوراه في المعلوماتية

2023

Syrian Arab Republic
Higher Institute for Applied Sciences and Technology
Informatics Department



Ph.D. Thesis

Anomaly and Trend Detection in Short Arabic Text in Social Networks

By
Medyan AbdelHamid

Supervised by
Dr. Assef Jafar **Dr. Yasser Rahal**

A thesis submitted for the degree of Doctor of Philosophy in Informatics

2023

تصريح

أنا الموقع أدناه **مدين عبد الحميد** مُعدّ أطروحة الدكتوراه التي تحمل العنوان:

كشف الشّذوذ والتّوجه في النصوص العربية القصيرة على شبكات التّواصل الاجتماعي

أصرح بأن:

- الأعمال والنتائج المعروضة في هذه الأطروحة هي نتيجة جهودي الشخصية وبتوجيه من السادة المشرفين، وأن ما عدا ذلك من معلومات ونتائج قد نُسبت إلى مصادرها ومؤلفيها، وأشير إلى ذلك في متن النص وفي قائمة المراجع.
- المعطيات والمعلومات المستخدمة في هذه الأطروحة جرى تحصيلها بطرائق سليمة ومشروعة ونُسبت إلى مصادرها في المواضع الملائمة.
- كلّ مكوّن من مكونات هذه الأطروحة (مقطع نصّي، صورة، مخطط، ...) مقتبس من عمل آخر جرى تمييزه بوضوح ونُسب إلى مصدره.
- الأعمال والنتائج المعروضة في هذه الأطروحة لم تُستخدم سابقًا وليست قيد الاستخدام للحصول على أي شهادة أكاديمية أخرى.

التوقيع

دمشق 2022

المعهد العالي للعلوم التطبيقية والتكنولوجيا
Higher Institute for Applied Sciences and Technology

المعهد العالي للعلوم التطبيقية والتكنولوجيا مؤسسة حكومية للتعليم العالي أحدثت بموجب المرسوم التشريعي رقم /24/ لعام 1983، وذلك بهدف إعداد كوادر علمية متميزة من مهندسين وباحثين للإسهام الفاعل في عملية التطوير العلمي والتنمية في الجمهورية العربية السورية. يمنح المعهد العالي درجة الإجازة في الهندسة في الاتصالات والمعلوماتية والنظم الإلكترونية والميكاترونيكس وعلوم وهندسة المواد وهندسة الطيران. يقبل المعهد العالي لدراسة هذه الاختصاصات شريحة منتقاة من المتفوقين في الشهادة الثانوية من الفرع العلمي. يتيح المعهد العالي أيضًا برامج ماجستير أكاديمي في نظم الاتصالات وفي التحكم والروبوتيك وفي نظم المعطيات الكبيرة ونظم المعلومات ودعم القرار وفي علوم وهندسة المواد وعلوم وهندسة البصريات. أخيرًا، يمنح المعهد العالي درجة الدكتوراه في الاتصالات والمعلوماتية ونظم التحكم والفيزياء التطبيقية. تُحدث في المعهد العالي اختصاصات جديدة بحسب متطلبات سوق العمل وتوجهات البحث والتطوير المحلية والعالمية.

إلى جانب النشاط التعليمي، يمارس المعهد العالي عبر جهود أطره وفعالياته العلمية المختلفة نشاطًا حثيثًا في البحث والتطوير، إذ ينفذ مشاريع ودراسات واستشارات متنوعة لصالح الجهات العامة والخاصة في القطر، كما يتعاون مع جهات خارج القطر في بعض المشاريع البحثية والتطويرية. يسعى المعهد أيضًا، عبر دورات تدريبية نظرية وعملية متاحة للقطاعين العام والخاص وللأفراد، إلى إفادة أوسع فئة من المهتمين من إمكانيات أطره العلمية ومختبراته. استكمالًا لدوره الرائد في مجال التعليم ونشر العلم، ينشر المعهد العالي كتبًا علمية عالية المستوى من نتاج أطره، منها ما هو تدريسي يوافق المناهج في المعهد العالي ويفيد شريحة واسعة من الطلاب الجامعيين عمومًا، ومنها ما هو علمي ثقافي. يتيح المعهد العالي بعضًا من أبحاثه على موقعه على الشابكة، كما يتيح إمكانية الاطلاع على رسائل الماجستير والدكتوراه المنفذة في المعهد العالي وعلى بعض أبحاث طلابه وأطره من المقالات العلمية.

المعهد العالي للعلوم التطبيقية والتكنولوجيا،

الجمهورية العربية السورية، دمشق، ص.ب 31983

Higher Institute for Applied Sciences and Technology – HIAST.

P. O. Box 31983, Damascus, Syrian Arab Republic

هاتف 00963115123819 فاكس 00963115140761

موقع إلكتروني www.hiast.edu.sy بريد إلكتروني contact@hiast.edu.sy

إهداء

إلى نبض القلب

إلى ابنتي ليال

شكر وتقدير

أتوجّه بالشّكر والتّقدير والامتنان إلى أستاذَيّ: الدّكتور آصف جعفر والدّكتور ياسر رحّال لتفضّلهما بالإشراف على هذا البحث واللذين لم يدّخرا أيّ جهدٍ في تقديم الدّعم والتّوجيه والمشورة التي كان لها الأثر الكبير في إعداد هذا البحث وإنجازه بهذا الشّكل.

كما أشكر السّادة أعضاء لجنة الحكم لتفضّلهم بقبول تحكيم هذا البحث، وتخصيص ما يتطلّبه ذلك من وقتهم القيّم.

الملخص

يتزايد استخدام الناس لمواقع التواصل الاجتماعي للتعبير عن مشاعرهم وأفكارهم والتواصل وتبادل المعلومات فيما بينهم. ومع توفر مساحة حرية كبيرة، يميل بعض الناس إلى نشر خطاب الكراهية والشتائم عبر هذه المواقع. إن الكشف المبكر عن مثل هذا المحتوى أمر بالغ الأهمية يمكن أن يساعد في التنبؤ بالصراعات ويمنع هذه العواطف من أن تصبح أفعالًا أو أن تنتشر على نطاق واسع. تعتبر الدراسات التي تتناول الكشف عن خطاب الكراهية في النصوص العربية ولا سيما المكتوبة باللهجة المشرقية متناثرة وقليلة مقارنة بلغات أخرى ولا سيما الإنجليزية.

نعرض في هذا البحث بناء مجموعة بيانات لتغريدات مكتوبة باللهجة المشرقية محصلة من إحدى شبكات التواصل الاجتماعي (twitter) بغية التعرف على خطاب الكراهية ووسمها ومن ثم تدريب واختبار مجموعة من المصنفات التقليدية ومصنفات التّعلم العميق من أجل الكشف عن خطاب الكراهية في النصوص القصيرة والمكتوبة باللهجة المشرقية. وقد حصلنا على أفضل النتائج مع المصنف GigaBERT الذي أعطى معدل 94.6٪ على منحني ROC مع معدل 0.816 لمقياس F1-Score.

قمنا أيضًا في هذا البحث، بتطوير نظام يدمج بين تقنيات الوسم الذاتي Pseudo-labeling والذي يسمى في بعض المراجع بالتّعلم الذاتي self-training وتقنيات التّعلم النشط Active learning. يسمح النّظام المُطوَّر بوسم البيانات آليًّا، ويُمكِّن من انتخاب مجموعة محدودة من العينات بحيث تراعي التوازن بين الصنوف من جهة، وتحتوي من جهة أخرى العينات التي تمثل البيانات بشكل مناسب والعينات التي تتضمن القدر الأكبر من عدم اليقين عند تصنيفها، وذلك من أجل استخدام هذه المجموعة للتدريب. يقوم النّظام المقترح أيضاً بعرض العينات التي تتضمن القدر الأكبر من عدم اليقين على مراقب خبير Oracle من أجل تثبيت الوسم النهائي لها مما يسمح بتتبع التغيرات التي تطرأ على خطاب الكراهية. جرى استخدام النّظام المقترح لتنقيح مجموعة البيانات المحصلة ووسمها مما حسن الأداء وفق مقياس F1-score إلى 0.856.

الكلمات المفتاحية: اللغة العربية، اللهجة المشرقية، خطاب الكراهية، معالجة اللغات الطبيعية، تصنيف النصوص، تعلم الآلة بإشراف، التّعلم العميق، التّعلم النشط، عدم اليقين.

Abstract

People use online social networks to express their feelings and thoughts, communicate and share information. With much freedom, some people tend to spread hate speech and insults through these sites. Early detection of such content is very important to help predict conflicts and prevent these emotions from becoming actions or spreading widely.

Studies on the detection of hate speech in Arabic texts, particularly written in the Levantine dialect, are scattered and few compared to other languages, particularly English.

In this work, we collected a new Levantine tweets dataset from Twitter and annotated it then we trained and tested a set of conventional and deep learning classifiers on the dataset in order to detect hate speech detection in short texts written in the Levantine dialect. We have achieved good results using GigaBERT classifier with a 94.8% on the ROC Curve and 0.816 F1-score.

In this research, we developed a system that merges active learning techniques and Pseudo-labeling techniques, which in some references are called self-training. The developed system allows the data to be labeled automatically, and enables sampling which mitigates the imbalance of dataset on one hand, and on the other hand contains samples that represent the data appropriately and the samples the most uncertain for classification, in order to use this dataset for training. The most uncertain samples would be passed to an oracle to give them a final label, allowing trends in hate speech to be tracked. The proposed system was used to refine and relabel the collected dataset, improving performance on the F1-score scale to 0.856.

Keywords: Arabic; Levantine; hate-speech; natural language processing; text classification; supervised machine learning; deep learning; active learning; uncertainty;

المقالات

AbdelHamid, M., Jafar, A., & Rahal, Y. (2022). Levantine Hate Speech Detection in Twitter. *Social Network Analysis and Mining*. Springer Nature. Doi: 10.1007/s13278-022-00950-4

م. مدين عبد الحميد، د. آصف جعفر، د. ياسر رحال (2022). كشف خطاب الكراهية في النصوص العربية باللهجة المشرقية. مجلة جامعة دمشق للعلوم الهندسية

م. مدين عبد الحميد، د. آصف جعفر، د. ياسر رحال (2022). كشف خطاب الكراهية في النصوص العربية باللهجة المشرقية باستخدام التّعلم العميق. المؤتمر الدوري الأول في الهندسة المعلوماتية. جامعة دمشق.

فهرس المحتويات

فهرس الأشكال

فهرس الجداول

قاموس المصطلحات

المصطلح/المختصر باللغة الإنكليزية	المصطلح باللغة العربية
Abusive	مسيء
Accuracy	الصحة
Active Learning	التّعلم النشط
Agent	وكيل
Annotate	وسم
Artificial Intelligence	الذكاء الصنعي
Bag of Words	حقيبة الكلمات
Bagging	التغليف
Boosting	التعزيز
Classifier	مُصنِّف
Clustering	العنقدة
Computer Vision	الرؤية الحاسوبيّة
Confusion Matrix	مصفوفة الالتباس
Convolutional Neural Networks	الشّبكات التّلافيفية
Corpora	المدونات
Cyberbullying	التنمر الإلكتروني
Deep Learning	التّعلم العميق
Embedding	التّضمين
Embedding vector	متجه التّضمين
Ensemble Learning	التّعلم المجمّع
False Negative	السلبية الخاطئة
Features	السّمات
Filter	مُرشِّح
Framework	إطار عمل
Function	دالة

Hate	الكراهية
Hyper parameters	الموسطات الفوقية
Imbalanced	غير متوازن
Informativeness	الإفادة
Jihadist	الجماعات الإرهابية
Levantine	المشرقي
Lexical	معجمي
Machine Learning	تعلم الآلة
Modeling	النّمذجة
Offensive	الخطاب العدائي
Online Social Networks	شبكات التّواصل الاجتماعي
Oracle	خبير – حكيم
Pool	مخزن
Posterior Predictive Distribution	التوزيع التنبؤي اللاحق
Precision	الدقة
Recall	الاسترجاع
Recurrent Neural Networks	الشّبكات التّكرارية
Representativeness	التمثيلية
Self-Labeling	ذاتيّ الوسم
Self-Learning	ذاتي التّعلم
Stream Data	معطيات دفقية
Support Vector Machine	أشعة دعم الآلة
Uncertainty	عدم اليقين
Word Embedding	تضمين الكلمات

قائمة المختصرات

الاختصار	الاسم الكامل
AI	Artificial Intelligence
AL	Active Learning
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
BOW	Bag-Of-Words
CNN	Convolutions Neural Network
EGL	Expected Gradient Length
FN	False Negative
FP	False Positive
ML	Machine Learning
MLP	Multi-Layer Perceptron
NLP	Natural Languages Processing
OSN	Online Social Networks
QBC	Query By Committee
ReLU	**Re**ctified **L**inear **U**nit
RF	RandomForest
RNN	Recurrent Neural Network
SMOTE	**S**ynthetic **M**inority **O**ver-sampling **TE**chnique
SVC	Support Vector Classifier
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
URL	Uniform Resource Locator
XGB	eXtreme Gradient Boosting

1- المقدمة

أعد هذا الفصل بحيث يحقق ثلاثة أهداف رئيسة: أولها، تقديم مدخل موجز لموضوع كشف خطاب الكراهية ضمن شبكات التّواصل الاجتماعي، وتوضيح سياق البحث. أما الهدف الثاني، فهو تعريف مسألة البحث وتحديد الإشكاليات الأساسية التي نعالجها وتفصيل أسئلة البحث التي نسعى لإيجاد الأجوبة الشافية لها في فصول هذه الأطروحة. أما الهدف الثالث، فهو استعراض المساهمات الأساسية المقدمة في هذا العمل والتي تعبر عن الحلول المقدمة في هذه الأطروحة.

1-1- مقدمة

تتيح شبكات التّواصل الاجتماعي لمستخدميها التّعبير عن آرائهم وتحقيق التّواصل فيما بينهم وتبادل المعلومات من خلال التّغريدات والتّعليقات والمحادثات. كما تقدم شبكات التّواصل الاجتماعي كمية هائلة من البيانات عن سلوك الأشخاص الذين يتفاعلون معها من خلال أسلوب تواصلهم فيما بينهم وتواصلهم مع المصادر المفتوحة كالكتب والصور والفيديو وغيرها لا سيما مع ازدياد استخدام الأجهزة الذكية حيث أصبح من السهل الحصول على معطيات كبيرة ومتنوعة. تعتبر مصادر التّواصل الاجتماعي مفتوحة، ما يجعلها عرضة للاستخدام السيء من قبل بعض الأشخاص المشبوهين، حيث باتت تستخدم للتخطيط للجرائم والأعمال الإرهابية وتهديد أمن المجتمع، لذلك أصبح من الضروري العمل على تطوير نظم تعتمد على البيانات الكبيرة من أجل كشف الأعمال المشبوهة وحذف التّهديدات التي تمثلها. تستخدم هذه النظم معطيات شبكات التّواصل الاجتماعي والبيانات المحصلة من أنظمة المراقبة وخاصة الموجودة على الحدود ومعطيات شبكات الاتصال الخليوي لكشف خطاب الكراهية الذي يعتبر أحد أهم أشكال المحتوى الشاذ وحذف التّهديد الذي يمثله وكذلك كشف التّوجهات الطبيعية ضمن المجتمع للتفاعل الصحيح معها واستخدامها من أجل تلبية حاجات المجتمع بأفضل طريقة ممكنة.

أصبحت شبكات التّواصل الاجتماعي المنصة المثالية لنشر خطاب الكراهية والمحتوى السيء [1]، وانتشر خطاب الكراهية خلال الشبكات الاجتماعية مثل تويتر وفيسبوك، وأصبح له تأثير سلبي واضح مما يقتضي كشفه وتحديد مصدره للحد من خطره في المجتمع، إن التّحديد المبكر للمستخدمين الذين يروجون لمثل هذا النوع من الخطاب يمكن أن يمنع التّصعيد في خطاب الكراهية من الكلام إلى الفعل. يوجد في منصات وسائل التّواصل الاجتماعي عدد لا يمكن السيطرة عليه من الرسائل الصادرة في كل ثانية مما يجعل من المستحيل تتبع أو التّحكم في محتوى هذه الرسائل يدويًا [2]، لذلك يجب مواجهته من خلال تسخير قوة الذكاء الصنعي وخوارزميات التّعلم الآلي

لأتمتة كشف خطاب الكراهية في وسائل التّواصل الاجتماعي وخاصةً عندما لا يحتوي النص على كلمات صريحة، حيث يجب الأخذ بالاعتبار المعنى الدلالي الذي تحمله هذه الرسائل.
تواجه المنصات الاجتماعية مشكلة في الحد من هذه الرسائل مع موازنة حرية التّعبير، ونظرًا لأن الأشخاص الذين ينشرون خطاب الكراهية قد يجري حظرهم أو معاقبتهم أو مراقبتهم لعدم تحويل تهديداتهم إلى أفعال، فإن وثوقية الكشف الآلي هامة في صنع القرار وتساعد المراقب البشري بتوضيح الرؤية.
تعتبر اللغة العربية واحدة من أكثر اللغات انتشارًا في العالم فهي رابع أكثر اللغات استخدامًا في العالم ورابع أكثر اللغات استخدامًا على الإنترنت [3]. يتحدث أكثر من 6.0٪ من سكان العالم اللغة العربية وهناك نمو ملحوظ في استخدام منصات التّواصل الاجتماعي في المنطقة العربية. يُنشر على هذه المنصات الكثير من خطاب الكراهية باللغة العربية على تويتر وغيرها ومع ذلك لا يوجد إلا عدد قليل من الدراسات لكشفه مقارنة باللغات الأخرى واسعة الانتشار كاللغة الإنجليزية [4]، ويعود ذلك لعدة عوامل منها وجود كثير من اللهجات فيها مثل (المصرية، الخليجية، المشرقية، المغربية، وغيرها...) بالإضافة إلى اللغة الكلاسيكية، واللغة الحديثة المنتشرة في وسائل الإعلام، وتتميز اللغة العربية بخواص تختلف بها عن اللغات الأخرى فلديها مورفولوجيا غنية وبنية نحوية مركبة ومما يميزها أيضًا الكم الهائل من المفردات المترادفة وتشكيل الأحرف وطبيعتها الخاصة في الاشتقاق والإعراب [5]. تتعامل معظم هذه الأبحاث مع الوثائق بشكل offline من خلال معطيات مخزنة مسبقًا ولا تتعامل معها كمعطيات دفقية stream data، ومع ذلك، فالأبحاث التي عالجت النصوص كبيانات دفقية، عالجت الموضوع على وثائق كبيرة الحجم، واستخدمت مفاهيم مختلفة وذلك بسبب طبيعة البيانات الدفقية فهي بيانات ذات تدفق مستمر وتتطلب استجابة سريعة، ما يجعل آليات التّنقيب التّقليديّة غير قابلة للتّطبيق في هذا الحالة.
تتعامل معظم الأبحاث المنفذة على الوثائق المكتوبة باللغة العربية من خلال خوارزميات التّعلم التقليدية Traditional Machine Learning، وجرى تناول هذه المسألة عبر عدد قليل الأبحاث من منظور التّعلم العميق Deep Learning.

1-2- دوافع البحث

تزداد نسبة تعرض المستخدمين لأشخاص أو مجموعات مختلفة بكتابات تدعو إلى العنف أو الكراهية لأسباب عدة، ولا سيما في المناطق التي تشهد نزاعات سياسية أو أهلية مع انتشار الفقر والجهل والتعصب، ومن بين هذه المناطق منطقتنا العربية. لذلك تظهر الحاجة لوجود أنظمة تساعد في الكشف عن خطاب الكراهية في النصوص العربية وهذا ما تؤكده الأعداد المتزايدة من الأبحاث التي تناولت هذه المسألة.

ينطلق عملنا في هذا البحث من خلال الحاجة المتزايدة لتطوير نظام يتعامل مع النصوص القصيرة المكتوبة باللغة العربية –وخاصة باللهجة المشرقية– المنتشرة على منصات التواصل الاجتماعي من أجل الكشف عن خطاب الكراهية دون المساس بحرية التعبير.

1-3- مشكلة البحث

تعتبر عملية الكشف عن خطاب الكراهية في النصوص المكتوبة من المسائل الصعبة في تعلم الآلة، وذلك بسبب عدة عوامل (نذكر منها: نصوص قصيرة، شخصية المؤلف والمتلقي، النوايا غير الواضحة، استخدام اللهجة العامية، والدمج بين اللغات، الأخطاء الإملائية، ...الخ)، ولا سيما عند اعتماد المعنى الدلالي على السياق بشكل شبه كامل، حيث يمكن أن تختلف معاني الكلمات إلى حد كبير باستخدام الفكاهة والسخرية والتّلميحات والاستعارة.
نعرض في الفقرات التالية أهم الإشكالات التي تعاني منها مسألة الكشف عن خطاب الكراهية في النصوص العربية.

1-3-1- خصوصية اللغة العربية

تعتبر اللغة العربية ذات طبيعة خاصة فهي تتميز بالغنى والتّعقيد على حد سواء في الصرف والمفردات والإملاء، كما يوجد عدد كبير من اللهجات المستخدمة المحكية أو المكتوبة على وسائل التّواصل الاجتماعي، وهذه اللهجات لا تختلف بين البلدان فحسب، وإنما تختلف أيضًا ضمن مناطق أو أقاليم البلد الواحد. يمكن ملاحظة وجود كلمات ذات لفظ أو كتابة واحدة بين اللهجات ولكنها مختلفة كليًا في المعنى.

1-3-2- عدم توفر مجموعة بيانات موسومة

بالرغم من توافر حجوم هائلة من المعطيات النصية ضمن مواقع شبكات التّواصل الاجتماعي وشبكات الأخبار وغيرها، إلا أن هذه البيانات النصية غير موسومة ولا سيما فيما يتعلق بمسألة خطاب الكراهية. يعود ذلك بشكل أساسي إلى أن عملية الوسم مكلفة جدًا من حيث الجهد والزمن.

1-3-3- عدم وجود تعريف واضح لخطاب الكراهية

لا يعتبر تعريف خطاب الكراهية من المسائل المتفق عليها كليًا سواء عالميًا أو فرديًا، كما أن الحدود الفاصلة بين خطاب الكراهية والتّعبير عن الرأي بحرية هي منطقة ضبابية، وبالتّالي يجب أخذ بعض الحذر عند إعطاء خطاب الكراهية تعريفًا دقيقًا، مع الأخذ بالاعتبار أن وضع تعريف

واضح لخطاب الكراهية يمكن أن يساعد في دراسة مسألة الكشف عن خطاب الكراهية من خلال جعل عملية الوسم أسهل وأكثر موثوقية [11].

1-3-4- عدم توازن مجموعات البيانات

تتصف مجموعات البيانات التي تتناول مسألة خطاب الكراهية بأنها غير متوازنة imbalanced عمومًا، مما يؤثر سلبًا على أداء خوارزميات التّصنيف المستخدمة، وذلك لأن تدريب مصنف ما على مجموعة بيانات غير متوازنة يمكن أن يقود إلى معدل عالٍ للسلبية الخاطئة false negative (تصنيف عدد أكبر من العينات على أنها خطاب عادي فيما هي خطاب كراهية)، بالرغم من الحصول على معدل دقة عالٍ.

1-3-5- عدم كفاية مجموعات البيانات

تكمن أهم تحديات مسائل تعلم الآلة في القدرة على الحصول على مجموعة بيانات مثالية كافية للمسالة المطروحة، فبالإضافة إلى حجوم مجموعات البيانات الصغيرة التي تؤثر سلبًا على أداء نظم التصنيف، نلاحظ أن معظم مجموعات البيانات تحوي على عينات تمثل ضجيجًا بالإضافة إلى مجموعات عينات متشابهة إلى حدٍ كبير. كما أنه لا يمكن اعتبار العينات على درجةٍ واحدةٍ من الأهمية خلال عملية الوسم.

لتحسين أداء نظم التعلم نحن بحاجة لتحصيل مجموعة كبيرة من البيانات ووسمها. على الرغم من توفر حجوم بيانات ضخمة في شبكات التواصل الاجتماعي إلا أن وسم هذه البيانات عملية مكلفة من حيث الجهد والزمن، كما أن عملية الوسم اليدوي معرضة للأخطاء.

غالبًا ما يجري الاعتماد على مجموعات بيانات مُحَصَّلَة بالاعتماد على محددات خاصة ما يجعل مجموعة البيانات منحازة لمنطقة محددة من فضاء العينات. كما أنها قد لا تكون الأفضل لتدريب النموذج المطلوب وتحسين أداء خوارزميات التّعلم، أي أن النموذج المدرب سينحاز إلى مجموعة التدريب هذه.

نولي في هذا البحث أهمية كبيرة لنوعية البيانات التي سيجري التدريب عليها، حيث سيتم انتخاب مجموعة محددة ذات نوعية مناسبة للتدريب، أي أنه سيتم التركيز على ما يعرف بمركزية البيانات data-centric approach.

1-3-6- عدم استقرار البيانات

تعاني الكثير من نظم تعلم الآلة من مشكلة عدم استقرار البيانات unstable data حيث أن خصائص تابع الكثافة الاحتمالية الذي يولد المعطيات تتغير مع الزمن، مما يجعل خصائص

البيانات التي يجري العمل عليها أثناء نشر النظام على أرض الواقع مختلفة عن خصائص البيانات المستخدمة في مرحلة التدريب، مما يؤدي لتراجع أداء هذه النظم مع الزمن. تتواجد هذه الظاهرة في عدة مجالات منها كشف خطاب الكراهية حيث يمكن أن تظهر مع الزمن تعابير ومفردات مختلفة في خصائصها عن التعابير والمفردات التي كانت موجودة أثناء تحصيل مجموعة التدريب. **نركز بشكل أساسي في هذا البحث على تطوير إطار عمل من أجل لتتبع التغيرات التي يمكن أن تحصل على تابع الكثافة الاحتمالية المولدة للبيانات وتطبيقه على تتبع التغيرات التي تطرأ على خطاب الكراهية.**

1-4- الهدف من البحث

يهدف البحث إلى تطوير نموذج خاص من أجل كشف خطاب الكراهية للنصوص العربية في شبكات التّواصل الاجتماعي، وبالتّحديد للنصوص القصيرة المكتوبة باللهجة المشرقية، بحيث يكون قادرًا على متابعة التغيرات التي يمكن أن تطرأ على التوزيع الخاص بخطاب الكراهية وذلك من خلال الكشف عن العينات الجديدة وإضافتها بشكل مستمر إلى مجموعة التدريب.

1-5- أسئلة البحث

نحدد في هذه الفقرة أسئلة البحث الأساسية التي نسعى لبحثها ودراستها واقتراح الحلول التي تجيب عنها خلال الفصول التّالية في هذه الأطروحة.

Q1- ما هو الوضع الراهن للأبحاث ذات الصلة؟

تناولت مجموعة من الدراسات السابقة [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] مسألة الكشف عن خطاب الكراهية للنصوص العربية في شبكات التّواصل الاجتماعي. اعتمدت دراسة واحدة فقط على مجموعة بيانات لنصوص عربية مكتوبة باللهجة المشرقية [29]، وهي تتناول هذه المسألة بالاعتماد على بيانات ذات طبيعة سياسية [29]. لذلك، تُعدُّ هذه المراجعة أحد الجوانب الأساسية في هذا البحث لتكون منطلقًا في تجاوز العقبات الأساسية التي عانت منها الدراسات السابقة.

Q2- هل تعتبر مجموعات البيانات المتوفرة على شبكة الإنترنت كافية؟

اعتمدت الدراسات السابقة التي تناولت مسألة الكشف عن خطاب الكراهية للنصوص العربية في شبكات التّواصل الاجتماعي على مجموعات بيانات خاصة بها، وقد أتاحت بعض هذه الدراسات

مجموعات البيانات الخاصة بها على شبكة الإنترنت. لكن هل تعتبر مجموعات البيانات هذه كافية من أجل استخدامها في نظام فعّال للكشف عن خطاب الكراهية باللهجة المشرقية؟

Q3- ما هو التمثيل الأفضل للنصوص في مجموعات بيانات خطاب الكراهية؟

جرى تمثيل البيانات في مسائل تصنيف النصوص بطرق مختلفة، كما عُولِجت هذه المسائل باستخدام نماذج تعلم الآلة المختلفة. يبقى السؤال: ما هو التمثيل الأفضل لنصوص مجموعات البيانات التي تتناول مسألة خطاب الكراهية؟ وما هي نماذج تعلم الآلة المناسبة لبناء نظام فعال للكشف عن خطاب الكراهية؟

Q4- هل يساعد بناء مجموعة بيانات وتدريب نظم تصنيف عليها في زيادة القدرة على التعميم؟

غالبًا ما تعطي نماذج تعلم الآلة نتائج جيدة عند الاختبار على مجموعة بيانات مشابهة لمجموعة التدريب، وتكمن قدرة نموذج التّعلم على التعميم في نجاحه في الاختبار على مجموعات مغايرة. نطرح هنا السؤال التالي: هل يعتبر تحصيل مجموعة بيانات جديدة ووسمها عملية مفيدة تساعد في تطوير نظام تصنيف قادر على التّعميم خارج إطار مجموعة البيانات المحصَّلة من خلال اختباره على مجموعات بيانات أخرى متوفِّرة؟

Q5- كيف يمكن انتخاب مجموعة محددة من عينات التدريب ذات تأثير أكبر على نظم التصنيف؟

تتأثر نظم التصنيف بشكل كبير بمجموعة التدريب، وعادة ما يتم العمل على انتخاب عينات للتدريب من خلال تقدير درجة عدم اليقين uncertainty أو تقدير درجة التمثيل representativeness للعينات المنتخبة بعدة طرق حسب نموذج التّعلم المستخدم. نطرح هنا السؤال التالي: هل يوجد طريقة جديدة فعالة قادرة على تقدير درجة عدم اليقين أو درجة التمثيل من أجل انتخاب العينات بفعالية أكبر؟

إضافة إلى ذلك، فإن الاعتماد على إحدى هاتين الطريقتين (تقدير درجة عدم اليقين أو درجة التمثيل) يؤدي بالنّظام إلى التّحيز bias باتجاه فضاء جزئي من فضاء العينات الكلي، كما أن الدمج بين الطريقتين بشكل اعتباطي لا يساهم في حل المشكلة. كما أن أغلب طرق الانتخاب المستخدمة لا تتناول مسألة البيانات غير المتوازنة.

Q6- كيف يساعد دمج تقنيات التّعلم النشط مع التّعلم الذاتي في تطوير نظام تصنيف قادر على ملاحقة التغيرات دون أن ينخفض الأداء مع الزمن؟

تعتمد نظم التصنيف بشكل كبير على حجم وجودة مجموعة البيانات المستخدمة في التدريب، لذلك هل يساعد دمج تقنيات التّعلم النشط مع التّعلم الذاتي في تطوير نظام تصنيف فعال باستخدام عدد محدد من العينات الموسومة؟ وهل يمتلك نظام التصنيف هذا القدرة على ملاحقة التغيرات التي يمكن أن تطرأ على خطاب الكراهية؟

1-6- المساهمات الأساسية

نلخص فيما يلي المساهمات الأساسية التي قدمناها في هذا البحث:

- دراسة مرجعية للآليات المعتمدة في الأبحاث التي تناولت موضوع كشف خطاب الكراهية في النصوص العربية في شبكات التّواصل الاجتماعي. جرى تحليل هذه الأبحاث ومقارنتها مع التّركيز على الآلية المعتمدة لتمثيل النصوص. تشكل هذه المراجعة نقطة انطلاق لطرح مقاربتنا، كما أنها تشكل أساسًا لأبحاث لاحقة في مجال كشف خطاب الكراهية.
- بناء مجموعات بيانات موسومة: تُشكِّل كل واحدة من هذه المجموعات عدة تغريدات مأخوذة من موقع تويتر موسومة وفقًا لاحتوائها على ما يدل أنها خطاب كراهية أو لا. تحتوي المجموعة الأولى على التّغريدات التي المحصلة بدون أي عملية تعزيز على البيانات، بينما تحتوي المجموعة الثانية بالإضافة إلى التغريدات في المجموعة الأولى مجموعة تغريدات مكونة يدويًا، بينما تحتوي المجموعة الثالثة بالإضافة إلى التغريدات في المجموعة الأولى مجموعة تغريدات مكونة بشكل آلي. أما المجموعة الرابعة، فهي تحتوي بالإضافة إلى التغريدات في المجموعة الأولى جميع التّغريدات المضافة إلى كلٍ من المجموعة الثانية والثالثة، بينما تحتوي المجموعة الأخيرة على نفس العينات في المجموعة الرابعة ولكن بوسم منقح لبعض العينات.
- نظام تصنيف آلي لكشف خطاب الكراهية اعتمادًا على تمثيل تضمين الكلمات غير السياقي من خلال استخدام تقنية التّصويت voting على مجموعة من المصنفات التّقليدية.
- نظام تصنيف آلي لكشف خطاب الكراهية اعتمادًا على تمثيل تضمين الكلمات السياقي من خلال استخدام شبكة عصبونية.
- خوارزمية جديدة لانتخاب العينات المرشحة لإجراءات التّعلم النشط من خلال تقدير عدم اليقين عبر احتساب التوزيع التنبؤي اللاحق posterior predictive distribution.
- خوارزمية جديدة لانتخاب العينات في نظم التّعلم النشط تراعي تحقيق التّوازن trade-off بين الصفوف ولا سيما غير المتوازنة imbalanced، كما تراعي التّوازن بين عينات عدم اليقين uncertainty والعينات الأكثر تمثيلًا لفضاء العينات.

- تطوير إطار عمل تكيفي لمتابعة التغيرات التي تطرأ على تابع الكثافة الاحتمالية المولدة للبيانات يأخذ بالاعتبار النقاط أعلاه.
- حزمة برمجية منجزة بلغة بايثون Python تتيح إمكانية الاستفادة من المساهمات السابقة.

1-7- مخطط الأطروحة

تشتمل هذه الأطروحة على الفصول التّالية:

الفصل الأول: المقدمة، يعرف هذا الفصل البحث ودوافعه والمشكلة البحثية والمسائل المتفرعة عنه البحث، كما يشمل هذا الفصل المساهمة العلمية والعملية.

الفصل الثاني: الدراسة المرجعية، خُصِّصَ هذه الفصل للدراسة المرجعية للأبحاث التي تناولت مسألة كشف خطاب الكراهية في النصوص العربية ضمن شبكات التّواصل الاجتماعي، بالإضافة إلى التّحديات التي تواجه مسألة كشف خطاب الكراهية.

الفصل الثالث: مجموعة البيانات المشرقية، يعرض هذا الفصل مجموعة بيانات جديدة باللغة العربية باللهجة المشرقية مستخرجة من موقع تويتر خاصة بخطاب الكراهية، بحيث تتجاوز هذه المجموعة بعض الثغرات الموجودة في مجموعات البيانات المتاحة، كما نعرض الإجراءات المنفذة على هذه المجموعة كالمعالجة الأولية وآلية الوسم وتعزيز البيانات وطرق التمثيل المعتمدة لاختبارها على مجموعة من المصنفات التقليدية المعروفة وعرض نتائج الاختبارات على هذه المجموعة.

الفصل الرابع: نموذج كشف خطاب الكراهية، يعرض هذا الفصل النموذج المقترح لكشف خطاب الكراهية ضمن شبكات التّواصل الاجتماعي، من خلال التدريب على مجموعة البيانات المحصلة عبر استخدام نماذج تضمين الكلمات السياقي وغير السياقي لتمثيل النصوص وتمريرها إلى مصنفات التّعلم المختلفة المعتمدة، بالإضافة إلى توصيف الاختبارات المنفذة ومناقشة النتائج وتحليلها.

الفصل الخامس: إطار عمل تكيفي لكشف خطاب الكراهية، يعرض هذا الفصل شرحًا عن إطار العمل الذي جرى تصميمه وبناؤه من أجل الكشف عن خطاب الكراهية ضمن شبكات التّواصل الاجتماعي، بالإضافة إلى توصيف الاختبارات المنفذة ومناقشة النتائج وتحليلها.

الفصل السادس: الخاتمة والآفاق المستقبلية، يعرض هذا الفصل ملخصًا عن الأعمال المنجزة والمساهمات العلمية المقدمة في هذا البحث، بالإضافة إلى لمحة عن الأعمال والأفاق المستقبلية التي سيجري العمل عليها لاحقًا.

2- الدراسة المرجعية

نتناول في هذا الفصل الدّراسات التي اهتمّت بموضوع الكشف عن خطاب الكراهية في النّصوص العربية، محاولين تقديم الإجابة على السؤال Q1 "**ما هو الوضع الراهن للأبحاث ذات الصلة؟**" من خلال دراسة مرجعية تركز على الجوانب التّالية: (1) تصنيف الأبحاث ذات الصلة وفق مجموعة البيانات المعتمدة (مصدرها، حجمها، طبيعتها، توزعها بين الصفوف)، أو وفق تعريف خطاب الكراهية المعتمد، أو طريقة تمثيل البيانات. (2) سبر أهم الثغرات والإشكالات التي تعاني منها هذه الأبحاث واستشراف التّوجهات المستقبلية.

كما نتناول في هذا الفصل أيضًا، الدّراسات التي اهتمّت بموضوع التّعلم النشط الذي يعتبر أحد الطرق لتحسين أداء نظم التصنيف بشكل عام، محاولين تقديم دراسة مرجعية تركز على الجوانب التّالية: (1) تصنيف الأبحاث ذات الصلة وفق آلية انتخاب العينات، أو وفق الاستراتيجيات المعتمدة. (2) سبر أهم الثغرات والإشكالات التي تعاني منها هذه الأبحاث واستشراف التّوجهات المستقبلية بما يشكل أساسًا للحلول المطروحة لاحقًا في هذه الأطروحة.

2-1- مقدمة

أعطى الاستخدام المتزايد لشبكات التّواصل الاجتماعي وتشارك المعلومات منافع جمة للإنسانية. بيد أن ذلك أدى أيضًا إلى نشوء تحديات متنوعة تشمل نشر رسائل خطاب الكراهية وتشاركها. يعتبر خطاب الكراهية آفةً عامةً في المجتمع الحديث، حيث يمكن للشخص في الوقت الحاضر أن يستعرض أو ينشر محتوى خطاب الكراهية بسهولةٍ أكثر من خلال شبكات التّواصل الاجتماعي والمواقع والمنتديات التي تضم محتوى من إنشاء الأشخاص أنفسهم. يحصل الأشخاص على فرصة أكبر لنشر محتوى يتضمن خطاب كراهية بسهولة، لا سيما مع الازدياد المضطرد لاستخدام هذه الشبكات والوصول إلى عدد من الأفراد أكثر من أي وقتٍ مضى. وفقًا لبعض الدراسات الاستقصائية الأخيرة، أدّى ارتفاع محتوى خطاب الكراهية على الإنترنت إلى جرائم تتعلق بالكراهية (انتخابات ترامب في الولايات المتحدة [30]، وهجمات مانشستر ولندن في المملكة المتحدة [31]، والهجمات الإرهابية في نيوزيلندا [32]).

نجد من ناحيةٍ أخرى، أنّ شبكات التّواصل الاجتماعي مثل فيسبوك أو تويتر تريد الامتثال للتشريعات التي تواجه خطاب الكراهية بدون التأثير على أداء هذه الشبكات. لذلك، فهي تحتاج إلى تتبع وإزالة هكذا محتوى من مواقعها بشكلٍ متقنٍ. إلّا أنّ تفويض مثل هذه المهمة للبشر غير فعّالٍ نهائيًّا نظرًا لكمية البيانات الضّخمة التي تنتقل عبر هذه المنصات. اعتمدت هذه المنصات حلًا وسطًا وهو الاعتماد على تقارير وتبليغات المستخدمين. هذا أيضًا غير فعّال، لأنّه يعتمد على موضوعية المستخدم وهل هو جديرٌ بالثّقة أم لا، فضلًا عن عدم قدرة المصنفات على التّتبع الدّقيق لمثل هذا المحتوى. يُضاف إلى ذلك تطور

المفردات أو الكلمات التي تعبر عن الكراهية، بالإضافة إلى الاختلاف بين المستخدمين على تعريف خطاب الكراهية، فما يعتبر خطاب كراهية بالنّسبة لشخص ما، ليس بالضّرورة أن يكون كذلك لشخصٍ آخر.
بالنظر إلى كلِّ ما سبق، يعتبر تطوير أدوات آليّة لكشف محتوى خطاب الكراهية حاجةً ماسّةً وضروريةً. صنفت العديد من الدراسات هذه المشكلة على أنها مهمة تصنيف ثنائي [33] [34]، ولكن بعض هذه الدراسات فشلت في معالجة حالات خاصة من خطاب الكراهية، مثل استخدام لغة مبتذلة ضد الفرد أو المجتمع أو وجود خطاب كراهية غير مباشر، كما فشلت هذه النماذج في التّمييز بين اللغة الهجومية العامة وخطاب الكراهية [35]. تعاني هذه المصنفات كذلك من مسألة جوهرية أخرى وهي عدم توازن البيانات، حيث يمثل خطاب الكراهية نسبة صغيرة من مجمل البيانات. بالتّالي فإننا بحاجة لمزيد من عينات خطاب الكراهية مع مراعاة عدم جرِّ النّماذج أو المصنفات إلى الملاءمة الزائدة overfitting.

2-2- الدراسة المرجعية لكشف خطاب الكراهية

2-2-1- هدف الدراسة المرجعية

نهدف في هذه الفقرة إلى تحديد وتصنيف وتحليل الأبحاث التي تتناول موضوع الكشف عن خطاب الكراهية **المكتوب باللغة العربية**. كما نهدف إلى الإجابة عن السؤال Q1 وتقديم الأجوبة على بعض الأسئلة الفرعية المندرجة تحته.
للإجابة على السؤال Q1 "**ما هو الوضع الراهن للأبحاث ذات الصلة؟**"، سنقوم بوضع بعض الأسئلة الفرعية لهذا السؤال وتقديم الإجابة عليها.

Q1.1. ما هو الاتجاه العام لتوزع هذه الأبحاث؟
Q1.2. ما هي أنواع خطاب الكراهية المستهدف؟
Q1.3. ما هي مصادر البيانات وأنواع اللهجات المستخدمة؟
Q1.4. ما هي تصنيفات مجموعات البيانات المستخدمة؟ وما هي نسب توزع الصفوف؟
Q1.5. ما هي التّمثيلات المقترحة لتمثيل النصوص؟
Q1.6. ما هي المصنفات المستخدمة؟
Q1.7. ما هي التوجهات العامة للأبحاث التي تناولت كشف خطاب الكراهية باللغات المختلفة؟
Q1.8. ما هي الثغرات والاتجاهات المستقبلية التي يمكن العمل عليها لتحسين النّتائج المحقّقة؟

2-2-2- لمحة عن الأبحاث المشابهة

اعتمدنا في هذه الدراسة على أربعين ورقة بحثية تناولت موضوع خطاب الكراهية باللغة العربية، وتوزعت هذه الأوراق بين عامي 2015 و2022، وجرى تحديد المقاربات المطروحة في كل هذه الأبحاث وتحديد هدفها ونطاقها ومساهمتها الأساسية، حيث جرى استخلاص أصناف المعلومات التالية:

- المعلومات العامة: أسماء المؤلفين، سنة النشر، نوع النشر، مكان النشر.
- نوع الخطاب المستهدف ضمن مجال كشف خطاب الكراهية.
- مجموعات البيانات: سواء مجموعات التدريب أو التحقق أو الاختبار.
- الحلول المقترحة: من خلال تبيان طرق تمثيل البيانات ونماذج تعلم الآلة ونتائج الاختبارات.
- المساهمات والقيود.

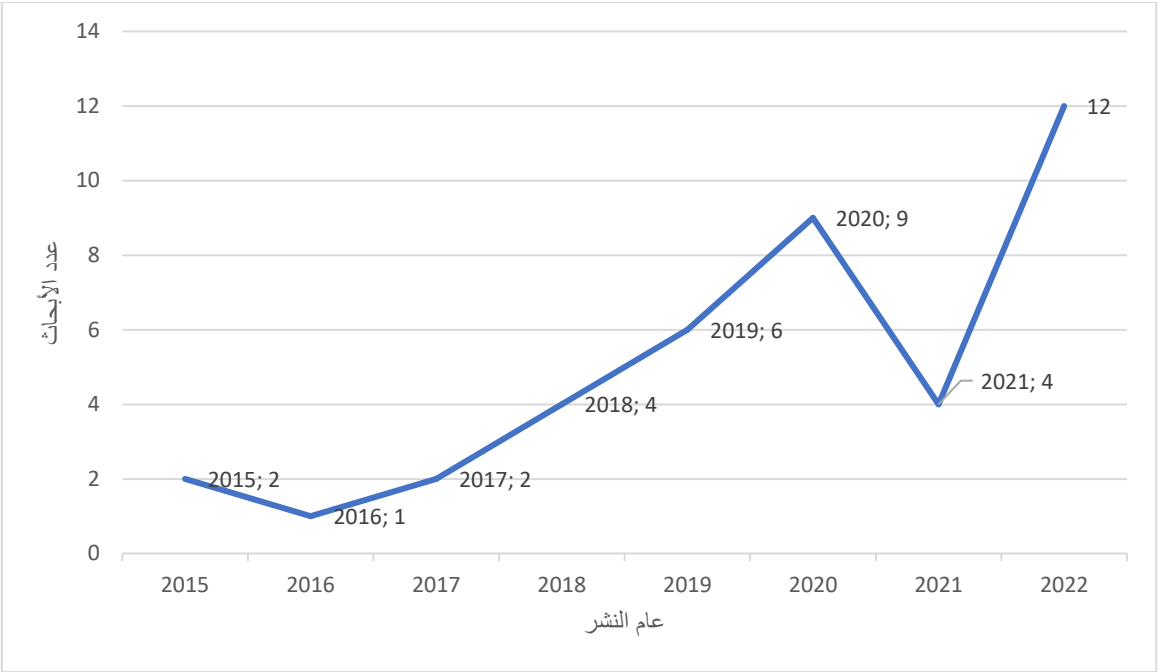
يحتوي الملحق /ملحق 1/ على قائمة بهذه الأوراق البحثية مع المعلومات العامة. كما يحتوي الملحق /ملحق 2/ على تفاصيل الحلول المقدمة في هذه الأبحاث.

2-2-3- نتائج الدراسة المرجعية

نقدم في هذه الفقرة النتائج التي توصّلنا إليها اعتمادًا على تحليل الأوراق البحثية، محاولين تقديم الإجابات على الأسئلة التي طرحناها سابقًا في الفقرة 2-2-.

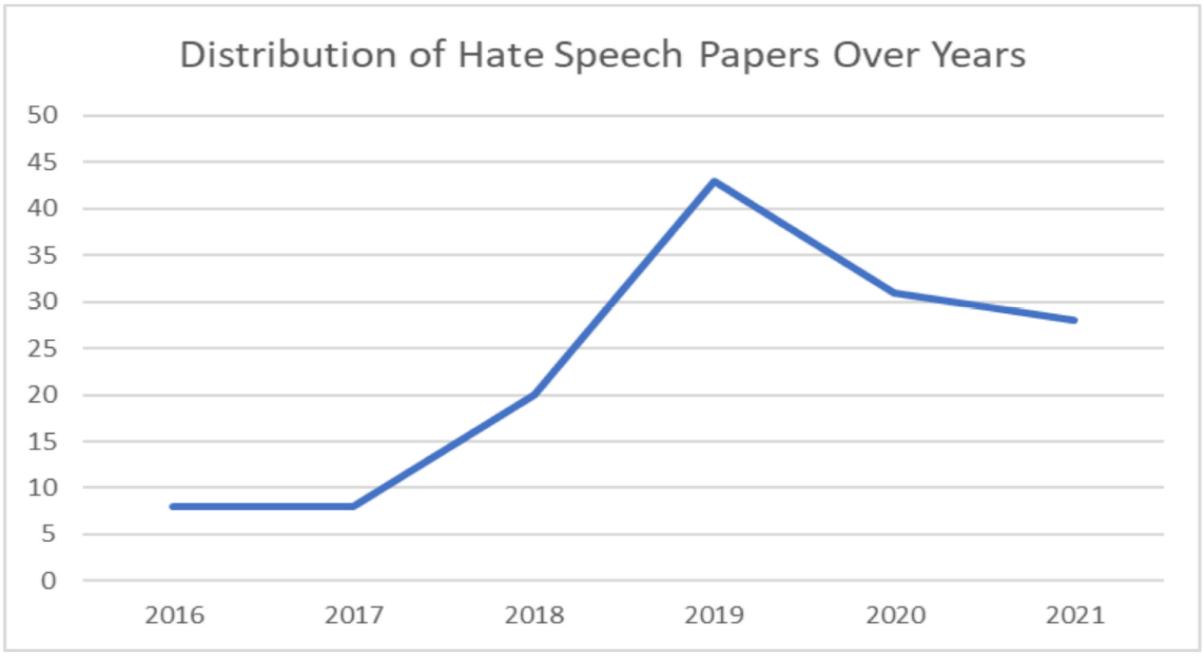
2-2-3-1- Q1.1 ما هو الاتجاه العام لتوزع هذه الأبحاث؟

يبين الشكل 2-1 توزع هذه الأبحاث وفقًا لعام النشر، حيث يبين عدد الأبحاث ذات الصلة في كل عام بين عامي 2015 و2022. يعكس المخطط البياني اهتمامًا متزايدًا بمسألة الكشف عن خطاب الكراهية ولا سيما منذ العام 2019. كما نلاحظ أنّ هذه المسألة جرى تناولها حديثًا بدءًا من العام 2017 إذا اعتبرنا أن الدراسات [36] [37] [38] تتناول مسألة دعم التّنظيمات الإرهابية.



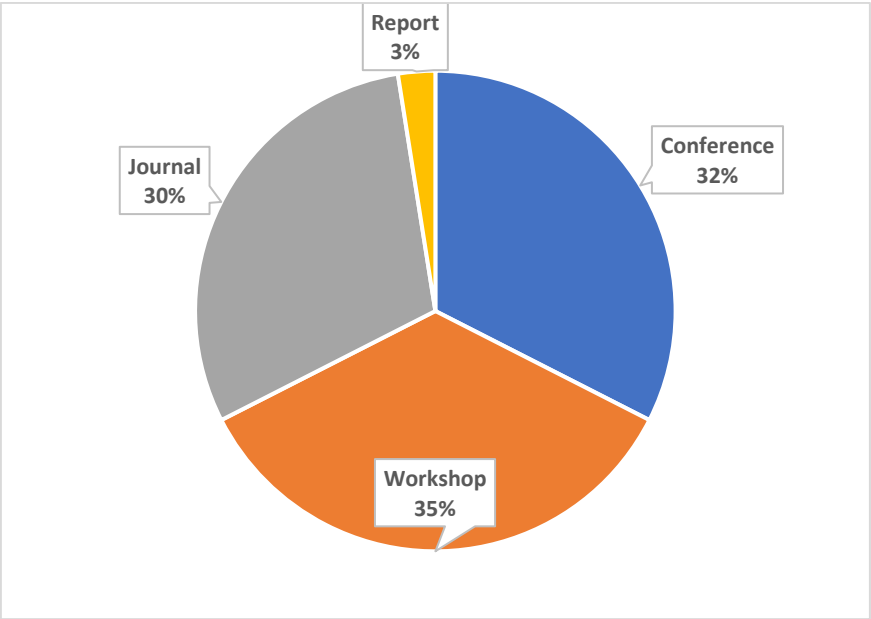
الشكل 2–1 أعداد الأبحاث التي تناولت موضوع كشف خطاب الكراهية حسب الأعوام

ويتوافق هذا مع الاهتمام العالمي لهذه المسألة وإن كان بدرجة أقل من حيث عدد الأبحاث، حيث يبين الشكل 2–2 توزع الأوراق البحثية عالميًّا التي تناولت مسألة كشف خطاب الكراهية باللغات المختلفة بين عامي 2016 و2021 [39]:



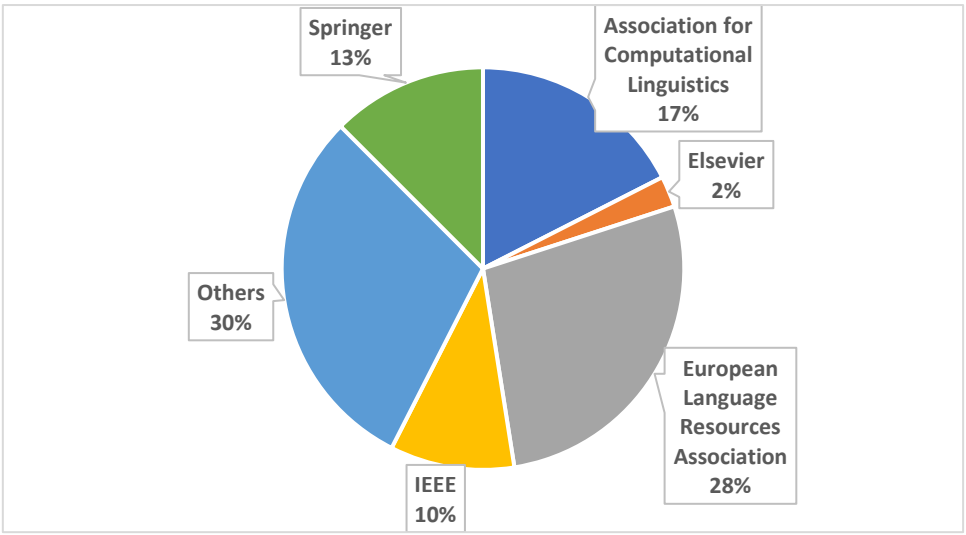
الشكل 2–2 توزع أعداد الأوراق البحثية عالميًّا التي تناولت موضوع خطاب الكراهية باللغات المختلفة بين عامي 2016 و2021

كما يبين الشكل 2–3 توزع هذه الأبحاث وفق طبيعة النشر، حيث نلاحظ توزع الأبحاث المنتقاة بين 32% في مؤتمرات محكمة و35% في ورش عمل مختصة بهذه المسألة و30% في مجلات محكمة.



الشكل 2–3 توزع أعداد الأوراق البحثية التي تناولت موضوع خطاب الكراهية باللغة العربية وفق طبيعة النشر

كما يبين الشكل 2–4 توزع هذه الأبحاث وفق الناشرين، حيث نلاحظ توزع الأبحاث المنتقاة بين أهم الناشرين.



الشكل 2–4 توزع أعداد الأوراق البحثية التي تناولت موضوع خطاب الكراهية باللغة العربية وفق الناشر

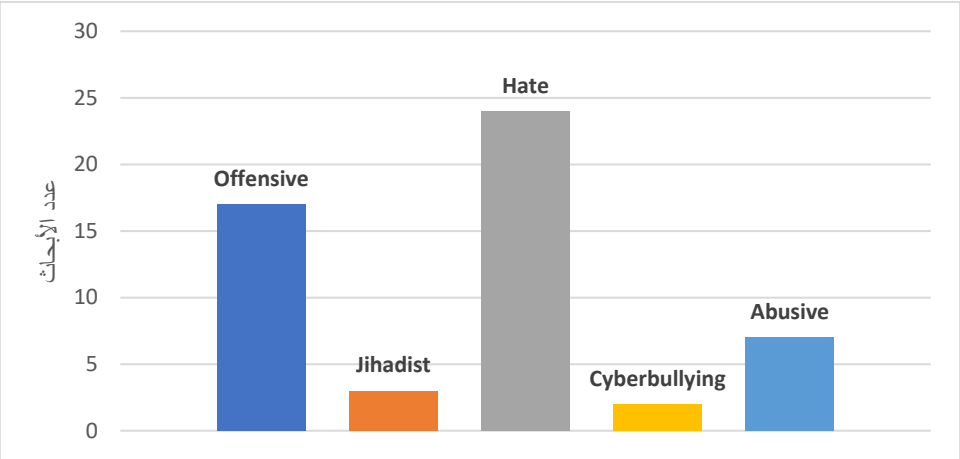
2–2–3–2– Q1.2 ما هي أنواع خطاب الكراهية المستهدف؟

جرى تصنيف الأبحاث السابقة وفقًا لنوع خطاب الكراهية المستهدف وفق الأصناف التالية:

- الكلام المسيء Abusive.

- التنمر الإلكتروني Cyberbullying.
- الكراهية Hate.
- دعم الجماعات الإرهابية Jihadist.
- الخطاب العدائي Offensive.

يبين الشكل 2–5 تصنيف الأبحاث وفق نوع الخطاب المستهدف بكل بحث، والذي يمكننا من خلاله ملاحظة الاهتمام المتزايد بكشف خطاب الكراهية.



الشكل 2–5 توزع أعداد الأوراق البحثية التي تناولت موضوع خطاب الكراهية باللغة العربية وفق نوع الخطاب

i. الكلام المسيء Abusive:

جرى تناول مسألة الكلام المسيء في البحث [36] من خلال تصنيف محتوى الحسابات المسيئة التي تستخدم أكثر من ثلاث hashtags وتنشر تغريدات أكثر من الحسابات العادية وتستخدم الكلمات المهينة insulting. كذلك، اعتبر الباحثون في [40] أن الكلام المسيء وخطاب الكراهية شيئان مختلفان، وجرى تعريف الكلام المسيء على أنه التّعليقات التي تضمّ محتوى عدائيًّا أو هجوميًّا، أو كلامًا نابيًا.

ii. دعم الجماعات الإرهابية Jihadist:

جرى تناول مسألة دعم الجماعات الإرهابية مثل "داعش" في البحثين [37] [38] من خلال الكشف عن التغريدات الداعمة أو المناهضة لهذه التنظيمات، وفي البحث [41] جرى تناول مسألة خطاب الكراهية الصادرة عن أشخاص منتمين أو مؤيدين لهذه الجماعات.

iii. التنمر الإلكتروني Cyberbullying:

جرى تناول مسألة التنمر الإلكتروني في بحثين فقط [42] [43] من خلال اعتبار استخدام الإنترنت أو الهواتف المحمولة أو غيرها لإرسال أو نشر نصوص أو صور مخصّصة لإيذاء أو إحراج شخص أو مجموعة أشخاص.

iv. الخطاب العدائي Offensive:

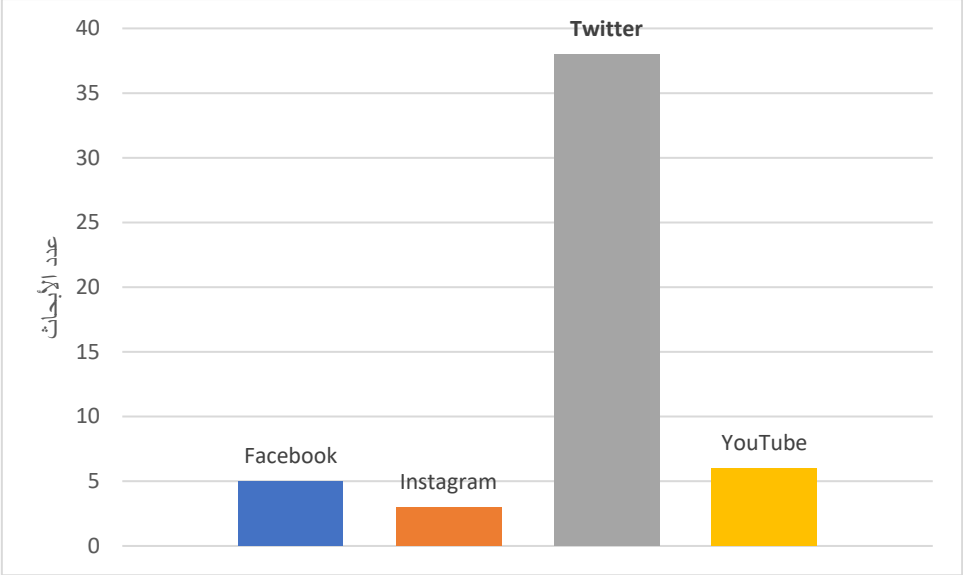
جرى تناول مسألة الخطاب العدائي في الأبحاث [44] [45] [46] [47] [48] [49] [50] من خلال الكشف عن الخطاب الذي يحوي الشتائم أو الإهانة أو الوعيد. بينما جرى تناول مسألة الخطاب العدائي في عدة أبحاث أخرى على أنه مختلف عن خطاب الكراهية الذي يحوي تحريضًا على القتل أو النبذ [18] [19] [20] [21] [22] [23] [24] [25] [26].

v. الكراهية Hate:

جرى تناول مسألة خطاب الكراهية في العديد من الدراسات. حيث تناول الباحثون في [6] [7] مسألة خطاب الكراهية الديني، بينما تناول الباحثون في [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] مسألة خطاب الكراهية بشكل عام. كما تناول الباحثون مسألة خطاب الكراهية مع الخطاب العدائي أو الكلام المسيء [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29].
نلاحظ من هذه الدراسات، تركز البحث حول مسألة الكشف على خطاب الكراهية بشكل أساسي وبدرجة أقل على الخطاب العدائي.

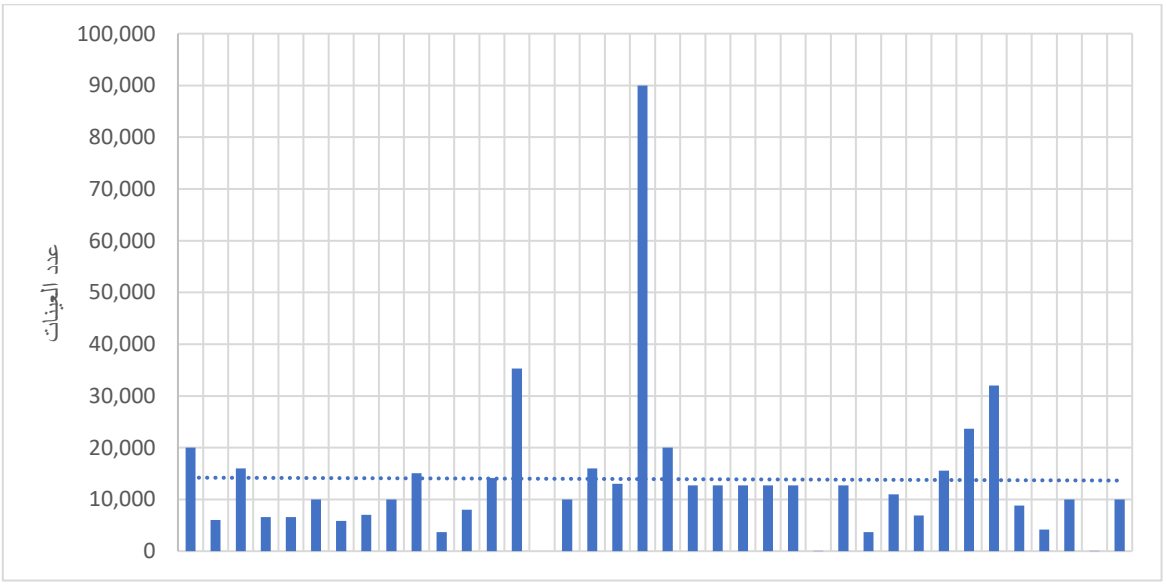
2-2-3-3- Q1.3 ما هي مصادر البيانات وأنواع اللهجات المستخدمة؟

تتيح معظم مواقع شبكات التواصل الاجتماعي الوصول إلى البيانات من خلال واجهة برمجة التطبيقات Application Programming Interface API خاصة بكل منها، ولكن نلاحظ اعتماد أغلب الأوراق البحثية في هذا المجال على موقع تويتر كما هو مبين في الشكل 2-6:



الشكل 2–6 أعداد الأبحاث التي تناولت خطاب الكراهية باللغة العربية وفق مصدر البيانات

يبين الشكل 2–7 حجوم مجموعات البيانات المستخدمة في الأوراق البحثية:



الشكل 2–7 حجوم مجموعات البيانات المستخدمة في الأوراق البحثية التي تناولت خطاب الكراهية باللغة العربية

نلاحظ أن حجم مجموعة البيانات في أغلب الحالات لا يتجاوز 20k عينة، ويعود ذلك إلى كلفة وسم مجموعات البيانات الكبيرة. نلاحظ وجود مجموعة بيانات وحيدة ذات حجم كبير نسبيًا [41] استُخدِمت للكشف عن دعم التنظيمات الإرهابية.

قام عدد قليل من الباحثين بتحديد اللهجة المستخدمة في مجموعة البيانات المستهدفة. فقد استخدم الباحثون [28] اللهجة التونسية، بينما استخدم الباحثون [29] اللهجة المشرقية Levantine. كذلك استخدم الباحثون [27] [40] اللهجة المصرية، بينما استخدم الباحثون [17] اللهجة السعودية. أما الباحثون في [10] [1]

فقد اعتمدوا على مجموعة بيانات من عدة لغات من بينها العربية والإنكليزية. يبين الشكل 2–8 توزع أعداد الأوراق البحثية وفق اللهجة المستخدمة.

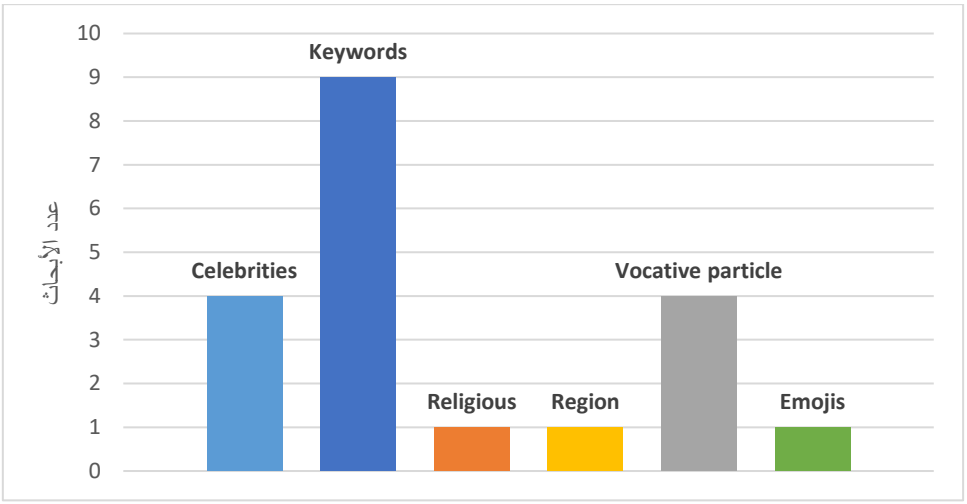
الشكل 2–8 توزع أعداد الأوراق البحثية وفق اللهجة المستخدمة

من جهة أخرى، قام الباحثون –كما ذكرنا سابقًا– بإنشاء مجموعات البيانات الخاصة بهم من مواقع شبكات التواصل الاجتماعي المختلفة، ولكن تباينت كيفية اختيار العينات ضمن مجموعات البيانات. فمثلًا، قام الباحثون بانتخاب العينات من مواقع وحسابات لمشاهير العرب في السياسة والفن والصحافة أو لحسابات شخصيات مثيرة للجدل [38] [47] [27] [29]. بينما اعتمد الباحثون [28] [46] [8] [36] [37] [10] [41] [23] على كلمات مفتاحية كأن تكون مفردات عدائية أو تخص مواضيع مثيرة للجدل. كذلك قام الباحثون في [6] [7] باستخدام نفس مجموعة البيانات والتي جرى انتخاب العينات فيها من حسابات لشخصيات دينية. بينما اعتمد الباحثون في [42] على المنطقة الجغرافية التي تشمل الشرق الأوسط بدائرة نصف قطرها 10,000كم. بينما اختار الباحثون [51] العينات التي تحتوي أداة النداء "يا". كذلك قام الباحثون [52] باعتماد التغريدات التي تحتوي الرموز التعبيرية Emojis، حيث جرى استخلاص الرموز التعبيرية التي ظهرت بكثرة ضمن الخطاب العدائي في مجموعتي البيانات [51] [7] وأُخِذَ تغيراتها من موقع https://emojipedia.org والتي جرى تلخيصها في الشكل 2–9.

Anger/Disgust	
Animals	
Inanimate things	
Disrespect/ Dislike	
Violence/Threat	

الشكل 2–9 الرموز التعبيرية المستخدمة ضمن خطاب الكراهية باللغة العربية

يبين الشكل 2–10 أعداد مجموعات البيانات المستخدمة في الأوراق البحثية وفق طريقة انتخاب العينات. نذكر هنا أن بعض هذه المجموعات قد استخدمت في أكثر ورقة بحثية ولكن مع اختلاف هدف البحث.

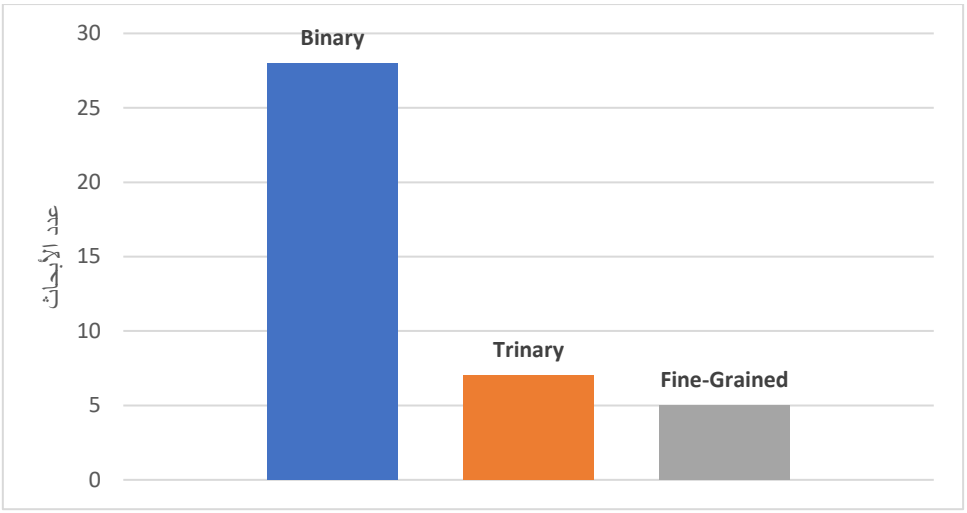


الشكل 2–10 أعداد مجموعات البيانات المستخدمة وفق طريقة انتخاب العينات للأبحاث التي تناولت مسألة كشف خطاب الكراهية باللغة العربية

2–2–3–4– Q1.4 ما هي تصنيفات مجموعات البيانات المستخدمة؟ وما هي نسب توزع الصفوف؟

تؤول مسألة الكشف عن خطاب الكراهية أو الكلام المسيء أو العدائي إلى مسألة تصنيف النصوص، منها ما هو تصنيف ثنائي كما في أغلب الحالات أو من ثلاث صفوف مثل [48] [11]، واتجه البعض إلى وضع صفوف فرعية خاصة تعرف خطاب الكراهية أو المسألة المطروحة مثل [14] [15].

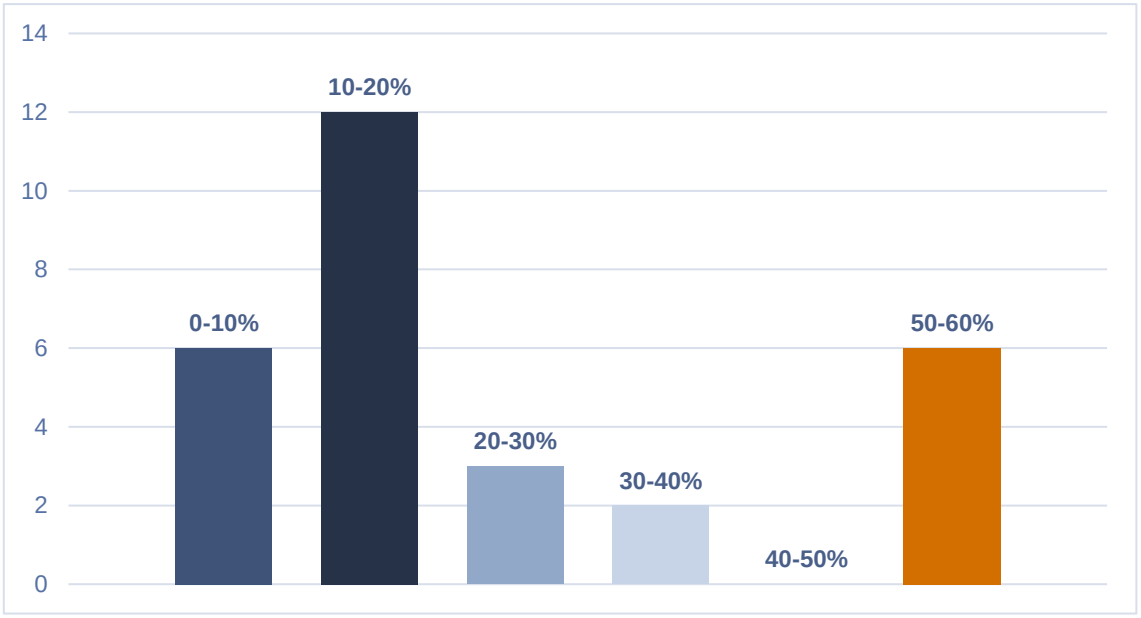
يبين الشكل 2–11 توزع أعداد الأبحاث المنتقاة وفق عدد صفوف النص.



الشكل 2–11 توزع أعداد الأبحاث التي تناولت خطاب الكراهية باللغة العربية وفق عدد الصفوف

تتسم معظم مجموعات البيانات التي تتناول مسألة كشف خطاب الكراهية بأنها غير متوازنة imbalanced، حيث نلاحظ بسهولة أن أغلب هذه المجموعات لا تتجاوز فيها نسبة صفوف الكراهية 11%.

يبين الشكل 2–12 توزع أعداد الأبحاث وفق نسبة توزع الصفوف.



الشكل 2–12 توزع أعداد الأبحاث التي تناولت خطاب الكراهية باللغة العربية وفق نسبة توزع الصفوف

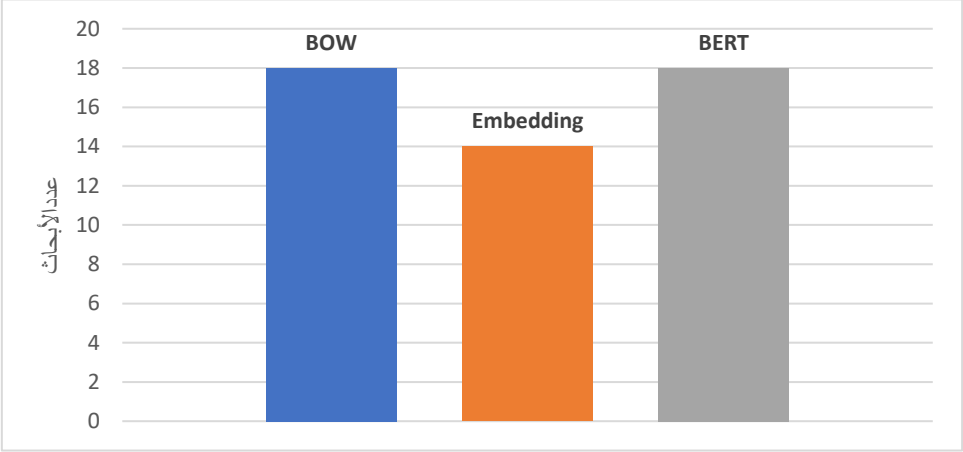
2–2–3–5– **Q1.5** ما هي التّمثيلات المقترحة لتمثيل النصوص؟

تجري عملية تمثيل النص عبر استخراج السمات feature extraction من خلال تحويل التغريدة إلى تمثيل رقمي.

يبين الشكل 2–13 توزع أعداد الأبحاث المنتقاة وفق نوع تمثيل النصوص، حيث جرى اعتماد تمثيل النصوص وفق ثلاثة أنواع أساسية وهي:

- **التمثيل المفرداتي:** ويضم نموذج حقيبة الكلمات BOW، تردد المفردة وتردد التغريدة المعكوسTF-IDF، n-gram، حيث استُخدِم هذا التمثيل بفعالية لمسألة الكشف عن خطاب الكراهية [53].
- **تضمين الكلمات غير السياقي:** مثل Word2vec ومشتقاتها.
- **تضمين الكلمات السياقي:** مثل BERT ومشتقاتها.

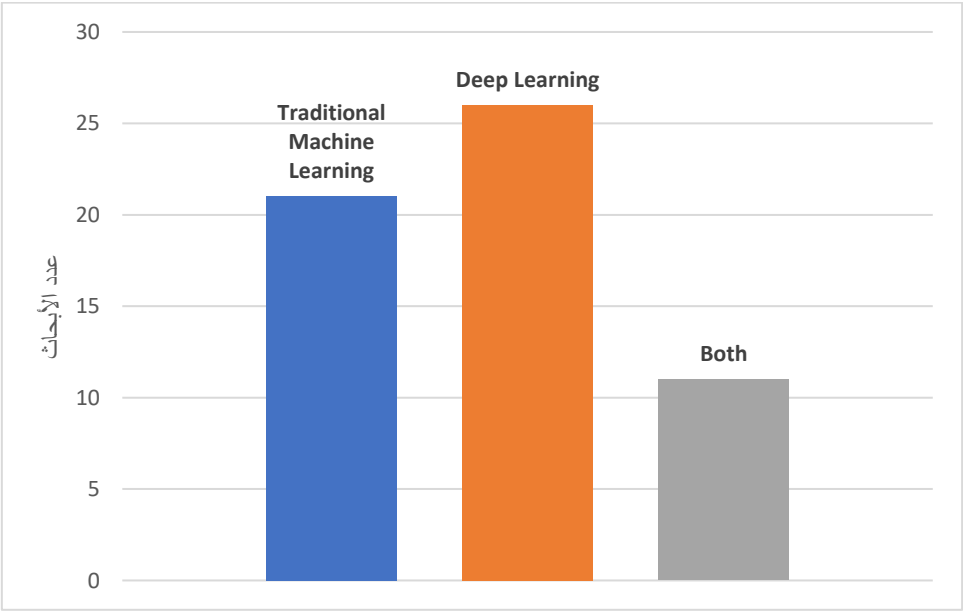
نلاحظ من الشكل 2–13 ازدياد الاهتمام باستخدام تضمين الكلمات السياقي.



الشكل 2–13 أعداد الأبحاث التي تناولت خطاب الكراهية باللغة العربية وفق طريقة تمثيل النصوص

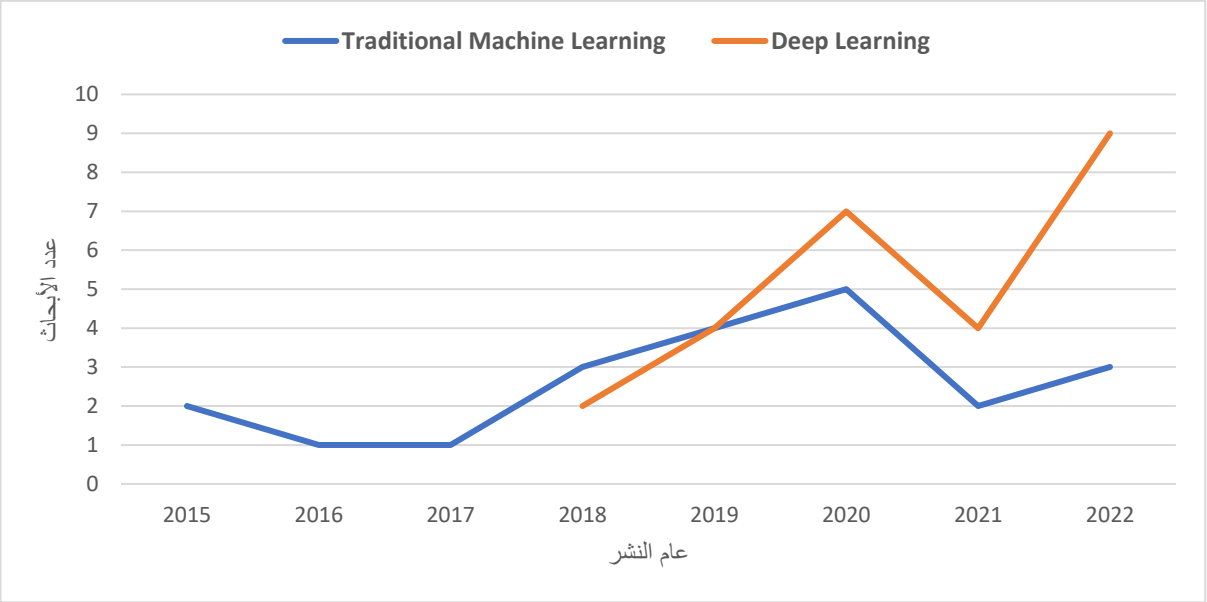
2–2–3–6– Q1.6 ما هي المصنفات المستخدمة؟

جرى استخدام العديد من خوارزميات تعلم الآلة التقليدية وخوارزميات التّعلم العميق على حد سواء، وقام بعض الباحثين باستخدام النوعين معًا وذلك لمقارنة أداء وتقييم النوعين، كما هو مبين في الشكل 2–14.



الشكل 2–14 توزع أعداد الأبحاث التي تناولت كشف خطاب الكراهية باللغة العربية بين تعلم الآلة التقليدي والتّعلم العميق

ونلاحظ الاهتمام المتزايد باستخدام خوارزميات التّعلم العميق في الأعوام الأخيرة، كما هو مبين في الشكل 2–15.



الشكل 2–15 توزع أعداد الأبحاث التي تناولت كشف خطاب الكراهية باللغة العربية بين تعلم الآلة التقليدي والتّعلم العميق حسب الأعوام

2–2–1–3– Q1.7 ما هي توجهات الأبحاث التي تناولت كشف خطاب الكراهية باللغات المختلفة؟

استخدمت الأبحاث التي تناولت مسألة كشف خطاب الكراهية باللغات الأخرى ولا سيما اللغة الإنكليزية خوارزميات التعلم التقليدي مثل Support Vector Machine (SVM), Logistic Regression وغيرها، وخوارزميات التعلم العميق مثل الشّبكات التّلافيفية Convolutional Neural Networks (CNN). يمكن القول إن الاتجاه العام في الأدبيات الحالية للكشف عن خطاب الكراهية يتجه نحو أساليب التعلم العميق [54]. استخدمت هذه الأبحاث عدة طرق لتمثيل النصوص [54]:

- التمثيل المفرداتي: مثل نموذج حقيبة الكلمات BOW، تردد المفردة وتردد التغريدة المعكوس-TF IDF، n-gram.
- تعميم الكلمات word generalization: حيث يجري تجميع الكلمات ضمن عناقيد clusters ومنح كل عنقود خاصية معينة.
- الموارد المعجمية lexical resources: بالاعتماد على قائمة بالمفردات التي تشير إلى خطاب كراهية معينة.
- سمات لغوية مهمة important linguistic features: والتي تمثل الترابط بين الكلمات.

- سمات مبينة على المعرفة knowledge-based features: والتي تتضمن معلومات السياقية للكلمات.
- المعلومات الوصفية meta-information: كالجنس وعدد المتابعين.
- معلومات من أنماط أخرى مثل الصور أو الصوت أو الفيديو.

تواجه الباحثين في هذه المسألة مجموعة من التحديات منها:

- قابلية التعميم generalizability: حيث أظهرت الدراسات السابقة أن نماذج كشف خطاب الكراهية لا تعمل بشكل جيد عند تطبيقها على مجموعات بيانات من خارج التوزيع الخاص بمجموعة البيانات المدرب عليها [55] [56] [57].
- عدم توازن البيانات: حيث تمثل العينات الممثلة لخطاب الكراهية نسبة قليلة جدًا من مجموعة البيانات ما يجعل هذه النماذج عرضة للملاءمة الزائدة overfitting [55] [56].
- الانحياز إلى كلمات محددة: حيث تكون النماذج تلائم بشكل أكبر الكلمات التي تتواتر بكثرة في مجموعة البيانات، ما يجعلها منحازة [57] [58].
- مجموعات بيانات التدريب: تؤثر مجموعات البيانات المستخدمة في تدريب نماذج الكشف بشكل كبير على التطبيق العملي لهذه النماذج [59] [58].

2-2-3-2- Q1.8 ما هي الثغرات والتوجهات المستقبلية التي يمكن العمل عليها لتحسين النّتائج؟

تظهر النتائج التي حصلنا عليها في هذه الدراسة المرجعية تقدمًا ملحوظًا للآليات المعتمدة في تمثيل النصوص في الأعوام السابقة وازدياد الاعتماد على خوارزميات التعلم العميق. يمكن القول إن استخدام تقنيات تضمين الكلمات السياقي وغير السياقي قادر على تحقيق نتائج واعدة لمعظم مسائل تصنيف النصوص بشكل عام ومسألة الكشف عن خطاب الكراهية بشكل خاص.

يمكن تلخيص الثغرات والمشاكل التي تعاني منها مسألة الكشف عن خطاب الكراهية بالنقاط التالية:

- **ندرة مجموعات البيانات الموسومة:** حيث نلاحظ وجود عدد قليل جدًا من مجموعات البيانات الخاصة بمسألة الكشف عن خطاب الكراهية باللهجة المشرقية.
- **خصوصية مجموعات البيانات الحالية:** نلاحظ أن مجموعات البيانات الحالية ذات طبيعة خاصة بكل مجموعة، فمنها ما هو سياسي فقط، أو مأخوذة من حسابات أو صفحات لبعض مشاهير العرب.
- **حجوم مجموعات البيانات الحالية:** تُعتبَر حجوم مجموعات البيانات الحالية صغيرة نسبيًا، ولا سيّما عند استخدام خوارزميات التّعلم العميق المتعطِّشة للبيانات.

- **عدم توازن مجموعات البيانات:** نلاحظ أن أغلب مجموعات البيانات الحالية غير متوازنة، وهو ما يؤثر على أداء الخوارزميات المستخدمة. يعود ذلك طبعًا إلى طبيعة المسألة المدروسة.

يمكننا تلخيص النتائج في النقاط التالية:

- لوحظت زيادة كبيرة في عدد الأبحاث التي تعالج مسألة تصنيف النصوص خلال الأعوام السابقة، مع زيادة أكثر وضوحًا في الأعوام الثلاثة الماضية.
- حقق استخدام تقنيات التّعلم العميق لمعالجة خطاب الكراهية نتائج واعدة في معظم المسائل التي تصدّت لها هذه الأبحاث.
- يوجد حاجة كبيرة إلى مقاربات أكثر مرونة قادرة على التعامل مع مجموعات بيانات غير متوازنة، وقادرة على التعامل مع مشاكل النصوص المكتوبة باللهجة المحلية.
- لا يزال عدد الأبحاث التي تتناول مسألة خطاب الكراهية باللغة العربية (وتحديدًا المكتوبة باللهجة المشرقية) محدودًا، ولا سيما عند مقارنته مع مثيلاتها باللغة الإنكليزية.

تظهر هذه النتائج الحاجة لمزيد من الأبحاث من أجل تحسين أداء نظم كشف خطاب الكراهية المكتوب باللغة العربية وخاصة باللهجة المشرقية، ويمكن العمل لاحقًا في عدة محاور:

- تأمين مجموعات بيانات كبيرة ومتوازنة نوعًا ما، الأمر الذي يقودنا إلى تقنيات التّعلم النشط في محاولة إضافة عينات أكثر إلى مجموعات البيانات وتحسين جودة أداء المصنفات المستخدمة. لذلك، سنتناول في القسم الثاني من هذا الفصل دراسة مرجعية لبعض الأبحاث الخاصة التي تناولت موضوع التعلم النشط.
- الأخذ بالاعتبار إمكانية استخلاص سمات أخرى من غير النص، كالاعتماد على سمات التغريدة نفسها، أو سمات صاحب التغريدة.

2-2-4- الخلاصة

قدّمنا في الفقرات السابقة مراجعةً لمسائل معالجة خطاب الكراهية للنصوص العربية، حيث جرى دراسة 40 ورقة بحثية وتحليلها وتصنيفها وفق مهمة الكشف والطريقة المعتمدة في تمثيل النصوص. تشير النتائج إلى وجود عدد متزايد من الأبحاث التي تعتمد التّعلم العميق، ولا سيّما في الأعوام القليلة السابقة. كذلك، جرى تلخيص اتجاهات الأبحاث الأساسية وتحديد الثغرات والتوجهات المستقبلية التي يمكن العمل عليها.
يمكن أن تساعد البيانات المستخلصة من الأوراق البحثية كمنطلق لأبحاث مستقبلية في مجال معالجة خطاب الكراهية للنصوص العربية، ويمكن اعتبار النتائج والخلاصات في هذا الفقرة نقطة انطلاق جيدة لهذه الأبحاث.

2-3- الدراسة المرجعية للتعلم النشط

مع ازدياد أعداد مستخدمي شبكات التواصل الاجتماعي، والحرية الممنوحة لهم لمشاركة أخبارهم ومشاعرهم، تنتج هذه الشبكات حجومًا ضخمةً من البيانات النصية وغيرها. ما يجعل مسألة تصنيف النصوص أكثر أهمية من ذي قبل، ولا سيما عند التعامل مع مسألة ككشف خطاب الكراهية. يعتمد نجاح أنظمة التصنيف بشكل كبير على جودة مجموعة التدريب المستخدمة. لكن بناء هذه المجموعة يتطلب عددًا كبيرًا من العينات الموسومة، وهو ما يمكن أن يكون مكلفاً من حيث الجهد والزمن.

يجري قياس أو تقييم هذه النماذج من خلال مصفوفة الالتباس confusion matrix وحساب معاملات التقييم مثل F1-score، ومع أن هذه المقاييس تزودنا بمعلومات مهمة عن أداء النموذج بشكل عام، إلا أنها لا تعطينا درجة الثقة بهذا النموذج بالنسبة لمنطقة محددة من فضاء توزع العينات الذي يمكن أن يضم في منطقة ما عينات بدرجة احتمالية عالية ومنطقة أخرى تضم عينات بدرجة احتمالية منخفضة. في كل الحالات، يمكن القول إن الحصول على مجموعات بيانات كبيرة وموسومة يعاني من المشاكل التالية:

- عملية الوسم اليدوي معرضة للخطأ البشري.
- عملية الوسم مكلفة من حيث الزمن والجهد.
- لا تتمتع جميع العينات بنفس الدرجة من الأهمية.

يعتبر التّعلم النشط حلًّا لهذه المشاكل [60] [61] [62]. نقدم في الفقرات التالية تعريفًا بالتّعلم النشط واستراتيجياته وطرقه، بالإضافة إلى دراسة بعض الأبحاث التي تناولت مسألة التّعلم النشط وتصنيفها وفق آلية انتخاب العينات، أو وفق الاستراتيجيات المعتمدة، بالإضافة إلى سبر أهم الثغرات والإشكالات التي تعاني منها هذه الأبحاث.

2-3-1- هدف التّعلم النشط

تهدف عملية تصنيف النصوص إلى إيجاد الدالة $y = f(X)$ التي تحقق أكبر دقة معينة من خلال ربط السمات مع الوسم [63]. لكن تعاني مفاهيم التّعلم التّقليديّة من معضلتين أساسيتين: تكمن الأولى في أن الوصول إلى عتبة أداء معينة يتطلب عددًا كبيرًا من العينات الموسومة وهو أمرٌ مكلف ولا سيما عندما تكون البيانات غير متوازنة، وتكمن الثانية في صعوبة إيجاد التّوزع distribution الذي يمثل البيانات بشكل فعال [64]. يضاف إلى ذلك صعوبة الوسم اليدوي وتكرار العينات المتشابهة، الأمر الذي لا يسهم في تحسين عملية التّعلم [60]. يمكن القول إن التّعلم النشط هو مفهوم تعلم آلة يحاول تقليل التدخل البشري في عملية وسم البيانات من خلال استخدام استراتيجيات انتخاب العينات المرشحة للوسم والتي تساهم في تحسين نموذج تعلم الآلة [65]. أثبتت الدراسات [66] في حالات عديدة، أن عدد العينات الموسومة التي نحتاجها في التّعلم النشط ينخفض بشكل لوغاريتمي مقارنةً مع مثيلها في التّعلم الخامل passive

learning. كما يساعد التّعلم النشط في التخلص من عينات الضجيج، الأمر الذي يساهم في تحسين دقة النموذج [67]، حيث أثبتت بعض الدراسات [68] أن تصميم طرق تعلم نشط بعناية يساهم في تحسين الصحة accuracy بشكل أفضل. كما يقدم التّعلم النشط الوسائل لتخفيف الكلفة والجهد المرتبط بالوسم اليدوي اللازم مع الحفاظ على نفس الدقة [60]، حيث أظهرت بعض الدراسات السابقة [61] [62] أن التّعلم النشط يمكن أن يقلّل من عدد العينات الموسومة اللازمة لبناء مصنفات نصوص دقيقة بنسبة تصل إلى 90%، ما يجعل عملية بناء نظم تصنيف نصوص تتطلب عددًا ضخمًا من العينات الموسومة أمرًا قابلًا للتحقيق [69].

تكمن الفرضية الأساس للتعلم النشط في أنه لو أتيح لخوارزمية التّعلم أن تنتقي بياناتها التي ستتعلم منها لكان أداؤها أفضل وبمجموعة تدريب أقل من خلال أخذ القرار في تحديد العينات المراد وسمها، لا سيما أن البيانات غير الموسومة غزيرة وسهلة المنال [65]. مع ذلك، فإن الحصول على استراتيجية تعلم نشط فعالة وشاملة لأية مهمة أمر مستحيل [70].

تبدأ نماذج التّعلم النشط بالتدرب على مجموعات صغيرة من البيانات، وتقرر دالة تحصيل acquisition function أي العينات التي ستمرر لخبير oracle ليقوم بوسمها وإضافتها إلى مجموعة التدريب لتعاد عملية التدريب من جديد ويزداد حجم مجموعة البيانات مع مرور الوقت [71]. تزوَّد خوارزمية التّعلم غالباً بمخزن كبير من العينات غير الموسومة مع القدرة على طلب وسم أية مجموعة عينات بطريقة تكرارية [65] [72]، على عكس التّعلم الخامل حيث يكون اختيار العينات عشوائيًّا الأمر الذي قد يؤدي إلى مجموعات بيانات غير متوازنة مما يؤثر على صحة accuracy المصنف بشكل كبير [60].

2-3-2- أطر عمل التّعلم النشط

يمكن تمييز وجود عدة أطر عمل للتعلم النشط لاقت انتشارًا في الأدبيات وفي التّطبيقات [73] [74]، وهي وفقًا لـ Settles [65] كالتّالي:

- **تأليف أو تركيب العينات Membership Query Synthesis:** حيث يقوم المتعلم بتركيب العينات من فضاء العينات الكلي [75]. تكمن مشكلة هذا النوع في أن العينات المركبة ليست بالضرورة عينة حقيقية.
- **انتخاب العينات الدفقية Stream-based Selective Sampling:** حيث يجري سحب عينة من توزيع العينات الأساسي، ويقوم المتعلم بتحديد فيما إذا كان سيقوم بوسم العينة أم لا [76]. تكمن مشكلة هذا النوع في أن العينات تُسحَب واحدة بعد أخرى ويجري التحقق من إمكانية انتخاب كل عينة على حدة، الأمر الذي يسبب كلفة زمنية كبيرة.

- **انتخاب العينات من مخزن Pool-based Sampling**: حيث يفترض وجود مخزن كبير من العينات المرشحة للانتخاب من أجل عملية الوسم [77]، وهي أكثر انتشارًا من سابقاتها كونها تساهم في اختصار كمية بيانات التّدريب اللازمة للوصول إلى درجة أداء معينة [77]. تبدأ نماذج التّعلم النشط –في هذا النوع– بالتدرب على مجموعة صغيرة من البيانات [78] وتضاف العينات المنتخبة تباعًا إلى مجموعة التدريب هذه.

2-3-3- استراتيجيات التّعلم النشط

يجري انتخاب العينات من خلال دالة تقيس أهمية العينة المحتملة وذلك من خلال قياس الإفادة informativeness أو التّمثيلية representativeness [79] [80]:

- **الإفادة informativeness**: حيث يجري انتخاب العينات التي يتوقع أن تحمل الإفادة الأكبر للنموذج الإحصائي Statistical model من خلال تحديد العينات غير الموسومة ذات درجة عالية من عدم اليقين أو تساعد في تخفيض التباين في النموذج.
  - عدم اليقين uncertainty: يجري تقدير قيمة عدم اليقين بعدة طرق:
    - الهامش الأصغري Smallest Margin [81] [82]: حيث تكمن الفكرة الأساسية في انتخاب العينات المرشحة التي يكون تمثيلها في فضاء العينات قريبًا من مستوي الفصل hyperplane في مصنفات أشعة دعم الآلة SVM، فكلما كانت العينة u_i قريبة من مستوي الفصل أي كلما نتيجة الدالة f قريبة من الصفر اعتبرنا العينة أكثر قبولًا من خلال الدالة I وفق المعادلة التّالية:

(E.1) $$I(u_i) = 1 - |f(u_i)|$$

    - بالاعتماد على الأنتروبية Entropy [83]: وتقاس الأنتروبية حسب الدالة En حيث يرمز P_i إلى احتمالية انتماء العينة Y إلى الصف i وفق المعادلة التّالية:

(E.2) $$En(Y) = -\sum_i P_i * \log(P_i)$$

أو وفق المعادلة التالية:

(E.3) $$En(Y) = \sum_i \|P_i - 0.5\|$$

    - الثّقة المُهمَّة الصّغرى Least Significant Confidence [84]: وتقاس هذه الثقة حسب الدالة ϕ من خلال احتمال انتماء العينة x للصف الأكثر احتمالية y^* وفق محددات النموذج θ حسب المعادلة التالية:

(E.4) $$\emptyset^{LSC}(x) = 1 - P(y^*|x;\theta)$$

- طول التدرج المتوقع Expected Gradient Length (EGL): حيث يجري انتخاب العينات التي يتوقع أن تؤدي إلى تغير أعظمي للنموذج الحالي [83].
- تقليص التباين Variance Reduction: حيث يجري انتخاب العينات التي نحصل من أجلها على أقل قيمة لدالة الفقد Loss Function [85].
- الانتخاب عبر لجنة Query-By-Committee (QBC) [86]: يجري انتخاب العينات من خلال قياس الاختلاف بين مجموعة المصنفات Committee's Classifiers:
  - التّباين الوسطي Average Kullback Leibler Divergence [87].
  - مقارنة التّصويت Vote Comparison: يجري انتخاب العينات التي حصلت على تصنيفات أو توقعات مختلفة.
  - أنتروبية التّصويت Vote Entropy [88] [89]: حيث يجري انتخاب العينات التي تحقِّق الأنتروبية العظمى وفق المعادلة التّالية حيث تشير القيمة C إلى عدد أعضاء اللجنة، بينما يشير $V(y_i)$ إلى عدد أعضاء اللجنة الذين أعطوا الصف i للعينة x:

$$x_{VE}^{*} = \underset{x}{\operatorname{argmax}} \left\{ -\sum_i \frac{V(y_i)}{C} * \log \frac{V(y_i)}{C} \right\} \quad \text{(E.5)}$$

- **التّمثيلية representativeness**: التي تقيس كيفية انتخاب العينات التي تعطي التمثيل الأفضل لفضاء العينات، والتي يمكن تحديدها من خلال مراكز كثافة التوزيع means of distribution density [67].
  - التشابه similarity: من خلال قياس وسطي تشابه العينات من خلال الدالة K التي تقيس البعد بين عينتين وفق المعادلة التالية:

$$R(u_i) = \frac{\sum_{j \neq i} K(u_i, u_j)}{n-1} \quad \text{(E.6)}$$

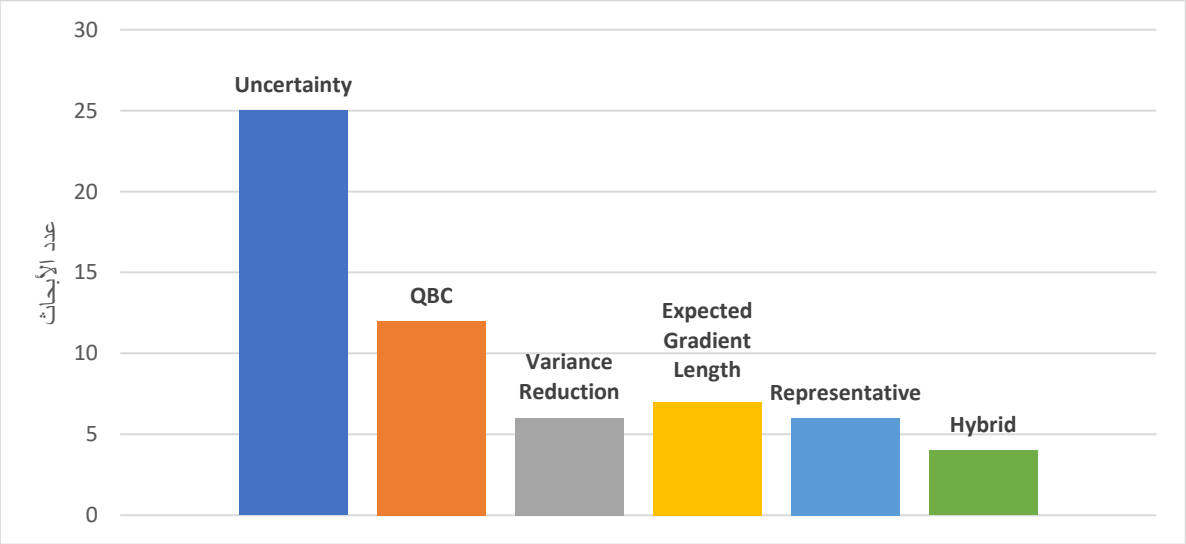
  - العنقدة Clustering: من خلال تقسيم مجموعة البيانات إلى عدة عناقيد وأخذ مراكز هذه العناقيد.

2-3-4- الأبحاث ذات الصلة

نعرض في هذه الفقرة بعض الأبحاث ذات الصلة التي تناولت مسألة التّعلم النشط ولا سيما ما يخص تصنيف النصوص والتي بلغت 60 بحثًا. يحتوي الملحق /3/ على قائمة بهذه الأوراق البحثية مع المعلومات العامة. عالجت الاستراتيجيات المختلفة وفق ما يلي:

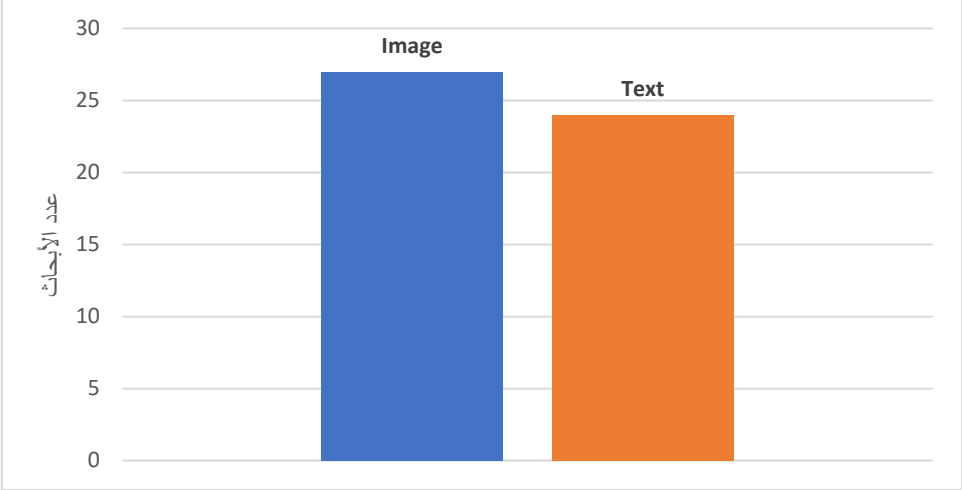
- عدم اليقين: حيث عالجت بعض الأبحاث انتخاب العينات التي تحقق أكبر درجة من عدم اليقين وبطرق مختلفة [68] [77] [83] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101] [102] [103] [104] [105] [106] [107] [62] [108] [109].
- الانتخاب عن طريق لجنة QBC: حيث عالجت بعض الأبحاث انتخاب العينات التي تحقق الاختلاف بين المصنفات بطرق مختلفة [87] [89] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120].
- طول التدرج المتوقع EGL: [121] [122] [123] [124] [125] [126] [127].
- التمثيلية: [128] [129] [130] [131] [132] [133].
- تقليل التباين: [83] [134] [135] [136] [137] [138]
- دمج عدة تقنيات: حيث عالجت بعض الأبحاث انتخاب العينات من خلال دمج تقنيتي الإفادة والتمثيلية [79] [139] [140] [141].

يبين الشكل 2–16 توزع أعداد هذه الأبحاث وفق الاستراتيجية المعتمدة.



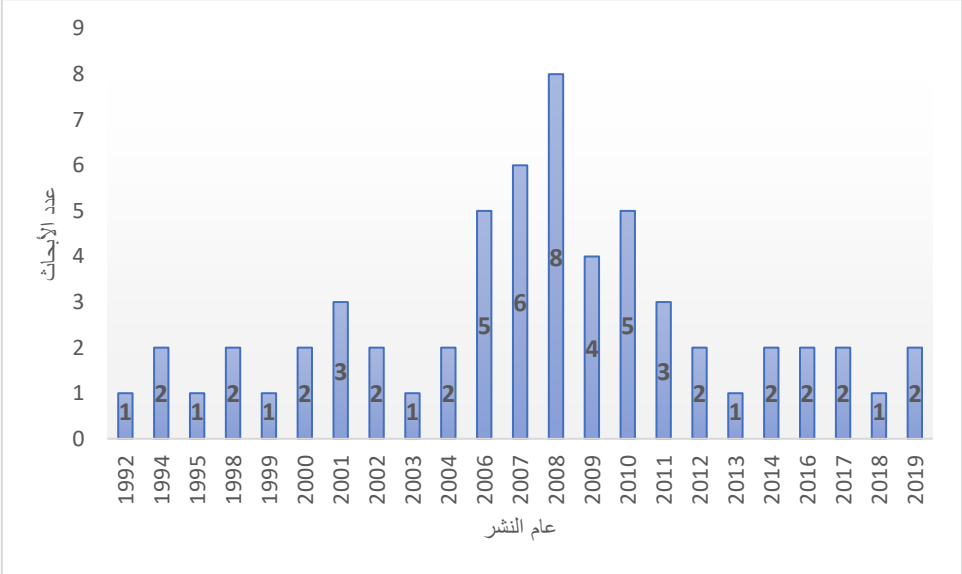
الشكل 2–16 توزع أعداد أبحاث التّعلم النشط وفق الاستراتيجية المعتمدة

نُفِّذت هذه الأبحاث على مجموعات بيانات ذات طبيعة نصّيّة أو صور وفق ما يبين الشكل 2–17.



الشكل 2–17 توزع أعداد أبحاث التّعلم النشط وفق نوع البيانات المعتمدة

يبين الشكل 2–18 توزع أعداد هذه الأبحاث وفق عام النشر.



الشكل 2–18 توزع أعداد أبحاث التّعلم النشط وفق عام النشر

2–3–5– ما هي الثغرات التي يمكن العمل عليها لتحسين عملية انتخاب العينات؟

تظهر النتائج التي حصلنا عليها في هذه الفقرة اهتمام الأبحاث بمسألة التّعلم النشط منذ عام 1992 وحتى تاريخه، الأمر الذي يدل على أهمية هذا الموضوع ولا سيما للمسائل التي تحتاج إلى بيانات بحجوم كبيرة. يمكن تلخيص بعض الثغرات والمشاكل التي تعاني منها مسألة التّعلم النشط عمومًا بالنقاط التالية:

- **عدم توازن مجموعات البيانات:** نلاحظ أن أغلب التقنيات المستخدمة تقوم بانتخاب العينات اعتمادًا على دالة قياس معينة كأن تقيس درجة عدم اليقين، ولكنها لا تنتخب العينات بحيث تخفف من عدم توازن مجموعة البيانات.

- **أيهما أفضل الإفادة أم التمثيلية؟:** إن التركيز على عينات الإفادة قد يجعل نموذج التصنيف ينحاز إلى منطقة ما أو أكثر من فضاء العينات. ينطبق الأمر ذاته عند التركيز على عينات التمثيلية. يلجأ الباحثون عادة إلى دمج التقنيتين.
- **تقدير عدم اليقين بطرق أكثر فعالية:** من خلال البحث عن طرق أخرى لتقدير عدم اليقين.

يمكن أن يكون العمل لاحقًا في المحاور التالية:

- انتخاب العينات وفق دالة القياس المعتمدة، ولكن مع توقع انتماء العينات إلى الصفوف بنسبة تحدد مسبقًا من أجل تخفيف نسبة عدم التوازن في كل مرة.
- الأخذ باعتبار دمج عينات الإفادة والتمثيلية بنسبة تحدد مسبقًا للحفاظ على التوزع المثالي لفضاء العينات قدر الإمكان.
- تقدير عدم اليقين بطرق أكثر فعالية من خلال تقدير التوزيع التنبؤي اللاحق $P(y|x,\theta)$ وليس من خلال قيمة وحيدة، حيث يدل $P(y|x,\theta)$ على احتمال خرج النموذج y للعينة x وفق محددات النموذج θ.

2-3-6- النتائج الرئيسة

نلخص في هذه الفقرة الاستنتاجات الرئيسة في النقاط التالية:

- ما تزال عملية التّعلم النشط منذ بداية هذا المجال مسرحًا هامًا للباحثين كي يستكشفوا المزيد والمزيد.
- حقق استخدام تقنيات التّعلم النشط نتائج واعدة في معظم المسائل التي تصدّت لها هذه الأبحاث من خلال تحسين أداء المصنفات.
- يوجد حاجة كبيرة إلى مقاربات أكثر مرونة قادرة على التعامل مع مجموعات بيانات غير متوازنة، ولا تقود البيانات إلى توزع مختلف عن التوزع الطبيعي لها.
- لا يزال عدد الأبحاث التي تتناول مسألة التّعلم النشط للنصوص العربية محدودًا، ولا سيما عند مقارنته مع مثيلاتها باللغة الإنكليزية.

2-4- الخلاصة

قدّمنا في هذه الفقرة مراجعةً لمسائل التّعلم النشط. جرى دراسة 60 ورقة بحثية وتحليلها وتصنيفها وفق آلية انتخاب العينات. تشير النتائج إلى استمرار الاهتمام بهذا الموضوع ولكنه يحتمل التوسع في العديد من القضايا. كذلك، جرى تلخيص اتجاهات الأبحاث الأساسية وتحديد الثغرات والتوجهات المستقبلية.

2-5- خاتمة

قدّمنا في هذا الفصل الإجابة على السؤال Q1 "**ما هو الوضع الراهن للأبحاث ذات الصلة؟**" من خلال إجراء دراسة مرجعية لمسألة الكشف عن خطاب الكراهية بالإضافة إلى تقديم مراجعة لمسائل التّعلم النشط، حيث جرى دراسة 40 ورقة بحثية تناولت مسألة الكشف عن خطاب الكراهية باللغة العربية، بالإضافة إلى 60 ورقة بحثية تناولت مسألة التّعلم النشط. تشير النتائج إلى استمرار الاهتمام بهذين الموضوعين مع وجود العديد من الثغرات والتوجهات المستقبلية التي يمكن العمل عليها.

3- بناء مجموعة البيانات المشرقية

نقدم في هذا الفصل مجموعة بيانات جديدة باللغة العربية باللهجة المشرقية مستخرجة من موقع تويتر خاصة بخطاب الكراهية، كما نعرض الإجراءات المنفذة على هذه المجموعة كالمعالجة الأولية وتعزيز البيانات وطرق التمثيل المعتمدة. كما نقوم باختبار هذه المجموعة باستخدام مصنفات تقليدية مُدرّبَة على مجموعة بيانات مشرقية متاحة مخصصة لخطاب الكراهية، ثم نعرض نتائج الاختبارات على هذه المجموعة من أجل الإجابة على السؤال Q2 "**هل تعتبر مجموعات البيانات المتوفرة على شبكة الإنترنت كافية؟**".

3-1- مقدمة

تعتبر خطوة تحصيل مجموعة بيانات موسومة الجزء الأكثر كلفةً زمنيًا في عمليات تصنيف النصوص [142]، لذلك يقوم الباحثون الذين يتناولون مسألة الكشف عن خطاب الكراهية غالبًا باعتماد مجموعة بيانات خاصة بهم تُجمَع من مصادر مختلفة، ويقوم بعضهم بوسم annotate هذه البيانات بأحد الصفين: (كراهية hate أو لا كراهية not hate)، أو بأحد الصفوف: (عادي normal، كراهية hate أو شتيمة abusive).

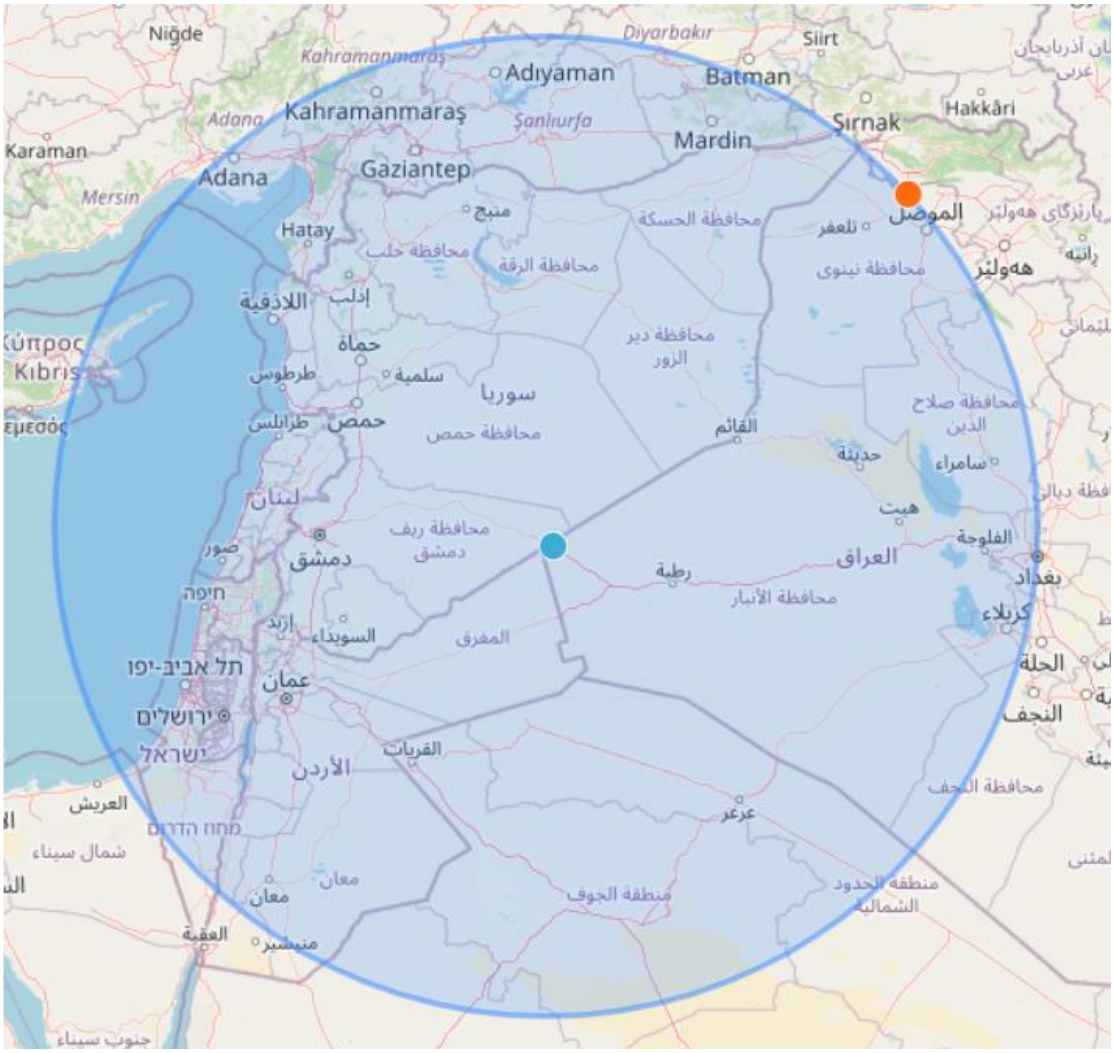
أشرنا سابقًا إلى عدد من مجموعات البيانات في الفقرة 3-3-2-2- ضمن إجابتنا على السؤال Q1.3. يمكن القول إن معظم مجموعات البيانات المتوفرة مُحَصَّلة من صفحات أو حسابات لشخصيات سياسية أو فنية أو دينية أو ناشطة سياسيًا أو مجتمعيًا أو مُحَصَّلة من خلال تحديد كلمات مفتاحية خلال مرحلة البحث. تتوفر من بين هذه المجموعات مجموعة واحدة فقط تخص منطقة سوريا وجوارها. نظرًا لوجود مجموعة بيانات واحدة فقط باللهجة المشرقية وهي ذات طبيعة سياسية ومُحَصَّلة من صفحات قليلة محددة، رأينا أنه لا بد من تأمين مجموعة بيانات نصية باللغة العربية مكتوبة باللهجة المشرقية تكون عامة غير مخصصة قدر الإمكان.

3-2- تحصيل البيانات

جرى تحصيل مجموعة بيانات بالاعتماد على خدمة تويتر لتأمين مجموعة كبيرة نوعًا ما من التغريدات باستخدام Twitter API، ولتحقيق شروط مجموعة البيانات المطلوبة -مجموعة بيانات نصية باللغة العربية مكتوبة باللهجة المشرقية، عامة قدر الإمكان- جرت عملية البحث من خلال إضافة المحددات التالية:

- ✓ نقطة إحداثيات: للحصول على بيانات تخص المنطقة الجغرافية الخاصة بسورية ودول الجوار. جرى تحديد النقطة المركز وبنصف قطر يبلغ 500كم بحيث تشمل السكان الذين يتحدثون اللهجات المشرقية. يمثل الشكل 3–1 المنطقة الجغرافية المستهدفة تقريبًا.
- ✓ اللغة: للحصول على التغريدات المكتوبة باللغة العربية.
- ✓ استبعاد المواقع الإخبارية: مؤقتًا وذلك من أجل استبعاد النصوص الإخبارية المتواترة بكثرة، وذلك من خلال استبعاد الحسابات التي لديها عدد متابعين أكبر من عتبة محددة (اعتمدنا القيمة 1000 كقيمة مبدئية).

لم نقم بتحديد أي كلمات مفتاحية ضمن عملية التحصيل التي تمت بين عامي 2017 و2020 على فترات متقطعة ومتباعدة وجرى تحصيل 21,440 تغريدة، حُفِظَت التغريدات ضمن مجموعة بيانات سميت **LHS-TRAIN-A**.



الشكل 3–1 خريطة تبين المنطقة الجغرافية المستهدفة في تحصيل البيانات من تويتر

3-3- وسم مجموعة البيانات

تعتبر عملية وسم البيانات الخطوة الأكثر صعوبة. جرى وسم مجموعة البيانات بعد تحصيلها أي بتصنيفها ضمن صنفين: إيجابي positive، كراهية من الصف /1/ أو لا كراهية، عادي أو سلبي negative من الصف /0/ من خلال قراءة هذه التغريدات ومحاولة معرفة طبيعة هذه التغريدة ووضعها في أحد الصفين.

يمكن أن يختلف الأشخاص (المُصَنِّفون) فيما بينهم على اعتبار تغريدة ما من هذا الصنف أو ذاك لعدة اعتبارات تعود إلى ميول كل شخص وشخصيته وثقافته أو لاختلاف تعريف خطاب الكراهية فيما بينهم.

اعتمدت الأبحاث التي تناولت مسألة خطاب الكراهية على تعريفات مختلفة لهذا الخطاب، ومنها من اعتبر الكلام البذيء خطاب كراهية ومنها من اعتبر الشتائم والسباب خطاب كراهية. ولما كان تعريف خطاب الكراهية أحد الاختلافات بين الدراسات وبين المصنفين أنفسهم، فقد جرى اعتماد التعريف التالي لخطاب الكراهية كأساس في هذا البحث على الشكل التالي:

خطاب الكراهية هو: "**أي فعل act يتضمن تحريضًا على العنف أو القتل أو النبذ أو وجود شتائم أو كلام بذيء أو تشويه سمعة شخص أو مجموعة أشخاص بناءً على العرق أو اللون أو الجنس أو العقيدة أو الدين**" [143].

بما أن عملية التصنيف يمكن أن تؤثر بشكل كبير على نتائج بناء وتدريب واختبار المصنفات، فقد اعتمدنا على فريق عمل مُكوَّن من ثلاثة أشخاص سوريين بمستوى تأهيل دراسي جيد ومن اتجاهات فكرية مختلفة، وحرصنا على وجود كلا الجنسين ضمن فريق العمل لضمان عدم انحياز الوسم قدر الإمكان.

طُلِبَ من فريق العمل قراءة هذه التغريدات وتصنيفها -بعيداً عن أي تحيز ممكن- كخطاب كراهية أو لا من خلال اعتماد التعريف السابق.

جرى الاتفاق بين المصنفين على اعتبار التغريدات التي يمكن أن تقع ضمن أحد الأنواع المبينة في الجدول 3-1 كخطاب كراهية.

نوع خطاب الكراهية	أمثلة
كراهية ضد دين ما	إسلام، مسيحية، يهودية
كراهية ضد مذهب ما	سني، شيعي، ماروني
كراهية ضد إحدى القوميات	عربي، كردي، فارسي
كراهية ضد أحد البلدان	سوري، لبناني، سعودي
كراهية ضد المرأة	العنف ضد المرأة

الجدول 3-1 أنواع خطاب الكراهية ضمن مجموعة البيانات

بعد الحصول على نتائج عملية الوسم من المصنفين، نلاحظ وجود حالتين وهما:

- **اتفاق بالإجماع Unanimous agreement**: اتفق جميع الفاعلين بالعملية على نفس الوسم. بالتالي، جرى قبول هذه التغريدات وحفظها في مجموعة البيانات. بلغت نسبة العينات المتفق عليها بالإجماع 79%..
- **اتفاق بالأغلبية Majority agreement**: اتفق اثنان فقط من الفاعلين بالعملية على أحد الوسمين. عُرِضَت هذه التغريدات على مراقب آخر الذي يقرر بدوره قبول وسم الأغلبية وحفظ هذه التغريدات في مجموعة البيانات أو لا.

يبين الجدول 3-2 أعداد التغريدات ونسب توزعها:

العدد الكلي	18,110
الصف /0/ عادي	17,249
الصف /1/ كراهية	861
نسبة عينات الكراهية	4.91%

الجدول 3-2 نسب توزع التغريدات بين الصفوف في مجموعة البيانات

سنقوم من الآن وصاعدًا بتسمية التغريدة المصنفة من الصف /0/ **بالتغريدة العادية، وبتغريدة كراهية** للتغريدات المصنفة من الصف /1/.

3-4- المعالجة المسبقة Preprocessing

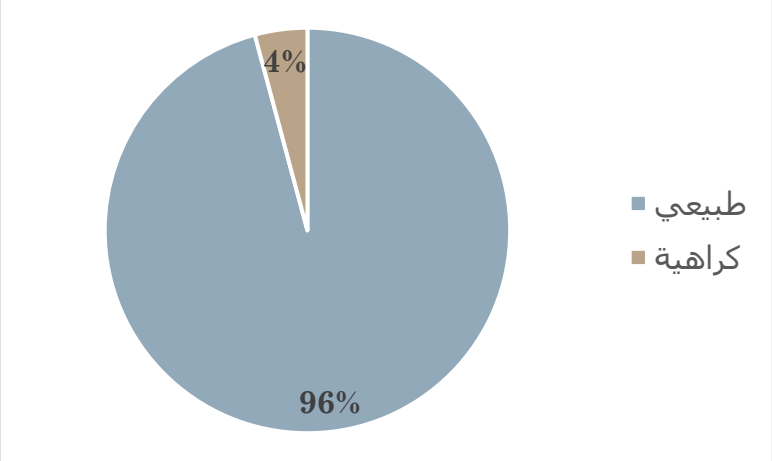
تحوي التغريدات رموزًا وأرقامًا بالإضافة لأخطاء إملائية وتكرار لحرف ضمن كلمة ما (للتأكيد على معنى ما) وكلمات بغير اللغة العربية وعناوين صفحات على شبكة الإنترنت URL. تتضمن مرحلة المعالجة الأولية إلغاء الكلمات غير العربية وعلامات الترقيم والأرقام وحذف عناوين الصفحات. تعد هذه المرحلة مهمة لأنها تنعكس على المراحل الأخرى فكلما كانت التغريدات أنظف وأوضح كانت عملية وضع علامات لها تحدد نوعها خطاب كراهية أو لا من قبل المصنفين أسهل، وقد أوضحت العديد من الدراسات البحثية أن المعالجة المسبقة للنص تؤدي إلى تحسين نتائج التصنيف [144]. تشمل هذه المرحلة الخطوات التالية:

- حذف الأجزاء المكتوبة بغير اللغة العربية.
- تسوية الأحرف Letters Normalization: هذه الخطوة ضرورية بسبب التشابه بين هذه الأحرف واستخدامها من المستخدمين بشكل خاطئ [145].
  - تحويل الأحرف (إ أ آ) إلى الحرف (ا).

- تحويل الحرف (ة) إلى الحرف (ه).
- تحويل الحرف (ى) إلى الحرف (ي).

- حذف علامات الترقيم والأرقام.
- حذف علامات التصنيف Hashtags.
- حذف الإشارات إلى الأشخاص Mentions.
- حذف التغريدات المكونة من كلمة واحدة فقط.
- إبقاء تغريدة واحدة من التغريدات المكررة.

يبين الشكل 3–2 توزع البيانات بين الصفوف.



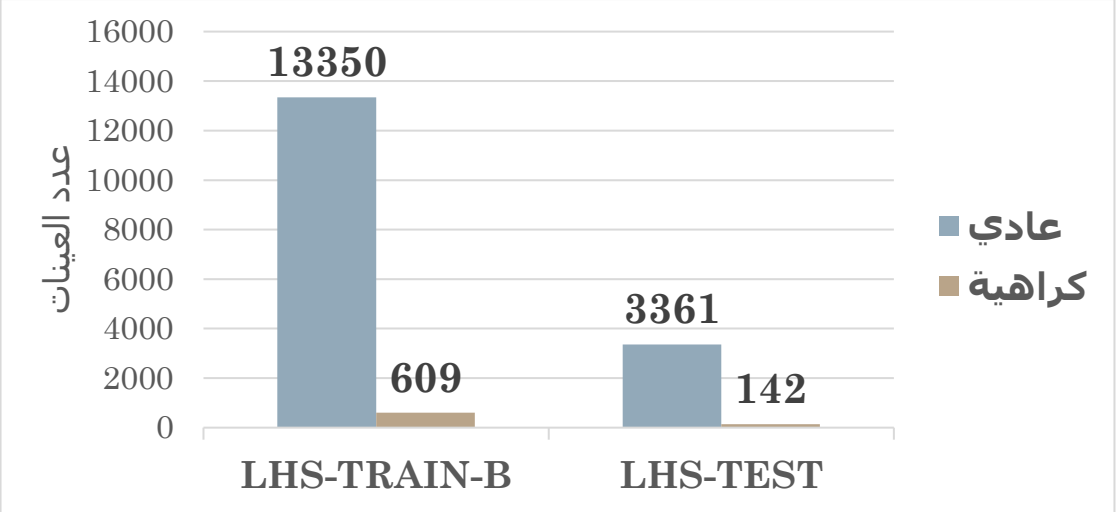
الشكل 3–2 نسب توزع العينات بين الصفوف في مجموعة البيانات المحصلة

للحصول على مجموعة اختبار، قسمنا مجموعة البيانات إلى مجموعتين وفق قاعدة 80/20 وحصلنا في نهاية المطاف على مجموعتي بيانات مصنفة يدويًا واحدة للتدريب وأخرى للاختبار، كما هو مبين في الجدول 3–3.

النوع	الاسم	العدد الكلي	الصف /0/	الصف /1/	عينات الكراهية
التدريب	**LHS-TRAIN-B**	13,959	13,350	609	4.36%
الاختبار	**LHS-TEST**	3,503	3,361	142	4.05%
	المجموع	17,462	16,711	751	4.30%

الجدول 3–3 توزع العينات بين مجموعتي التدريب والاختبار

يبين الشكل 3–3 توزع العينات بين مجموعتي التدريب والاختبار.



الشكل 3–3 تقسيم مجموعة البيانات المحصلة إلى مجموعة تدريب ومجموعة اختبار وفق مبدأ 20/80

3–5– خصائص التغريدات

تزود واجهة التخاطب مع المستخدمين API الخاصة بموقع تويتر مجموعة من الخصائص لكل تغريدة تتضمن مواصفات للتغريدة نفسها أو للحساب صاحب التغريدة.

يبين الجدول 3–4 بعض خصائص التغريدات.

اسم الخاصية	البيان
ID	وهو رقم وحيد مميز للتغريدة.
TRUNCATED	للدلالة على أن طول التغريدة أكثر من الطول المسموح للتغريدة وجرى اجتزاء التغريدة أم لا.
IN_REPLY_TO_STATUS_ID	يشير إلى الرمز المعرف الخاص بالتغريدة الأساس في حال كانت التغريدة ردًّا على واحدة أخرى.
IN_REPLY_TO_USER_ID	يشير إلى الرمز المعرف الخاص بالحساب مالك التغريدة الأساس في حال كانت التغريدة ردًّا على واحدة أخرى.
PLACES	يدل على المكان الذي أرسلت منه التغريدة في حال وجوده.
RETWEET_COUNT	يشير إلى عدد إعادة نشر التغريدة.
FAVORITE_COUNT	يشير إلى عدد المرات التي جرى فيها تفضيل التغريدة.
LANG	يدل على لغة التغريدة.

CREATED_AT	يشير إلى تاريخ نشر التغريدة.
HASHTAGS	تحتوي قائمة بالتصنيفات الموجودة ضمن نص التغريدة.
MENTIONS	تحتوي قائمة بالحسابات المذكورة ضمن نص التغريدة.

الجدول 3-4 أنواع خصائص التغريدات

كما يبين الجدول 3-5 بعض خصائص حسابات توتير.

اسم الخاصية	**البيان**
ID	وهو رقم وحيد مميز للحساب.
NAME	اسم الحساب
SCREEN_NAME	اسم الحساب المستخدم
LOCATION	الموقع
DESCRIPTION	توصيف الحساب
FOLLOWERS_COUNT	عدد المتابعين
FRIENDS_COUNT	عدد الأصدقاء
CREATED_AT	تاريخ إنشاء الحساب
VERIFIED	هل تحقق تويتر من صاحب الحساب؟
LANG	لغة الحساب

الجدول 3-5 أنواع خصائص حسابات تويتر

جرى دراسة توزع التغريدات وفق بعض الخصائص. فمثلًا، نلاحظ وجود 9,205 تغريدة عادية تضم ذكر لحسابات أخرى mention ما يعادل 66.72% من إجمالي التغريدات العادية. بينما في المقابل، نجد 481 تغريدة كراهية تضم ذكر لحسابات أخرى ما يعادل 69.71% من إجمالي تغريدات الكراهية. كذلك نلاحظ -على سبيل المثال- وجود 1868 تغريدة عادية جرى تفضيلها ما يمثل 13.54% من إجمالي التغريدات العادية. بينما في المقابل، نجد 151 تغريدة كراهية جرى تفضيلها ما يعادل 21.88% من إجمالي تغريدات الكراهية. نلاحظ أن نسبة تغريدات الكراهية التي تحقق خاصية ما هي أكبر من مثيلتها العادية.

نرمز بالاختصار FIRM للتغريدات التي جرى تفضيلها **F**avorite_count، أو كانت ردًا على مستخدم آخر **I**n_reply_to_user_id، أو أُعيد نشرها **R**etweet_count، أو تضم ذكر لحساب ما **M**entions، كذلك إذا أضفنا الخاصيتين: متجزأة **T**runcated أو تصنيفات

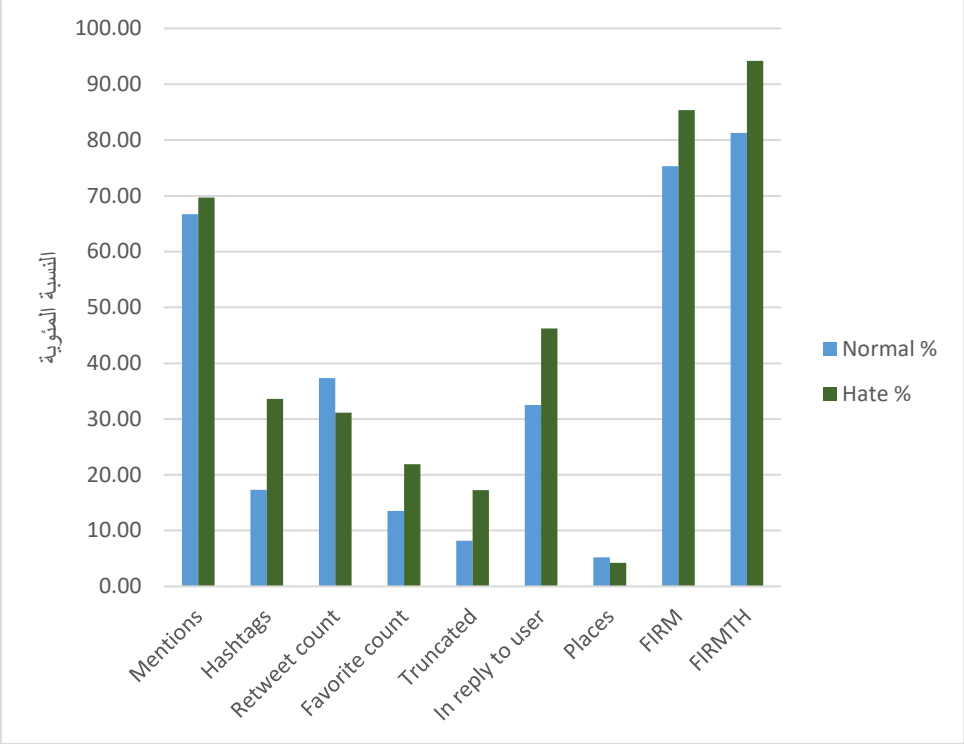
Hashtags، لنحصل على الاختصار FIRMTH. نلاحظ أن التغريدات العادية من النمط FIRM تبلغ 10,388 ما يعادل 75.29% من إجمالي التغريدات العادية، بينما يصل هذا المعدل إلى 85.36% من خلال 589 تغريدة كراهية. كذلك، نلاحظ أن التغريدات العادية من النمط FIRMTH تبلغ 11,212 ما يعادل 81.26% من إجمالي التغريدات العادية، بينما يصل هذا المعدل إلى 94.20% من خلال 650 تغريدة كراهية. ما نريد قوله هنا، هو أنه بالرغم من قلة نسبة تغريدات الكراهية من إجمالي التغريدات –أقل من 5%– فإن تغريدات الكراهية تعرض نسبة أكبر من هذه الخصائص مقارنة مع التغريدات العادية. مثلًا، تغريدات الكراهية التي تضم ذكر لأشخاص mentions تمثل 69.71% من إجمالي تغريدات الكراهية، بينما هي 66.72% في التغريدات العادية. كذلك، تزداد هذه النسبة بين تغريدات الكراهية عند الدمج بين عدة خصائص مثل FIRMTH التي ترتفع من 81.26% في التغريدات العادية إلى 94.20% في تغريدات الكراهية.

يبين الجدول 3–6 أعداد ونسب التغريدات وفق كل خاصية.

م	الخاصية	تغريدات عادية		تغريدات كراهية		الإجمالي	
		العدد	%	العدد	%	العدد	%
1	Totals	13797	95.24	690	4.76	14487	100.00
2	**M**entions	9205	66.72	481	69.71	9686	66.86
3	**H**ashtags	2392	17.34	232	33.62	2624	18.11
4	**R**etweet_count	5149	37.32	215	31.16	5364	37.03
5	**F**avorite_count	1868	13.54	151	21.88	2019	13.94
6	**T**runcated	1132	8.20	119	17.25	1251	8.64
7	**I**n reply to user	4484	32.50	319	46.23	4803	33.15
8	**P**laces	718	5.20	29	4.20	747	5.16
9	FIRM	10388	75.29	589	85.36	10977	75.77
10	FIRMTH	11212	81.26	650	94.20	11862	81.88

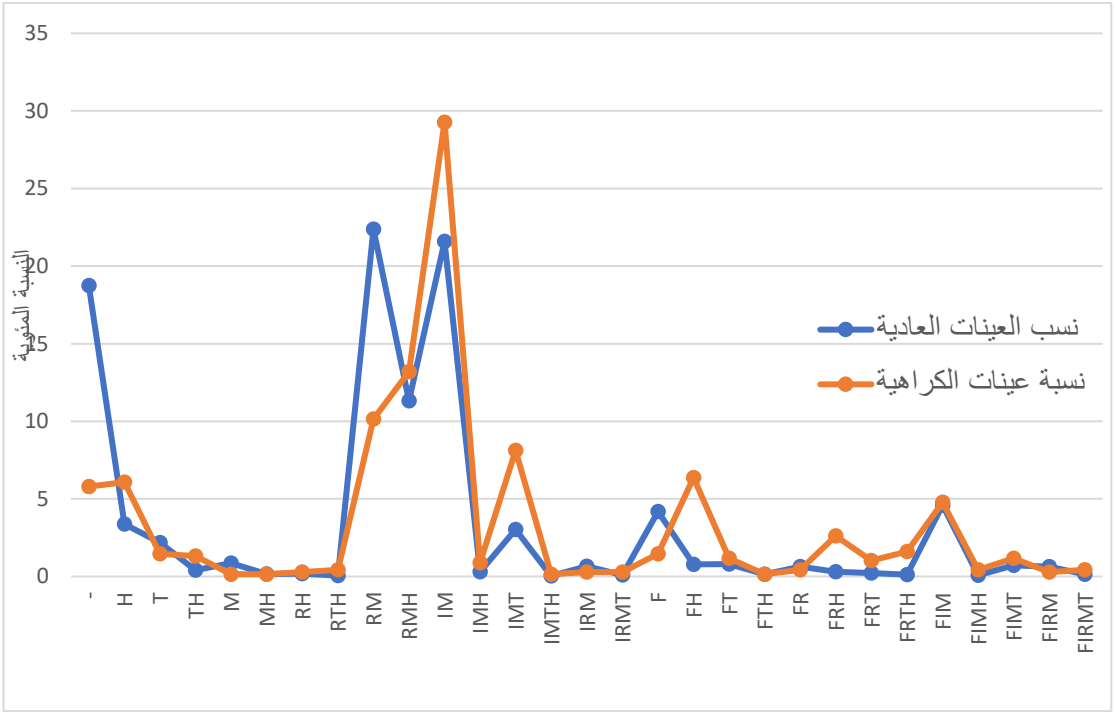
الجدول 3–6 أعداد ونسب التغريدات ضمن الصفوف وفق كل خاصية

يبين الشكل 3–4 أعداد ونسب التغريدات وفق كل خاصية.



الشكل 3–4 نسب التغريدات وفق كل خاصية حسب الصفوف

يبين الشكل 3–5 توزع الصفوف وفق مجموعات الخصائص، حيث يرمز كل حرف إلى الخاصية المقابلة. نلاحظ في هذا الشكل أيضًا، ارتفاع نسبة عينات تغريدات الكراهية المحققة لمجموعة خصائص ما مقارنة مع العينات العادية لنفس مجموعة الخصائص في أغلب الحالات.



الشكل 3–5 نسب التغريدات وفق سمات التغريدة حسب كل صف

نريد هنا التأكيد على الحاجة لأخذ هذه الخصائص بالاعتبار، سواء خلال مرحلة تحصيل البيانات أو خلال دراسة مسائل كشف خطاب الكراهية. كما نريد التأكيد على أن التغريدات التي لا تحتوي على أي من الخصائص السابقة هي تغريدات مهملة وبدون قيمة حتى وإن كانت تحمل في طياتها الكراهية والمشاعر السلبية.

بناءً على ذلك، إذا طرحنا السؤال التالي: "**هل يمكن اعتبار المعلومات التي يمكن تحصيلها حول التغريدة معلومات مفيدة في مسألة كشف خطاب الكراهية؟**" نرى أهمية هذه المعلومات والخصائص، ولكن لا نعمم هذه الأهمية لحين دراسة توزع العينات وفق هذه الخصائص على مجموعات أخرى من البيانات.

كذلك، جرى دراسة المفردات الأكثر تكرارًا بين عينات الكراهية وكانت النتيجة وفق ما هو مبين في الشكل 3–6.

الشكل 3–6 المفردات الأكثر تكرارًا في عينات الكراهية

3–6– تعزيز البيانات

تتصف مجموعات البيانات التي نهتم بها في بحثنا بشكل عام بأنها غير متوازنة imbalanced، وهو ما نلاحظه في مجموعة البيانات الأخيرة بنسبة أقل من 5% للعينات المصنفة كخطاب كراهية. لذلك، سنحاول في هذه الفقرة تجاوز مسألة عدم توازن البيانات بطريقة أو بأخرى؟

تعتمد فعالية مصنفات تعلم الآلة إلى حد كبير على حجم ونوعية بيانات التدريب [146]. علمًا أنه حتى مع وجود مجموعات بيانات كبيرة، لا تزال العديد من خوارزميات التّعلم حساسة لتوزع

الصفوف غير المتوازن [147]. تعالج مشكلة البيانات غير المتوازنة بإضافة المزيد من عينات البيانات إلى الصفوف الأقل أو ما يعرف بتعزيز البيانات Data Augmentation.

3-6-1- التعزيز اليدوي Manual Augmentation

تنفذ تقنية التعزيز اليدوي بإجراء تحويلات على مستوى الكلمة word-level transformations، حيث تولد جمل جديدة من جمل موجودة مع الحفاظ على السمات الدلالية للجملة الأساس. ويعتبر التحويل على مستوى الكلمة أو الاستبدال بالمرادفات synonym replacement من التقنيات الأكثر استخدامًا [148].

جرى تحديد مجموعة من تغريدات الكراهية التي تتناول دينًا ما أو مذهبًا أو عرقًا، ثم استبدلنا الكلمات التي تدل على الدين أو المذهب أو العرق بمرادفات تدل على دينٍ أو مذهبٍ أو عرقٍ مختلفٍ في كل مرة بما يتماشى مع خطاب الكراهية المنتشرة في المنطقة.

يبين الجدول 3-7 بعض الأمثلة عن التعزيز اليدوي.

التغريدة الأصل	التغريدات المضافة
لا تحكي عن اليهود لان طلعو اشرف منكم يا شيعه	لا تحكي عن اليهود لان طلعو اشرف منكم يا سنه
	لا تحكي عن اليهود لان طلعو اشرف منكم يا علويه
	⋮
لعنه الله على الاكراد	لعنه الله على العرب
	لعنه الله على الفرس
	⋮
هو لبناني نجس	هو سوري نجس
	هو سعودي نجس
	⋮

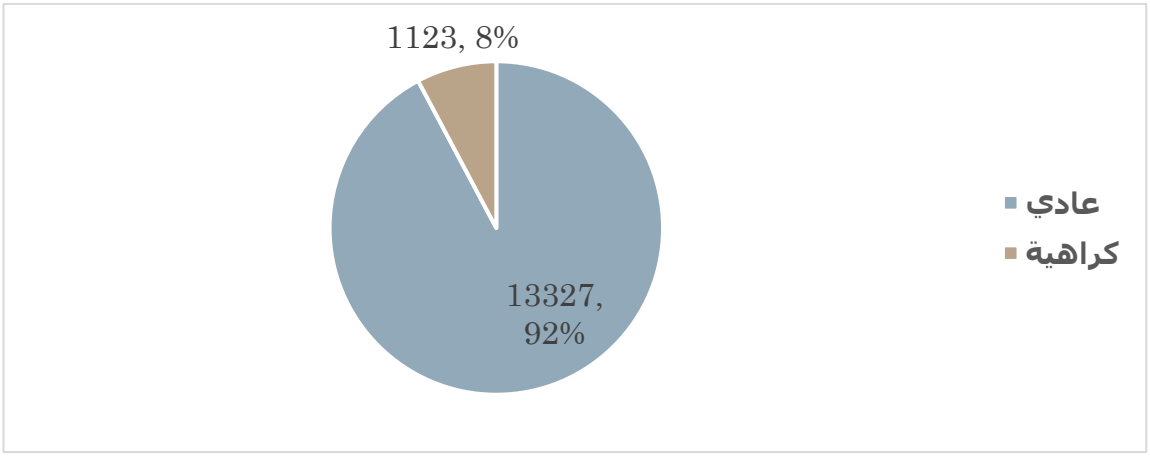
الجدول 3-7 أمثلة عن بعض التغريدات التي تمت إضافتها يدويًا إلى مجموعة البيانات

يبين الجدول 3-8 أعداد التغريدات ضمن مجموعة التدريب المعززة يدويًا **LHS-TRAIN-C** ونسب توزعها بين الصفوف:

العدد الكلي	الصف /0/ عادي	الصف /1/ كراهية	نسبة عينات الكراهية
14,450	13,327	1123	7.77%

الجدول 3-8 أعداد ونسب توزع العينات بين الصفوف بعد عملية التعزيز اليدوي

يبين الشكل 3–7 توزع البيانات بين الصفوف بعد عملية التعزيز اليدوي في مجموعة التدريب **LHS-TRAIN-C**.



الشكل 3–7 توزع البيانات بين الصفوف بعد عملية التعزيز اليدوي في مجموعة البيانات LHS-TRAIN-C

3–6–2– التعزيز الآلي Auto Augmentation

اعتمدنا في تقنية التعزيز الآلي للبيانات على تحديد مجموعة تغريدات الكراهية وإجراء بعض العمليات [149] على هذه العينات مثل:

- عكس ترتيب الكلمات.
- حذف الكلمة الأولى.
- حذف الكلمة الأخيرة.
- إضافة كلمة عشوائية.

بما أن هذه الإجراءات قد تؤدي إلى تغير في المعنى للتغريدة، فقد جرى إعادة التحقق من أن النص الجديد بعد التعديل ما زال يتضمن خطاب كراهية. بالتالي، قمنا بحذف التغريدات الناتجة التي لم تعد تشير إلى خطاب كراهية.

يبين الجدول 3–9 بعض الأمثلة عن التعزيز الآلي.

التغريدة الأصل	التغريدات المضافة
وجيه عباس طائفي إيراني الله يلعنهم	وجيه عباس طائفي إيراني الله
	عباس طائفي إيراني الله يلعنهم
	يلعنهم الله إيراني طائفي عباس وجيه
	وجيه عباس شجره طائفي إيراني الله يلعنهم

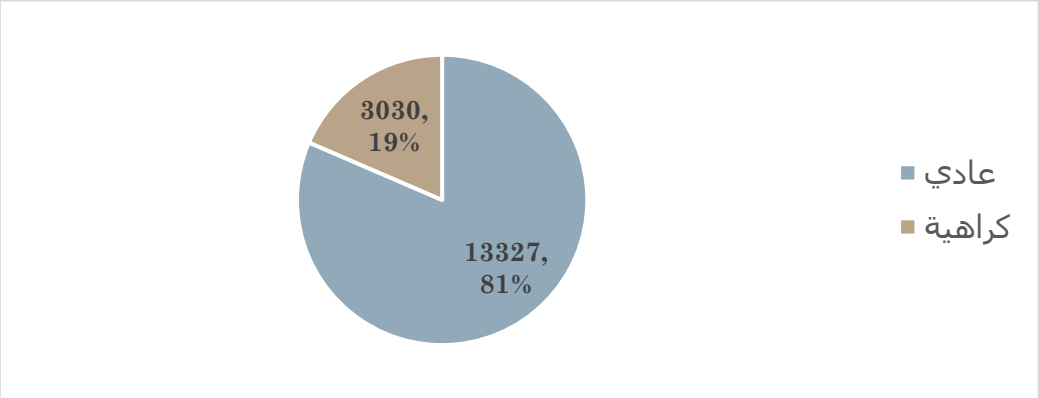
الجدول 3–9 أمثلة عن بعض التغريدات التي تمت إضافتها آليًا إلى مجموعة البيانات

يبين الجدول 3–10 أعداد التغريدات ضمن مجموعة التدريب المعززة آليًا **LHS-TRAIN-D** ونسب توزعها بين الصفوف:

العدد الكلي	الصف /0/ عادي	الصف /1/ كراهية	نسبة عينات الكراهية
16,357	13,327	3030	18.5%

الجدول 3–10 أعداد ونسب توزع العينات بين الصفوف بعد عملية التعزيز الآلي

يبين الشكل 3–8 توزع العينات بين الصفوف بعد التعزيز الآلي ضمن مجموعة التدريب **LHS-TRAIN-D**.



الشكل 3–8 توزع البيانات بين الصفوف بعد عملية التعزيز الآلي في مجموعة البيانات LHS-TRAIN-D

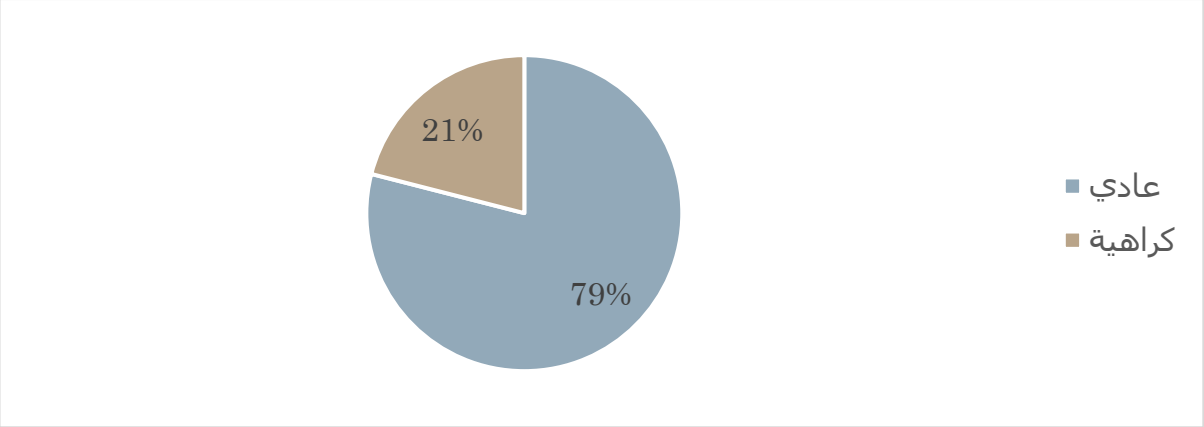
3–6–3– دمج التقنيتين

جرى دمج العينات الناتجة عن التعزيز اليدوي مع مثيلاتها الناتجة عن التعزيز الآلي لتكوين مجموعة أخرى من البيانات. يبين الجدول 3–11 أعداد التغريدات ونسب توزعها:

العدد الكلي	الصف /0/ عادي	الصف /1/ كراهية	نسبة عينات الكراهية
16,871	13,327	3544	21.01%

الجدول 3–11 أعداد ونسب توزع العينات بين الصفوف بعد دمج التعزيز اليدوي والآلي

يبين الشكل 3–9 توزع العينات بين الصفوف ضمن مجموعة التدريب المعززة يدويًا وآليًا **LHS-TRAIN-E** بعد دمج التقنيتين.



الشكل 3–9 توزع البيانات بين الصفوف بعد دمج تقنيتي التعزيز اليدوي والآلي في مجموعة البيانات LHS-TRAIN-E

3–7– اختبار مجموعة البيانات

كما ذكرنا سابقًا، استخدم الباحثون مجموعات بيانات مختلفة في الأبحاث التي تناولت مسألة الكشف عن خطاب الكراهية، وقد أتاح بعض الباحثين مجموعات البيانات المستخدمة. أجرينا الاختبارات على بعض من مجموعات البيانات المتاحة، وهي:

- مجموعة البيانات المشرقية L-HSAB [29] وهي مجموعة بيانات سياسية كما صنفها مؤلفوها مأخوذة من صفحات بعض الساسة اللبنانيين، وهي مجموعة البيانات الوحيدة الخاصة باللهجة المشرقية. تضم هذه المجموعة 6,000 عينة.
- مجموعة بيانات ضمن ورشة العمل الرابعة للمحتوى العربي مفتوح المصدر وأدوات المعالجة [45] The 4th Workshop on Open-Source Arabic Corpora and Processing Tools ونرمز لها اختصارًا OSACT، وتضم حوالي 10,000 تغريدة مكتوبة باللغة العربية من مختلف اللهجات.
- مجموعة بيانات متاحة من Google [51] ضمن ورشة العمل /14/ للتقييم الدلالي عام 2020، ونرمز لها اختصارًا OffensEval. تضم هذه المجموعة حوالي 8,000 تغريدة مكتوبة باللغة العربية من مختلف اللهجات.

لاختبار مجموعة البيانات المحصلة، استخدمنا مجموعة من المصنفات التي جرى تدريبها على مجموعة البيانات L-HSAB لأنها المجموعة الوحيدة المتاحة باللهجة المشرقية، مع اعتماد مُضَمِّن الكلمات غير السياقي Aravec [150] لتمثيل النص.

3–7–1– المصنفات المستخدمة

استخدمنا مجموعة المصنفات التالية:

- RandomForest [151] [152].
- SGD Classifier.
- SVC [153].
- XGB Classifier [154].
- CatBoost [155].
- Logistic Regression [156].
- Multi-Layer Perceptron (MLP) [157].
- GuassianNB [158].

جرى تدريب هذه المصنفات على مجموعة البيانات L-HSAB واختبارها على مجموعة البيانات LHS-TEST.

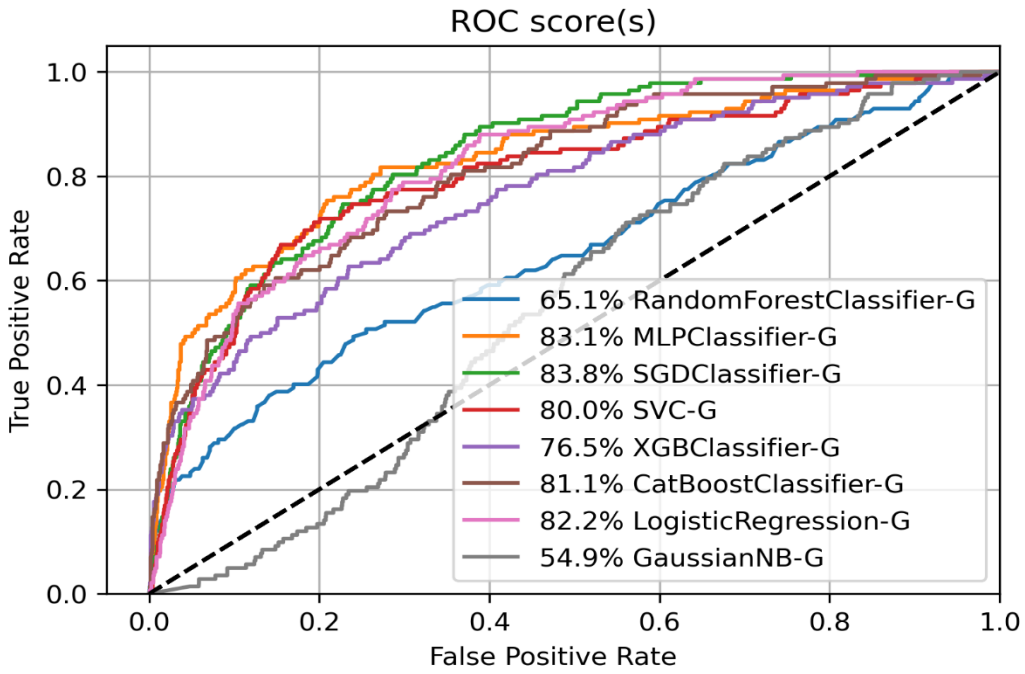
3-7-2- نتائج الاختبارات

أجرينا اختبار مجموعة المصنفات على مجموعة البيانات **LHS-TEST**. حصلنا على النتائج التالية لمعايير الدقة Precision والإرجاع Recall ومقياس F1 والصحة Accuracy المبينة في الجدول 3-12. نشير هنا إلى أننا اعتمدنا المعيار f1 المحسوب على أساس macro في جميع النتائج ما لم يذكر خلاف ذلك.

ROC	Accuracy	F1	Recall	Precision	Classifier
65.1	**0.93**	0.58	0.59	0.57	RandomForest
83.8	**0.93**	0.626	0.65	**0.61**	SGD Classifier
80.0	0.85	0.58	**0.74**	0.57	SVC
76.5	0.92	0.62	0.66	0.60	XGB Classifier
81.1	**0.93**	**0.632**	0.67	**0.61**	**CatBoost**
82.2	**0.93**	0.61	0.64	0.60	Logistic Regression
83.1	0.90	0.62	0.73	0.59	Multi-Layer Perceptron
54.9	0.71	0.44	0.47	0.49	GuassianNB

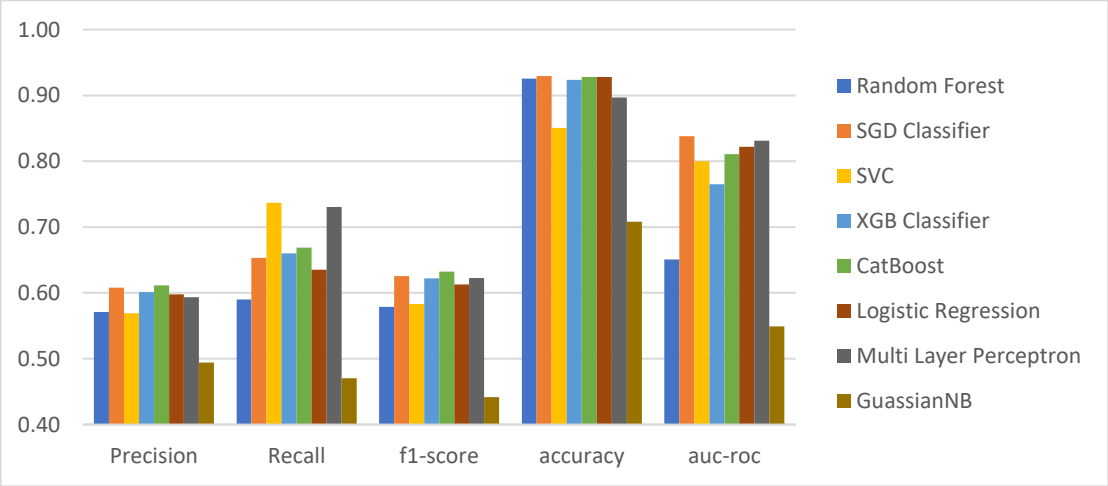
الجدول 3-12 نتائج اختبار المصنفات المدربة على المجموعة L-HSAB

كما يبين الشكل 3-10 نتائج الاختبار على منحنى AUC-ROC.



الشكل 3–10 نتائج اختبار النموذج المدرب على مجموعة البيانات L-HSAB على مجموعة الاختبار LHS-TEST وفق منحنى AUC-ROC

يبين الشكل 3–11 نتائج اختبار النموذج المدرب على مجموعة البيانات L-HSAB وفق مجموعة الاختبار LHS-TEST.



الشكل 3–11 نتائج اختبار النموذج المدرب على مجموعة البيانات L-HSAB على مجموعة الاختبار LHS-TEST

نلاحظ من النتائج والمخططات السابقة، أن المصنف CatBoost حقق أفضل معدل وفق مقياس F1-score بلغ 63.2%. إذا ما قورنت هذه القيمة مع نتيجة أفضل مصنف في نفس الدراسة [29] التي اعتمدت مجموعة البيانات هذه والتي بلغت 74.4%، نلاحظ أن المعدل الذي حصلنا عليه قليل نسبيًا لا سيما إذا أخذنا بالاعتبار أن معظم الدراسات البحثية التي تناولت مسألة كشف

خطاب الكراهية حصلت على f1-score أكبر من 70%. بالتالي، لا يمكن اعتبار هذه القيمة جيدة بشكل كاف، ويعود ذلك إلى خصوصية مجموعة البيانات المدرب عليها كونها ذات طبيعة سياسية. بما أننا نهدف لكشف خطاب الكراهية للنصوص العربية القصيرة المكتوبة باللهجة المشرقية، نعتبر الأداء على مجموعة بيانات خاصة بالمنطقة هي المعيار. من خلال النتيجة السابقة ومن أجل الإجابة على السؤال Q2 "**هل تعتبر مجموعات البيانات المتوفرة على شبكة الإنترنت كافية؟**"، نتبيّن عدم كفاية المجموعة L-HSAB وضرورة التوسع في مسألة خطاب الكراهية عبر تحصيل مجموعة بيانات خاصة بنا واستخدامها لتدريب نماذج للحصول على نتيجة مقبولة.

3-8- خاتمة

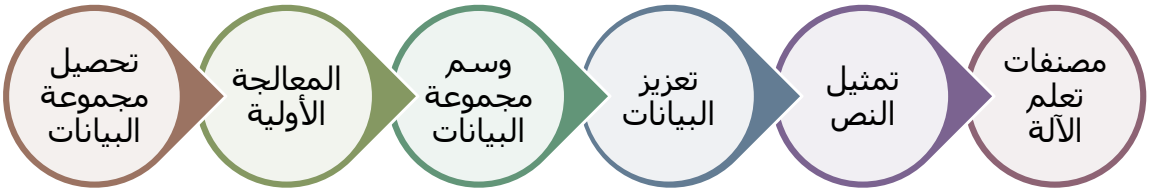
قدّمنا في هذا الفصل شرحًا تفصيليًّا لمجموعة البيانات المحصلة من موقع تويتر، والإجراءات المسبقة التي تطبيقها عليها وآليات تعزيز البيانات المطبقة. كما تعرضنا في هذا الفصل إلى أهمية خصائص التغريدات بعيدًا عن النص. كما جرى تدريب مجموعة من المصنفات على مجموعة البيانات المشرقية L-HSAB واختبار المصنفات على مجموعة الاختبار **LHS-TEST**.

4- نظام كشف خطاب الكراهية

نقدم في هذا الفصل نظامًا لكشف خطاب الكراهية مدربًا على مجموعة بيانات جديدة باللغة العربية باللهجة المشرقية مستخرجة من موقع تويتر خاصة بخطاب الكراهية، حيث نبدأ بشرح لطرق تمثيل النصوص المستخدمة كدخل لخوارزميات التصنيف المعتمدة. نعرض لاحقًا في هذا الفصل أهمية استخدام خوارزميات التّعلم العميق ومدى التحسن الذي يمكن أن يضيفه إلى نظام الكشف، حيث سنحاول في هذا الفصل تقديم الإجابة على السؤال Q3- "**ما هو التمثيل الأفضل للنصوص في مجموعات بيانات خطاب الكراهية؟**" بالإضافة إلى السؤال Q4- "**هل يساعد بناء مجموعة بيانات وتدريب نظم تصنيف عليها في زيادة القدرة على التعميم؟**"

4-1- مقدمة

يتكون نظام كشف خطاب الكراهية المقترح كغيره من نظم التصنيف من المكونات الأساسية الموضحة في الشكل 4-1.



الشكل 4-1 البنية العامة لنظام تصنيف النصوص المقترح

كنا عرضنا في الفصل 3- بشكل مفصل للمراحل الأربع الأولى، وسنعرض في الفقرات اللاحقة تمثيل النص والمصنفات المستخدمة ونتائج الاختبارات.

4-2- تمثيل النص

يعرف تمثيل النص بأنه الربط بين اللغة المكتوبة ومجموعة خصائص مفيدة بتنسيق قابل للقراءة من الآلة. تجري عملية تمثيل النص عبر استخراج السمات feature extraction من خلال تحويل التّغريدة إلى تمثيل رقمي قابل للقراءة والفهم من قبل الآلة [159]، ويوجد عدة طرق لتمثيل النصوص نذكر منها:

- التّمثيل المفرداتي:

- نموذج حقيبة الكلمات Bag Of Words (BOW)، حيث تُمثَّل كل كلمة بوجودها أو تكرارها في النص.
- تردد المفردة TF (Term Frequency) وتردّد التّغريدة المعكوس ويعبر عنها بالرمز IDF (Inverse Document Frequency)، ومن الشائع استخدام تركيبة هاتين القيمتين والمعروفة باسم TF-IDF [160]. بالرغم من سهولة هذا التّمثيل إلا أنه يعاني من عدة مشاكل منها التّرادف وتعدّد المعاني.
- تتالي المفردات n-gram: يتحول التركيز في تقنية n-gram [161] من الكلمات إلى مفردات متتالية عددها n، ما يجعل هذا التمثيل يحتفظ بمعلومات الترتيب والمكان للكلمات.

- تضمين الكلمات: حيث ظهرت حديثًا طرق أخرى لاستخراج السمات تستخدم تضمين الكلمات:
  - تضمين الكلمات غير السياقي: حيث يُجيز هذا التّمثيل أن يكون للكلمات المتشابهة في المعنى تمثيلات رقمية قريبة من بعضها، مما يحسن كفاءة نماذج تعلم الآلة [162] مثل Word2Vec. يعتبر تدريب متجهات تضمين الكلمات سهلًا لكنه يحتاج لكمية كبيرة من البيانات، وهناك متجهات للكلمات المدربة مسبقًا (قليلة في اللغة العربية) ومن أهمها Aravec وهو تمثيل موزع للكلمات خاص باللغة العربية [150]، وقد دربت Aravec مسبقًا باستخدام بيانات كبيرة (67 مليون تغريدة) من (تويتر، ويكبيديا، وِب) ونفذ بواسطة Word2Vec [163] مع الأخذ بالاعتبار (unigram, bigram, trigram) وبأطوال مختلفة لمتجه التّضمين.
  - تضمين الكلمات السياقي: في التمثيل السابق تُمنَح كل كلمة تمثيلًا واحدًا بغض النظر عن إمكانية اختلاف المعنى بين جملة وأخرى. هذا ما جرى تداركه في تضمين الكلمات وفق السياق، حيث تمنح الكلمةُ الواحدةُ تمثيلاتٍ مختلفةً وفق سياقاتها المختلفة [164] مثل BERT.

بسبب الاهتمام المتزايد لخوارزميات التّعلم العميق خلال السنوات الأخيرة [165] [166]، اعتمدنا على استخدام نموذجي التضمين: السياقي وغير السياقي.

- **تضمين الكلمات غير السياقي:** اقترح الباحثون في [151] استخدام Aravec في مسائل كشف الخطاب العدائي المكتوب باللغة العربية. كما ذكرنا يوجد عدة نماذج من Aravec مدربة مسبقًا باستخدام بيانات كبيرة من (تويتر، ويكبيديا، وِب)، ولكل مجموعة بيانات

جرى بناء وتدريب نموذجين (CBOW and Skip-gram). بما أن مجموعة البيانات المستخدمة موضوع البحث مستخرجة من تويتر فقد قررنا استخدام النموذج المدرب على مجموعة بيانات تويتر. بما أن الباحثين في [167] اقترحوا استخدام النموذج Skip-gram باعتباره يعطي دقة دلالية semantic accuracy أكبر من النموذج الآخر، فقد اعتمدنا تطبيق النموذج full-skip-gram للبعد d=300.

- **تضمين الكلمات السياقي:** جرى استخدام عدة نماذج من BERT وهي:
  - AraBERT [168]: وهو نموذج BERT [164] مدرَّب على اللغة العربية.
  - ArabicBERT [169]: وهو مجموعة من نماذج BERT المدرَّبة على اللغة العربية.
  - GigaBERT [146]: وهو نموذج BERT مخصَّص للتحويل من اللغة الإنكليزية إلى اللغة العربية مدرب على نصوص إخبارية من GigaBERT [170] وويكيبيديا وبيانات وب.

4-3- المصنفات المستخدمة

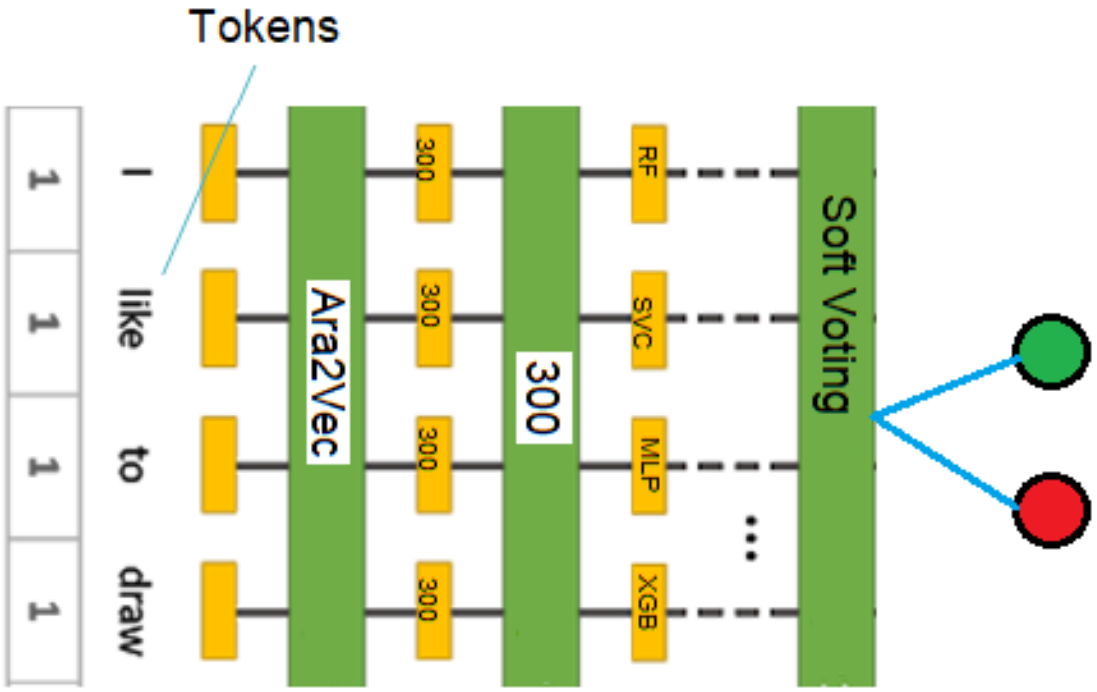
اعتمدنا على بناء شبكة عصبونية لاستخدامها مع تمثيل النص السياقي (تضمين الكلمات السياقي)، بينما اعتمدنا على مجموعة من المصنفات مع تمثيل النص غير السياقي (تضمين الكلمات غير السياقي)، وهي:

- RandomForest [151] [152].
- SVC [153].
- XGB Classifier [154].
- CatBoost [155].
- Multi-Layer Perceptron [157].

كذلك، اعتمدنا على مفهوم التّعلم المجمّع Ensemble Learning [165] حيث تقوم مجموعة من النماذج الأساسية base models بتنفيذ نفس المهمة. يعبر عن النماذج الأساسية بالمتعلم الضعيف weak learner. يُعتمَد المصنف الي يقوم بجمع توقع كل متعلم وأخذ التوقع الحاصل على أكبر عدد من الأصوات. يدعى هذا النهج التصنيف بالتصويت Voting Classification. نميز طريقتين في هذا النهج:

- التصويت الصلب Hard Voting: من خلال تجميع التنبؤات لكل مصنف والتنبؤ بالصف الذي يحصل على أكبر عدد من الأصوات. وهذا ما يسمى "تصويت الأغلبية" أو "التصويت الصلب".
- التصويت الناعم Soft Voting: في هذا النموذج، يمكن لجميع المصنفات تقدير احتمالية كل صف، ثم يمكننا التنبؤ بالصف ذي الاحتمال الأعلى من خلال حساب متوسط احتمالية كل صف على جميع المصنفات الفردية.

يبين الشكل 4–2 مخططًا لنموذج كشف خطاب الكراهية المعتمد على تمثيل النص غير السياقي.



الشكل 4–2 مخطط نموذج تعلم مع تمثيل الكلمات غير السياقي

4–4– النموذج المرجعي baseline model

في البداية، اعتمدنا على تمثيل النص غير السياقي ثم قمنا بتدريب النموذج على مجموعة بيانات التدريب بدون أي تعزيز LHS-TRAIN-B باستخدام تمثيل البيانات Aravec الذي يمرر إلى مجموعة المصنفات التالية:

- RandomForest (RF).
- SVC.
- XGB Classifier (XGB).
- CatBoost.

- Multi-Layer Perceptron (MLP).
- Hard Voting.
- Soft Voting.

اعتمدنا هذا النموذج كنموذج مرجعي لمقارنته مع النماذج والحالات الأخرى.

4-4-1- اختبارات النموذج المرجعي

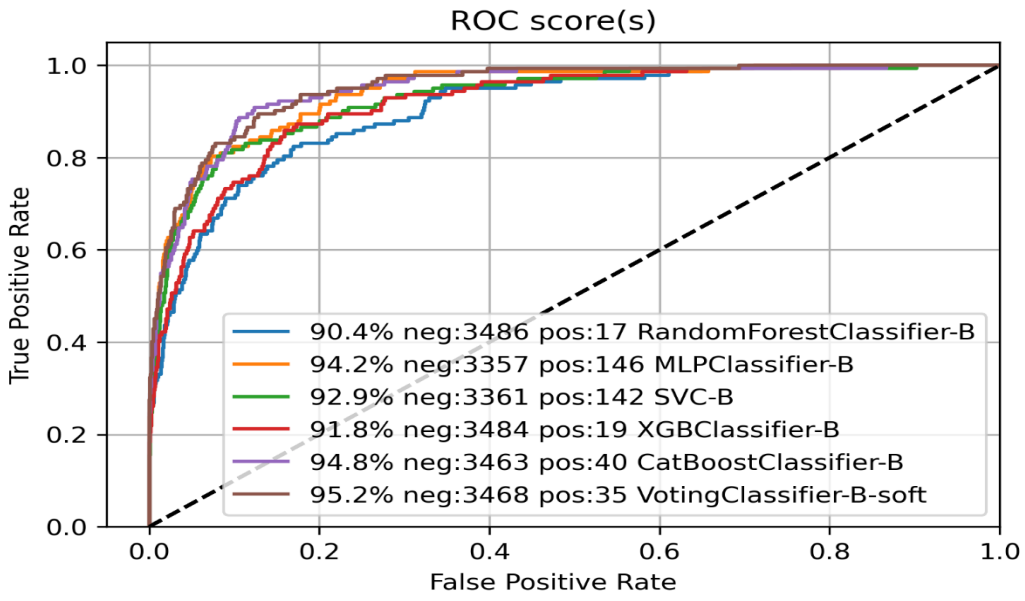
جرى اختبار النموذج المرجعي على مجموعة البيانات LHS-TEST، وحصلنا على القيم التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4-1.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.904	0.96	0.60	0.56	**0.98**	RF
0.929	0.96	0.76	**0.76**	0.76	SVC
0.918	0.96	0.61	0.57	**0.98**	XGB
0.948	**0.97**	0.70	0.63	0.95	CatBoost
0.942	**0.97**	**0.77**	**0.76**	0.78	MLP
0.952	**0.97**	**0.77**	0.71	0.89	**Soft Voting**
	0.97	0.69	0.62	0.95	Hard Voting

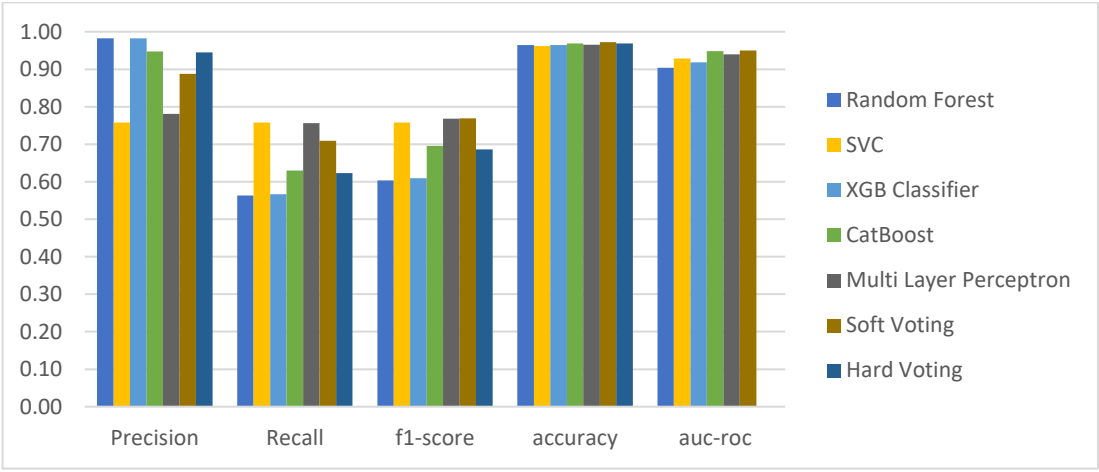
الجدول 4-1 نتائج اختبار النموذج المرجعي على مجموعة البيانات LHS-TEST

نلاحظ من هذه النتائج تفوق المصنف Soft Voting على بقية المصنفات وفق المعيار F1، وهذا أمر طبيعي عند استخدام التعلم المجمع Ensemble Learning. كذلك، يمكننا ملاحظة التحسن الواضح في قيم المعيار f1-score بشكل عام وذلك عند مقارنة هذه النتائج مع نتائج اختبارات النموذج المدرب على مجموعة البيانات المشرقية L-HSAB المذكورة في الجدول 3-12، حيث نجد أن المصنف CatBoost (الذي حصل على أفضل نتائج اختبارات النموذج المدرب على مجموعة البيانات L-HSAB) حصل على معدل f1-score بلغت 0.70 بزيادة 7 نقاط مئوية تقريبًا أي ما يعادل 11% تقريبًا. أما أفضل نتيجة –إذا أخذنا مقياس ROC كمعيار ثان- فكانت للمصنف Soft Voting بلغت 0.77. يمكن القول إن النموذج المرجعي أعطى قدرة على التعميم أعلى من النموذج المدرب على المجموعة L-HSAB.

يبين الشكل 4-3 نتائج الاختبار على منحنى AUC-ROC. كذلك يبين الشكل 4-4 معايير الأداء للمصنفات في النموذج المرجعي على مجموعة البيانات LHS-TEST.



الشكل 4–3 نتائج اختبار النموذج المرجعي على مجموعة البيانات LHS-TEST وفق منحني ROC



الشكل 4–4 معايير الأداء للمصنفات في النموذج المرجعي على مجموعة البيانات LHS-TEST

4–4–2– اختبارات النموذج المرجعي على مجموعات خارج المجال

لقياس قدرة النموذج المرجعي على التعميم، جرى اختبار هذا النموذج على عدة مجموعات خارج المجال out of domain –ذُكِرَت سابقًا في الفقرة 7–3– وهي: L-HSAB، OSACT، OffensEval. كذلك، أجرينا جميع الاختبارات اللاحقة على هذه المجموعات الثلاث بالإضافة طبعًا إلى مجموعة البيانات المحلية LHS-TEST.

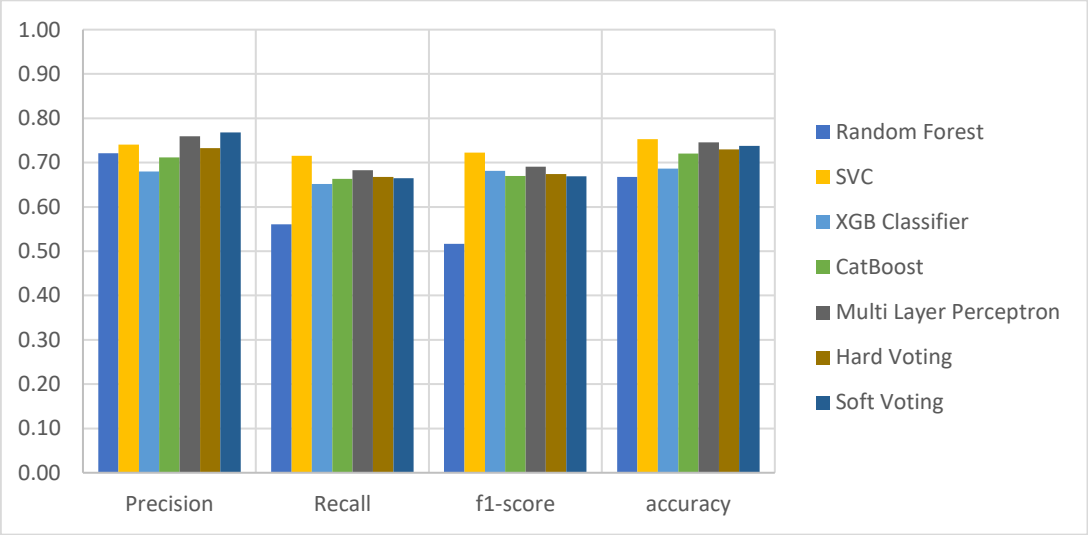
4–4–2–1– الاختبارات على مجموعة البيانات اللبنانية

جرى اختبار النموذج على مجموعة البيانات اللبنانية L-HSAB، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–2.

Accuracy	F1	Recall	Precision	Classifier
0.67	0.52	0.56	0.72	RF
0.75	**0.72**	**0.72**	0.74	SVC
0.69	0.68	0.65	0.68	XGB
0.72	0.67	0.66	0.71	CatBoost
0.75	0.69	0.68	0.76	MLP
0.74	0.67	0.66	**0.77**	Soft Voting
0.73	0.67	0.67	0.73	Hard Voting

الجدول 4–2 نتائج اختبار النموذج المرجعي على مجموعة البيانات L-HSAB

كما يبين الشكل 4–5 نتائج الاختبار بيانيًا.



الشكل 4–5 مخطط بياني لنتائج اختبار النموذج المرجعي على مجموعة البيانات L-HSAB

عند مقارنة هذه النتائج مع نتائج اختبارات النموذج المدرب على مجموعة البيانات L-HSAB المذكورة في الجدول 3–12، حيث نلاحظ التحسن الواضح في قيم المقياس f1-score لأغلب المصنفات. نجد أن المصنف CatBoost حصل على معدل f1-score بلغت 0.67 بزيادة 3 نقاط تقريبًا أي ما يعادل 5% تقريبًا. أما أفضل نتيجة فكانت للمصنف SVC بلغت 0.77 بزيادة عن المصنف CatBoost 14 نقطة مئوية تقريبًا أي ما يعادل 22% تقريبًا.

بالتالي، يمكن القول إن النموذج المدرب على مجموعة البيانات LHS-TRAIN-B أعطى قدرة على التعميم أكبر من النموذج المدرب على مجموعة البيانات L-HSAB.

4–4–2–2– الاختبارات على مجموعة بيانات ورشة عمل المحتوى العربي

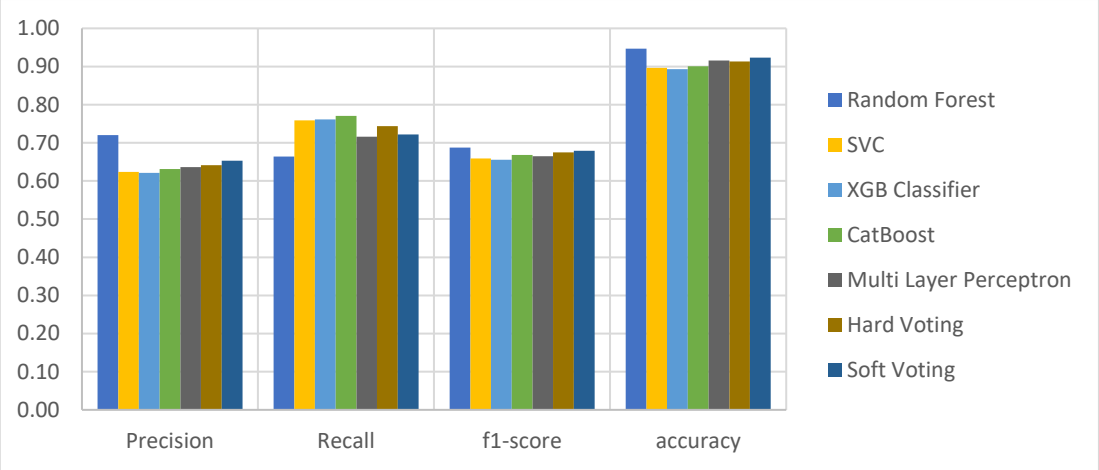
جرى اختبار النموذج على مجموعة بيانات ورشة عمل المحتوى العربي OSACT، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–3.

Accuracy	F1	Recall	Precision	Classifier

0.95	**0.69**	0.66	**0.72**	RF
0.90	0.66	0.76	0.62	SVC
0.89	0.66	0.76	0.62	XGB
0.90	0.67	**0.77**	0.63	CatBoost
0.92	0.66	0.72	0.64	MLP
0.92	0.68	0.72	0.65	Soft Voting
0.91	0.67	0.74	0.64	Hard Voting

الجدول 4–3 نتائج اختبار النموذج المرجعي على مجموعة البيانات OSACT

كما يبين الشكل 4–6 نتائج الاختبار بيانيًا.



الشكل 4–6 مخطط بياني لنتائج اختبار النموذج المرجعي على مجموعة البيانات OSACT

عند مقارنة هذه النتائج مع اختبارات النموذج المدرب المذكورة في الفقرة 2–7–3–، نلاحظ التحسن الواضح في قيم المقياس f1-score لأغلب المصنفات. نجد أن المصنف CatBoost حصل على معدل f1-score بلغت 0.67 بزيادة 4 نقاط تقريبًا أي ما يعادل 6% تقريبًا. أما أفضل نتيجة فكانت للمصنف RandomForest بلغت 0.69 بزيادة عن المصنف CatBoost 6 نقاط تقريبًا أي ما يعادل 10% تقريبًا.

بالتالي، يمكن القول إن النموذج المدرب على مجموعة البيانات LHS-TRAIN-B أعطى قدرة جيدة على التعميم بالنسبة لمجموعة البيانات OSACT.

4–4–2–3– الاختبارات على مجموعة بيانات ورشة عمل التقييم الدلالي

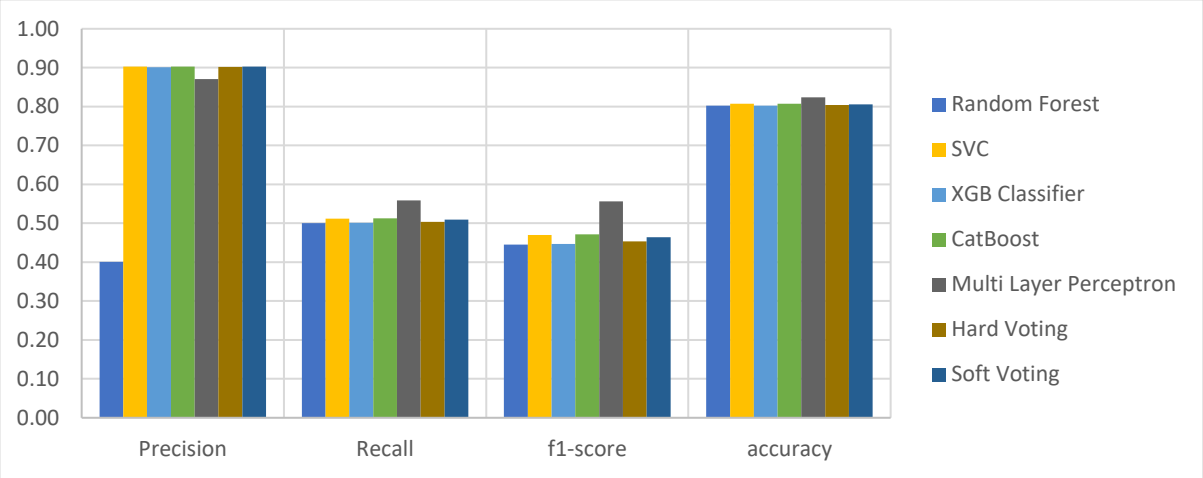
جرى اختبار النموذج على مجموعة بيانات ورشة عمل التقييم الدلالي OffensEval، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–4.

Accuracy	**F1**	**Recall**	**Precision**	**Classifier**
0.80	0.45	0.50	0.40	RF
0.81	0.47	0.51	**0.90**	SVC

0.80	0.45	0.50	**0.90**	XGB
0.81	0.47	0.51	**0.90**	CatBoost
0.84	**0.61**	**0.56**	0.89	**MLP**
0.81	0.46	0.51	**0.90**	Soft Voting
0.80	0.45	0.50	**0.90**	Hard Voting

الجدول 4–4 نتائج اختبار النموذج المرجعي على مجموعة البيانات OffensEval

كما يبين الشكل 4–7 نتائج الاختبار بيانيًا.



الشكل 4–7 مخطط بياني لنتائج اختبار النموذج المرجعي على مجموعة البيانات OffensEval

قمنا لاحقًا بدراسة تأثير تقنيات تعزيز البيانات الثلاث (التعزيز اليدوي، التعزيز الآلي، ودمج التقنيتين) على المصنفات المستخدمة.

نجد في ملحق نتائج الاختبارات بعد التدريب على التعزيز اليدوي والآلي نتائج دراسة تأثير تقنيات تعزيز البيانات اليدوي والتعزيز الآلي. كما نجد في الملحق نتائج الاختبارات بعد استخدام تقنية تعزيز البيانات برمجيًا باستخدام تقنية synthetic minority over-sampling technique (SMOTE) التي تحقق توازن مجموعة البيانات من خلال دمج تقنية over-sampling للصف الأقل نسبة وتقنية under-sampling للصف الآخر [171].

تمت دراسة تأثير دمج تقنيتي التعزيز اليدوي والآلي من خلال تدريب المصنفات على مجموعة البيانات LHS-TRAIN-E والتي أضفنا إليها العينات التي نتجت عن التعزيز اليدوي والتعزيز الآلي. جرى اختبار النموذج على مجموعات البيانات الأربعة المذكورة سابقًا.

4–4–3– الاختبارات على مجموعة البيانات المحلية

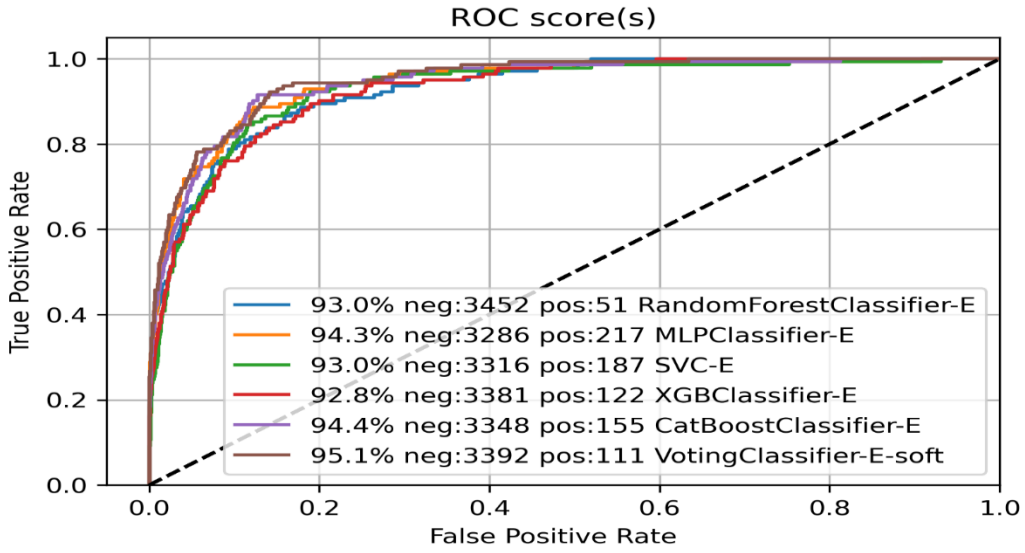
جرى تدريب المصنفات على مجموعة البيانات LHS-TRAIN-E والتي أضفنا إليها العينات الناتجة عن تقنيات التعزيز اليدوي والآلي، ثم اختبار النموذج على مجموعة البيانات LHS-

TEST، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–5.

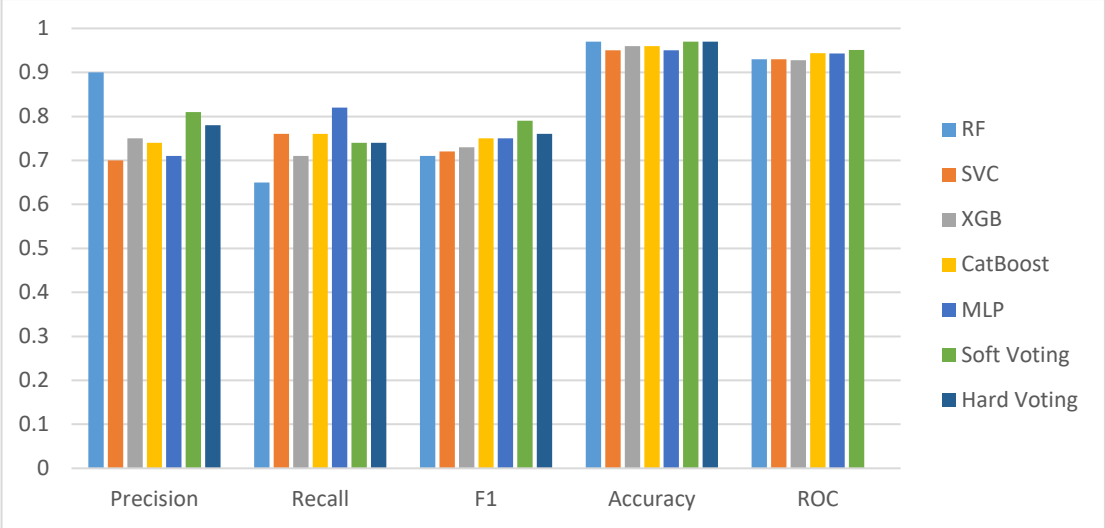
ROC	Accuracy	F1	Recall	Precision	Classifier
0.93	**0.97**	0.71	0.65	**0.90**	RF
0.93	0.95	0.72	0.76	0.70	SVC
0.928	0.96	0.73	0.71	0.75	XGB
0.944	0.96	0.75	0.76	0.74	CatBoost
0.943	0.95	0.75	**0.82**	0.71	MLP
0.951	**0.97**	**0.79**	0.74	0.81	**Soft Voting**
	0.97	0.76	0.74	0.78	Hard Voting

الجدول 4–5 نتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات LHS-TEST

كما يبين الشكل 4–8 نتائج الاختبار على منحنى AUC-ROC. كذلك، يبين الشكل 4–9 مخططًا بيانيًا لنتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج تقنيتي التعزيز.



الشكل 4–8 مخطط ROC لنتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات LHS-TEST



الشكل 4–9 مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات LHS-TEST

عند مقارنة هذه النتائج مع اختبارات النموذج المرجعي في الفقرة 4–4–، نلاحظ التحسن الواضح في قيم المقياس f1-score حيث كانت أقل قيمة 0.71 للمصنف RandomForest، وأما **أفضل نتيجة فكانت للمصنف Soft Voting بقيمة 0.79**.

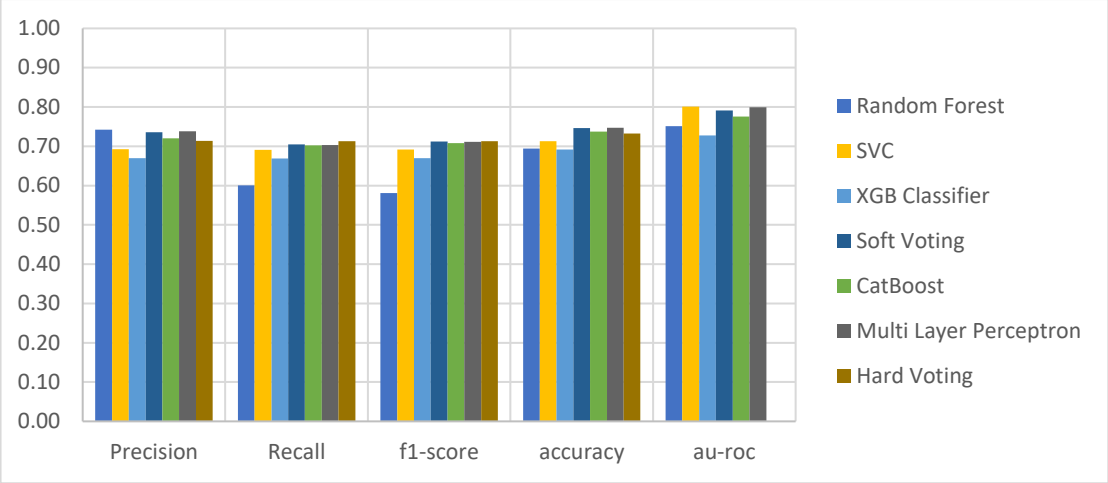
4–4–4– الاختبارات على مجموعة البيانات اللبنانية

جرى اختبار النموذج على مجموعة البيانات اللبنانية L-HSAB، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–6.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.751	0.69	0.58	0.60	**0.74**	RF
0.801	0.71	0.69	0.69	0.69	SVC
0.728	0.69	0.67	0.67	0.67	XGB
0.776	0.74	**0.71**	0.70	0.72	CatBoost
0.799	**0.75**	**0.71**	0.70	**0.74**	MLP
0.791	**0.75**	**0.71**	0.70	**0.74**	Soft Voting
	0.73	**0.71**	**0.71**	0.71	Hard Voting

الجدول 4–6 نتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات L-HSAB

كما يبين الشكل 4–10 نتائج الاختبار بيانيًا.



الشكل 4–10 مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات L-HSAB

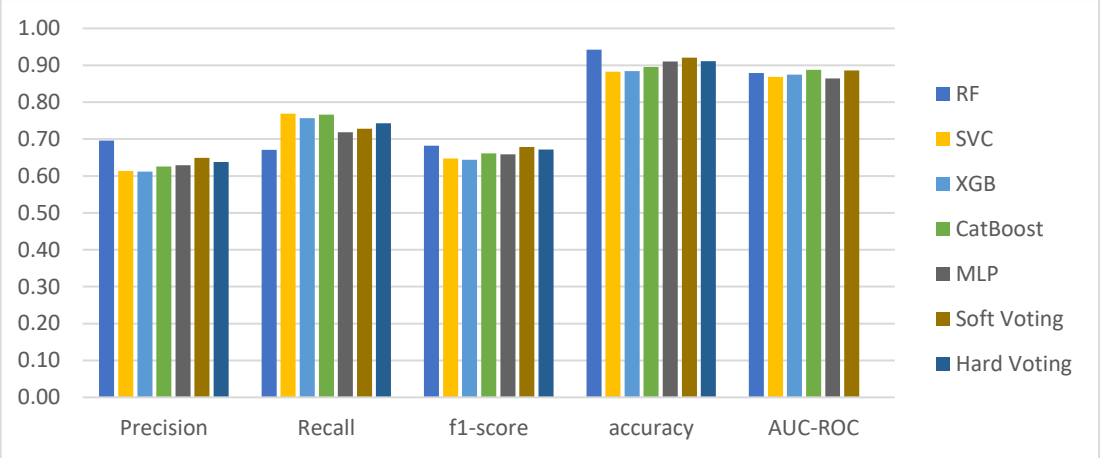
4–4–5– الاختبارات على مجموعة بيانات ورشة عمل المحتوى العربي

جرى اختبار النموذج على مجموعة بيانات ورشة عمل المحتوى العربي OSACT، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–7.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.879	0.94	0.68	0.67	0.70	RF
0.869	0.88	0.65	0.77	0.61	SVC
0.875	0.88	0.64	0.76	0.61	XGB
0.888	0.90	0.66	0.77	0.63	CatBoost
0.864	0.91	0.66	0.72	0.63	MLP
0.886	0.92	0.68	0.73	0.65	Soft Voting
	0.91	0.67	0.74	0.64	Hard Voting

الجدول 4–7 نتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات OSACT

كما يبين الشكل 4–11 نتائج الاختبار بيانيًا.



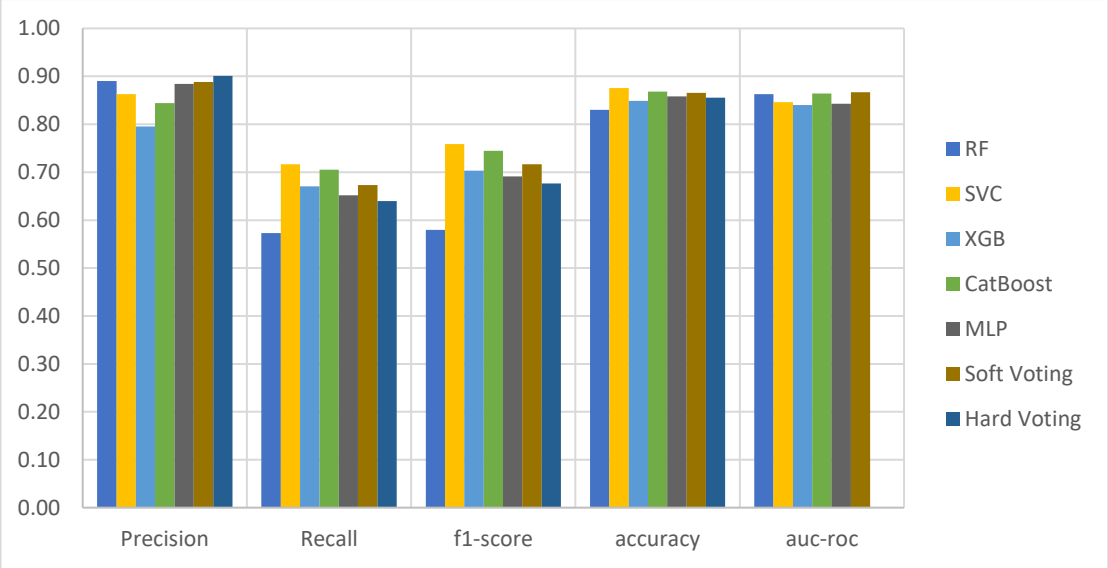
الشكل 4–11 مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات OSACT

جرى اختبار النموذج على مجموعة بيانات ورشة عمل التقييم الدلالي OffensEval، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–8.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.863	0.83	0.58	0.57	0.89	RF
0.846	0.88	0.76	0.72	0.86	SVC
0.84	0.85	0.70	0.67	0.80	XGB
0.864	0.87	0.74	0.71	0.84	CatBoost
0.843	0.86	0.69	0.65	0.88	MLP
0.867	0.86	0.68	0.64	0.90	Soft Voting
	0.87	0.72	0.67	0.89	Hard Voting

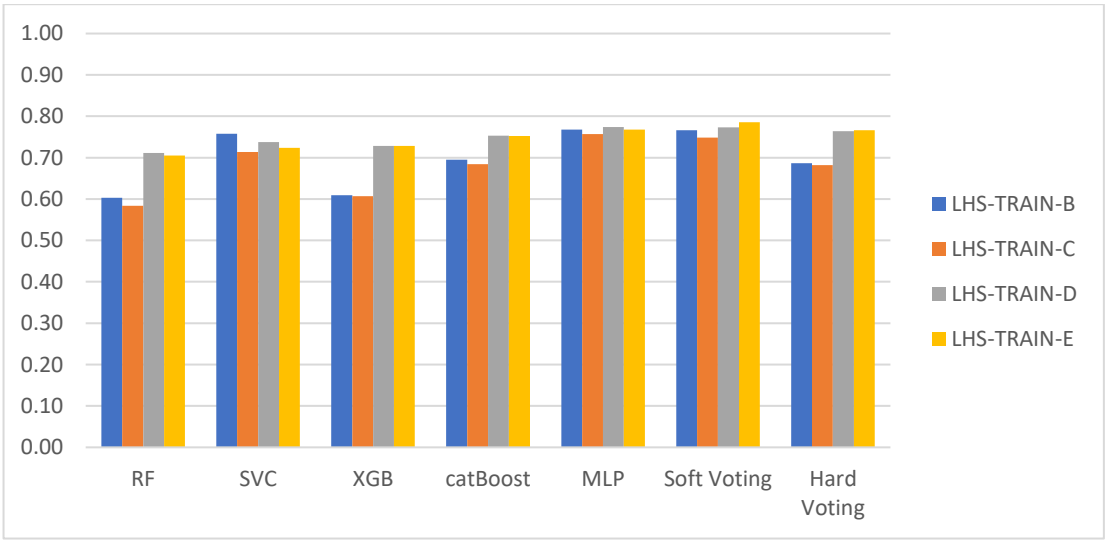
الجدول 4–8 نتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات OffensEval

كما يبين الشكل 4–12 نتائج الاختبار بيانيًا.



الشكل 4–12 مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة البيانات بعد دمج التعزيز اليدوي والآلي على مجموعة البيانات OffensEval

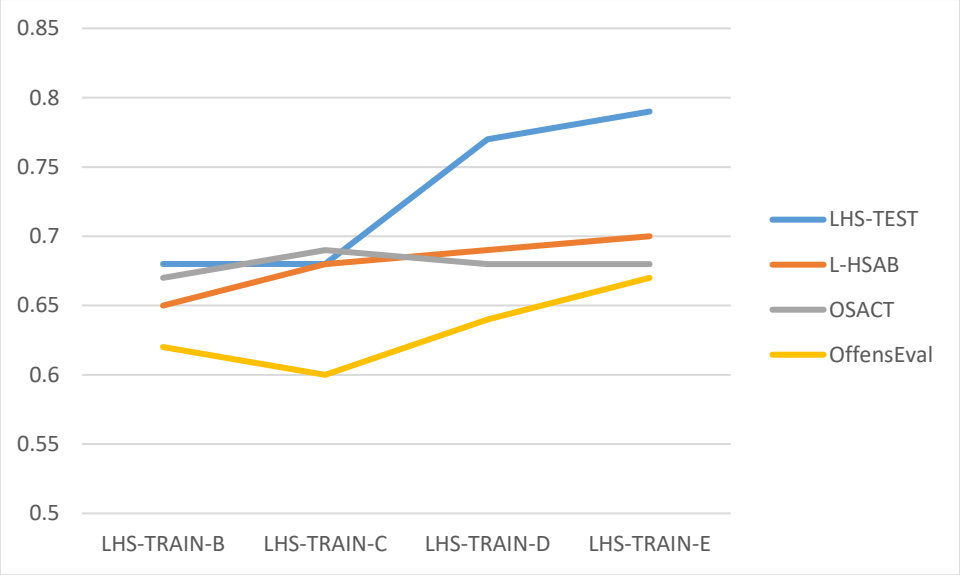
نعرض في هذه الفقرة ملخصًا لنتائج المعيار F1-score التي حصلنا عليها من مجموعات البيانات التي تختلف فيما بينها بتقنية التعزيز المستخدمة. تبين هذه النتائج تحسن أداء النموذج ولا سيما في حال دمج تقنيات تعزيز البيانات المستخدمة، كما هو مبين في الشكل 4–13.



الشكل 4–13 مقارنة تأثير تعزيز البيانات على نموذج الكشف

4–4–6– مقارنة نتائج اختبار النموذج بين المجموعات المختبرة

نعرض في هذه الفقرة ملخصًا لنتائج اختبار هذا النموذج بين المجموعات المختبرة وفق تقنيات التعزيز المختلفة المستخدمة. يبين الشكل 4–14 مقارنة تأثير تعزيز البيانات على نموذج الكشف عن خطاب الكراهية بين عدة مجموعات بيانات مختلفة.



الشكل 4–14 مقارنة تأثير تعزيز البيانات على نموذج الكشف واختباره على عدة مجموعات بيانات

يمكننا بسهولة ملاحظة مدى تأثير تعزيز البيانات على جودة نموذج الكشف، ويمكننا أيضًا ملاحظة هذا الأثر لدى اختباره على مجموعات البيانات الأخرى.

4–5– دراسة تأثير تعزيز البيانات على نموذج التضمين السياقي

جرى تطبيق بعض خوارزميات التّعلم العميق من خلال التدريب على مجموعة البيانات LHS-TRAIN-E واختبارها على مجموعات البيانات المختلفة.

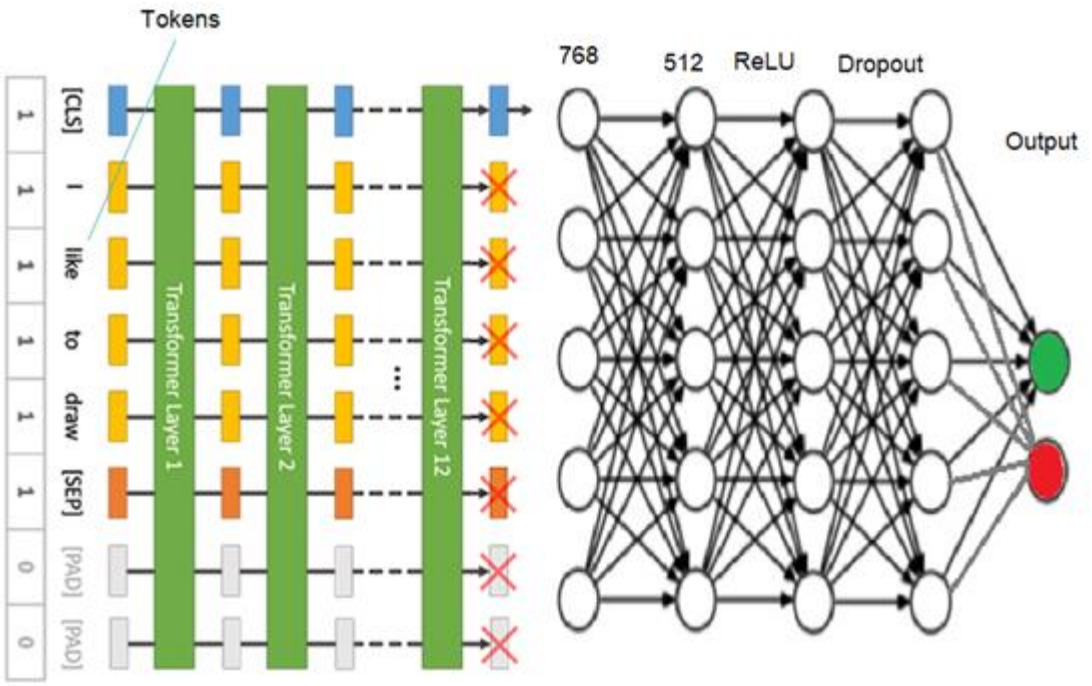
4–5–1– بنية النموذج

جري اعتماد ثلاث طرق تستخدم تضمين الكلمات السياقي وهي:

- AraBERT.
- ArabicBERT.
- GigaBERT.

يمرر ناتج هذا التمثيل إلى نموذج شبكة عصبونية كدخل. تتألف الشبكة العصبونية من طبقة تحويل خطي من متجهات بطول 768 إلى متجهات بطول 512. جرى تطبيق دالة تنشيط ReLU [172] على هذه الطبقة، ثم أضفنا مبدأ الإسقاط dropout [173] لتجنب مشكلة الملاءمة الزائدة over-fitting، حيث يُلغى تفعيل بعض العقد في كل مرحلة تدريب. أخيرًا تأتي طبقة تحويل خطي أخرى من متجهات بطول 512 إلى متجهات بطول 2.

يبين الشكل 4–15 بنية النموذج المعتمد على تضمين الكلمات السياقي.



الشكل 4–15 بنية نموذج التضمين السياقي للكشف عن خطاب الكراهية

4–5–2– نتائج الاختبار

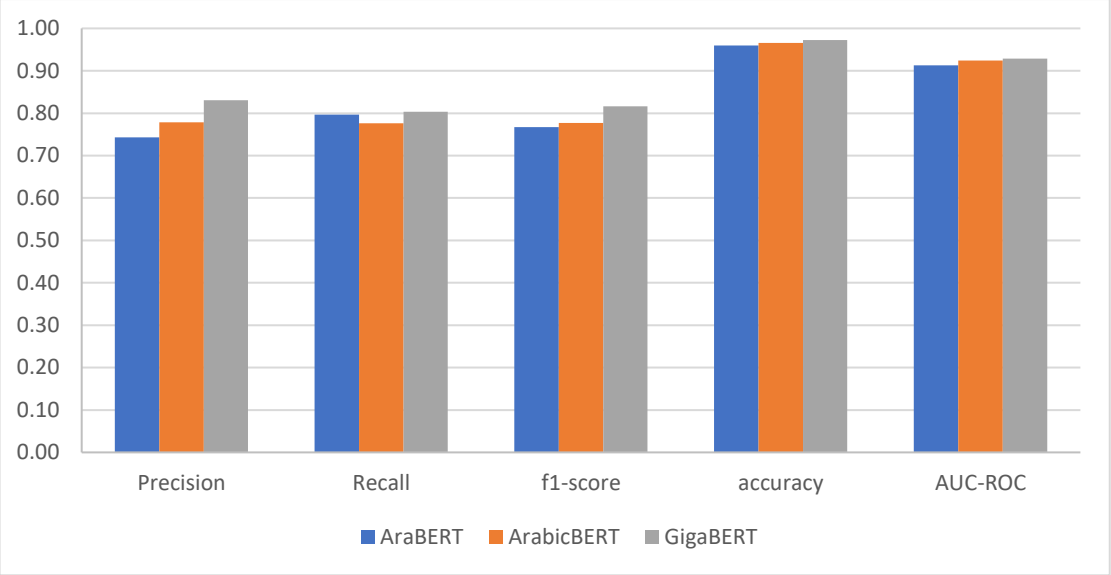
قمنا بتدريب النموذج على مجموعة بيانات التدريب مع دمج تقنيتي التعزيز LHS-TRAIN-E باستخدام طرق تمثيل البيانات التي ذكرناها، وجرى اختبار النموذج على مجموعة البيانات LHS-TEST، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–9.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.913	0.96	0.77	**0.80**	0.74	AraBERT
0.924	**0.97**	0.78	0.78	0.78	ArabicBERT
0.929	**0.97**	**0.82**	**0.80**	**0.83**	**GigaBERT**

الجدول 4–9 نتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات LHS-TEST

نلاحظ من هذه النتائج تفوق النموذج الذي اعتمد GigaBERT لتمثيل النصوص على بقية النماذج الأخرى، وحصلنا على معدل f1-score بلغ 0.82.

يبين الشكل 4–16 المخطط البياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات LHS-TEST. ونلاحظ من هذا المخطط تفوق النموذج الذي اعتمد GigaBERT لتمثيل النصوص على النموذجين الآخرين في كافة المقاييس.



الشكل 4–16 مخطط بياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات LHS-TEST

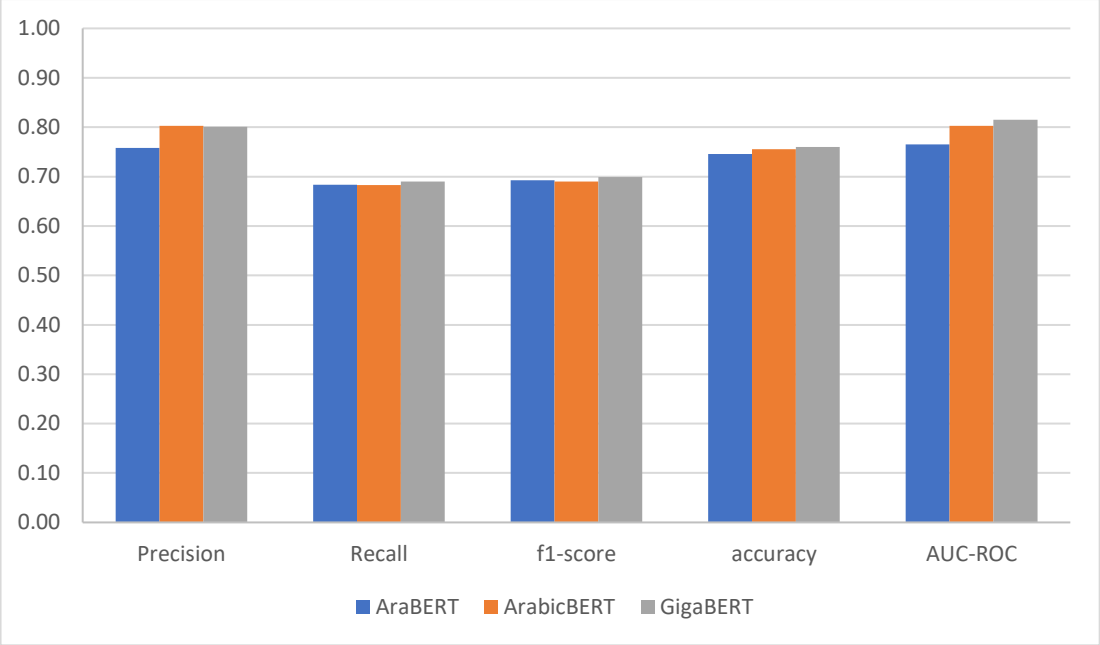
لاحقًا، جرى اختبار النموذج على مجموعة البيانات L-HSAB، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–10.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.765	0.75	0.69	0.68	0.76	AraBERT
0.803	**0.76**	0.69	0.68	**0.80**	ArabicBERT
0.815	**0.76**	**0.70**	**0.69**	**0.80**	**GigaBERT**

الجدول 4–10 نتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات L-HSAB

نلاحظ كذلك من هذه النتائج تفوق نموذج GigaBERT على بقية النماذج الأخرى، وحصلنا على معدل f1-score بلغ 0.70.

يبين الشكل 4–17 مخطط بياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات L-HSAB المخطط البياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات L-HSAB. ونلاحظ من هذا المخطط تفوق النموذج GigaBERT على النموذجين الآخرين في كافة المقاييس.



الشكل 4–17 مخطط بياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات L-HSAB

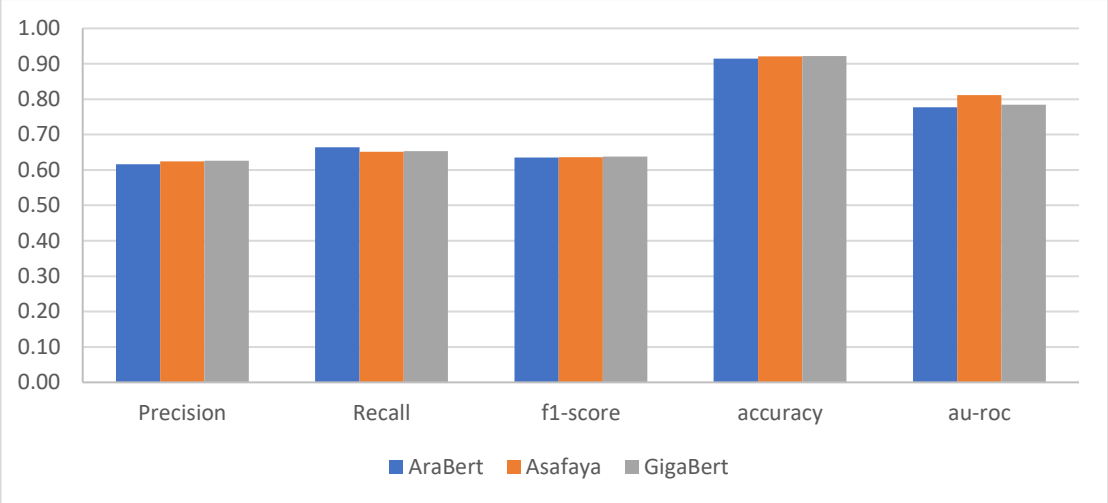
كذلك عند اختبار النموذج على مجموعة البيانات OSACT حصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–11.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.777	0.91	0.63	0.66	0.62	AraBERT
0.812	0.92	0.64	0.65	0.62	ArabicBERT
0.784	0.92	0.64	0.65	0.63	GigaBERT

الجدول 4–11 نتائج اختبار نموذج التضمين السياقي على مجموعة البيانات OSACT

نلاحظ كذلك من هذه النتائج حصول نموذجي GigaBERT وArabicBERT على نتائج متقاربة مع أفضلية نسبية لنموذج GigaBERT من مقياس الدقة، وحصلنا على معدل f1-score بلغ 0.64.

يبين الشكل 4–18 المخطط البياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات OSACT. ونلاحظ من هذا المخطط تفوق النموذج GigaBERT على النموذجين الآخرين في كافة المقاييس.



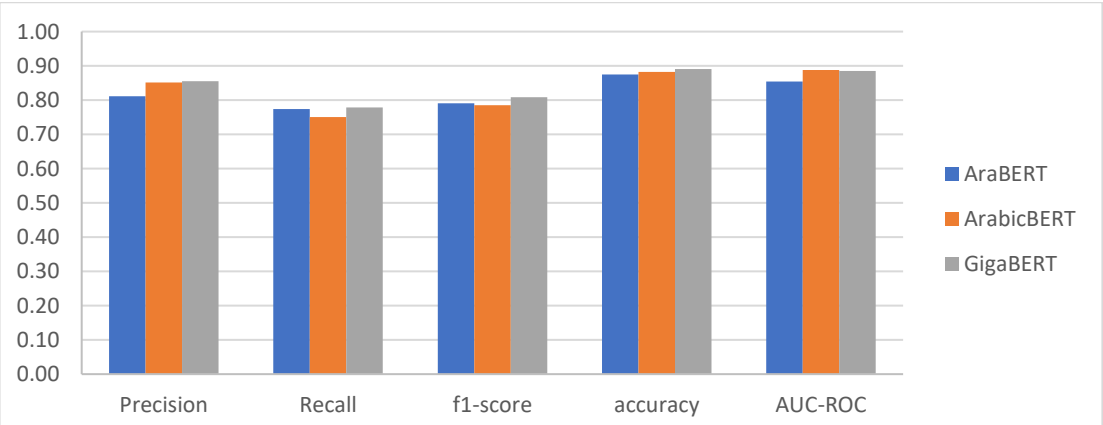
الشكل 4–18مخطط بياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات OSACT

كذلك جرى اختبار النموذج على مجموعة البيانات OffensEval، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول 4–12.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.854	0.87	0.79	0.77	0.81	AraBERT
0.888	0.88	0.79	0.75	0.85	ArabicBERT
0.885	**0.89**	**0.81**	**0.78**	**0.86**	**GigaBERT**

الجدول 4–12 نتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات OffensEval

يبين الشكل 4–19 المخطط البياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات OffensEval. ونلاحظ من هذا المخطط تفوق النموذج GigaBERT على النموذجين الآخرين في معظم المقاييس.



الشكل 4–19 مخطط بياني لنتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا على مجموعة البيانات OffensEval

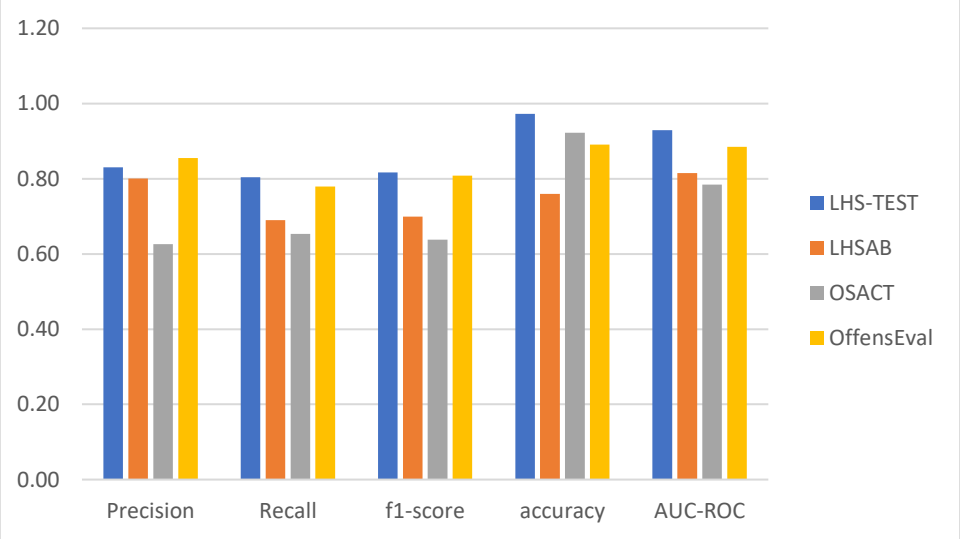
4-5-3- ملخص النتائج

نعرض في هذه الفقرة ملخصًا للنتائج التي حصلنا عليها نتيجة الاختبارات السابقة على مجموعات البيانات المختلفة في نموذج التضمين السياقي. تبين هذه النتائج أفضلية التمثيل GigaBERT على بقية التمثيلات الأخرى سواءً على مجموعة البيانات المحلية LHS-TEST أو على مجموعات البيانات الأخرى، مع أداء أفضل بالنسبة لمجموعة البيانات المحلية. يبين الجدول 4-13 مقارنة الاختبارات على مجموعات البيانات المختلفة وفق نموذج GigaBERT.

ROC	Accuracy	F1	Recall	Precision	Dataset
0.929	**0.97**	**0.82**	**0.80**	0.83	**LHS-TEST**
0.815	0.76	0.70	0.69	0.80	L-HSAB
0.784	0.92	0.64	0.65	0.63	OSACT
0.885	0.89	0.81	0.78	**0.86**	OffensEval

الجدول 4-13 مقارنة نتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا بين مجموعات البيانات

كذلك، يبين الشكل 4-20 مقارنة نتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا بين مجموعات البيانات.



الشكل 4-20 مخطط بياني يبين مقارنة نتائج اختبار نموذج التضمين السياقي المدرب على مجموعة البيانات المعززة يدويًا وآليًا بين مجموعات البيانات

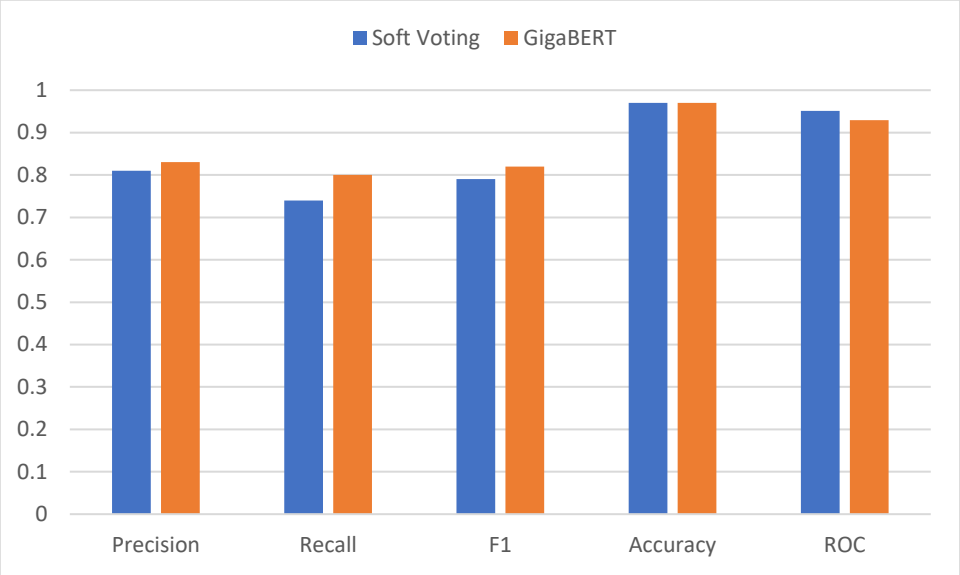
كذلك نلاحظ تحسن أداء نموذج التضمين السياقي بشكل واضح مقارنة مع المصنفات التي اعتمدنا فيها على التمثيل غير السياقي.

يبين الجدول 4-14 مقارنة أفضل نموذج (Soft Voting) باستخدام التمثيل غير السياقي مع أفضل نموذج تمثيل سياقي GigaBERT.

ROC	Accuracy	F1	Recall	Precision	Classifier
0.951	**0.97**	0.79	0.74	0.81	Soft Voting
0.929	**0.97**	**0.82**	**0.80**	**0.83**	**GigaBERT**

الجدول 4–14 مقارنة نتائج اختبار نموذجي التعلم السياقي وغير السياقي على مجموعة البيانات LHS-TEST

كذلك، يبين الشكل 4–21 مقارنة نتائج اختبار نموذجي التعلم السياقي وغير السياقي على مجموعة البيانات LHS-TEST.



الشكل 4–21 مخطط بياني يبين مقارنة نتائج اختبار نموذجي التعلم السياقي وغير السياقي على مجموعة البيانات LHS-TEST

4–6– خاتمة

قدّمنا في هذا الفصل شرحًا تفصيليًّا لنماذج التّعلم المعتمدة على تضمين الكلمات غير السياقي التي اختبرناها على مجموعات البيانات المحلية المحصلة، وأعطت نتائج جيدة مقارنة مع اختبارها على عدة مجموعات بيانات أخرى. لاحظنا في هذه الاختبارات أثر تعزيز البيانات على جودة هذه النماذج. كما قدمنا شرحًا تفصيليًا لنماذج التّعلم العميق المعتمدة على تمثيل تضمين الكلمات السياقي واختبارها أيضًا على عدة مجموعات بيانات مع أفضلية بالنسبة لمجموعة البيانات المحلية. لاحظنا في هذه الاختبارات تفوق نماذج التّعلم العميق المعتمدة على تمثيل تضمين الكلمات السياقي على النماذج الأخرى، ما يمثل الإجابة على السؤال Q3– **ما هو التمثيل الأفضل للنصوص في مجموعات بيانات خطاب الكراهية؟** كذلك قدمنا الإجابة على السؤال Q4– **"هل يساعد بناء مجموعة بيانات وتدريب نظم تصنيف عليها في زيادة القدرة على التعميم؟".**

بالرغم من حصولنا على معدل f1-score بلغ 0.82 وهي نسبة جيدة خصوصًا في مسألة شائكة كمسألة كشف خطاب الكراهية، إلا أن هذا النموذج كغيره من نماذج التعلم الخامل passive learning يعاني من عدة إشكاليات:

- ثبات أداء النموذج: حيث أننا لزيادة أو تحسين أداء النموذج فنحن بحاجة لإضافة مجموعة كبيرة من العينات وإعادة تدريب النموذج من جديد واختباره، وكما ذكرنا سابقًا تعتبر عملية الوسم مكلفة من حيث الجهد والزمن.
- احتمالية الخطأ البشري: بما أن عملية الوسم تنفذ من قبل أشخاص، فهي معرضة لوجود أخطاء في الوسم لأسباب متعددة، الأمر الذي يؤثر سلبًا على أداء النموذج.
- عدم توازن مجموعة البيانات: حيث يتأثر أداء نظم التصنيف بشكل كبير في حال كانت مجموعات البيانات التي تجري عليها عمليات التدريب غير متوازنة.
- عدم ملاحقة التغيرات: تعتبر أنظمة التعلم الخامل غير قادرة على تتبع أو ملاحقة التغيرات التي يمكن أن تطرأ على التوزيع الخاص بالعينات، أي أن أداء نظام الكشف المقترح سينخفض أداؤه مع تغير التوزيع الخاص بعينات خطاب الكراهية.

نبين في الفصل التالي إطار العمل framework الذي اقترحناه والذي يعالج الإشكالات المذكورة سابقًا.

5- إطار عمل تكيفي لكشف خطاب الكراهية

نقدم في هذا الفصل شرحًا تفصيليًا لإطار العمل framework الخاص بكشف خطاب الكراهية. يعتمد إطار العمل على تقنيات التّعلم النشط بناءً على النماذج جرى اختبارها وتقييمها في الفصل السابق وذلك من أجل التحسين المستمر لمجموعة البيانات المعتمدة وتوسعتها وموازنتها من أجل رفع جودة هذه النماذج. سنحاول في هذا الفصل الإجابة على الأسئلة التالية: Q5- "**كيف يمكن انتخاب العينات في إجراءات التّعلم النشط بحيث تساهم في تحسين جودة نظم التّصنيف؟**" وكذلك السؤال Q6- "**كيف يساعد دمج تقنيات التّعلم النشط مع التّعلم الذاتي في تطوير نظام تصنيف قادر على متابعة التغيرات التي يمكن أن تطرأ على نظم التصنيف؟**"

5-1- مقدمة

تعتمد نماذج التّعلم التقليدية ونماذج التّعلم العميق بشكل كبير على مجموعة البيانات الموسومة، إلا أن الحصول على مجموعات بيانات كبيرة وموسومة يعتبر عملية مكلفة من حيث الزمن والجهد. إذا أخذنا بعين الاعتبار أن العينات لا تتمتع جميعها بنفس الدرجة من الأهمية، بالإضافة إلى إمكانية وجود توزع غير متجانس بين الصفوف، هنا تبرز أهمية التّعلم النشط في التغلب على هذه المشاكل.

جرى تطوير نظام يدمج بين تقنيات الوسم الذاتي Pseudo-labeling والذي يسمى في بعض المراجع بالتّعلم الذاتي self-training وتقنيات التّعلم النشط Active learning. يسمح النّظام المُطوَّر بوسم البيانات آليًّا، ويُمكِّن من انتخاب مجموعة محدودة من العينات بحيث تراعي من جهة التوازن بين الصفوف من خلال انتخاب عينات بعددٍ متساوٍ من كل صف، وتحتوي من جهة أخرى العينات التي تمثل البيانات بشكل مناسب والعينات التي تتضمن القدر الأكبر من عدم اليقين عند تصنيفها، وذلك من أجل إضافة هذه المجموعة لمجموعة التدريب. بالتالي، سيتم إضافة العينات بشكل تلقائي لمجموعة التدريب، وتباعًا سوف تميل مجموعة التدريب إلى التوازن بين الصفوف. يقوم النّظام المقترح أيضاً بعرض العينات التي تتضمن القدر الأكبر من عدم اليقين على مراقب خبير Oracle من أجل الحصول على الوسم النهائي لها مما يسمح بمتابعة التغيرات التي تطرأ على التوزيع الخاص المولد لعينات خطاب الكراهية، حيث يتوقع أن تكون العينات الجديدة ذات درجة عالية من عدم اليقين.

5-2- إجرائية التّعلم النشط

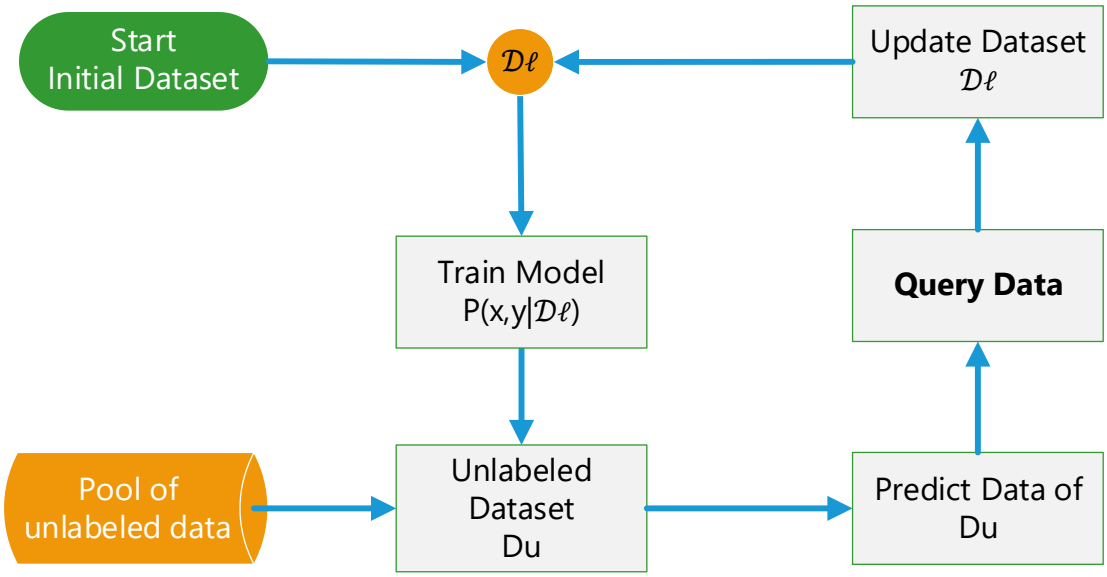
سنحاول في هذه الفقرة الإجابة على السؤال Q5 "**هل يمكن انتخاب عينات التدريب الأكثر تأثيرًا على أداء نظم التعلم النشط؟**".

تعتمد إجرائية التّعلم النشط بشكل عام على المكونات التالية:

- **مجموعة التدريب الأولية:** تبدأ عملية التّعلم النشط المستندة إلى مخزن pool-based بمجموعة صغيرة من العينات الموسومة، حيث تُحدَّد هذه المجموعة بشكل عشوائي في معظم تطبيقات التّعلم النشط. في حالتنا، اعتمدنا LHS-TRAIN-E كمجموعة بيانات أولية.
- **استراتيجية انتخاب العينات:** في هذا النموذج، جرى استخدام مفهوم الإسقاط dropout في الشبكات العصبونية خلال مرحلة الاختبار وإعادة الاختبار عددًا معينًا من المرات الأمر الذي يسمح باحتساب التوزيع التنبؤي اللاحق posterior predictive distribution. يجري انتخاب العينات من هذا التوزع من خلال طريقة جديدة في تقدير درجة عدم اليقين. سنقوم بشرح مفصل للآلية المعتمدة في انتخاب العينات في الفقرة التالية.
- **معيار التوقف:** تبقى الإجرائية في حالة عمل طالما يوجد مجموعة بيانات ضمن مخزن العينات غير الموسومة $\mathcal{P}_u$.
- **مقياس التقييم Evaluation Measure:** يوجد عدة طرق لتقييم نماذج التّعلم النشط. ففي نظم التصنيف classification، يجري اختبار مجموعة بيانات موسومة لقياس فعالية النموذج التي يمكن أن تقاس بصحة التعميم (أو خطأ التعميم) generalization accuracy [174] أو من خلال معدل f-score [175] أو بالاعتماد على مقياسي الضبط والإرجاع [62]. في حالتنا، اعتمدنا على مقياس f1-score كونه المعتمد لدينا في قياس أداء النماذج المستخدمة.

تبدأ عملية التعلم النشط بتدريب نظام التصنيف على مجموعة بيانات موسومة صغيرة $P(x,y|\mathcal{D}\ell)$، ثم يجري انتقاء مجموعة بيانات Du من مخزن العينات غير الموسومة $\mathcal{P}_u$ حيث تمرر هذه المجموعة إلى نظام التصنيف الذي يقوم باختبارها. يجري انتخاب بعض العينات وفق استراتيجية انتخاب العينات المعتمدة كتقدير درجة عدم اليقين. تضاف العينات المنتخبة إلى مجموعة البيانات الموسومة $\mathcal{D}_\ell$. يعاد تدريب نظام التصنيف على مجموعة البيانات الجديدة، ثم تكرر العملية حتى الوصول إلى عتبة معينة في أحد المقاييس كالصحة accuracy مثلًا أو حتى تحقق شرط توقف ما.

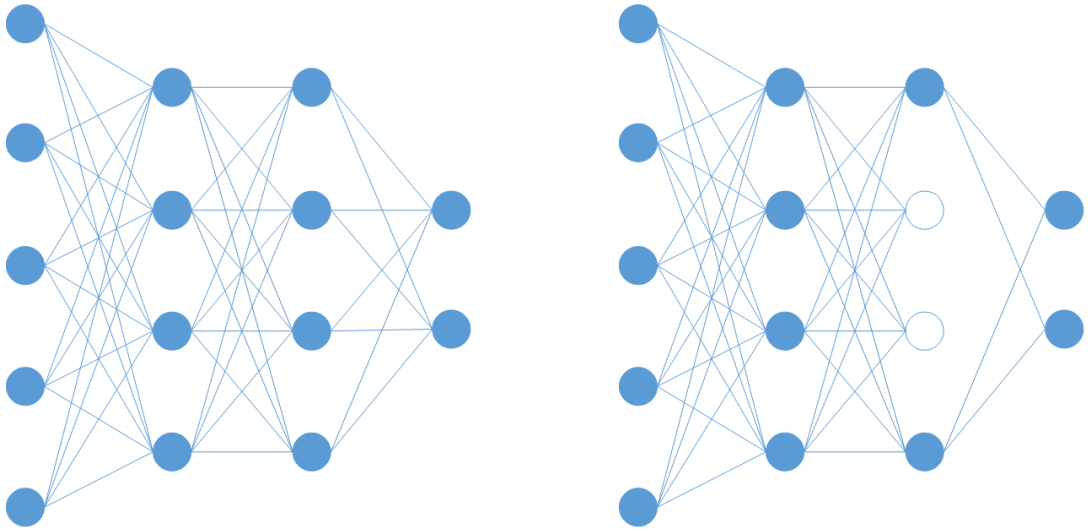
يبين *الشكل* 5–1 المخطط العام لنظم التعلم النشط.



الشكل 5–1 البنية العامة لنموذج التعلم النشط

5–3– استراتيجية انتخاب العينات

يجري استخدام مفهوم الإسقاط في الشبكات العصبونية لتجنب مشكلة الملاءمة الزائدة، حيث يُلغى تفعيل بعض العقد –بنسبة محددة– في كل مرحلة تدريب. يبين الشكل 5–2 مثالًا عن آلية الإسقاط.



الشكل 5–2 رسم توضيحي لتقنية الإسقاط

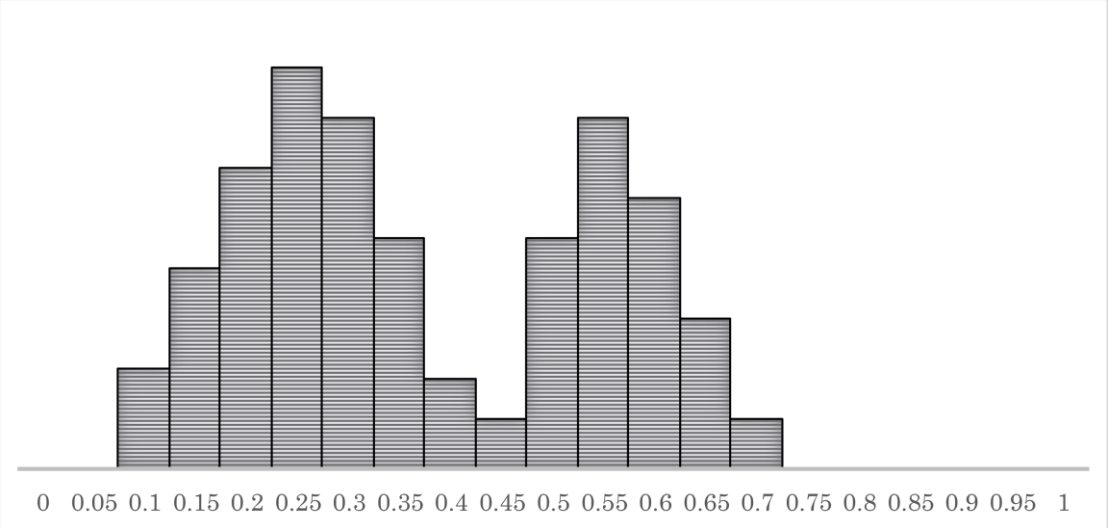
في الجزء اليساري من الشكل 5–2، يجري تدريب الشبكة العصبونية بدون استخدام تقنية الإسقاط، بينما تُستخدم تقنية الإسقاط بمعدل 0.5 في الشبكة العصبونية المبينة في الجزء اليميني من الشكل المذكور.

يجري عادةً استخدام مفهوم الإسقاط أثناء مرحلة التدريب، ولكن مع تطبيق الإسقاط أثناء الاختبار وتكراره لعدد كافٍ (100 مرة بناءً على بعض التجارب) فإننا نحصل على قيم توقع مختلفة في كل تكرار. تتيح هذه العملية الحصول على مجموعة كبيرة من الاحتمالات لكل عينة، الأمر الذي يسمح باحتساب التوزيع التنبؤي اللاحق posterior predictive distribution لكل عينة دخل x والحصول على $P(y|x, \theta)$. يمكننا القول أيضًا، أننا أمام مقاربة جديدة في التعلم النشط بالاستعلام عن طريق لجنة Query-By-Committee، ففي حالتنا هذه أصبح لدينا لجنة مكونة من 100عضو أو مصنف ويمكننا تطبيق مقاربات مختلفة لانتخاب العينات المستخدمة في هذه الطريقة.

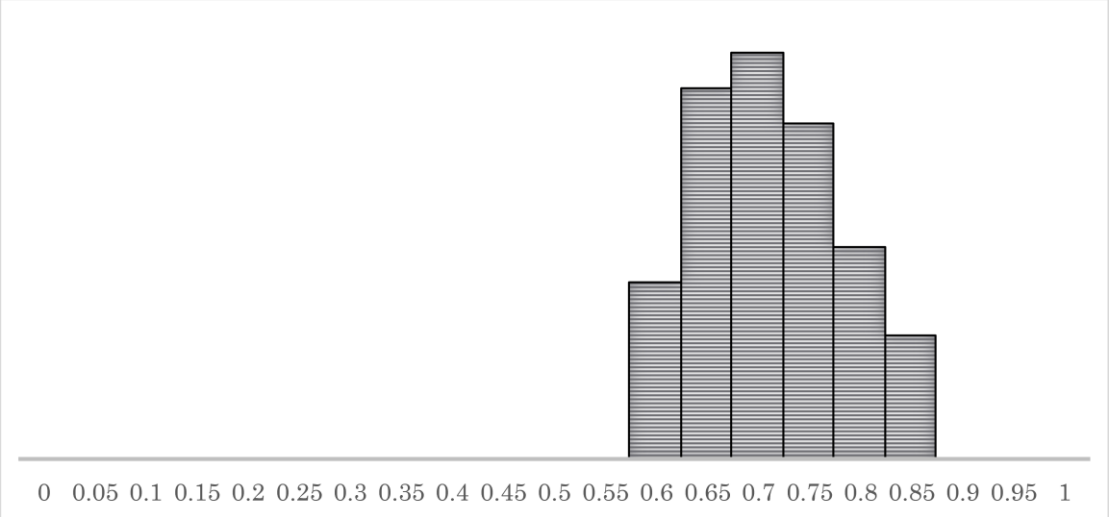
5-3-1- تقدير درجة عدم اليقين

تتيح معرفة التوزيع التنبؤي اللاحق مجموعة من البيانات المفيدة حول توقع كل عينة من خلال مجموعة من القراءات كالقيمة الوسطية والقيمة العظمى والقيمة الصغرى والانحراف المعياري. تتيح هذه القراءات تحديد درجة اليقين أو عدم اليقين في احتمال انتماء العينة لصف ما. نلاحظ وجود ثلاث حالات أساسية وهي:

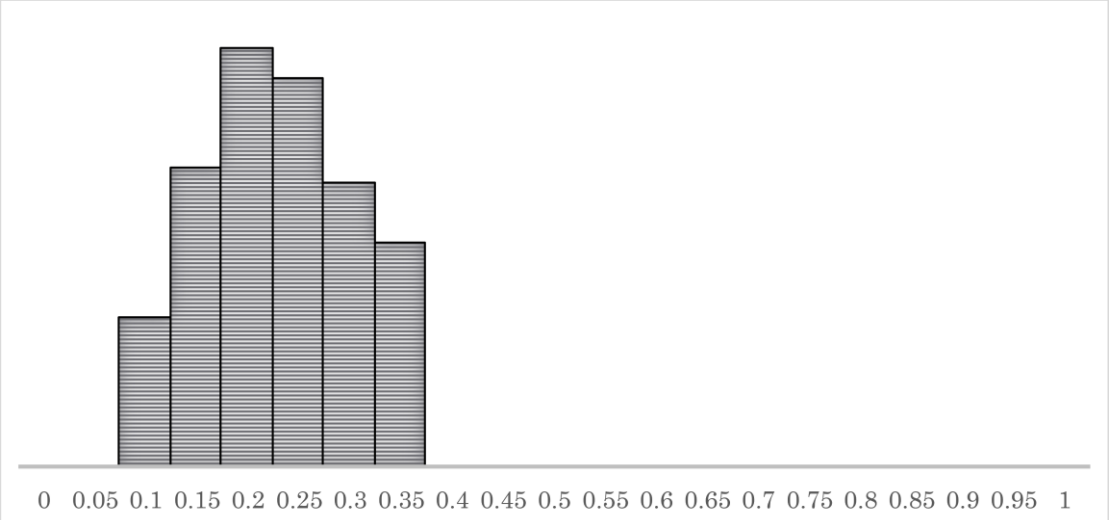
- درجة عالية من حالة عدم اليقين: اعتمدنا على تحديد العينات التي يكون فيها الفارق بين أكبر احتمال وأصغر احتمال أكبر من عتبة تباين v، وتكون هذه العتبة أكبر من 0.5 وذلك لضمان وجود الاحتمالات بين الصفين. يبين الشكل 5-3 مثالًا عن التوزيع التنبؤي اللاحق لإحدى العينات.
- درجة عالية من حالة اليقين لانتماء العينات للصف /1/ أو عينات الكراهية: اعتمدنا على تحديد العينات التي يكون فيها قيمة أصغر احتمال أكبر من عتبة t1، وتكون هذه العتبة أكبر من 0.5. يبين الشكل 5-4 مثالًا عن التوزيع التنبؤي اللاحق لإحدى العينات.
- درجة عالية من حالة اليقين لانتماء العينات للصف /0/ أو العينات العادية: اعتمدنا على تحديد العينات التي يكون فيها قيمة أكبر احتمال أصغر من عتبة t0، وتكون هذه العتبة أصغر من 0.5. يبين الشكل 5-5 مثالًا عن التوزيع التنبؤي اللاحق لإحدى العينات.



الشكل 5–3 التوزيع التنبؤي اللاحق لإحدى العينات التي تعبر عن عدم اليقين



الشكل 5–4 التوزيع التنبؤي اللاحق لإحدى العينات التي تعبر عن اليقين من الصف /1/



الشكل 5–5 التوزيع التنبؤي اللاحق لإحدى العينات التي تعبر عن اليقين من الصف /0/

يمكننا من خلال معرفة التوزيع التنبؤي اللاحق حول توقع كل عينة تحديد قيمة احتمال انتماء العينة لصف ما عبر عدة طرق:

- القيمة الاحتمالية الوسطية Average Probabilities: حيث نقوم بحساب وسطي جميع الاحتمالات الناتجة. تعبر p^* عن القيمة الاحتمالية الوسطية للعينة x وفق محددات النموذج θ، ومنها نقوم بحساب الوسم النهائي، كما تعبر n عن عدد مرات تكرار الاختبار (مثلًا، 100 مرة) وفق المعادلة التالية:

$$p^* = \frac{1}{n}\sum_i p_i(y|x,\theta) \quad \text{(E.7)}$$

- التنبؤ الوسطي Average Prediction: حيث نقوم بحساب وسطي جميع التنبؤات الناتجة. تعبر p^* عن القيمة الوسطية للعينة x وفق محددات النموذج θ، ومنها نقوم بحساب الوسم النهائي، كما تعبر n عن عدد مرات تكرار الاختبار وفق المعادلة التالية:

$$p^* = \frac{1}{n}\sum_i y_i(x,\theta) \quad \text{(E.8)}$$

- القيمة المتوسطة Median: حيث نقوم بأخذ قيمة التوزيع عند النقطة التي تفصل المساحة تحت المنحنى إلى النصف، أو القيمة التي تقع في المنتصف عند ترتيب جميع التنبؤات تصاعديًا أو تنازليًا:

$$p^* = \begin{cases} p_{\frac{n+1}{2}}(y|x,\theta) & n\ is\ odd \\ \frac{p_{\frac{n}{2}-1}(y|x,\theta)+p_{\frac{n}{2}+1}(y|x,\theta)}{2} & n\ is\ even \end{cases} \quad \text{(E.9)}$$

- القيمة العظمى Maximum: حيث نأخذ أكبر قيمة وفق المعادلة التالية:

$$p^* = \max_i p_i(y|x,\theta) \quad \text{(E.10)}$$

- القيمة الصغرى Minimum: حيث نأخذ أصغر قيمة وفق المعادلة التالية:

$$p^* = \min_i p_i(y|x,\theta) \quad \text{(E.11)}$$

تمنح العينة الوسم المبدئي y بغض النظر عن الطريقة المتبعة وفق المعادلة التالية:

$$y = \begin{cases} 1 & p^* \geq 0.5 \\ 0 & p^* < 0.5 \end{cases} \quad \text{(E.12)}$$

تتوافق القيم الناتجة عن هذه الطرق مع التنبؤ الخاص بالعينات ذات الدرجة العالية من اليقين، وتمنح كل عينة الوسم الناتج عن ذلك. بالنسبة للعينات ذات الدرجة العالية من حالة عدم اليقين،

تعطي هذه الطرق وسمًا مختلفًا لكل طريقة. اعتمدنا طريقة القيمة الاحتمالية الوسطية Average من أجل تحديد وسم أولي للعينات في حال عدم توفر خبير Oracle.

5-3-2- انتخاب العينات

عند تمرير مجموعة بيانات إلى إجرائية التعلم النشط المعتمدة من أجل انتخاب بعض العينات، نلاحظ توزع العينات في مجموعة البيانات هذه وفق التوزيع التنبؤي اللاحق الناتج إلى الأنواع المبينة في الجدول 5-1:

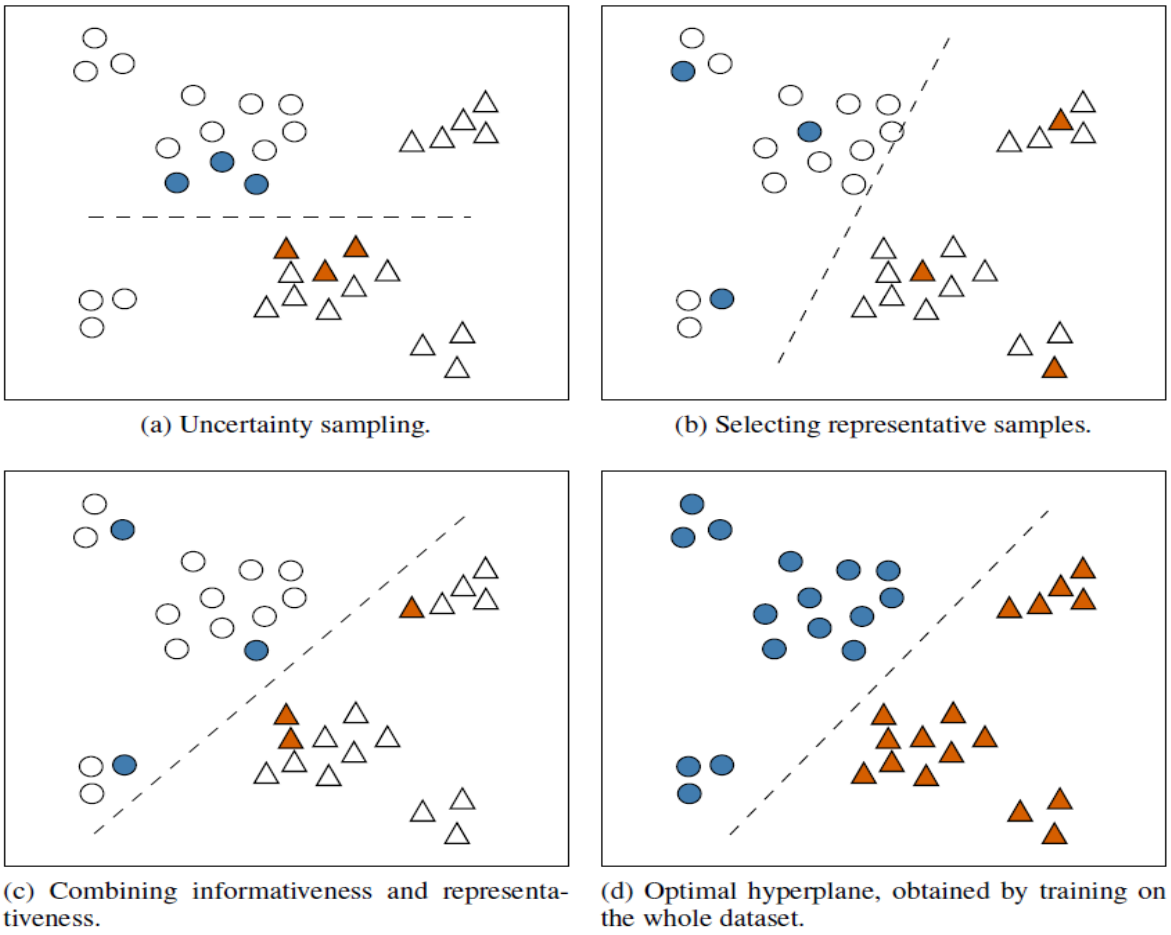
النوع	الرمز	النسبة التقديرية
درجة عالية من حالة عدم اليقين	S-uncertain	5%
درجة عالية من حالة اليقين مع تصنيف كعينة كراهية	S-certain-1	5%
درجة عالية من حالة اليقين مع تصنيف كعينة عادية	S-certain-0	90%

الجدول 5-1 أنواع العينات وفق التوزيع التنبؤي اللاحق

جرى تقدير هذه النسب بشكل تقريبي بناءً على تجارب عديدة وهي متوافقة مع مقياس صحة نموذج التعلم accuracy ونسبة عينات الكراهية إلى إجمالي العينات.

يعتبر انتخاب العينات من النوع S-uncertain هو الانتخاب بالإفادة informativeness، وأما انتخاب العينات من النوعين S-certain-1 وS-certain-0 هو الانتخاب بالتمثيل representativeness.

يؤدي الاعتماد على إحدى تقنيتي الانتخاب إلى انحياز النموذج نحو منطقة معينة من البيانات، حيث يعرض الباحثون Kremer وآخرون [176] أهمية الدمج بين تقنيتي الانتخاب: الإفادة informativeness والتمثيلية representativeness وفق الشكل 5-6 الذي يبين أن عملية الدمج بين التقنيتين تساعد أكثر في الوصول إلى نموذج أكثر واقعية بعدد أقل من العينات.



الشكل 5–6 أهمية دمج طريقتي الانتخاب: الإفادة والتمثيلية – المرجع [167]

لذلك، دمجنا طريقتي الانتخاب الإفادة والتمثيلية.

بالنسبة لعينات الإفادة، استخدمنا تقنية مقارنة التصويت Vote Comparison من خلال تنوع الوسم المتوقع diversity of predicted labels، حيث عبرنا عن درجة عدم اليقين من خلال قياس التباين بين الاحتمالات المتوقعة الذي يجب أن يكون أكبر أو يساوي عتبة معينة $v \in [0.5, 1[$ وذلك لضمان وجود اختلاف في التوقع النهائي بين هذه الاحتمالات. يتم حفظ العينات المنتقاة في مجموعة عينات الإفادة $\mathcal{D}_{inf}$: حيث يعبر $P_{i,c}(x)$ عن احتمالية انتماء العينة x إلى الصف i وفق المصنف c.

$$\mathcal{D}_{inf} = \underset{x,v}{\operatorname{argmax}} \left\{ \max_{c} P_{i,c}(x) - \min_{c} P_{i,c}(x) \geq v \right\} \quad \text{(E.13)}$$

تُمنَح هذه العينات وسمًا مبدئيًا قبل المرور على الخبير كما ذكرنا وفق المعادلة التالية:

$$f(x) = \begin{cases} 1 & \frac{1}{C}\sum_{c} P_c(x) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad \text{(E.14)}$$

أما بالنسبة لمجموعة عينات الانتخاب بالتمثيل $\mathcal{D}_{rep}$، فقد انتخبنا العينات التي حصلت على إجماع المصنفات وجرى استخدام دالة التمثيلية وفق المعادلة التالية:

$$\mathcal{D}_{rep} = \underset{x}{argmax} \left\{ \sum_i \prod_c P_{i,c}(x) \right\} \quad \text{(E.15)}$$

تُمنَح هذه العينات وسمًا نهائيًا كونها لا تمر على الخبير وفق المعادلة التالية:

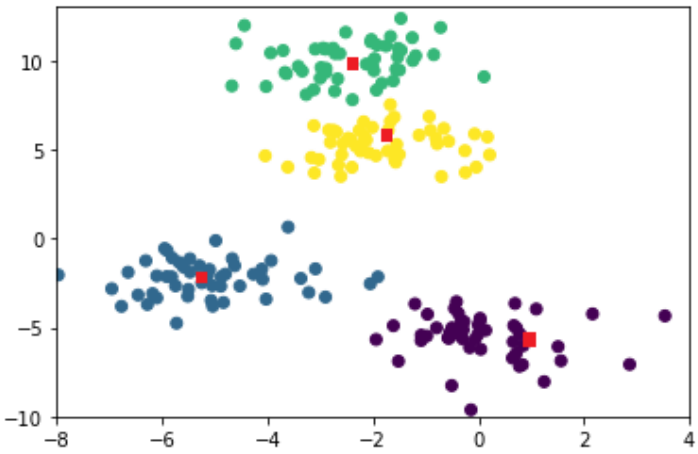
$$f(x) = \begin{cases} 1 & \min_c P_c(x) \geq 0.5 \\ 0 & otherwise \end{cases} \quad \text{(E.16)}$$

في كل تكرار iteration لمرحلة انتخاب العينات، تحتوي المجموعة المنتقاة $\mathcal{D}_{rep}$ على عدد من العينات من الصفين C_0, C_1 والتي توافق النوعين S-certain-0, S-certain-1 على الترتيب وليكن r_0, r_1 على الترتيب وعادةً ما تتبع r_0, r_1 لتوزع البيانات.

بما أن مجموعة البيانات غير متوازنة، فمن الطبيعي أن تتبع المجموعة المنتقاة لنفس التوزع distribution ويكون غالبًا $r_0 >> r_1$. بالتالي، فإن انتخاب جميع العينات سيؤدي دومًا إلى الانزياح باتجاه الصف المهيمن. لتجاوز هذه المشكلة، قمنا بتعريف موسط فوقي hyper-parameter بغية التحكم في عدد العينات المنتقاة من كل صف $n \in [-1, 1]$ تعبر القيم السالبة للموسط n عن انتخاب عينات أكثر من الصف C_0، وبالمقابل تعبر القيم الموجبة للموسط n عن انتخاب عينات أكثر من الصف C_1

اخترنا القيمة 0 لانتخاب عدد عينات متساوي من الصفين والتخفيف من مشكلة عدم التوازن بشكل تدريجي. تواجهنا هنا مشكلة كيفية انتخاب هذه العينات مع الحفاظ على توزع البيانات العام وعدم انتخاب عينات متشابهة قدر الإمكان. لحل هذه المشكلة، جرى اللجوء إلى خوارزميات العنقدة clustering بحيث تُقسَّم عينات الصف الأكبر –وهو في حالتنا C_0– إلى r_1 عنقود وانتخاب أقرب نقطة إلى مركز العنقود.

يبين الشكل 5–7 رسمًا توضيحيًا لإحدى الحالات.



الشكل 5–7 رسم توضيحي لآلية انتخاب العينات: لدينا 4 عينات كراهية و200 عينة عادية

لدينا في هذا المثال: r_1=4 r_0=200، نقوم هنا بتجزئة العينات من الصف C_0 إلى r_1 عنقود باستخدام خوارزمية k-Means [177] [178]. نقوم بتحديد مركز كل عنقود centroid وانتخاب أقرب نقطة إلى المركز، وبالتالي الحصول على r_1 عينة من الصف C_0. بالتالي تصبح المجموعة $\mathcal{D}_{rep}$ مكونة من r_1 عينة من الصف C_1 ومثلها من الصف C_0.

تطبق عملية موازنة العينات بين الصفين C_0, C_1 على عينات المجموعة $\mathcal{D}_{inf}$ لحفظ عملية التوازن بين الصفوف بنفس الطريقة.

كذلك الأمر، في كل تكرار iteration لمرحلة انتخاب العينات، لدينا المجموعتان $\mathcal{D}_{rep}$ و $\mathcal{D}_{inf}$ ولكل منهما عدد من العينات وليكن d_r, d_i على الترتيب وعادةً ما تتبع d_r, d_i لأداء النموذج المدرب. فمن الطبيعي أن يكون لدينا عدد عناصر أكبر في المجموعة $\mathcal{D}_{rep}$ ويكون $d_r >> d_i$.

لذلك، قمنا بتعريف موسط فوقي hyper-parameter بغية التحكم في عدد العينات المنتقاة من كل مجموعة $u \in [-1, 1]$ كي لا يحدث تركيز أو انزياح تجاه نوع أكثر من الآخر. تعبر القيم السالبة للموسط u عن انتخاب عينات أكثر من المجموعة $\mathcal{D}_{inf}$، وبالمقابل تعبر القيم الموجبة للموسط u عن انتخاب عينات أكثر من المجموعة $\mathcal{D}_{rep}$. اخترنا القيمة 0 لانتخاب عدد عينات متساوي من المجموعتين.

تدمج العينات التي المنتخبة في المجموعتين $\mathcal{D}_{rep}$ و $\mathcal{D}_{inf}$ –كل عينة مع وسمها المبدئي– ضمن مجموعة بيانات جديدة وتُحفَظ في مخزن خاص بالعينات الموسومة $\mathcal{P}_\ell$. تتاح هذه المجموعة لاحقًا لخبير يمكن أن يقوم بعملية وضع وسم نهائي لعينات المجموعة $\mathcal{D}_{inf}$.

5-3-3- تقييم العينات المنتخبة

تُسحَب مجموعة من المخزن الخاص بالعينات الموسومة $\mathcal{P}_\ell$، ويفضل انتقاء مجموعة موسومة من قبل الخبير -إن وجدت-، ويضاف محتوى المجموعة المنتقاة إلى مجموعة التدريب $\mathcal{D}_\ell$ ويعاد تدريب النموذج على المجموعة الجديدة ثم يجري اختبار النموذج $\boldsymbol{\mathcal{M}}$ على مجموعة بيانات الاختبار LHS-TEST. في حال تحسن أداء النموذج وفق معيار f1-score، يجري تثبيت التعديلات على مجموعة البيانات $\mathcal{D}_\ell$ وحفظ الموسطات الجديدة لنموذج الكشف $\boldsymbol{\mathcal{M}}$. في الحالة المعاكسة، تُلغى التعديلات على مجموعة البيانات $\mathcal{D}_\ell$ واستعادة الموسطات القديمة لنموذج الكشف.

5-4- إطار عمل تكيفي لكشف خطاب الكراهية

نقدم في هذه الفقرة شرحًا تفصيليًا لإطار العمل الذي يقوم على كشف خطاب الكراهية في النصوص القصيرة المكتوبة باللغة العربية وباللهجة المشرقية، ويقوم باستخدام تقنيات التّعلم النشط لتحقيق هدفين أساسيين وهما:

- زيادة حجم مجموعة البيانات الموسومة مع التقليل المستمر من نسبة عدم التوازن بين الصفوف.
- تحسين أداء نموذج الكشف عن خطاب الكراهية عبر آلية انتخاب جديدة للعينات التي يمكن أن تسهم في رفع أدائه.

جرى تعريف مجموعة من الموسطات والإجرائيات ضمن إطار العمل، نذكر منها:

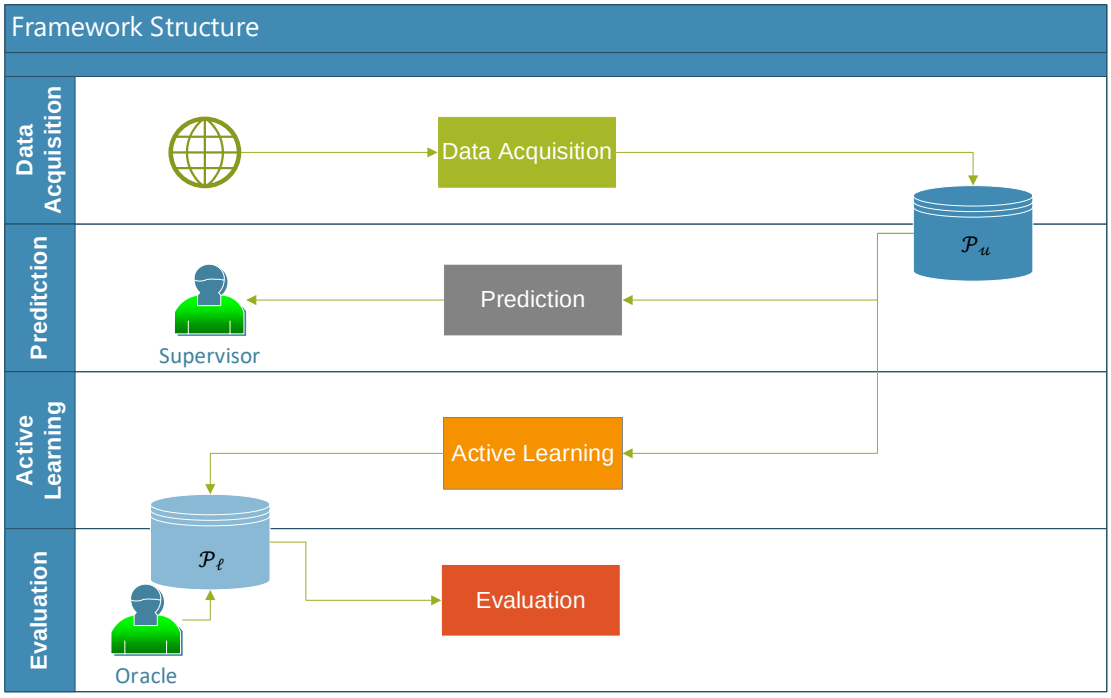
- β: موسط خاص لتحديد حجوم مجموعات البيانات المحصلة أو المخزنة في مخزن العينات.
- $\mathcal{P}_u$: مخزن مخصص للعينات غير الموسومة التي ستمرر لاحقًا لإجرائية التّعلم النشط.
- $\mathcal{P}_\ell$: مخزن مخصص للعينات الموسومة.
- $\mathcal{P}_{pre}$: إجرائية مخصصة للمعالجة المسبقة للنصوص.
- $\mathcal{M} = \{\text{GigaBERT}\}$: نموذج كشف خطاب الكراهية المعتمد.
- $\mathcal{P}_{rep}$: إجرائية مخصصة لتمثيل النصوص متضمنة في نموذج الكشف.
- $\mathcal{C}rawler$: تطبيق جمع العينات من موقع تويتر، ويمكن توسعته ليقوم بجمع العينات من مواقع التواصل الاجتماعي المختلفة.

- $\mathcal{A}_{oracle}$: وكيل agent مهمته إعطاء وسم للعينات المُمرَّرة إليه. قد يقوم الوكيل بإجراء تعديل على النص كأن يستبدل كلمة ما أو يجري تصحيحًا إملائيًا. كذلك، قد يقوم الوكيل بإضافة بعض العينات التي يراها مناسبة.
- $\mathcal{D}_\ell$: مجموعة البيانات الموسومة المعتمدة.
- $\mathcal{AL}$: إجرائية التّعلم النشط وتقوم بأخذ خرج نموذج الكشف وانتخاب العينات المناسبة، ولها ثلاث موسطات وهي:
  - ν: عتبة التباين بين التوقعات الاحتمالية للصفوف.
  - u: للموازنة بين عينات الإفادة وعينات التمثيلية.
  - n: للموازنة بين صفوف العينات المتوقعة.

يتكون إطار العمل من أربعة أقسام أساسية وهي:

- تحصيل العينات.
- اختبار العينات المحصلة.
- تطبيق آلية التّعلم النشط على العينات لانتخاب مجموعة منها.
- تقييم نتائج انتخاب العينات المنتقاة.

يمكن لكل من هذه الأقسام العمل بشكل منفصل عن المكونات الأخرى. يبين *الشكل 5–8* البنية العامة لإطار العمل المقترح.

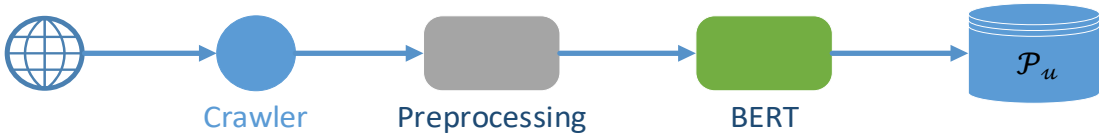


الشكل 5–8 البنية العامة لإطار العمل التكيفي لكشف خطاب الكراهية

5-4-1- تحصيل العينات

يقوم هذا القسم بشكل مستمر بتحصيل وجمع العينات من موقع تويتر عبر تطبيق crawler ويقوم هذا التطبيق بحفظ كل β عينة (مثلًا 1000 عينة) ضمن مجموعة في مخزن pool مخصص للعينات غير الموسومة $\mathcal{P}_u$ للاستخدام اللاحق.

يبين *الشكل* 5-9 بنية هذا القسم.



الشكل 5-9 القسم الخاص بتحصيل البيانات من مواقع شبكات التواصل الاجتماعي

يمكن لهذا التطبيق تحصيل العينات من مواقع شبكات التواصل الاجتماعي الأخرى بشرط أخذ العينات التي تحتوي على نصوص قصيرة فقط. يهتم هذا التطبيق بجمع العينات المكتوبة باللغة العربية فقط. يجري حاليًا تحصيل العينات من تويتر عبر API متاح من قبل الموقع، ويمكن تعديل خوارزمية التحصيل لتأمين بيانات وعينات من شبكات التواصل الاجتماعي الأخرى. عند الوصول إلى عدد محدد من العينات المحصلة β، تُحفظ هذه العينات في مجموعة جزئية S. تُمرَّر هذه المجموعة إلى إجرائية المعالجة المسبقة، ثم إلى إجرائية تمثيل النصوص (وهي في حالتنا GigaBERT). تُدمج العينات المعالجة مع تمثيلها ضمن مجموعة جديدة U، ثم تضاف المجموعة الجديدة إلى مخزن البيانات غير الموسومة $\mathcal{P}_u$.

تبين الخوارزمية Algorithm 1 آلية العمل في هذا القسم.

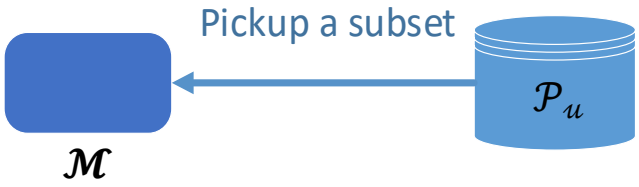
Algorithm 1: Data Acquisition - $\mathcal{C}rawler$

```
β = subset size parameter
Create subset S
while true do
    Try to get a sample from OSN
    if an error occurs wait 15 minutes
    if no more samples wait 15 minutes
    Add sample to subset S
    if size(S) = β
        S := Preprocessing(S)
        B := BERT(S)
        U := S + B
        Move new subset U to pool of unlabeled datasets 𝒫u
        Create a new subset S
end
```

5-4-2- اختبار العينات

يعمل هذا القسم بشكل مستمر طالما يوجد مجموعة عينات ضمن مخزن العينات غير الموسومة $\mathcal{P}_u$. يُمرَّر تمثيل النصوص في هذه المجموعة إلى نموذج التّعلم $\mathcal{M}$ أو بالأحرى إلى الشبكة العصبونية المدربة مسبقًا، والتي تقوم بدورها بتوقع تصنيف هذه العينات والحصول على النتيجة النهائية.

يبين *الشكل* 5-10 بنية قسم اختبار العينات.



الشكل 5-10 القسم الخاص باختبار العينات المحصلة

تبين الخوارزمية Algorithm 2 آلية العمل في هذا القسم.

Algorithm 2: Data Prediction

```
while true do
    Pick out a subset S of P_u
    Pass representation of S to prediction model M
    Show results
end
```

5-4-3- التّعلم النشط وانتخاب العينات

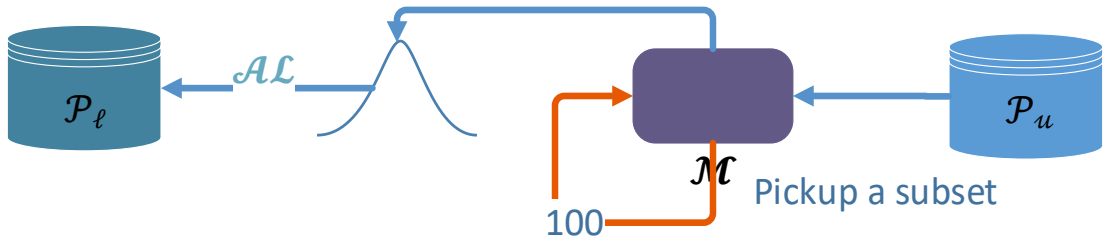
يعمل هذا القسم بشكل مستمر طالما يوجد مجموعة عينات ضمن المخزن $\mathcal{P}_u$. تُأخذ مجموعة عينات عشوائيًا من المخزن وتمرر إلى إجرائية المعالجة المسبقة $\mathcal{P}_{pre}$، ثم يُمرر خرج هذه الإجرائية إلى إجرائية تمثيل البيانات $\mathcal{P}_{rep}$. في هذه المرحلة، هنا يمكن إضافة مرشح للعينات المقبولة من خلال قياس جودتها وفق خصائص التغريدة.

اعتمدنا على تمثيل النص السياقي GigaBERT الذي أعطى أفضل النتائج مقارنة مع النماذج الأخرى. يُمرَّر خرج هذه الإجرائية إلى الشبكة العصبونية $\mathcal{M}$ التي جرى تدريبها مسبقًا، والتي تقوم بدورها بتوقع تصنيف هذه العينات. تُمرَّر هذه التوقعات إلى إجرائية التّعلم النشط $\mathcal{AL}$ والتي

شرحنا آلية عملها في الفقرة 3-5-. تقوم هذه الإجرائية بتحديد العينات التي يُتوقَع أن تساهم في تحسين أداء نظام الكشف عن خطاب الكراهية. تنتمي العينات المنتقاة من خلال إجرائية التّعلم النشط إلى إحدى فئتين:

- فئة $\mathcal{D}_{inf}$ ذات درجة عالية من عدم اليقين uncertainty: وتُمنَح كل عينة وسمًا مبدئيًا بناءً على وسطي التوقعات الناتجة عن نموذج التّعلم المعتمد.
- فئة $\mathcal{D}_{rep}$ ذات درجة عالية من اليقين: وهي مؤلفة من بعض العينات التي حصلت على توقع واحد في مجمل التوقعات الناتجة عن نموذج التّعلم المعتمد. يعتبر التوقع الناتج هو الوسم المعتمد لهذه العينات.

تُحفظ العينات المنتقاة في مخزن العينات الموسومة $\mathcal{P}_\ell$ تحضيرًا لإجرائية قياس صحة الانتخاب. يبين *الشكل* 5-11 بنية هذا القسم.



الشكل 5-11 القسم الخاص بالتّعلم النشط لانتخاب العينات

تبين الخوارزمية Algorithm 3 آلية العمل في هذا القسم.

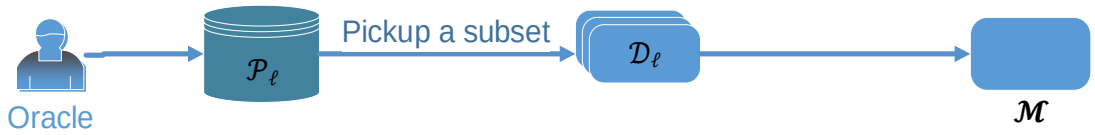
Algorithm 3: Data Sampling

while true **do**
  Pick out a subset S of $\mathcal{P}_u$
  Preprocessing of S
  Pass preprocessed S to prediction model $\mathcal{M}$
  $\mathcal{AL}$ produces two sets: $\mathcal{D}_{inf}$ and $\mathcal{D}_{rep}$
  Save samples with most uncertainty to $\mathcal{D}_{inf}$
  Attach $\mathcal{D}_{inf}$ samples with their average prediction
  Save samples with most certainty to $\mathcal{D}_{rep}$
  Attach $\mathcal{D}_{rep}$ samples with their certain prediction
  Save $\mathcal{D}_{inf}$ and $\mathcal{D}_{rep}$ into a subset in $\mathcal{P}_\ell$ pool
end

5-4-4- تقييم نتائج انتخاب العينات المنتقاة

يعمل هذا القسم بشكل مستمر طالما يوجد مجموعة عينات ضمن المخزن $\mathcal{P}_\ell$. في حال كان الخبير oracle متاحًا، يقوم باختيار مجموعة عينات من المخزن وتحديد الوسم الملائم لكل عينة من عينات الفئة $\mathcal{D}_{inf}$. ووضع مؤشر flag على هذه العينة بأنها خضعت لعملية الوسم النهائي. على التوازي، يقوم هذا القسم بانتخاب مجموعة بيانات من المخزن $\mathcal{P}_\ell$ ويُفضّل من العينات التي خضعت لعملية الوسم النهائي في حال وجودها. تضاف العينات في المجموعة المنتقاة إلى مجموعة البيانات الموسومة $\mathcal{D}_\ell$ والتي جرى تدريب النموذج عليها. يعاد تدريب النموذج على مجموعة البيانات الموسومة $\mathcal{D}_\ell$ واختبار النموذج الجديد على مجموعة بيانات الاختبار LHS-TEST. في حال تحسن أداء النموذج وفق معيار f1-score، تُحفَظ التعديلات على مجموعة البيانات $\mathcal{D}_\ell$ وحفظ الموسطات الجديدة لنموذج الكشف $\mathcal{M}$. في حال عدم وجود تحسن، تُلغى التعديلات التي تمت على مجموعة البيانات $\mathcal{D}_\ell$ واستعادة الموسطات القديمة لنموذج الكشف $\mathcal{M}$.

يبين *الشكل* 5-12 بنية القسم الخاص باختبار جودة العينات المنتقاة.



الشكل 5-12 القسم الخاص بتقييم نتائج العينات المنتقاة

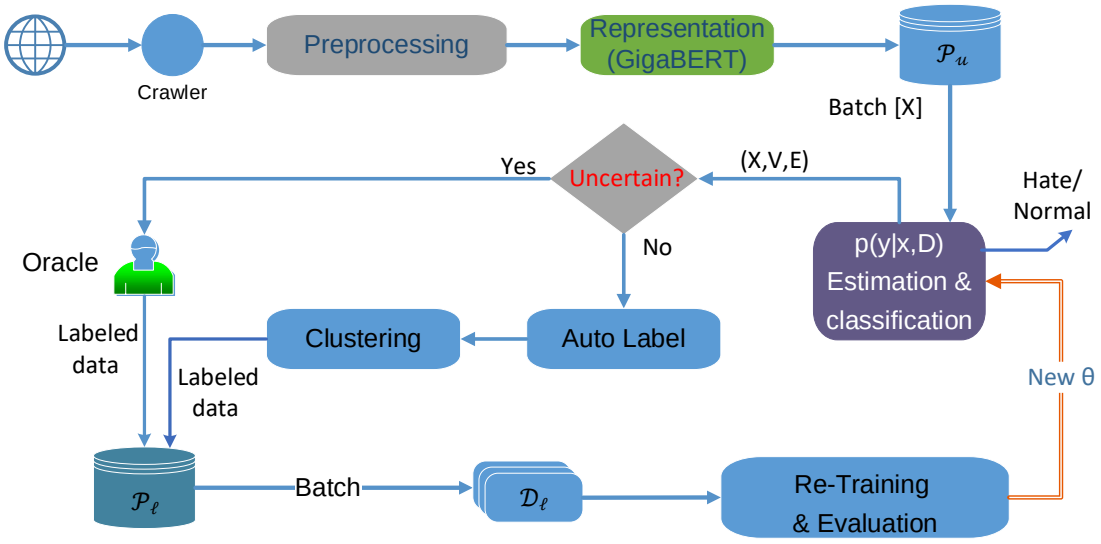
تبين الخوارزمية Algorithm 4 آلية العمل في هذا القسم.

Algorithm 4: Sampling Evaluation

Oracle:
  while available **do**
    Pick out a subset S of $\mathcal{P}\ell$
    Attach every sample to a new label
    Mark subset S as Checked by Oracle
  end
Evaluator:
while true **do**
  Pick out a subset S of $\mathcal{P}_\ell$ (Checked if available)
  Preprocessing of S
  Pass preprocessed S to prediction model $\mathcal{M}$
  $\mathcal{AL}$ produces two sets: $\mathcal{D}_{inf}$ and $\mathcal{D}_{rep}$
  Save samples with most uncertainty to $\mathcal{D}_{inf}$
  Attach $\mathcal{D}_{inf}$ samples with their average prediction

Save samples with most certainty to $\mathcal{D}_{rep}$
Attach $\mathcal{D}_{rep}$ samples with their certain prediction
Save $\mathcal{D}_{inf}$ and $\mathcal{D}_{rep}$ into a subset in $\mathcal{P}_\ell$ pool
end

يبين *الشكل* 5-13 البنية التفصيلية لإطار العمل.



الشكل 5-13 البنية التفصيلية لإطار عمل كشف خطاب الكراهية

5-5- كشف التوجه

نعتبر من الطبيعي حدوث تغيرات في خطاب الكراهية مع مرور الزمن. ونعتبر أنه عند ورود عينة تمثل خطاب الكراهية الجديد، سيجري تمييز هذه العينة وفق إطار العمل المقترح ضمن حالتين:

- تصنيف هذه العينة كخطاب كراهية بدرجة عدم يقين منخفضة (أو بمعنى أخر، بدرجة يقين عالية)، وبالتالي لا داعي لاتخاذ أي إجراء إضافي.
- تصنيف هذه العينة كخطاب كراهية بدرجة عدم يقين عالية، وبالتالي يجب تمرير هذه العينة إلى حكيم ومنحها وسم نهائي كعينة كراهية وإضافتها إلى مجموعة البيانات الموسومة وتدريب النموذج من جديد. عند ورود عينات مماثلة لهه العينة أو قريبة منها وتمريرها إلى الحكيم وإضافتها إلى الحكيم وتدريب النموذج من جديد، سوف تنخفض درجة عدم اليقين تباعًا إلى أن تقع ضمن منطقة اليقين العالية كعينة كراهية. أي أن إطار العمل المقترح قادر على تتبع التغيرات التي يمكن أن تطرأ على خطاب الكراهية وبالتالي كشف التوجه في هذا الخطاب.

ينطبق الأمر نفسه على التغير في الخطاب العادي، أي أن العينات العادية ستكون ذات درجة يقين عالية (أي لا داعي لاتخاذ إجراءات إضافية)، أو ذات درجة عدم يقين عالية وبالتالي ستمرر إلى حكيم لمنحها الوسم النهائي وإضافتها إلى مجموعة التدريب ولاحقًا ستنخفض درجة عدم اليقين تباعًا إلى أن تقع ضمن منطقة اليقين العالية كعينة عادية.

5-6- تنقيح البيانات

تبدأ نماذج التّعلم النشط عمومًا بالتدرب على مجموعة بيانات أولية ويجري إضافة العينات إليها تباعًا. في حالتنا، سنقوم باعتماد مجموعة البيانات LHS-TRAIN-E كمجموعة البيانات الأولية في التّعلم النشط، ولكن عملية الوسم قد تكون معرضةً للخطأ البشري، حتى وإن تمت عملية الوسم من قبل مجموعة من الأشخاص. لذلك، سنقوم بعملية تنقيح مجموعة البيانات قبل الخوض في اختبار نموذج التّعلم المعتمد. اعتمدنا في عملية تنقيح البيانات على نموذج تمثيل الكلمات السياقي GigaBERT.

وضعنا مجموعة من المقاربات التالية:

5-6-1- تحديد عتبة الفصل بين الصفوف

غالبًا ما تستخدم القيمة 0.5 للفصل بين الصفوف في نظم التصنيف الثنائي، ولكننا وضعنا المقاربات التالية:

- إذا كان احتمال أن تكون العينة عينة كراهية أكبر أو تساوي عتبة محددة Ω تكون عينة كراهية وإلا فهي عينة عادية، وهي الحالة الافتراضية عند عتبة مساوية للقيمة 0.5.

$$f(x) = \begin{cases} 1 & p_1(x) \geq \Omega \\ 0 & p_1(x) < \Omega \end{cases} \quad \text{(E.17)}$$

- إذا كان احتمال أن تكون العينة عينة عادية أكبر أو تساوي عتبة محددة تكون عينة عادية وإلا فهي عينة كراهية.

$$f(x) = \begin{cases} 1 & p_0(x) \leq \Omega \\ 0 & p_0(x) > \Omega \end{cases} \quad \text{(E.18)}$$

- إذا كان احتمال أن تكون العينة عينة كراهية أكبر من احتمال أن تكون العينة عينة عادية تكون العينة عينة كراهية وإلا فهي عينة عادية.

$$f(x) = \begin{cases} 1 & p_1(x) \geq p_0(x) \\ 0 & p_1(x) < p_0(x) \end{cases} \quad \text{(E.19)}$$

- إذا كان احتمال أن تكون العينة عينة كراهية أكبر أو تساوي عتبة محددة تكون عينة كراهية، أو إذا كان احتمال أن تكون العينة عينة عادية أكبر أو تساوي عتبة محددة تكون عينة عادية، وإلا تكون العينة غير محددة التصنيف.

$$f(x) = \begin{cases} 1 & p_1(x) \geq \Omega \\ 0 & p_0(x) > \Omega \\ undefined & otherwise \end{cases} \quad \text{(E.20)}$$

في حال كانت العتبة مساوية للقيمة 0.5، تصبح المقارنات السابقة متكافئة. اختبرنا المقاربات السابقة على عدة قيم للعتبة $\Omega \in \{0.5, 0.7, 0.9\}$، وكانت أفضل النتائج للقيمة 0.7 حيث حصلنا على تحسن طفيف في معدل f1-score بلغ 81.8 كما هو مبين في الجدول 5−2.

قيمة العتبة	0.5	0.7	0.9
معدل f1-score	81.6	81.8	80.6

الجدول 5−2 نتائج معدل F1-score عند عتبات عمل مختلفة

5−6−2− تنقيح مجموعة البيانات LHS-TRAIN-E

نهدف في هذه الفقرة إلى عزل العينات المشكوك في صحة وسمها بالاستفادة من إطار العمل، وإعادة عرض هذه العينات على خبير لمنحها الوسم النهائي.

جرى تقسيم مجموعة البيانات LHS-TRAIN-E بشكل عشوائي إلى أربعة أجزاء متساوية في الحجم وتتبع نفس توزع الصفوف. سُميت هذه الأجزاء كما يلي: TRAIN-E0, TRAIN-E1, TRAIN-E2, TRAIN-E3.

اختبرنا كل جزء عدة مرات من خلال تدريب النموذج على توليفة من الأجزاء الأخرى، كما هو مبين في الجدول 5−3:

م	الجزء المختبر	الأجزاء الداخلة في عملية التدريب			
		TRAIN-E0	TRAIN-E1	TRAIN-E2	TRAIN-E3
1	TRAIN-E0		✓		
2				✓	
3					✓
4				✓	✓
5			✓	✓	✓
6	TRAIN-E1	✓			
7				✓	

8					✓
9				✓	✓
10		✓		✓	✓
11	TRAIN-E2	✓			
12			✓		
13					✓
14		✓	✓		
15		✓	✓		✓
16	TRAIN-E3	✓			
17			✓		
18				✓	
19		✓	✓		
20		✓	✓	✓	

الجدول 5-3 تقسيم مجموعة البيانات إلى أربعة أرباع واختبار كل جزء على نموذج مدرب على توليفة من الأرباع الأخرى

جرت عملية تنقيح بيانات النصف الأول المكون من الجزأين TRAIN-E0 وTRAIN-E1 وفق الترتيب التالي:

1- اختبار الجزء TRAIN-E0 بعد تدريب النموذج وفق الجدول السابق.
2- عزل العينات التي أظهرت الاختبارات اختلافها عن الوسم اليدوي.
3- اختبار الجزء TRAIN-E1 بعد تدريب النموذج وفق الجدول السابق.
4- عزل العينات التي أظهرت الاختبارات اختلافها عن الوسم اليدوي.
5- اختبار الجزأين TRAIN-E0 وTRAIN-E1 بعد تدريب النموذج على الجزأين الآخرين، حيث جرى تنفيذ الاختبار مئة مرة وأخذ نتيجة التوقعات بحساب وسطي الاحتمالات.
6- عزل العينات التي أظهرت الاختبارات اختلافها عن الوسم اليدوي.
7- جمع العينات المعزولة من الخطوات السابقة وحذف المكرر منها.
8- عرض العينات المعزولة على خبير لوضع الوسم النهائي (تثبيت الوسم أو تغييره).
9- تدريب النموذج على مجموعة البيانات الجديدة بعد التنقيح وإجراء الاختبار.

تعاد نفس الخطوات السابقة على النصف الثاني المكون من الجزأين TRAIN-E2 و-TRAIN E3.

في النهاية، حصلنا على 1474 عينة من أصل 14017 ما يعادل 10.5%، ولدى عرضها على الخبير لتثبيت الوسم تبين وجود 717 عينة موسومة بشكل خاطئ ما يعادل 5.1%.

كذلك، جرى تنقيح مجموعة الاختبار LHS-TEST وفق الترتيب التالي:

1- تدريب النموذج على الجزأين TRAIN-E0 وTRAIN-E1واختباره على مجموعة الاختبار.
2- عزل العينات التي أظهر الاختبار اختلافها عن الوسم اليدوي.
3- تدريب النموذج على الجزأين TRAIN-E2 وTRAIN-E3واختباره على مجموعة الاختبار.
4- عزل العينات التي أظهر الاختبار اختلافها عن الوسم اليدوي.
5- تدريب النموذج على كامل مجموعة البيانات المنقحة وإجراء الاختبار على مجموعة التدريب مئة مرة وأخذ نتيجة التوقعات بحساب وسطي الاحتمالات.
6- عزل العينات التي أظهرت الاختبارات اختلافها عن الوسم اليدوي.
7- جمع العينات المعزولة من الخطوات السابقة وحذف المكرر منها.
8- عرض العينات المعزولة على خبير لوضع الوسم النهائي.

في النهاية، حصلنا على 201 عينة من أصل 3503 ما يعادل 5.7%، ولدى عرضها على الخبير لتثبيت الوسم تبين وجود 117 عينة موسومة بشكل خاطئ ما يعادل 3.3%. أخيرًا جرى حذف العينات التي تحتوي على كلمتين فقط، مع التذكير بأننا حذفنا سابقًا العينات التي تحتوي على كلمة في فقرة المعالجة الأولية، لنحصل على ملف اختبار مكون من 3287 عينة. جرى تدريب النموذج من جديد على كامل مجموعة البيانات المنقحة، ثم اختبرنا النموذج من جديد على مجموعة الاختبار المنقحة وحصلنا على معدل f1-score بلغ 0.851.

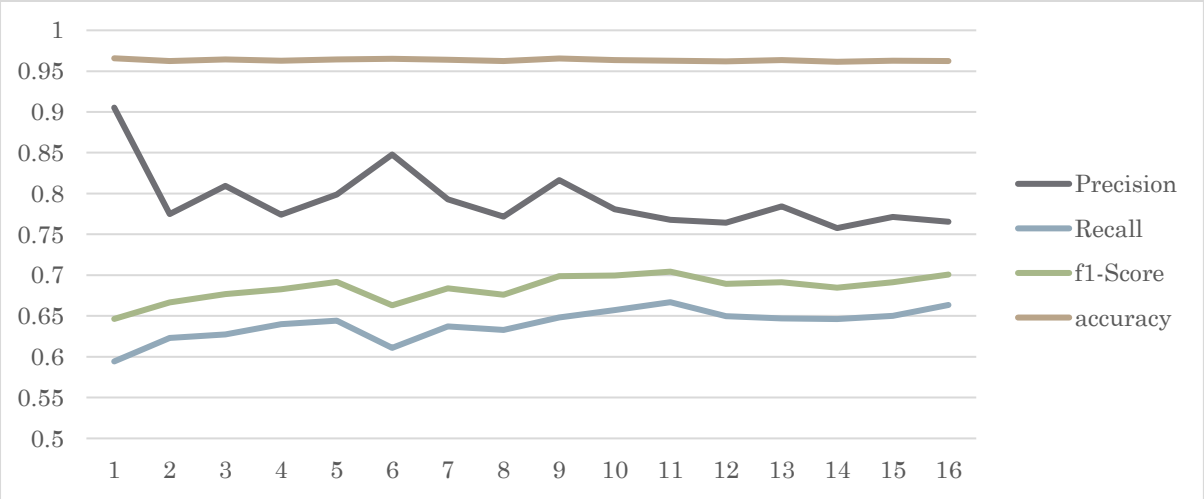
5-7- اختبارات

أجرينا اختبارًا أوليًّا للنموذج من خلال تقسيم مجموعة البيانات LHSS-TRAIN-B إلى 16 مجموعة جزئية. اخترنا إحدى هذه المجموعات عشوائيًا لتكون مجموعة التدريب الأولية، وكانت قيم الموسطات على الشكل التالي:

$$\{\beta = 867;\ v = 0.5;\ u = 0.8;\ n = 0\}$$

تمكنّا من الوصول إلى معدل F1-score بلغ 0.71 من مجموعة عينات بلغ عددها 1334 عينة.

يبين الشكل 5–14 تطور معدل f1-score مع تتالي تنفيذ الإجرائية.



الشكل 5–14 نتائج اختبار آلية الانتخاب على مجموعة مصنفات تقليدية

نلاحظ من خلال هذا الشكل التحسن التدريجي لنظام الكشف عن خطاب الكراهية من خلال إضافة العينات المنتخبة إلى مجموعة التدريب وإعادة تدريب النموذج من جديد والاختبار مرة ثانية، حيث ارتفع معدل F1-score من 0.64 عند اختبار النموذج المدرب على مجموعة مكونة من 876 عينة إلى 0.71 عند اختبار النموذج المدرب على مجموعة مكونة من 1334 عينة، أي من خلال إضافة 458 عينة منتخبة.

5–8– خاتمة

قدمنا في هذا الفصل شرحًا تفصيليًا لإطار عمل framework خاص بكشف خطاب الكراهية يعتمد تقنيات التّعلم النشط مبني على النماذج التي المختبرة والمقيمة في الفصل السابق بالاعتماد على التحسين المستمر لمجموعة البيانات المعتمدة وتوسعتها وموازنتها من أجل رفع جودة هذه النماذج. كذلك قدمنا الإجابة على الأسئلة الأساسية للبحث: Q5– "**كيف يمكن انتخاب مجموعة محددة من عينات التدريب ذات تأثير أكبر على نظم التصنيف؟**" والسؤال Q6– "**كيف يساعد دمج تقنيات التّعلم النشط مع التّعلم الذاتي في تطوير نظام تصنيف قادر على ملاحقة التغيرات دون أن ينخفض الأداء مع الزمن؟**"

6- الخاتمة والآفاق المستقبلية

نستعرض في هذا الفصل أهم المساهمات والخلاصات التي توصلنا إليها في نهاية هذا البحث، كما نبين أهم الجوانب التي يجب العمل عليها من أجل سد الثغرات التي لا تعالجها المقاربات المقترحة في هذه الأطروحة، أو من أجل رفع سوية أداء نظام كشف خطاب الكراهية في النصوص العربية في شبكات التواصل الاجتماعي.

6-1- المساهمات العلمية

نلخص فيما يلي المساهمات العلمية التي قدمناها في هذا البحث:

- دراسة مرجعية للآليات المعتمدة في الأبحاث التي تناولت موضوع كشف خطاب الكراهية في النصوص العربية في شبكات التّواصل الاجتماعي. جرى تحليل هذه الأبحاث ومقارنتها مع التّركيز على الآلية المعتمدة لتمثيل النصوص. قمنا باستخلاص الثغرات والإشكاليات التي عانت منها هذه الدراسات والتي حاولنا إيجاد الحلول لبعضها. تشكل هذه المراجعة نقطة انطلاق لطرح مقاربتنا، كما أنها تشكل أساسًا لأبحاث لاحقة في مجال معالجة اللغات الطبيعية لدعم كشف خطاب الكراهية.
- مجموعات بيانات موسومة يدويًا: تُشكِّل كل واحدة من هذه المجموعات عدة تغريدات مأخوذة من موقع تويتر موسومة وفقًا لاحتوائها على ما يدل أنها خطاب كراهية أو لا. تحتوي المجموعة الأولى على التّغريدات المحصلة بدون أي عملية تعزيز على البيانات، بينما تحتوي المجموعة الثانية بالإضافة إلى المجموعة الأولى مجموعة تغريدات أضيفت يدويًا، بينما تحتوي المجموعة الثالثة بالإضافة إلى المجموعة الأولى مجموعة تغريدات جرى إضافتها بشكل آلي. أما المجموعة الرابعة، فهي تحتوي بالإضافة إلى المجموعة الأولى جميع التّغريدات التي جرى إضافتها إلى كلٍ من المجموعة الثانية والثالثة، بينما تحتوي المجموعة الأخيرة على نفس العينات في المجموعة الرابعة ولكن بوسم منقح لبعض العينات. أظهرت الاختبارات التي قمنا بها قدرة نظامي كشف خطاب الكراهية المقترحين على التعميم بشكل جيد، ما يتيح للباحثين إمكانية الاعتماد على هذه المجموعات لدراسة مسألة كشف خطاب الكراهية في اللهجة المشرقية.
- نظام تصنيف آلي لكشف خطاب الكراهية مدرب على مجموعة البيانات المعززة يدويًا وآليًا اعتمادًا على تمثيل تضمين الكلمات غير السياقي من خلال استخدام تقنية التّصويت voting على مجموعة من المصنفات التّقليدية.
- نظام تصنيف آلي لكشف خطاب الكراهية مدرب على مجموعة البيانات المعززة يدويًا وآليًا اعتمادًا على تمثيل تضمين الكلمات السياقي من خلال استخدام شبكة عصبونية.

- خوارزمية جديدة لانتخاب العينات المرشحة لإجراءات التّعلم النشط من خلال تقدير عدم اليقين عبر احتساب التوزيع التنبؤي اللاحق posterior predictive distribution بالاعتماد على تقنية الإسقاط.
- مقاربة آلية جديدة لانتخاب العينات في نظم التّعلم النشط تراعي تحقيق التّوازن trade-off بين الصفوف وذلك للتقليل ما أمكن من عدم توزع العينات بين الصفوف، كما تراعي التّوازن بين عينات عدم اليقين uncertainty والعينات الأكثر تمثيلًا لفضاء العينات.
- تطوير إطار عمل تكيفي لمتابعة التغيرات التي تطرأ على تابع الكثافة الاحتمالية المولدة للبيانات يأخذ بالاعتبار النقاط أعلاه. يصلح إطار العمل لاستخدامه في النظم التي تتناول مجموعات بيانات غير مستقرة unstable data في التطبيقات المختلفة.
- حزمة برمجية أُنجِزت بلغة بايثون Python تتيح إمكانية الاستفادة من المساهمات السابقة.

6-2- الآفاق المستقبلية

تتنوع الخطط المستقبلية لهذا العمل في التوجهات التالية:

- **توسعة مجموعة البيانات الموسومة:** زيادة أعداد العينات وإضافتها إلى مجموعة البيانات الحالية مع التقليل المستمر من نسبة عدم التوازن بين الصفوف، من خلال استخدام إطار العمل التكيفي.
- **أهمية خصائص التغريدات:** مع تأمين تغريدات وبيانات أكثر، يمكن الاقتراب بشكل أفضل من توزع العينات وفق هذه الخصائص والاعتماد على تمثيل العينات من خلال فضاء جديد، مما يؤكد على أهمية دراسة خصائص التغريدات. كما يمكن التوسع في الأمر، من خلال دراسة خصائص الحسابات التي تنشر تغريدات الكراهية.
- **انتخاب العينات التي تمثل اللهجة المشرقية:** جرى الاعتماد على خاصية الموقع الجغرافي للتغريدات لاختيار العينات المكتوبة باللهجة المشرقية –أثناء مرحلة التحصيل–، مما أدى إلى احتواء مجموعة البيانات على تغريدات مكتوبة بلهجات متعددة كالخليجية والعراقية والمصرية. بالتالي، من الضروري إيجاد آلية ترشيح لتحديد لهجة التغريدة أو النص.
- **استخدام المقاربات المختلفة في انتخاب العينات:** يشمل ذلك انتخاب العينات من خلال تحديد درجة عدم اليقين بالطرق المعروفة الأخرى، مثل الهامش الأصغري أو أنتروبية التصويت وغيرها. كذلك يشمل انتخاب العينات من خلال تحديد العينات التي تمثل التوزع الطبيعي لفضاء العينات بشكل أفضل.

المراجع

[1] A.-M. Founta, D. Chatzakou, N. Kourtellis, J. Blackburn, A. Vakali and I. Leontiadis, "A unified deep learning architecture for abuse detection," in *WebSci 2019 – Proceedings of the 11th ACM Conference on Web Science*, 2019.

[2] C. Themeli, G. Giannakopoulos and N. Pittaras, "A study of text representations for Hate Speech Detection," in *In Proceedings of the 20th International Conference on Computational Linguistics and Intelligent Text Processing,*, La Rochelle, France, 2019.

[3] "INTERNET WORLD USERS BY LANGUAGE," 02 10 2022. [Online]. Available: https://www.internetworldstats.com/stats7.htm.

[4] M. Alrefai, H. Faris and I. Aljarah, "Sentiment analysis for Arabic language: A brief survey of approaches and techniques," *International Journal of Advanced Science and Technology,* vol. 119(1), pp. 13-24, 2018.

[5] G. Badaro, R. Baly, H. M. Hajj, W. El-Hajj, K. B. Shaban, N. Habash, A. Al-Sallab and A. Hamdi, "A survey of opinion mining in Arabic: A comprehensive system perspective covering challenges and advances in tools, resources, models, applications, and visualizations," 2019.

[6] N. Albadi, M. Kurdi and S. Mishra, "Are they our brothers? analysis and detection of religious hate speech in the Arabic twitter sphere," in *n 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2018.

[7] A. G. Chowdhury, A. Didolkar, R. Sawhney and R. R. Shah, "ARHNet – leveraging community interaction for detection of religious hate speech in Arabic," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, Florence, Italy, 2019.

[8] H. Faris, I. Aljarah, M. Habib and P. A. Castillo, "Hate Speech Detection using Word Embedding and Deep Learning in the Arabic

Language Context," in *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2020)*, 2020.

[9] A. Abuzayed and T. Elsayed, "Quick and Simple Approach for Detecting Hate Speech in Arabic Tweets," in *The 4thWorkshop on Open-Source Arabic Corpora and Processing Tools with a Shared Task on Offensive Language Detection*, 2020.

[10] N. Ousidhoum, Z. Lin, H. Zhang, Y. Song and D.-Y. Yeung, "Multilingual and Multi-Aspect Hate Speech Analysis," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Hong Kong, China, 2019.

[11] I. Aljarah, M. Habib, N. Hijazi, H. Faris, R. Qaddoura, B. Hammo, M. Abushariah and M. Alfawareh, "Intelligent detection of hate speech in Arabic social network: A machine learning approach," *Journal of Information Science,* vol. 47, no. 4, pp. 483-501, 2020.

[12] A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in Arabic tweets using deep learning," *Multimedia Systems,* 2021.

[13] R. Duwairi, A. Hayajneh and M. Quwaider, "A Deep Learning Framework for Automatic Detection of Hate Speech Embedded in Arabic Tweets," *Arabian Journal for Science and Engineering,* vol. 46, p. 4001–4014, 2021.

[14] F. Y. Al-Anezi, "Arabic Hate Speech Detection Using Deep Recurrent Neural Networks," *Applied Sciences,* vol. 12, 2022.

[15] M. A. B. Nessir, M. Rhouma, H. Haddad and C. Fourati, " iCompass at Arabic Hate Speech 2022 Detect Hate Speech Using QRNN and Transformers," in *Proceedings of the OSACT 2022 Workshop*, 2022.

[16] N. Elkaref and M. Abu-Elkheir, "GUCT at Arabic Hate Speech 2022 Towards a Better Isotropy for Hate Speech Detection," in *Proceedings of the OSACT 2022 Workshop,* 2022.

[17] R. Alshalan and H. Al-Khalifa, "A Deep Learning Approach for Automatic Hate Speech Detection in the Saudi Twittersphere," *Applied Sciences,* 2020.

[18] I. Abu-Farha and W. Magdy, "Multitask Learning for Arabic Offensive Language and Hate-Speech Detection," in *Proceedings of*

the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, 2020.

[19] M. J. Althobaiti, "BERT-based Approach to Arabic Hate Speech and Offensive Language Detection in Twitter Exploiting Emojis and Sentiment Analysis," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 13, no. 5, pp. 972-980, 2022.

[20] H. Mubarak, H. Al-Khalifa and A. Al-Thubaity, "Overview of OSACT5 Shared Task on Arabic Offensive Language and Hate Speech Detection," in *Proceedings of the OSACT 2022 Workshop*, 2022.

[21] K. H. Makram, K. G. Nessim, M. E. Abd-Almalak, S. Z. Roshdy, S. H. Salem, F. F. Thabet and E. H. Mohamed, "CHILLAX - at Arabic Hate Speech 2022 A Hybrid Machine Learning and Transformers based Model to Detect Arabic Offensive and Hate Speech," in *Proceedings of the OSACT 2022 Workshop,* 2022.

[22] A. Mostafa, O. Mohamed and A. Ashraf, "GOF at Arabic Hate Speech 2022 Breaking The Loss Function Convention For Data-Imbalanced Arabic Offensive Text Detection," in *Proceedings of the OSACT 2022 Workshop*, 2022.

[23] S. Hassan, H. Mubarak, A. Abdelali and K. Darwish, "ASAD: Arabic Social media Analytics and unDerstanding," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, 2021.

[24] A. Shapiro, A. Khalafallah and M. Torki, "AlexU-AIC at Arabic Hate Speech 2022: Contrast to Classify," in *Arabic Hate Speech 2022 Shared Task Workshop (OSACT5 2022)*, 2022.

[25] W. Aldjanabi, A. Dahou, M. A. A. Al-qaness, M. A. Elaziz, A. M. Helmi and R. Damaševicius, "Arabic Offensive and Hate Speech Detection Using a Cross-Corpora Multi-Task Learning Model," *Informatics,* vol. 8, no. 69, 2021.

[26] A. F. M. d. Paula, P. Rosso, I. Bensalem and W. Zaghouani, "UPV at the Arabic Hate Speech 2022 Shared Task: Offensive Language and Hate Speech Detection using Transformers and Ensemble Models," in *Proceedings of the OSACT 2022 Workshop*, Marseille, 2022.

[27] A. Omar, T. M. Mahmoud and T. A. El-hafeez, "Comparative Performance of Machine Learning and Deep Learning Algorithms for Arabic Hate Speech Detection in OSNs," 2020.

[28] H. Haddad, H. Mulki and A. Oueslati, "T-HSAB: A Tunisian hate speech and abusive dataset," *Springer International Publishing,* 2019.

[29] H. Mulki, H. Haddad, C. B. Ali and H. Alshabani, "L-HSAB: A Levantine Twitter Dataset for Hate Speech and Abusive Language," in *Proceedings of the Third Workshop on Abusive Language Online*, Florence, Italy, 2019.

[30] J. Rosa and Y. Bonilla, "Deprovincializing Trump, decolonizing diversity, and unsettling anthropology," *American Ethnologist,* vol. 44, no. 2, pp. 201-208, 2017.

[31] A. Travis, "Anti-Muslim hate crime surges after Manchester and London Bridge attacks," 2017.

[32] S. MacAvaney, "Hate speech detection: Challenges and solutions," *PloS one,,* vol. 14, no. 8, 2019.

[33] Z. Zhang, D. Robinson and J. Tepper, "Detecting hate speech on twitter using a convolution-gru based deep neural network," 2018.

[34] H. Watanabe, M. Bouazizi and T. Ohtsuki, "Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection," vol. 6, p. 13825–13835, 2018.

[35] T. Davidson, D. Warmsley, M. Macy and I. Weber, "Automated hate speech detection and the problem of offensive language," in *in Proceedings of the 11th International AAAI Conference on Web and Social Media*, 2017.

[36] E. A. Abozinadah, A. V. Mbaziira and J. Jones, "Detecting abusive arabic language twitter accounts using a multidimensional analysis model," *Int. J. Knowl. Eng,* p. 113–119, 2017.

[37] W. Magdy, K. Darwish and I. Weber, "“# failedrevolutions: Using twitter to study the antecedents of isis support," *First Monday,* vol. 21, 2016.

[38] L. Kaati, E. Omer, N. Prucha and A. Shrestha, "Detecting multipliers of jihadism on twitter," in *in Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*, 2015.

[39] F. Alkomah and X. Ma, "A Literature Review of Textual Hate Speech Detection Methods and Datasets," *Information*, vol. 13, no. 273, 2022.

[40] H. Mubarak, K. Darwish and W. Magdy, "Abusive language detection on Arabic social media," in *Proceedings of the First Workshop on Abusive Language Online*, Stroudsburg, 2017.

[41] T. D. Smedt, G. D. Pauw and P. V. Ostaeyen, "Automatic detection of online jihadist hate speech," 2018.

[42] B. Haidar, M. Chamoun and A. Serhrouchni, "A multilingual system for cyberbullying detection: Arabic content detection using machine learning," *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 6, p. 275–284, 2017.

[43] M. A. Al-Ajlan and M. Ykhlef, "Optimized twitter cyberbullying detection based on deep learning," in *In Proceedings of the 21st Saudi Computer Society National Computer Conference (NCC)*, 2018.

[44] H. Mohaouchane, a. Mourhir and N. S. Nikolov, "Detecting offensive language on Arabic social media using deep learning," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security*, 2019.

[45] H. Mubarak, K. Darwish, W. Magdy, T. Elsayed and H. Al-Khalifa4, "Overview of OSACT4 Arabic Offensive Language Detection Shared Task," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2020.

[46] H. Mubarak and K. Darwish, "Arabic offensive language classification on twitter," in *International Conference on Social Informatics*, 2019.

[47] A. Alakrot, L. Murray and N. S. Nikolov, "Towards accurate detection of offensive language in online communication in Arabic," in *Procedia computer science*, 2018.

[48] F. Husain and O. Uzuner, "Leveraging offensive language for sarcasm and sentiment detection in Arabic," in *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, Kyiv, Ukraine, 2021.

[49] B. Haddad, Z. Orabe, A. Al-Abood and N. Ghneim, "Arabic Offensive Language Detection with Attention-based Deep Neural

Networks," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools*, 2020.

[50] S. Alzu'bi, T. C. Ferreira, L. Pavanelli and M. Al-Badrashiny, "aiXplain at Arabic Hate Speech 2022 An Ensemble Based Approach to Detecting Offensive Tweets," in *Proceedings of the OSACT 2022 Workshop*, 2022.

[51] M. Zampieri, P. Nakov, S. Rosenthal, P. Atanasova, G. Karadzhov, H. Mubarak, L. Derczynski, Z. Pitenis and Ç. Çöltekin, "SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020)," in *In Proceedings of the Fourteenth Workshop on Semantic Evaluation*, Barcelona (online), 2020.

[52] H. Mubarak, S. Hassan and S. A. Chowdhury, "Emojis as Anchors to Detect Arabic Offensive Language and Hate Speech," *CoRR*, vol. abs/2201.06723, 2022.

[53] P. Fortuna and S. Nunes, "A Survey on Automatic Detection of Hate Speech in Text," in *ACM Computing Surveys (CSUR)*, 2018.

[54] T. X. Moy, M. Raheem and R. Logeswaran, "Hate Speech Detection in English and Non-English Languages A Review of Techniques and Challenges," *Webology,* vol. 18, 2021.

[55] A. Arango, J. Pérez and B. Poblete, "Hate speech detection is not as easy as you may think: A closer look at model validation (extended version).," in *Proceedings of the 42nd International Acm Sigir Conference on Research and Development in Information Retrieval*, 2020.

[56] T. Gröndahl, L. Pajola, M. Juuti, M. Conti and N. Asokan, "All You Need is "Love": Evading Hate Speech Detection," in *Proceedings of the ACM Conference on Computer and Communications Security,*, 2018.

[57] J. H. Park, J. Shin and P. Fung, "Reducing gender bias in abusive language detection," p. 2799–2804, 2018.

[58] M. Sap, D. Card, S. Gabriel, Y. Choi and N. A. Smith, "The Risk of Racial Bias in Hate Speech Detection," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019.

[59] A. Schmidt and M. Wiegand, "A Survey on Hate Speech Detection using Natural Language Processing," in *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, Valencia, Spain, 2017.

[60] B. Miller, F. Linder and W. Mebane, "Active Learning Approaches for Labeling Text," *Political Analysis*, vol. 28, pp. 1-20, 2018.

[61] D. D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," *Machine learning proceedings*, pp. 148-156, 1994.

[62] S. Tong and D. Koller, "Support Vector Machine Active Learning with Applications to Text Classification," *Journal of Machine Learning Research*, pp. 45-66, 2001.

[63] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, vol. 1, New York: Springer, 2001.

[64] "Active and adaptive ensemble learning for online activity recognition from data streams," *Knowledge-Based Systems*, vol. 138, pp. 69-78, 2017.

[65] B. Settles, Active Learning, Morgan & Claypool, 2012.

[66] S. Dasgupta, A. T. Kalai and C. Monteleoni, "Analysis of perceptron-based active learning," *Journal of Machine Learning Research*, p. 249–263, 2009.

[67] C. C. Aggarwal, X. Kong, Q. Gu, J. Han and P. S. Yu, "Active Learning: A Survey," in *Data Classification: Algorithms and Applications*, Chapman & Hall/CRC, 2014, pp. 571-606.

[68] G. Schohn and D. Cohn, "Less is more: Active learning with support vector machines," in *Proceedings of the Seventeenth International Conference on Machine Learning*, 2000.

[69] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen and X. Wang, "A Survey of Deep Active Learning," *ACM Computing Surveys*, vol. 54, no. 9, pp. 1-40, 2021.

[70] S. Dasgupta, "Analysis of a greedy active learning strategy," *Advances in neural information processing systems*, p. 337–344, 2005.

[71] Y. Gal, R. Islam and Z. Ghahramani, "Deep Bayesian Active Learning with Image Data," *CoRR*, 2017.

[72] S. Hanneke, Theoretical foundations of active learning, Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2009.

[73] J. Kranjc, J. Smailović, V. Podpečan, M. Grčar, M. Žnidaršič and N. Lavrač, "Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the ClowdFlows platform," *Information Processing & Management,* vol. 51, no. 2, pp. 187-203, 2015.

[74] L. Feng, Y. Wang and W. Zuo, "Quick online spam classification method based on active and incremental learning," *Journal of Intelligent & Fuzzy Systems,* 2016.

[75] D. Angluin, "Queries and Concept Learning," *Machine Learning,* p. 319–342, 1988.

[76] L. Atlas, D. Cohn, R. Ladner, M. A. El-Sharkawi, R. J. Marks, M. E. Aggoune and D. C. Park, "Training connectionist networks with queries and selective sampling," in *Proceedings of the 2nd International Conference on Neural Information Processing Systems*, 1989.

[77] D. D. Lewis and W. A. Gale, "A Sequential Algorithm for Training Text Classifiers," in *Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, 1994.

[78] M.-F. Balcan, A. Blum and K. Yang, "Co-training and expansion: towards bridging theory and practice," in *Proceedings of the 17th International Conference on Neural Information Processing Systems*, 2004.

[79] S.-J. Huang, R. Jin and Z.-H. Zhou, "Active learning by querying informative and representative examples," *Advances in neural information processing systems,* pp. 892-900, 2010.

[80] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review,* pp. 44-53, 2018.

[81] T. Scheffer, C. Decomain and S. Wrobel, "Active Hidden Markov Models for Information Extraction," in *Conference on Advances in Intelligent Data Analysis (IDA)*, (London, UK, UK, 2001.

[82] Y. Wu, I. Kozintsev, J.-Y. Bouguet and C. Dulong, "Sampling strategies for active learning in personal photo retrieval," in *IEEE Int. Conference on Multimedia and Expo (2006)*, 2006.

[83] B. Settles and M. Craven, "An Analysis of Active Learning Strategies for Sequence Labeling Tasks," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Honolulu, 2008.

[84] A. CULOTTA and A. MCCALLUMZY, "Reducing labeling effort for structured prediction tasks," in *Conference on Artificial Intelligence (AAAI)*, 2005.

[85] A. M, "Active learning with scarcely labeled instances via bias variance reduction," in *Proceedings of international conference on artificial intelligence and machine learning (ICAIML 2005)*, Cairo, 2005.

[86] N. A. H. MAMITSUKA, "Query learning strategies using boosting and bagging," in *International Conference on Machine Learning (ICML)*, 1998.

[87] A. K. McCallumzy and K. Nigamy, "Employing em and pool-based active learning for text classification," in *In Proceeding of International Conference on Machine Learning (ICML)*, 1998.

[88] C. E. SHANNON, "A Mathematical Theory of Communication," *Bell System Technical Journal*, 1948.

[89] S. A. Engelson and I. Dagan, "Committee-based sampling for training probabilistic classifiers," in *International Conference on Machine Learning (ICML)*, 1995.

[90] A. Joshi, "Multi-class active learning for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[91] K. Wang, D. Zhang, Y. Li, R. Zhang and L. Lin, "Cost-effective active learning for deep image classification," in *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[92] D. Roth and K. Small, "Margin-based active learning for structured output spaces," in *In European Conference on Machine Learning*, 2006.

[93] W. Luo, A. Schwing and R. Urtasun., "Latent structured active learning," *Advances in Neural Information Processing Systems,* pp. 728-736, 2013.

[94] S. Vijayanarasimhan and K. Grauman, "Large-scale live active learning: Training object detectors with crawled data and crowds," *International Journal of Computer Visio,* pp. 97-114, 2014.

[95] J. Zhu, H. Wang, B. Tsou and M. Ma, "Active Learning With Sampling by Uncertainty and Density for Data Annotations," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 18, pp. 1323 - 1331, 09 2010.

[96] N. Escudeiro and A. Jorge, "D-Confidence: an active learning strategy which efficiently identifies small classes," in *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, Los Angeles, California, 2010.

[97] M. Li and I. K. Sethi, "Confidence-Based Active Learning," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* vol. 28, no. 8, pp. 1251-1261, 08 2006.

[98] X. Li and Y. Guo, "Multi-level adaptive active learning for scene classification," in *In European Conference on Computer Vision*, 2014.

[99] C. Campbell, N. Cristianini and A. Smola, "Query Learning with Large Margin Classifiers," *International Conference on Machine Learning (ICML): 2000,* 05 2000.

[100] H. Cheng, R. Zhang, Y. Peng, J. Mao and P.-N. Tan, "Maximum Margin Active Learning for Sequence Labeling with Different Length," in *The 8th IEEE International Conference on Data Mining*, 2008.

[101] M. C. Burl and E. Wang, "Active Learning for Directed Exploration of Complex Systems," in *Proceedings of the 26 th International Conference on Machine Learning*, Montreal, Canada, 2009.

[102] Y. Song, K. Kim, J. Cha and G. Lee, "MMR-based Active Machine Learning for Bio Named Entity Recognition," in *Human language*

technology and the North American association for computational, 2006.

[103] J. Weber and M. Pollack, "Entropy-Driven online active learning for interactive calendar management," in *Proceedings of the 12th International Conference on Intelligent User Interfaces, IUI 2007*, Honolulu, Hawaii, USA, 2007.

[104] A. Holub, P. Perona and M. Burl, "Entropy-based active learning for object recognition," in *IEEE computer society conference on computer vision and pattern recognition workshop anchorage (CVPR 2008),*, 2008.

[105] G. Mann and A. Mccallum, "Efficient Computation of Entropy Gradient for Semi-Supervised Conditional Random Fields," in *Proceedings of the conference of the North American chapter of the association for computational linguistics (NAACL 2007)*, 2007.

[106] Y. Guo and D. Schuurmans, "Discriminative Batch Mode Active Learning," in *Proceedings of the 20th International Conference on Neural Information Processing Systems. NIPS'07*, Vancouver, British Columbia, Canada, 2007.

[107] M. Sharma and M. Bilgic, "Evidence-Based Uncertainty Sampling for Active Learning," *Data Mining and Knowledge Discovery,* vol. 31.1, p. 164–202, 2016.

[108] N. Houlsby, F. Huszar, Z. Ghahramani and M. Lengyel, "Bayesian Active Learning for Classification and Preference Learning," *arXiv preprint arXiv:1112.5745,* 2011.

[109] D. Cohn, L. Atlas and R. Ladner, "Improving Generalization with Active Learning," *Machine Learning,* vol. 15.2, pp. 201-221, 1994.

[110] N. Abe and H. Mamitsuka, "Query learning strategies using boosting and bagging," in *International Conference of Machine Learning*, Madison, USA, 1998.

[111] P. Melville and R. J. Mooney, "Diverse ensembles for active learning," in *International Conference of Machine Learning*, Banff, Canada, 2004.

[112] H. S. Seung, M. Opper and H. Sompolinsky, "Query by committee," in *ACM Workshop on Computational Learning Theory,*, 1992.

[113] J. E. Iglesias, E. Konukoglu, A. Montillo, Z. Tu and A. Criminisi, "Combining generative and discriminative models for semantic segmentation of ct scans via active learning," in *In Biennnial International Conference on Information Processing in Medical Imaging*, 2011.

[114] L. Copa, D. Tuia, M. Volpi and M. Kanevski, "Unbiased query-by-bagging active learning for VHR image classification," in *Proceedings of conference on image and signal processing for remote sensing XVI (ISPRS 2010)*, Toulouse, 2010.

[115] S. Shi, Y. Liu, Y. Huang, S. Zhu and Y. Liu, "Active learning for kNN based on bagging features," in *Fourth International Conference on Natural Computation*, 2008.

[116] Z. Wang, Y. Song and C. Zhang, "Efficient Active Learning with Boosting," in *Proceedings of the SIAM data mining conference (SDM 2009)*, Nevada, 2009.

[117] J. Huang, S. Ertekin, Y. Song, H. Zha and C. L. Giles, "Efficient multiclass boosting classification with active learning," in *The SIAM international conference on data mining (SDM 2007)*, Minnesota, 2007.

[118] Y. Zhao, C. Xu and Y. Cao, "Research on query-by-committee method of active learning and application," in *Lecture notes on artificial intelligence (LNAI 2006)*, 2006.

[119] S. Fine, R. Gilad-Bachrach and E. Shamir, "Query by committee, linear separation and random walks," *Theoretical Computer Science*, vol. 284, p. 25–51, 2002.

[120] Y. Freund, H. S. Seung, E. Shamir and N. Tishby, "Selective sampling using the query by committee algorithm,," *Machine Learning*, vol. 28, pp. 133-168, 1997.

[121] S. Jiang, G. Pang, M. Wu and L. Kuang, "An improved K-nearest-neighbor algorithm for text categorization," *Expert Systems with Applications*, vol. 39, pp. 1503-1509, 2012.

[122] S. Sohn, D. C. Comeau, W. Kim and W. J. Wilbur, "Term-Centric Active Learning for Naïve Bayes Document Classification," *The Open Information Systems Journal*, vol. 3, pp. 54-67, 2009.

[123] B. Settles, M. Craven and S. Ray, "Multiple-Instance Active Learning," *Advances in Neural Information Processing Systems (NIPS),* vol. 20, p. 1289–1296, 2008.

[124] B. Long, O. Chapelle, Y. Zhang, Y. Chang, Z. Zheng and B. Tseng, "Active learning for ranking through expected loss optimization," in *Proceedings of the 33rd international ACM SIGIR conference on Research;development in information retrieval*, 2010.

[125] Y. Zhang, M. Lease and B. C. Wallace, "Active Discriminative Text Representation Learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI'17*, 2017.

[126] A. Vezhnevets, J. M. Buhmann and V. Ferrari, "Active Learning for Semantic Segmentation with Expected Change," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[127] N. Roy and A. McCallum, "Toward Optimal Active Learning through Sampling Estimation of Error Reduction," in *Proceedings of the Eighteenth International Conference on Machine Learning. ICML '01*, 2001.

[128] H. T. Nguyen and A. Smeulders, "Active Learning Using Pre-Clustering," in *Proceedings of the Twenty-First International Conference on Machine Learning. ICML '04*, New York, NY, USA, 2004.

[129] S. Dasgupta and D. Hsu, "Hierarchical Sampling for Active Learning," in *Proceedings of the 25th International Conference on Machine Learning. ICML '08. Helsinki, Finland: Association for Computing Machinery,*, Helsinki, Finland, 2008.

[130] F. Poursabzi-Sangdeh, J. Boyd-Graber, L. Findlater and K. Seppi, "ALTO: Active Learning with Topic Overviews for Speeding Label Induction and Document Labeling," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2016.

[131] O. Sener and S. Savarese, "Active Learning for Convolutional Neural Networks: A Core-Set Approach," in *6th International Conference on Learning Representations, ICLR 2018*, 2018.

[132] A. Prabhu, C. Dognin and M. Singh, "Sampling Bias in Deep Active Classification: An Empirical Study," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*

and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.

[133] D. Gissin and S. Shalev-Shwartz, "Discriminative Active Learning," *arXiv preprint arXiv:1112.5745,* 2019.

[134] H. Hassanzadeh and M. Keyvanpour, "A variance based active learning approach for named entity," *Intelligent computing and information science,* vol. 135, pp. 347-352, 2011.

[135] S. C. H. Hoi, R. Jin, J. Zhu and M. R. Lyu, "Batch mode active learning and its application to medical," in *Proceedings of the 23 rd International Conference on Machine Learning*, Pittsburgh, 2006.

[136] S. Vijayakumar, M. Sugiyama and H. Ogawa, "Training data selection for optimal generalization with noise variance reduction in neural networks," *Neural Nets WIRN Vietri-98,* pp. 153--166, 1999.

[137] X. Zhu, P. Zhang, X. Lin and Y. Shi, "Active Learning from Data Streams," in *Seventh IEEE International Conference on Data Mining*, Nebraska,, 2007.

[138] D. Li, F. Qian and P. Fu, "Variance minimization approach for a class of dual control problems," in *Proceedings of the 2002 American control conference (ACC 2002)*, Alaska, 2002.

[139] P. Donmez, J. G. Carbonell and a. P. N. Bennett, "Dual strategy active learning," in *ECML conference*, 2007.

[140] S. C. H. Hoi, R. Jin, J. Zhu and M. Lyu, "Semi-supervised SVM batch mode active learning for image retrieval," in *IEEE Conference on CVPR*, 2008.

[141] Z. Xu, K. Yu, V. Tresp, X. Xu and J. Wang, "Representative sampling for text classification using support vector machines," in *ECIR Conference*, 2003.

[142] S. Stieglitz, M. Mirbabaie, B. Ross and C. Neuberger, "Social media analytics – challenges in topic discovery, data collection, and data preparation," *International Journal of Information Management,* vol. 39, pp. 156-168, 2018.

[143] W. Warner and J. Hirschberg, "Detecting hate speech on the World Wide Web," in *the Second Workshop on Language in Social Media 2012 Jun 7*, 2012.

[144] S. Shaikh and S. M. Doudpotta, "Aspects Based Opinion Mining for Teacher and Course Evaluation," *Sukkur IBA Journal of Computing and Mathematical Sciences,* vol. 3(1), pp. 34-43, 2019.

[145] E. A. Abozinadah, A. V. Mbaziira and J. H. J. Jr, "Detection of Abusive Accounts with Arabic Tweets," *International Journal of Knowledge Engineering*, vol. 1, no. 2, pp. 113-119, 2015.

[146] W. Lan, Y. Chen, W. Xu and A. Ritter, "GigaBERT: Zero-shot Transfer Learning from English to Arabic," in *Proceedings of the International Conference on Language Resources and Evaluation*, 2020.

[147] S. Qiu, B. Xu, J. Zhang, Y. Wang, X. Shen, G. Melo, C. Long and X. Li, "EasyAug: An Automatic Textual Data Augmentation Platform for Classification Tasks," in *Companion Proceedings of the Web Conference 2020*, 2020.

[148] S. Kobayashi, "Contextual augmentation: Data augmentation by words with paradigmatic relations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, Louisiana, 2018.

[149] C. Shorten, T. M. Khoshgoftaar and B. Furht, "Text Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 8, no. 101, 2021.

[150] A. B. S. Mohammad, K. Eissa and S. R. El-Beltagy, "AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP," in *the 3rd International Conference on Arabic Computational Linguistics (ACLing 2017)*, 2017.

[151] A. Liaw and M. Wiener, "Classification and regression by randomforest," *Forest*, vol. 23, 2002.

[152] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan and B. P. Feuston, "Random Forest, A Classification and Regression Tool for Compound Classification and QSAR Modeling," *Journal of Chemical Information and Computer Sciences*, vol. 43, pp. 1947-1958, 2003.

[153] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Language*, vol. 20, no. 3, p. 273–297, 1995.

[154] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[155] A. V. Dorogush, A. Gulin, G. Gusev, N. Kazeev, L. O. Prokhorenkova and A. Vorobev, "Fighting biases with dynamic boosting," *CoRR*, vol. abs/1706.09516, 2017.

[156] S. W. Menard, Applied logistic regression analysis, ThousandOaks,: Sage university paper series on quantitative application in the social sciences, series no. 106) (2nd ed.), 1995.

[157] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall PTR, 1998.

[158] D. D. Lewis, "Naive (bayes) at forty: The independence assumption in information retrieval," in *European conference on machine learning*, 1998.

[159] G. Mujtaba, L. Shuib, R. G. Raj, R. Rajandram and K. Shaikh, "Prediction of Cause of Death from Forensic Autopsy Reports using 1 Text Classification Techniques: A Comparative Study," *Journal of Forensic and Legal Medicine*, 2017.

[160] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*, 2003.

[161] S. S. M. Rahman, K. Biplob, M. Rahman, K. Sarker and T. Islam, An Investigation and Evaluation of N-Gram, TF-IDF and Ensemble Methods in Sentiment Classification, Springer Nature, 2020, pp. 391-402.

[162] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *In Proceedings of the 1st International Conference on Learning representations*, 2013.

[163] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013.

[164] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational*

Linguistics: Human Language Technologies, Minneapolis, Minnesota, 2019.

[165] S. J. Russell and P. Norvig, Artificial Intelligence - A Modern Approach, Fourth Edition ed., Pearson, 2021.

[166] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016.

[167] M. Ashi, M. A. Siddiqui and F. Nadeem, "Pre-trained Word Embeddings for Arabic Aspect-Based Sentiment Analysis of Airline Tweets," *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018,* 2019.

[168] W. Antoun, F. Baly and H. M. Hajj, "AraBERT: Transformer-based model for arabic language understanding," in *Proceedings of the International Conference on Language Resources and Evaluation*, 2020.

[169] A. A. Elmadany, W. Magdy and H. Mubarak, "ArSAS: An Arabic speech-act and sentiment corpus of tweets," in *Proceedings of the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT 3)*, 2018.

[170] D. Graff, J. Kong, K. Chen and K. Maeda, "English Gigaword Linguistic Data Consortium," 2003.

[171] M. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research,* vol. 16, no. 1, pp. 321-357, 2002.

[172] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010.

[173] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Machine Learning Research,* vol. 15, no. 1, pp. 1929-1958, 2014.

[174] A. I. Schein and L. H. Ungar, "Active learning for logistic regression: an evaluation," *Machine Learning,* vol. 68, p. 235–265, 2007.

[175] R. K. Ando and T. Zhang, "A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data," *Machine Learning Researches,* vol. 6, pp. 1817-1853, 2005.

[176] J. Kremer, K. Steenstrup Pedersen and C. Igel, "Active learning with support vector machines," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 4, no. 4, pp. 313-326, 2014.

[177] M. Girolami, "Mercer kernel-based clustering in feature space," *IEEE Trans. Neural Networks,* vol. 13, no. 3, pp. 780-784, 2002.

[178] J. Shawe-Taylor and N. Cristianini, "Kernel Methods for Pattern Analysis,," *Cambridge University Press*, 2004.

[179] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825--2830, 2011.

ملحق 1

م	العام	الناشر	طبيعة النشر	المرجع
1	2020	Springer	Conference	[27]
2	2019	Springer	Journal	[28]
3	2019	IEEE	Conference	[44]
4	2018	IEEE	Conference	[6]
5	2019	Association for Computational Linguistics	Workshop	[7]
6	2020	Association for Computational Linguistics	Workshop	[45]
7	2019	Association for Computational Linguistics	Workshop	[29]
8	2020	European Language Resources Association	Conference	[18]
9	2019	Springer	Conference	[46]
10	2018	Elsevier	Conference	[47]
11	2020	SCITEPRESS	Conference	[8]
12	2020	European Language Resources Association	Workshop	[9]
13	2020	European Language Resources Association	Conference	[49]
14	2017	Advances in Science, Technology and Engineering Systems Journal	Journal	[42]
15	2015	International Journal of Knowledge Engineering	Journal	[36]
16	2016	AAAI Spring Symposium Series	Journal	[37]
17	2015	IEEE	Conference	[38]
18	2019	Association for Computational Linguistics	Conference	[10]
19	2018	CLiPS Technical Report Series (CTRS)	Report	[41]
20	2018	IEEE	Conference	[43]
21	2022	International Journal of Advanced Computer Science and Applications	Journal	[19]
22	2022	European Language Resources Association	Workshop	[20]
23	2022	European Language Resources Association	Workshop	[15]
24	2022	European Language Resources Association	Workshop	[21]
25	2022	European Language Resources Association	Workshop	[22]
26	2022	European Language Resources Association	Workshop	[50]
27	2022	European Language Resources Association	Workshop	[16]
28	2020	SAGE	Journal	[11]
29	2021	Springer	Journal	[12]
30	2021	Association for Computational Linguistics	Conference	[23]
31	2021	Virtual	Workshop	[48]
32	2021	Springer	Journal	[13]

[40]	Workshop	Association for Computational Linguistics	2017	33
[24]	Workshop	European Language Resources Association	2022	34
[14]	Journal	MDPI	2022	35
[25]	Journal	MDPI	2022	36
[26]	Workshop	European Language Resources Association	2022	37
[51]	Conference	Association for Computational Linguistics	2020	38
[17]	Journal	MDPI	2020	39
[52]	Journal	CoRR	2022	40

ملحق 2

FD[8]	FM[7]	Rep[6]	DL[5]	ML[4]	Hate%	Size	Class	Type	Dialect[3]	OSN[2]	Task[1]	Ref	#
98.7	97.6		2	12	50	20,000	2	controversial	E	FTYI	HA	[27]	1
	83.618	G		2	18	6,039	3	Seed words	T	FTYI	HA	[28]	2
		E	4			16,000	3			Y	O	[44]	3
77	72	GE	1	2	50	6,600	2	Religious		T	H	[6]	4
78		E	12		50	6,600	2	Religious		T	H	[7]	5
94.4		GEB	5	3	4.5	10,000	2	Vocative particle		T	O	[45]	6
	74.4	G		2	8	5,846	3	Political	L	T	HA	[29]	7
73.7	39	EGB	2	1	5	7,000	2			T	HO	[18]	8
90	64	G	1	1		10,000	2	Seed words		T	O	[46]	9
	82	G		1	38.7	15,050	2	Celebrities		Y	O	[47]	10
71.7		E	1		52	3,696	2	Keywords		T	H	[8]	11
69		EG	8	7	4.5	10,000	2	Vocative particle		T	H	[9]	12

[1] Task of paper: H: Hate, A: Abusive, O: Offensive, J: Jihadi or Terrorism, C: Cyberbullying.
[2] Dataset's source: F: Facebook, Y: YouTube, I: Instagram, T: Twitter.
[3] Dataset's Dialect or Language: A: Arabic, F: French, En: English, E: Egyptian, S: Saudi, T: Tunisian, D: Danish, G: Greek, Tr: Turkey, L: Levantine.
[4] ML: Number of traditional machine learning models.
[5] DL: Number of deep learning models.
[6] Text Representation: E: Word Embedding, G: n-gram or BOW, B: BERT.
[7] FM: F1-score for machine learning models.
[8] DM: F1-score for deep learning models.

75		E	4		53	14,125	2	Vocative particle		T	O	[49]	13
	68.7			2	6	35,273	2	ME		FT	C	[42]	14
	96			3			2	Keywords		T	A	[36]	15
	87	G		1		10,000	3	Keywords		T	J	[37]	16
	68.6	G		1	17.2	16,000	2	HA		T	J	[38]	17
		GE	1	1	22.5	13,000	6	Seed Words	En-F-A	T	H	[10]	18
	82	G		1	50	90,000	2	Keywords		T	J	[41]	19
		E	1			20,000	2			T	C	[43]	20
84.3	81.9	GB	1	2	10.5	12,698	2	Emojis		T	HO	[19]	21
85.2		B	10		10.5	12,698	2			T	HO	[20]	22
83.1		B	4		10.5	12,698	6			T	H	[15]	23
	76	B		2	10.5	12,698	2			T	HO	[21]	24
85.2		B	8		10.5	12,698	2			T	HO	[22]	25
84.9		GEB			10.5	12.698	2			T	O	[50]	26
74.5		B	1		10.5	12,698	2			T	H	[16]	27
91.2		G		4	22.8	3,696	3			T	H	[11]	28
73	65	GE	4	1	29	11,000	5			T	H	[12]	29
75.1	75.2	GB	1	1	4.5	6,900	2	Seed words		T	HAO	[23]	30
69.2		B	2			15,548	3			T	O	[48]	31
81		E	4			23,678	5			T	H	[13]	32
	60	G				32,000	2		E	T	A	[40]	33
81.7		B	1		10.8	8,800	2			T	HO	[24]	34
	65	EB	1	4		4,203	7			FYTI	H	[14]	35
88.7		B	7		4.5	10,000	3			T	HO	[25]	36
79.5		B	6		10.5	12.698	2			T	HO	[26]	37

90.2		B			19.9	10,000	2	vocative particles	A,En,D,G,Tr	T	A	[51]	38
79		B	7			9,316	2	Keywords	S	T	H	[17]	39
91.57		B			35	12,698	2	Emojis		FTY	O	[52]	40

ملحق 3

#	العام	المرجع	التقنية المستخدمة	ملاحظات	البيانات
1	2010	[79]	Hybrid		Image
2	2008	[83]	uncertainty	Entropy: posterior probabilities	Text
3	1998	[87]	QBC	Expectation-Maximization	Text
4	1995	[89]	QBC		Text
5	1998	[110]	QBC	bagging and boosting	Image
6	2004	[111]	QBC	Committee diversity	Image
7	2009	[90]	uncertainty	margin-based uncertainty	Image
8	2017	[91]	uncertainty	Entropy: posterior probabilities	Image
9	2006	[92]	uncertainty	Margin between posterior probabilities	Text
10	2013	[93]	uncertainty	Entropy: posterior probabilities	Image
11	2014	[94]	Uncertainty;	Distances to the decision boundaries	Image
12	2014	[98]	uncertainty	Distances to the decision boundaries	Image
13	1992	[112]	QBC	Measures disagreement	
14	2011	[113]	QBC		Image
15	2007	[139]	Hybrid		Image
16	2008	[140]	Hybrid	Margin-based	Image
17	2003	[141]	Hybrid		Text
18	2010	[95]	Uncertainty and density	To solve the outlier problem	Text
19	2010	[96]	Uncertainty	imbalanced class distribution	Text

Image	to accurately estimate sample conditional error	Uncertainty	[97]	2006	20
Image	is most efficient when the dataset can be learnt using few support vectors	Uncertainty	[99]	2000	21
Text	SVM: Sequence labeling	Uncertainty	[100]	2008	22
	variances of Entropy metrics	Uncertainty	[101]	2009	23
Text	N-best sequence entropy metric	Uncertainty	[102]	2006	24
Image		Uncertainty	[103]	2007	25
Image	Improving the efficiency of web image-search queries and open-world visual learning by an autonomous agent.	Uncertainty	[104]	2008	26
	For calculating the entropy gradient	Uncertainty	[105]	2007	27
Text	web search	EGL	[121]	2012	28
Text	recommendation, text classification	EGL	[122]	2009	29
Image	multiple-instance active learning	EGL	[123]	2008	30
Text	Expected Loss Optimization metric; Web search	EGL	[124]	2010	31
Text	sequence labeling tasks	VR	[83]	2008	32
Text	Conditional Random Field	VR	[134]	2011	33
Image	Largest reduction in the Fisher information.	VR	[135]	2006	34
	Improve generalization ability and reduce noise variance	VR	[136]	1999	35
	Minimal Variance (MV) principle	VR	[137]	2007	36
	variance minimization approach	VR	[138]	2002	37
Text	Entropy Query by Bagging	QBC	[114]	2010	38
Image	Bagging features active learning (ALBF) for kNN	QBC	[115]	2008	39
Image	query by incremental committee	QBC	[116]	2009	40

Image	a novel multiclass boosting algorithm, Gentle Adaptive Multiclass Boosting Learning (GAMBLE)	QBC	[117]	2007	41
Image	Combining Vote Entropy with Kullback-Leibler divergence	QBC	[118]	2006	42
		QBC	[119]	2002	43
Image	a discriminative batch mode active learning approach	Uncertainty	[106]	2007	44
Text	avoid repeatedly labeling samples in the same cluster	Representative	[128]	2004	45
IT		Representative	[129]	2008	46
Text	CLUSTERING	Representative	[130]	2016	47
Image	Core-set selection	Representative	[131]	2018	48
Text	active set selection using the posterior entropy	Representative	[132]	2019	49
Image	MODEL UNCERTAINTY	Uncertainty	[107]	2016	50
Text		EGL	[125]	2017	51
Image	maximizing an upper-bound on accuracy gain	EGL	[126]	2012	52
Text		uncertainty	[77]	1994	53
Text	version space	MARGIN-BASED	[62]	2001	54
Text	closest to hyperplane	Uncertainty	[68]	2000	55
Image	expresses information gain in terms of predictive entropies	ENTROPY	[108]	2011	56
Image	PREDICTION-BASED	Representative	[133]	2019	57
Text	expected error reduction	EGL	[127]	2001	58
		uncertainty-based	[109]	1994	59
					60

ملحق نتائج الاختبارات بعد التدريب على التعزيز اليدوي والآلي

التعزيز اليدوي

تمت دراسة تأثير التعزيز اليدوي من خلال تدريب المصنفات على مجموعة البيانات LHS-TRAIN-C والتي أضفنا إليها بعض العينات بتقنيات التعزيز اليدوي. جرى اختبار النموذج على مجموعات البيانات الأربعة.

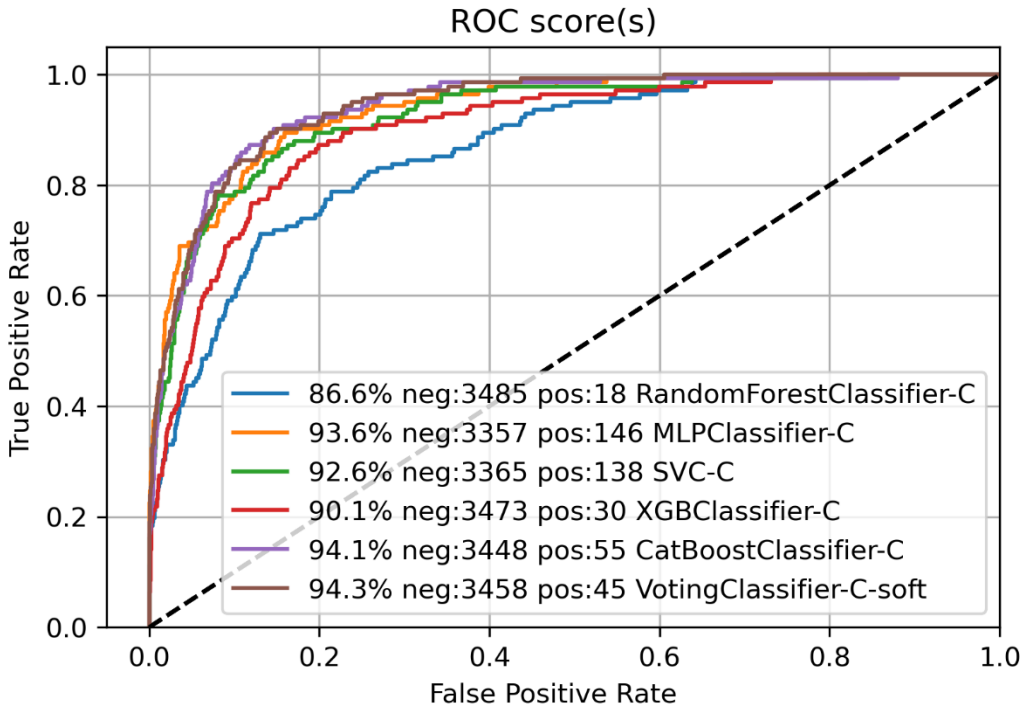
الاختبارات على مجموعة البيانات المحلية

جرى اختبار النموذج على مجموعة البيانات المحلية LHS-TEST، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

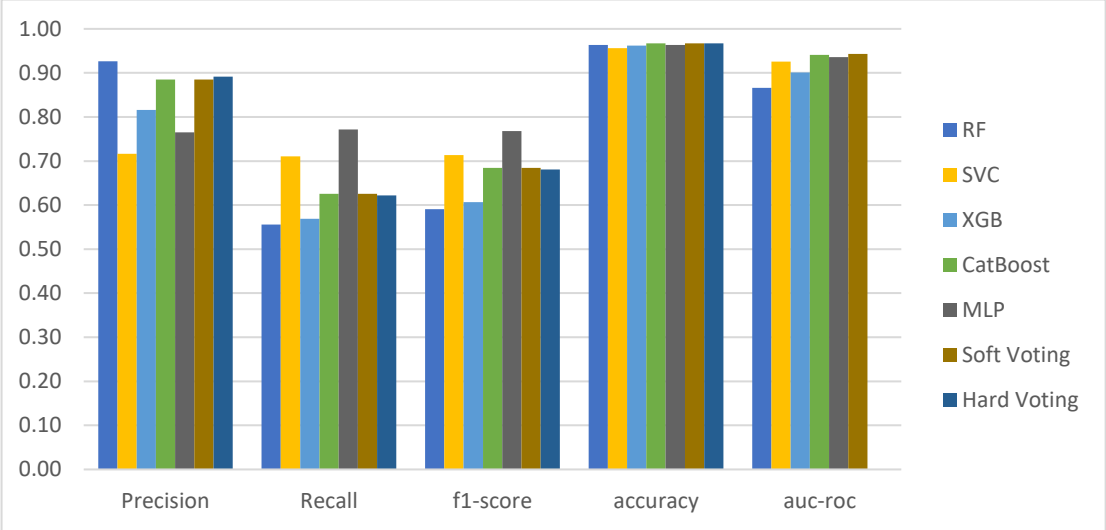
ROC	Accuracy	F1	Recall	Precision	Classifier
0.866	0.96	0.59	0.56	**0.93**	RF
0.926	0.96	0.71	0.71	0.72	SVC
0.901	0.96	0.61	0.57	0.82	XGB
0.941	**0.97**	0.68	0.63	0.88	CatBoost
0.936	0.96	**0.77**	**0.77**	0.76	MLP
0.943	**0.97**	0.68	0.63	0.88	**Soft Voting**
	0.97	0.68	0.62	0.89	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات LHS-TEST

كما يبين الشكل التالي نتائج الاختبار على منحنى AUC-ROC. كذلك، يبين الشكل التالي مخططًا بيانيًا لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة يدويًا وفق مجموعة الاختبار LHS-TEST.



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات LHS-TEST وفق منحني ROC



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات LHS-TEST

الاختبارات على مجموعة البيانات اللبنانية

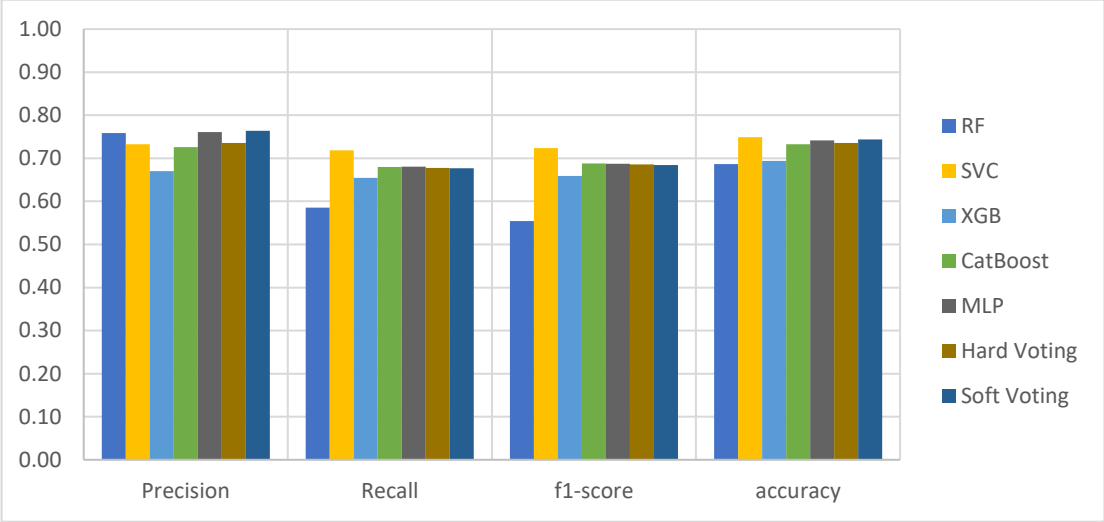
جرى اختبار النموذج على مجموعة البيانات اللبنانية L-HSAB، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

Accuracy	F1	Recall	Precision	Classifier
0.69	0.55	0.59	**0.76**	RF

0.75	**0.72**	**0.72**	0.73	SVC
0.69	0.66	0.65	0.67	XGB
0.73	0.69	0.68	0.73	CatBoost
0.74	0.69	0.68	**0.76**	MLP
0.74	0.69	0.68	0.74	Soft Voting
0.74	0.68	0.68	**0.76**	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات L-HSAB

كما يبين الشكل التالي نتائج الاختبار بيانيًا.



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات L-HSAB

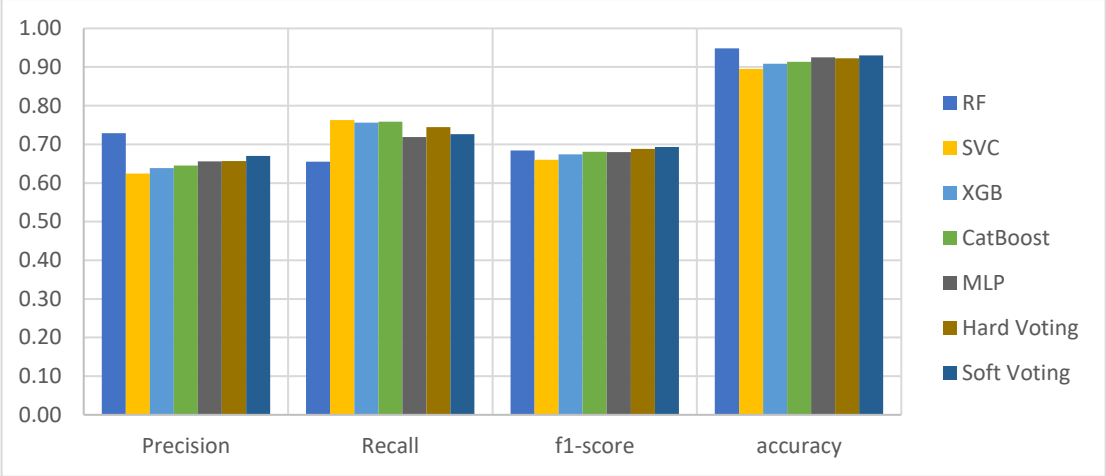
الاختبارات على مجموعة بيانات ورشة عمل المحتوى العربي

جرى اختبار النموذج على مجموعة بيانات ورشة عمل المحتوى العربي OSACT، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

Accuracy	**F1**	**Recall**	**Precision**	**Classifier**
0.95	0.68	0.66	**0.73**	RF
0.90	0.66	**0.76**	0.62	SVC
0.91	0.67	**0.76**	0.64	XGB
0.91	0.68	**0.76**	0.65	CatBoost
0.93	0.68	0.72	0.66	MLP
0.92	**0.69**	0.74	0.66	Soft Voting
0.93	**0.69**	0.73	0.67	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات OSACT

كما يبين الشكل التالي نتائج الاختبار بيانيًا.



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات OSACT

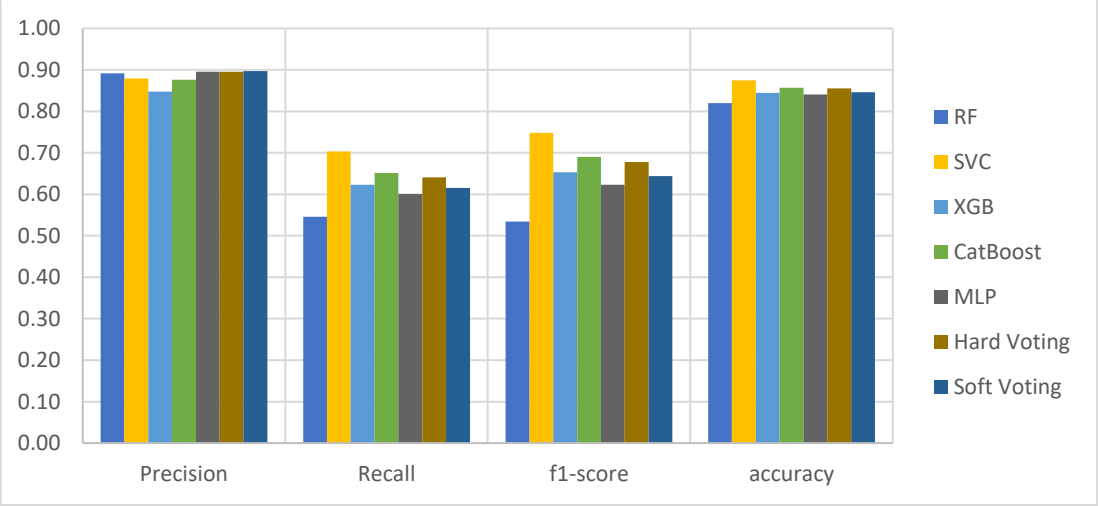
الاختبارات على مجموعة بيانات ورشة عمل التقييم الدلالي

جرى اختبار النموذج على مجموعة بيانات ورشة عمل التقييم الدلالي OffensEval، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

Accuracy	F1	Recall	Precision	Classifier
0.82	0.53	0.55	0.89	RF
0.87	**0.75**	**0.70**	0.88	SVC
0.84	0.65	0.62	0.85	XGB
0.86	0.69	0.65	0.88	CatBoost
0.84	0.62	0.60	**0.90**	MLP
0.86	0.68	0.64	**0.90**	Soft Voting
0.85	0.64	0.62	**0.90**	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات OffensEval

كما يبين الشكل التالي نتائج الاختبار بيانيًا.



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا على مجموعة البيانات OffensEval

التعزيز الآلي

تمت دراسة تأثير التعزيز الآلي من خلال تدريب المصنفات على مجموعة البيانات LHS-TRAIN-D والتي أضفنا إليها بعض العينات بتقنيات التعزيز الآلي. جرى إجراء اختبار النموذج على مجموعات البيانات الأربعة.

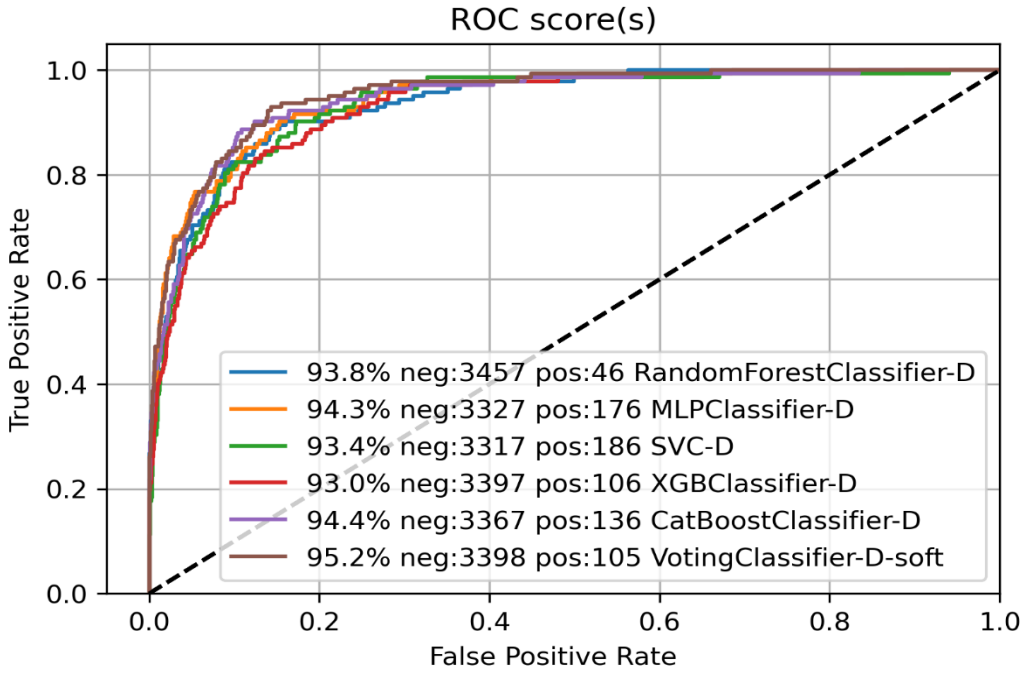
الاختبارات على مجموعة البيانات المحلية

جرى اختبار النموذج على مجموعة البيانات LHS-TEST، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

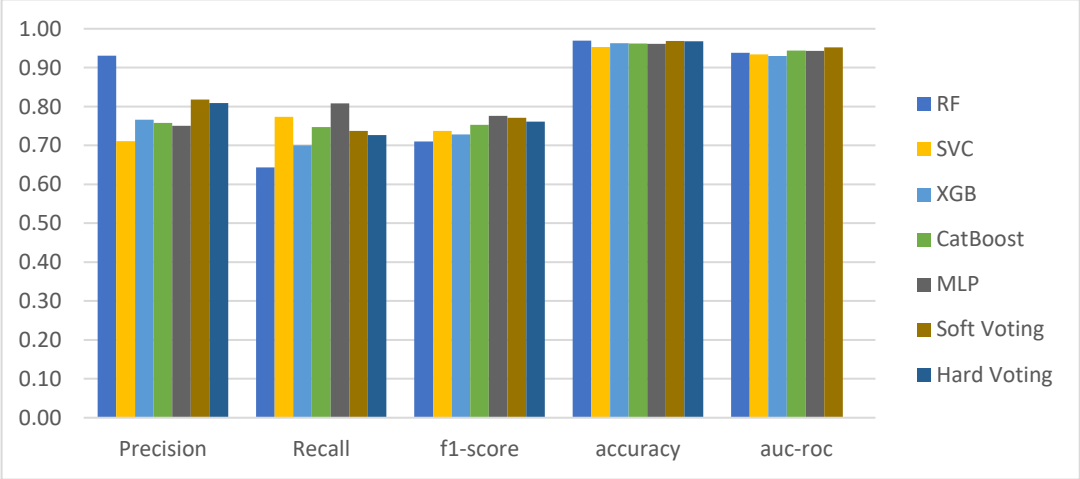
ROC	Accuracy	F1	Recall	Precision	Classifier
0.938	**0.97**	0.71	0.64	**0.93**	RF
0.934	0.95	0.74	0.77	0.71	SVC
0.93	0.96	0.73	0.70	0.77	XGB
0.944	0.96	0.75	0.75	0.76	CatBoost
0.943	0.96	**0.78**	**0.81**	0.75	MLP
0.952	**0.97**	0.77	0.74	0.82	Soft Voting
	0.97	0.76	0.73	0.81	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات LHS-TEST

كما يبين الشكل التالي نتائج الاختبار على منحنى AUC-ROC. كذلك، يبين الشكل التالي مخططًا بيانيًا لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا وفق مجموعة الاختبار LHS-TEST.



مخطط ROC لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات LHS-TEST



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات LHS-TEST

الاختبارات على مجموعة البيانات اللبنانية

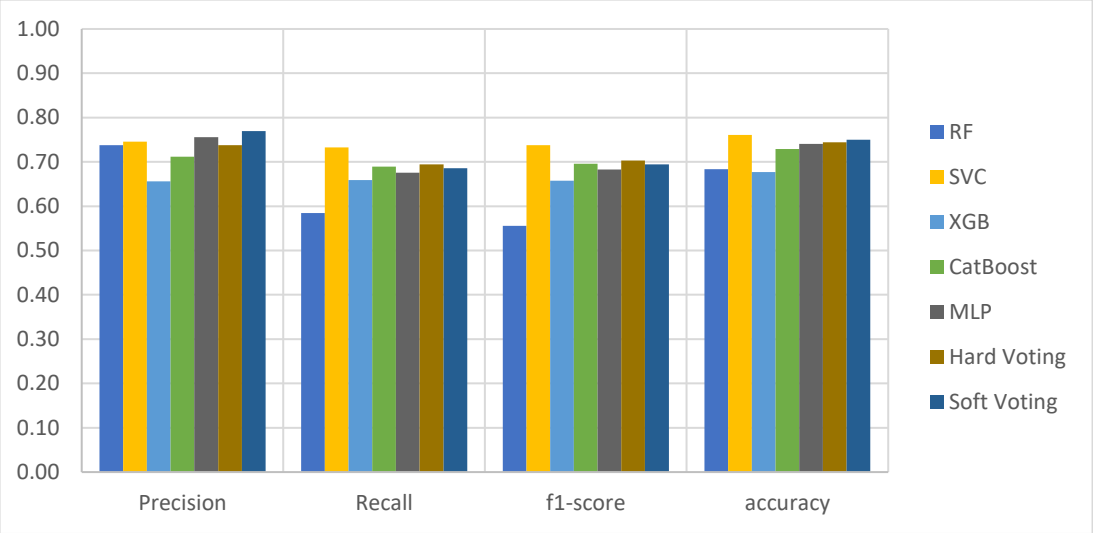
جرى اختبار النموذج على مجموعة البيانات اللبنانية L-HSAB، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

Accuracy	F1	Recall	Precision	Classifier
0.68	0.56	0.58	0.74	RF
0.76	**0.74**	**0.73**	0.75	SVC
0.68	0.66	0.66	0.66	XGB
0.73	0.70	0.69	0.71	CatBoost
0.74	0.68	0.68	0.76	MLP

0.74	0.70	0.69	0.74	Soft Voting
0.75	0.69	0.69	**0.77**	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات L-HSAB

كما يبين الشكل التالي نتائج الاختبار بيانيًا.



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات L-HSAB

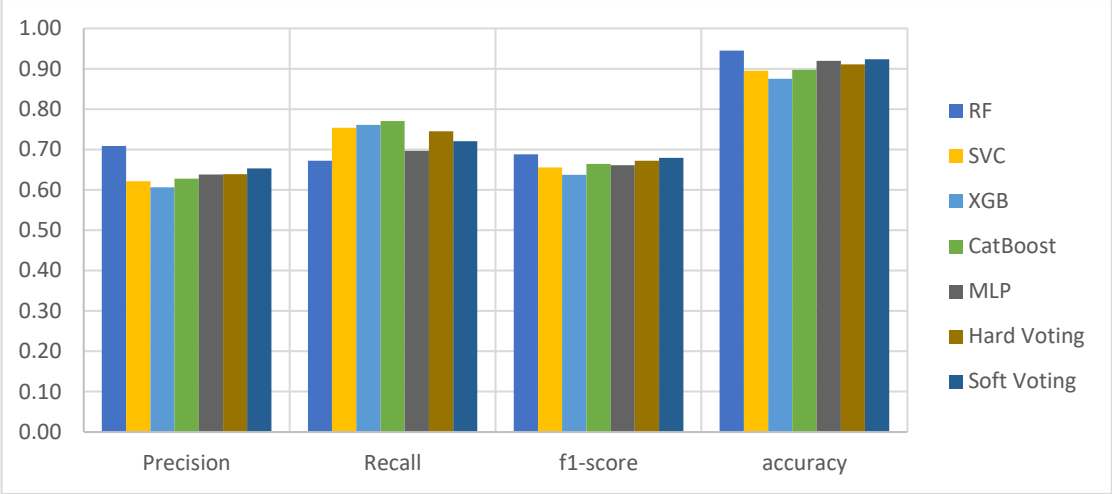
الاختبارات على مجموعة بيانات ورشة عمل المحتوى العربي

جرى اختبار النموذج على مجموعة بيانات ورشة عمل المحتوى العربي OSACT، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

Accuracy	F1	Recall	Precision	Classifier
0.94	**0.69**	0.67	**0.71**	RF
0.90	0.66	0.75	0.62	SVC
0.88	0.64	0.76	0.61	XGB
0.90	0.66	**0.77**	0.63	CatBoost
0.92	0.66	0.70	0.64	MLP
0.91	0.67	0.75	0.64	Soft Voting
0.92	0.68	0.72	0.65	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات OSACT

كما يبين الشكل التالي نتائج الاختبار بيانيًا.



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات OSACT

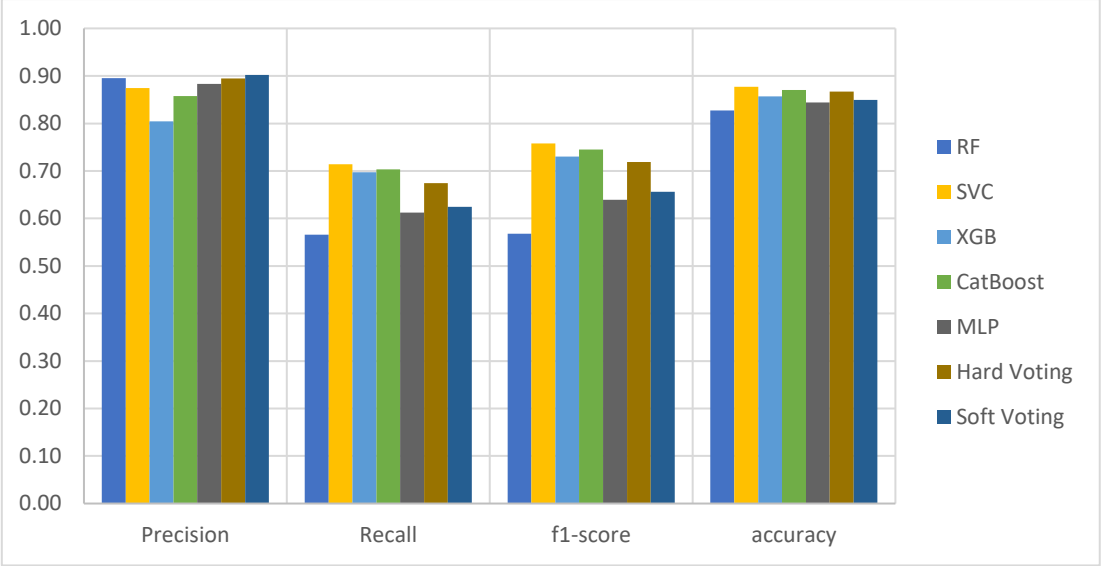
الاختبارات على مجموعة بيانات ورشة عمل التقييم الدلالي

جرى اختبار النموذج على مجموعة بيانات ورشة عمل التقييم الدلالي OffensEval، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

Accuracy	F1	Recall	Precision	Classifier
0.83	0.57	0.57	**0.90**	RF
0.88	**0.76**	**0.71**	0.87	SVC
0.86	0.73	0.70	0.80	XGB
0.87	0.74	0.70	0.86	CatBoost
0.84	0.64	0.61	0.88	MLP
0.87	0.72	0.67	0.89	Soft Voting
0.85	0.66	0.62	**0.90**	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات OffensEval

كما يبين الشكل التالي نتائج الاختبار بيانيًا.



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة آليًا على مجموعة البيانات OffensEval

دراسة تأثير توازن البيانات

سنقوم في الفقرات التالية باختبار تأثير تقنيات تعزيز البيانات مع إضافة تقنية تعزيز البيانات برمجيًا وذلك على المصنفات المستخدمة.

إجراءات إضافية

بالرغم من ارتفاع نسبة عينات الكراهية إلى 21.01، إلا أنه لا يمكن اعتبار أن مجموعة البيانات هذه أصبحت متوازنة بما فيه الكفاية، ولذلك استخدمنا تقنية تعزيز البيانات برمجيًا باستخدام synthetic minority over-sampling technique (SMOTE) التي تحقق توازن مجموعة البيانات من خلال دمج تقنية over-sampling للصف الأقل نسبة وتقنية under-sampling للصف الآخر [171].

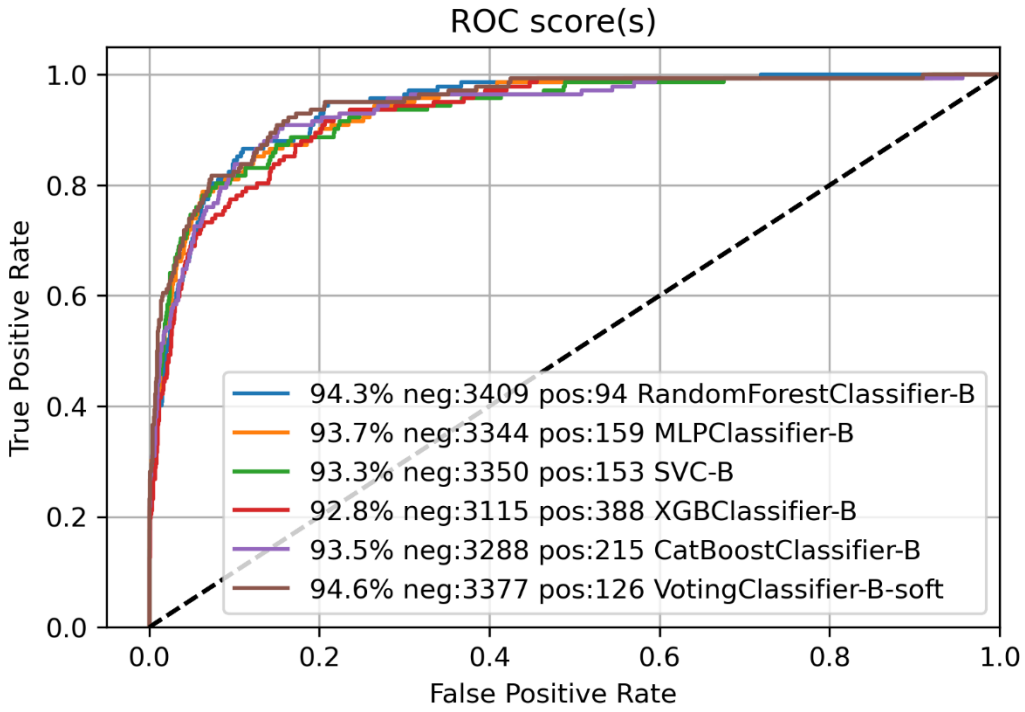
مجموعة البيانات LHS-TRAIN-B

جرى إجراء تدريب المصنفات على مجموعة البيانات LHS-TRAIN-B، ثم اختبار النموذج على مجموعة البيانات LHS-TEST، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

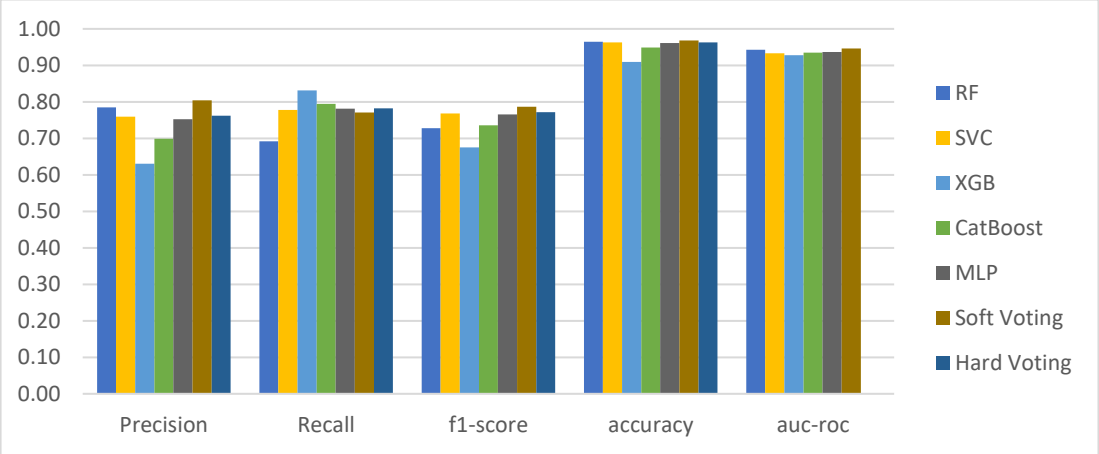
ROC	Accuracy	F1	Recall	Precision	Classifier
0.943	0.96	0.73	0.69	0.79	RF
0.933	0.96	0.77	0.78	0.76	SVC
0.928	0.91	0.68	**0.83**	0.63	XGB
0.935	0.95	0.74	0.79	0.70	CatBoost
0.937	0.96	0.77	0.78	0.75	MLP
0.946	**0.97**	**0.79**	0.77	**0.80**	**Soft Voting**
	0.96	0.77	0.78	0.76	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة البيانات بدون تعزيز مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST

كما يبين الشكل التالي نتائج الاختبار على منحنى AUC-ROC. كذلك، يبين الشكل التالي مخططًا بيانيًا لنتائج اختبار النموذج المدرب على مجموعة البيانات بدون تعزيز مع استخدام تقنية SOMTE وفق مجموعة الاختبار LHS-TEST.



مخطط ROC لنتائج اختبار النموذج المدرب على مجموعة البيانات بدون تعزيز مع استخدام تقنية SMOTE على مجموعة الاختبار



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة البيانات بدون تعزيز مع استخدام تقنية SMOTE على مجموعة الاختبار

مجموعة البيانات LHS-TRAIN-C

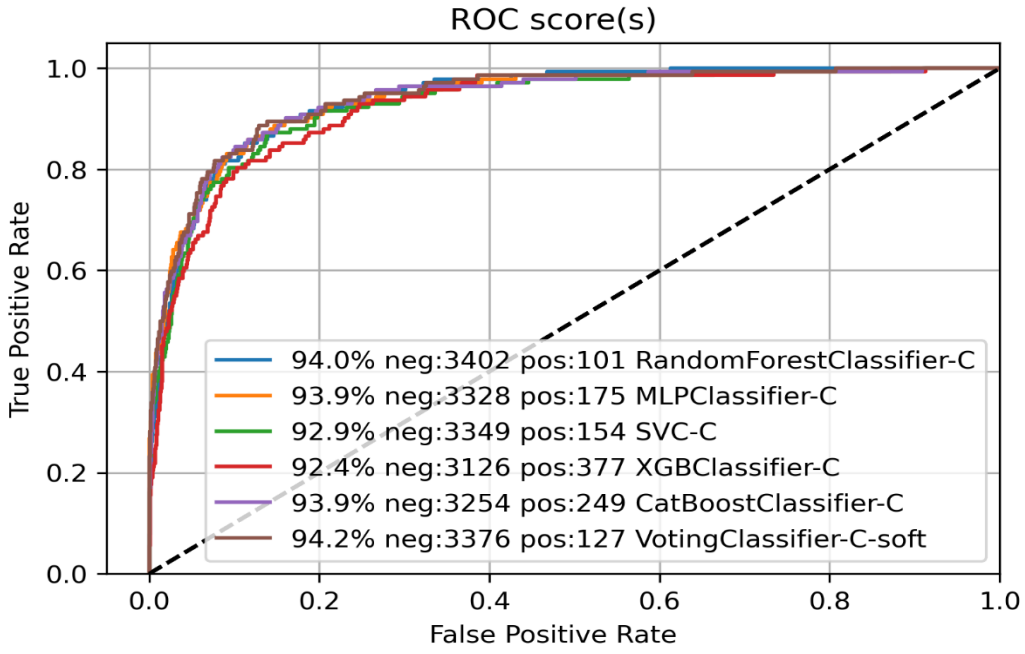
جرى تدريب المصنفات على مجموعة البيانات LHS-TRAIN-C والتي أضفنا إليها بعض العينات بتقنيات التعزيز اليدوي، ثم اختبار النموذج على مجموعة البيانات LHS-TEST، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

ROC	Accuracy	F1	Recall	Precision	Classifier

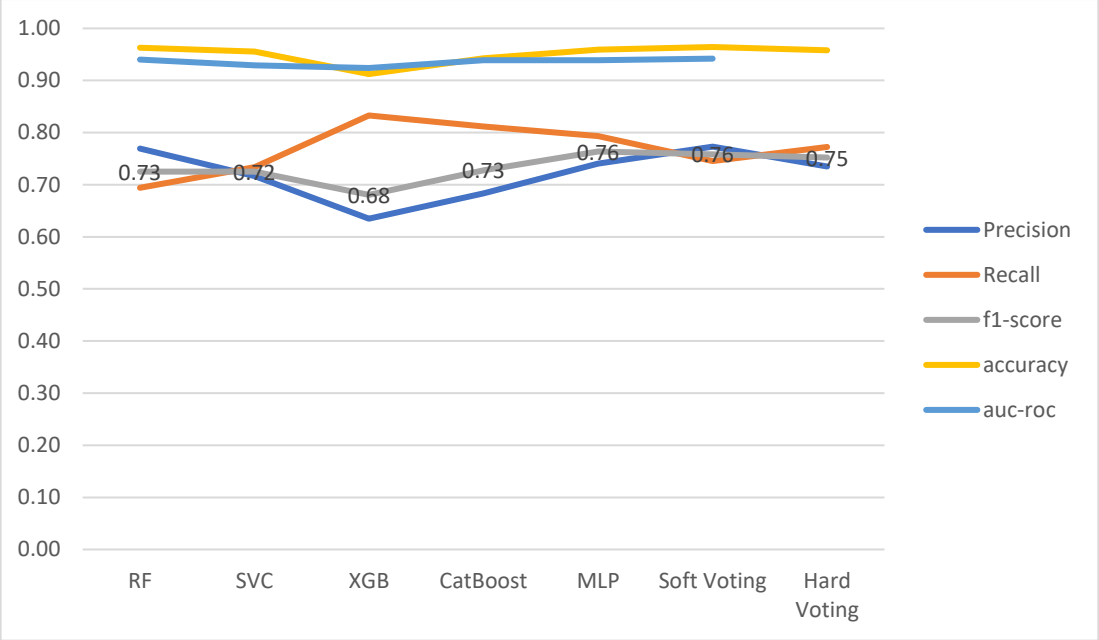
0.94	**0.96**	0.73	0.69	**0.77**	RF
0.929	**0.96**	0.72	0.73	0.72	SVC
0.924	0.91	0.68	**0.83**	0.63	XGB
0.939	0.94	0.73	0.81	0.68	CatBoost
0.939	**0.96**	**0.76**	0.79	0.74	MLP
0.942	**0.96**	**0.76**	0.75	**0.77**	**Soft Voting**
	0.96	0.75	0.77	0.73	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة البيانات المعززة يدويًا مع استخدام تقنية SMOTE على مجموعة الاختبار

يبين الشكل التالي نتائج الاختبار على منحنى AUC-ROC. كذلك، يبين الشكل التالي مخططًا بيانيًا لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة يدويًا مع استخدام تقنية SMOTE على مجموعة الاختبار.



مخطط ROC لنتائج اختبار النموذج المدرب على مجموعة بيانات معززة يدويًا مع استخدام تقنية SMOTE على مجموعة الاختبار



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة يدويًا مع استخدام تقنية SMOTE على مجموعة الاختبار

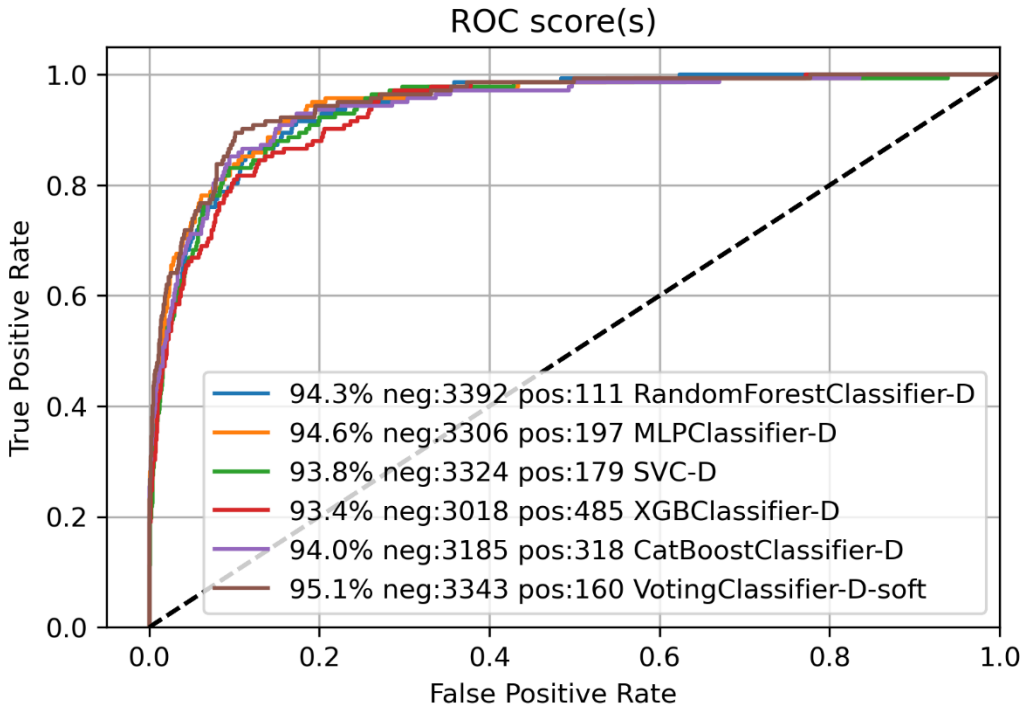
مجموعة البيانات LHS-TRAIN-D

جرى تدريب المصنفات على مجموعة البيانات LHS-TRAIN-D والتي أضفنا إليها بعض العينات بتقنيات التعزيز الآلي، ثم اختبار النموذج على مجموعة البيانات LHS-TEST، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

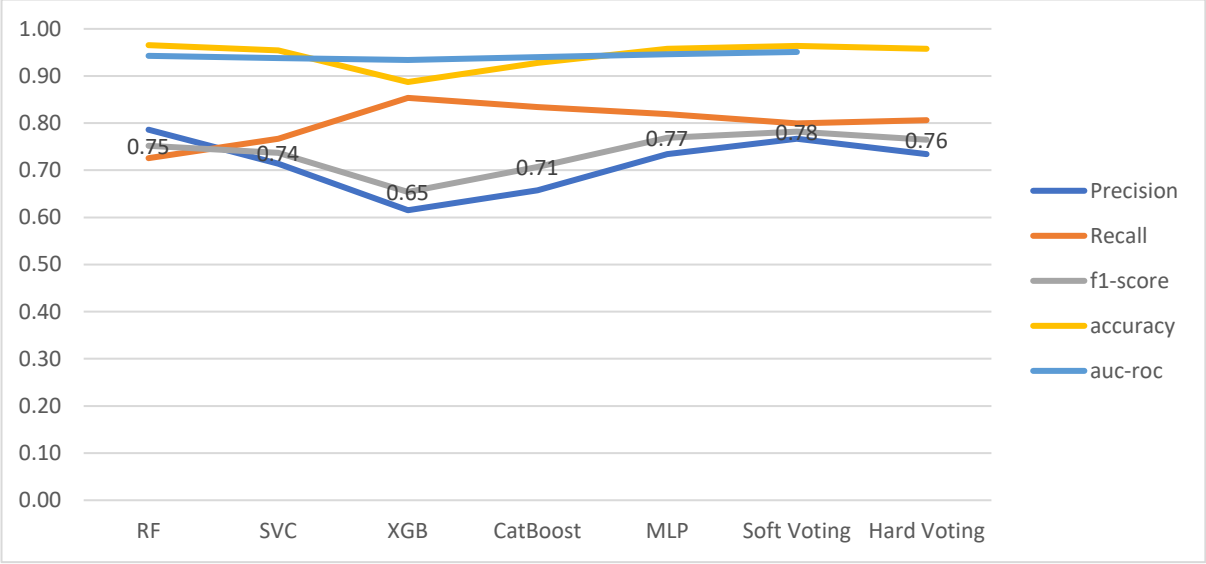
ROC	Accuracy	F1	Recall	Precision	Classifier
0.943	**0.97**	0.75	0.73	**0.79**	RF
0.938	0.95	0.74	0.77	0.71	SVC
0.934	0.89	0.65	**0.85**	0.62	XGB
0.94	0.93	0.71	0.83	0.66	CatBoost
0.946	0.96	0.77	0.82	0.73	MLP
0.951	0.96	**0.78**	0.80	0.77	**Soft Voting**
	0.96	0.76	0.81	0.73	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة البيانات المعززة آليًا مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST

كما يبين الشكل التالي نتائج الاختبار على منحنى AUC-ROC. كذلك، يبين الشكل التالي مخططًا بيانيًا لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة آليًا مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST.



مخطط ROC لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة آليًا مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة آليًا مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST

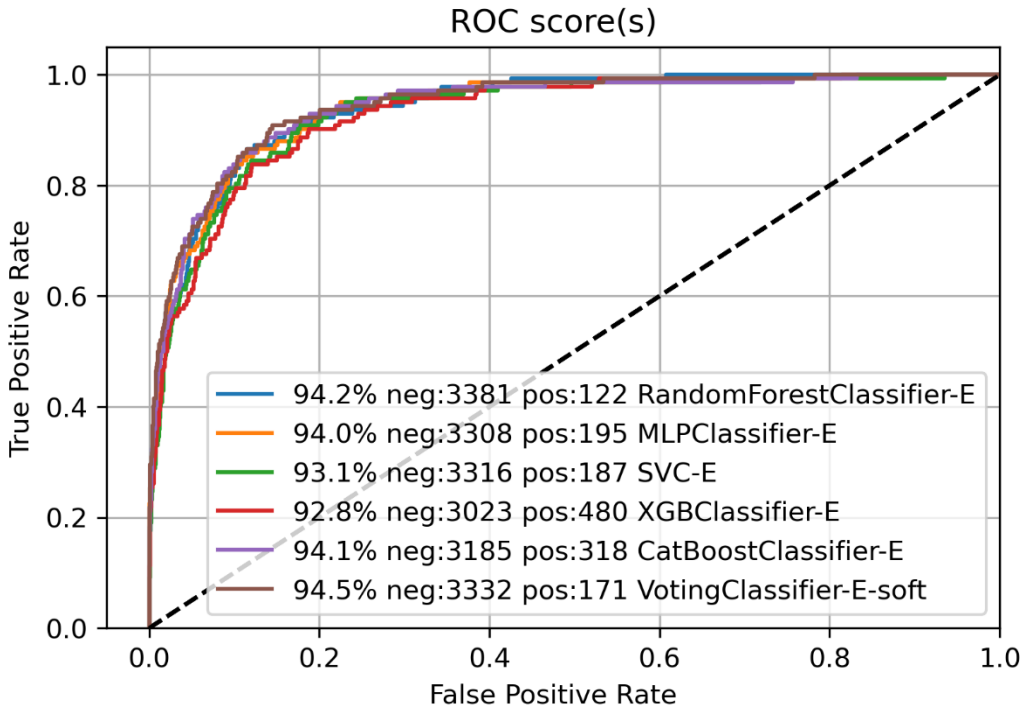
مجموعة البيانات LHS-TRAIN-E

جرى تدريب المصنفات على مجموعة البيانات LHS-TRAIN-E والتي أضفنا إليها العينات الناتجة عن تقنيات التعزيز اليدوي والآلي، ثم اختبار النموذج على مجموعة البيانات LHS-TEST، وحصلنا على النتائج التالية لمعايير الدقة والإرجاع ومقياس F1 والصحة المبينة في الجدول التالي:

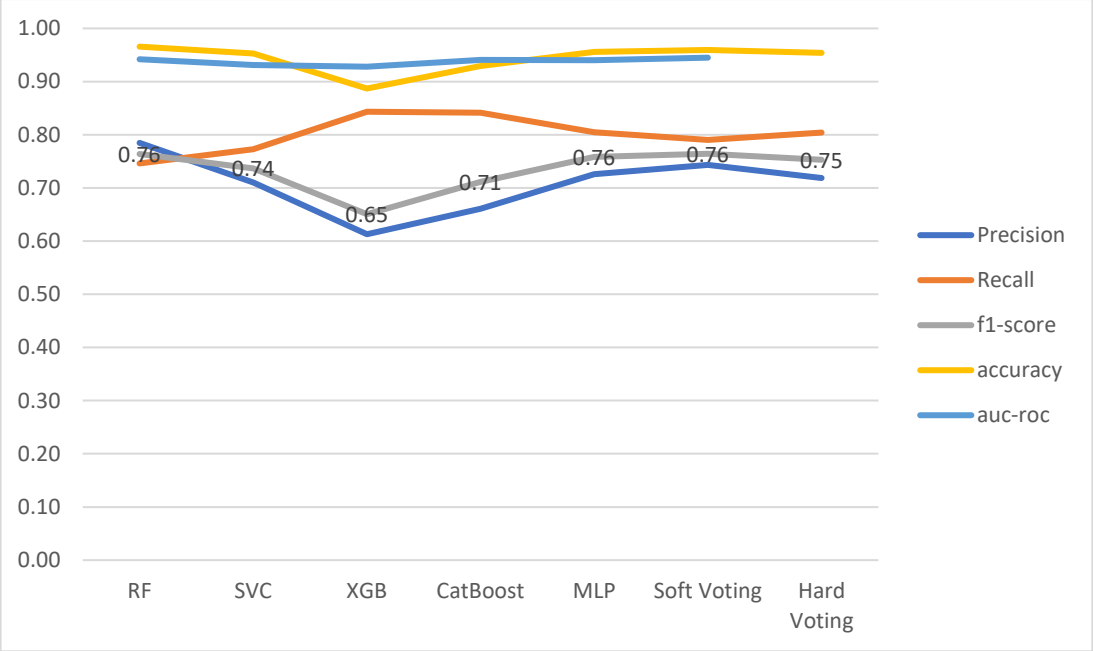
ROC	Accuracy	F1	Recall	Precision	Classifier
0.942	**0.97**	**0.76**	0.75	**0.78**	RF
0.931	0.95	0.74	0.77	0.71	SVC
0.928	0.89	0.65	**0.84**	0.61	XGB
0.941	0.93	0.71	**0.84**	0.66	CatBoost
0.94	0.96	**0.76**	0.80	0.73	MLP
0.945	0.96	**0.76**	0.79	0.74	**Soft Voting**
	0.95	0.75	0.80	0.72	Hard Voting

نتائج اختبار النموذج المدرب على مجموعة البيانات المعززة يدويًا وآليًا مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST

كما يبين الشكل التالي نتائج الاختبار على منحنى AUC-ROC. كذلك، يبين الشكل التالي مخططًا بيانيًا لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة يدويًا وآليًا مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST.



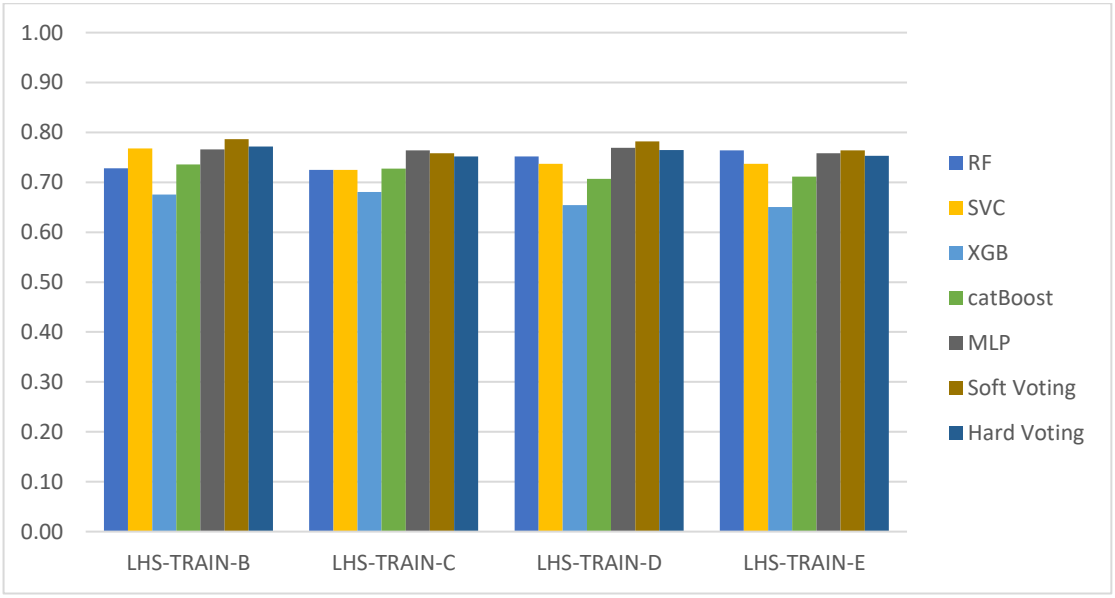
مخطط ROC لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة يدويًا وآليًا مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST



مخطط بياني لنتائج اختبار النموذج المدرب على مجموعة البيانات المعززة يدويًا وآليًا مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST

ملخص النتائج

سنقوم في هذه الفقرة ملخص للنتائج التي حصلنا عليها من مجموعات البيانات التي تختلف فيما بينها بتقنية التعزيز المستخدمة مع تقنية SMOTE.



مقارنة نتائج اختبار النموذج مع استخدام تقنية SMOTE على مجموعة البيانات LHS-TEST حسب تقنية تعزيز البيانات