تحليل مشاعر روّاد الشّبكات الاجتماعيّة التي تستخدم اللهجة الجزائريّة

إبراهيم لجدل[1]، غريسي نورالدّين[2]، عقبة كزار[3]، أحمد قريشي[4]

[1]جامعة الشّهيد حمة لخضر بالوادي، الجزائر.

[2]مخبر نظريّة المؤثرات والمعادلات التّفاضليّة الجزئيّة، جامعة الشّهيد حمة لخضر بالوادي، الجزائر.

[3] جامعة بسكرة، الجزائر.

[4]جامعة ورقلة، الجزائر.

تاريخ الارسال: 17جانفي 2021 /تاريخ القبول: 24 ماي 2021

الملخّص (Abstract)

إنّ تحليل النّص اليوم، له أهميّة كبيرة، خاصّة في مجالات مثل السّياسة والإنتاج والخدمات ... إلخ. كما أنّ، الشّبكات الاجتماعيّة مليئة بالنّصوص التي يعبر فيها مستخدمو الإنترنت عن مواضيع مختلفة، حيث يكون الاهتمام بآرائهم كبيرًا، ويعد فهم المحتوى الذي تنقله هذه النّصوص عنصرًا أساسيًا. يمكننا القول أن المدير النّاجح هو الذي يستمع جيدا لآراء المواطنين. ومن هنا، فإن تحليل المشاعر مهم جدًا لتلبيّة احتياجات المواطنين. في هذا العمل، سننفذ خوارزميّات محددة مسبقًا تحلل وتصنف مجموعة من المنشورات المأخوذة من الشّبكات الاجتماعيّة. الفئات التي حددناها هي: إيجابيّة سلبيّة أو محايدة. وعملنا هو من بين الأعمال الأولى التي تستخدم وتقارن عدة خوارزميّات تصنيف التّعليقات على فايسبوك، باستخدام اللهجة الجزائريّة.

Abstract : Today, text analysis has a great importance, especially in areas such as politics, productions, and services...etc. Currently, the social networks full of thetexts in which, the Internet-users express in different subjects, the interest of their opinions is considerable, where the comprehension of the content conveyed by these texts is an essential element. We can say that the good manager who listens well to the opinions of citizens. In this sense, the Sentiment Analysis is very important to meet the needs of citizens. In this work, we will implementpredefined algorithms that analyze and classify a set of publications derived from social networks. The classes that we have defined are: positive, negative or neutral. Our work is among the first one that uses and compares several comment classification algorithms on Facebook, using Algerian dialect.

Keywords- Sentiment Analysis; social network; annotated corpus;Lexicon of Sentiment.

الكلمات المفتاحيّة (Keywords): تحليل المشاعر؛ الشّبكات الاجتماعيّة؛ نصوص (corpus) موسوم؛ معجم المشاعر.

1. المدخل Introduction

يتمّ استخدام الإنترنت بشكل متكرّر كوسيلة لتبادل المعلومات. ويمكن للأشخاص نشر المعلومات بسهولة بما في ذلك آرائهم الشّخصيّة حول أيّ موضوع على الإنترنت. حيث يقوم المستخدم بإنشاء محتوى على الويب بلغات طبيعيّة على شكل نصوص حرّة غير منظّمة. اليوم، يتوفّر قدر هائل من المعلومات عبر الإنترنت حيث يمكننا العثور على أنواع مختلفة من مستندات الويب مثل: صفحات الويب، الصّور ملفات الصّوت والفيديو ومجموعة كبيرة ومتنوّعة من الملفات المختلفة. أيضًا، يمكننا العثور على مجموعات الأخبار والمنتديات والمدوّنات ومنشورات الشّبكات الاجتماعيّة. الآراء التي يعبر عنها الأشخاص اتّجاه موضوع ما هي من بين أنواع المعلومات التي لا تعدّ ولا تحصى المتاحة على الإنترنت. في هذا السّياق، سوف ندرس استخدام اللهجات في وسائل التّواصل الاجتماعي. يجب النّظر إلى اللهجة باللغة العربيّة لأن تحديد اللهجة العربيّة يساعد في تحديد السّياق. اللغة العربيّة لها نسخة قياسيّة مفهومة جيدًا في جميع أنحاء العالم العربي. وهي معروفة باسم اللغة العربيّة القياسيّة الحديثة (MSA). يتمّ استخدامها جنبًا إلى جنب مع اللهجات العاميّة العربيّة في المحتوى عبر الإنترنت. يميل معظم مستخدمي الشّبكات الاجتماعيّة إلى استخدام اللهجة العربيّة (DA). المشاكل النّاشئة عن استخدام هذه الأخيرة هي أبعد من تلك النّاشئة عن اللغة العربيّة القياسيّة (MSA) بسبب عدم توحيد اللهجة العربيّة (DA) وندرة أدوات معالجتها (Harrat, Meftouh, Smali, 2017) [1].

Syiam et al. (2006) [2] اعتبر الرّأي كمعتقد شخصي ، أو نتيجة عاطفة أو تفسير للحقائق هو نتيجة منظور الشّخص ، وفهمه ، ومشاعره الخاصّة ، ومعتقداته ، ورغباته. ويشير إلى معلومات غير مؤكّدة تختلف عن المعرفة

والمعتقدات القائمة على الحقائق. تختلف اللهجة العربيّة (DA) اختلافًا كبيرًا عن اللغة العربيّة التي تُدرَّس في المدارس وتستخدم في التّواصل الكتابي والخطاب الرّسمي (الأخبار المذاعة، والدّين والسّياسة، وما إلى ذلك).

تمّ اقتراح العديد من الأساليب الحاليّة لتحليل المشاعر باللغة العربيّة (SA)؛ ومع ذلك، اقتصرت بشكل عامّ على اللغة العربيّة الفصحى الحديثة (MSA) أو بعض اللهجات ذات الأهميّة الاقتصاديّة أو السّياسيّة.

في هذه الورقة، سوف نستكشف الأعمال المختلفة المتعلّقة بتحليل المشاعر في الشّبكات الاجتماعيّة باستخدام اللهجات العربيّة المختلفة (القسم 2). في القسم 3، سنقدّم مساهمتنا في تحليل المشاعر في الشّبكات الاجتماعيّة باستخدام اللهجة الجزائريّة. ثمّ نناقش نتائج التّجربة (القسم 4). أخيرًا، نقدّم خاتمة وبعض الأعمال المستقبليّة.

2. نظرة عامّة

في هذا القسم، سنقدّم الأعمال الرّئيسيّة التي تستخدم بشكل أساسي اللهجة العربيّة (AD) لتحليل المشاعر في المراجعات أو التّعليقات أو التّغريدات.

يُقدِّم Shoukry وRafea (2012) [3] تطبيقًا على تحليل المشاعر العربيّة من خلال القيام بتصنيف المشاعر للتّغريدات العربيّة. يتمّ تحليل التّغريدات المتحصّل عليها من أجل تزويدها بالمشاعر القطبيّة (إيجابيّة، أو سلبيّة). حيث يتمّ جمع هذه البيانات من الشّبكة الاجتماعيّة Twitter؛ التي لها أهميّة في منطقة الشّرق الأوسط، حيث يتحدّث معظمها اللغة العربيّة. لقد جمعا 1000 تغريدة مقسّمة بالتّساوي إلى 500 إيجابيّة و500 سلبيّة. بعد تصفيّة التّغريدات لإزالة الكلمات غير العربيّة وعلامات HTML والصّور وما إلى ذلك، استخدما ميزات n-gram القياسيّة وجربا العديد من المصنّفات (SVM وNB) من خلال مجموعة أدوات Weka. في عمل آخر، اقترحا طريقة بسيطة لدمج النّهج القائم على النّصوص مع النّهج القائم على المعجم. ركزا على اللهجة المصريّة وقاما بالتّجربة على مجموعة بياناتهما المكوّنة من 4800 تغريدة (مقسّمة بالتّساوي عبر الفئات الإيجابيّة والسّلبيّة والحياديّة).

اقترح Al-Subaihin ، Al-Khalifa وAl-Salman (2011) [4] وAl-Subaihin وAl-Khalifa (2014) [5] تقنيّة جديدة تعتمد على المعجم للتعامل مع اللهجة العربيّة. تكمن الحداثة في نهجهم في استخدام لعبة على الإنترنت لخلق معجم المشاعر من خلال ما يسمى بالحساب البشري. في عمل آخر من قبل نفس المجموعة، ناقش Albraheem وAl-Khalifa (2012) [6] بالتّفصيل القضايا / التّحديات التي تواجهها مناهج تحليل المشاعر للغة العربيّة القائمة على المعجم. حيث جمعوا 1000 تغريدة مقسمة بالتّساوي إلى 500 إيجابيّة و500 سلبيّة. بعد تصفيّة التّغريدات لإزالة الكلمات غير العربيّة وعلامات HTML والصّور وما إلى ذلك، استخدموا ميزات n-gram القياسيّة وجربوا العديد من المصنّفات (SVM وNB) من خلال مجموعة أدوات Weka.

تناول Nawaf et al. (2013) [7] كلا النّهجين في تحليل المشاعر للغة العربيّة. نظرًا لوجود عدد محدود من مجموعات البيانات العربيّة والمعاجم العربيّة المتاحة لتحليل المشاعر، يبدأ عملهم ببناء مجموعة بيانات موسومة يدويًّا ثمّ يأخذ القارئ من خلال الخطوات التّفصيليّة لبناء المعجم. ويتم إجراء التّجارب خلال المراحل المختلفة من هذه العمليّة لمراقبة التّحسينات المكتسبة على دقّة النّظام ومقارنتها بالنّهج القائم على النّصوص (coprus) .

حاول Al-Kabi et al. (2013) [8] دراسة تحليل الرّأي في وسائل التّواصل الاجتماعي باللغة العربيّة. في محتوى هذه الشّبكات الاجتماعيّة يتمّ خلط اللغة العربيّة الفصحى الحديثة (MSA) المستخدمة في وسائل الإعلام والأدب مع اللغة العاميّة. قاموا بتحديد استقطاب الرّأي واسناد كل رأي أو تعليق الى فئة معيّنة. ثمّ اختبار وتقييم فئات المجال المختلفة لمعرفة الاختيار الذي يمكن أن يؤدّي إلى أفضل النّتائج من حيث الدّقة مقارنة بالأحكام اليدويّة. وطوّروا أداة لتحليل الرّأي والتّصنيف مخصّصة للغة العربيّة. تمّ تجميع قاموس للكلمات الإيجابيّة والسّلبيّة والحياديّة باللغة العربيّة على أساس مسح عدد كبير من الوثائق والمنشورات. بناءً على قاموس الاستقطاب هذا، قاموا بجمع مجموعة كبيرة من الآراء أو المنشورات من الشّبكات الاجتماعيّة. وفحص الكلمات الموجودة في هذه المنشورات مقابل القاموس بحثًا عن القطبيّة ثم إعطاء فئة القطبيّة (على سبيل المثال، إيجابي قويّ، إيجابي متوسّط، إيجابي ضعيف، إلخ) لكل منشور بناءً على عدد الكلمات الإيجابيّة أو السّلبيّة أو المحايدة.

اقترح Mataoui et al. (2016) [9] نهجاً جديداً لتحليل المشاعر قائم على المعجم لمعالجة الجوانب المميّزة للغة العربيّة الجزائريّة العاميّة المستخدمة بالكامل في الشّبكات الاجتماعيّة. حيث تمّ إنشاء مجموعة بيانات موسومة يدويًا وثلاثة معاجم عربيّة جزائريّة لدراسة المراحل المختلفة لنهجهم. يتكوّن هذا النّهج من أربع وحدات: وحدة حساب تشابه العبارات الشّائعة؛ وحدة المعالجة المسبّقة؛ وحدة الاشتقاق والكشف عن اللغة؛ ووحدة حساب القطبيّة. يتكون قاموسهم من ثلاثة أجزاء: معجم الكلمات الرّئيسيّة؛ معجم كلمات النّفي؛ معجم كلمات التّكثيف والحدّة. تم إثراء هذه المعاجم الثلاثة بقاموس الرّموز وقاموس آخر للعبارات الشّائعة. وأخيرًا، قاموا ببناء مجموعة اختبار للأغراض التّجريبيّة. حيث تمّت تصفيّة هذه المجموعة ووسمها لتسهيل عمليّة تقييم الاقتراح. تظهر النّتائج التّجريبيّة أن نظامهم حصل على أداء بنسبة 79.13٪ من الدّقة.

ركّز Medhaffar et al. (2017) [10] على تحليل المشاعر باللهجة التّونسيّة. حيث استخدموا تقنيّات التّعلم الآلي لتحديد قطبيّة التّعليقات المكتوبة باللهجة التّونسيّة. أوّلاً، قاموا بتقييم أداء أنظمة تحليل المشاعر من خلال نماذج مدرّبة باستخدام مجموعات بيانات متاحة مجّانًا للعربيّة الفصحى والمتعدّدة اللهجات. حيث جمعوا ووسموا مجموعة تعليقات باللهجة التّونسيّة تحتوي على 17.000 تعليقاً من Facebook. أظهرت هذه

المجموعة تحسّنا ملحوظا مقارنة بأفضل نموذج تم تدريبه على بيانات اللهجات العربيّة الأخرى أو العربيّة الفصحى.

أنشأ Baly et al. (2017) [11] مجموعة بيانات Twitter المتعددة اللهجات العربيّة والخاصّة بالمشاعر الأولى من نوعها (MD-ArSenTD) التي تتكون من تغريدات تم جمعها من 12 دولة عربيّة، مع تعليقات توضيحيّة للمشاعر واللهجة. استخدموا مجموعة البيانات هذه لتحليل التّغريدات التي تمّ جمعها من مصر والإمارات العربيّة المتّحدة، بهدف اكتشاف السّمات المميزة التي قد تسهل تحليل المشاعر. كما قاموا بإجراء تقييم مقارن لنماذج المشاعر المختلفة على التّغريدات المصريّة والإماراتيّة. أشارت النّتائج إلى الأداء المتفوّق لنماذج التّعلم العميق، وأهميّة السّمات المورفولوجيّة في معالجة اللغات الطبيعيّة العربيّة، وأن التّعامل مع اللهجات العربيّة يؤدّي إلى نتائج مختلفة اعتمادًا على الدّولة التي يتمّ جمع التّغريدات منها.

يقدّم Alomari et al. (2017) [12] مجموعة بيانات تويتر عربيّة أردنيّة حيث يتمّ وسم التّغريدات على أنّها إيجابيّة أو سلبيّة. قاموا أوّلا بدراسة مناهج مختلفة للتّعلم الآلي الخاضع للإشراف لتحليل المشاعر عند تطبيقها على المستخدمين العرب لوسائل التّواصل الاجتماعي في المواضيع العامّة الموجودة إمّا باللغة العربيّة الفصحى (MSA) أو اللهجة الأردنيّة، ثمّ إجراء تجارب لتقييم استخدام تقنيّات وسيناريوهات مختلفة لأنظمة الوزن، الاشتقاق وN-grams. تكشف النّتائج التّجريبيّة عن أفضل سيناريو لكل مصنّف؛ وتشير إلى أنّ مصنّف SVM باستخدام معامل (TF-IDF) للوزن مع الاشتقاق من خلال ميزة Bigrams يتفوّق على أفضل سيناريو لمصنّف Nave Bayesian من حيث نتائج الأداء.

قدّم Al-Twairesh (2018) [13] ثلاثة مصنِّفات هجينة لتحليل المشاعر للتّغريدات العربيّة. تعمل المصنّفات على مستويات مختلفة من التّصنيف: تصنيف ثنائي (إيجابي او سلبي) تصنيف ثلاثي (إيجابي، سلبي او محايد) وتصنيف رباعي الاتّجاه (إيجابي، سلبي، محايد او مختلط). يعتمد المنهاج على دمج المعرفة المستخرجة من الطريقة القائمة على المعجم كميزات في الطريقة المستندة إلى النّصوص من اجل تطوير الطريقة الهجينة. تم استخلاص مجموعة ميزات من البيانات ثمّ اقتراح خوارزميّة اختيار عكسيّة للقيام باختيار الميزة للوصول إلى أفضل أداء للتّصنيف.

استخدم Elouardighi et al. (2018) [14] نهجاً قائماً على التّعلم الآلي. واقترحوا نهجا يحلّل المشاعر من التّعليقات على فيسبوك خاصّةً باللهجة المغربيّة. تحليل المشاعر هي عمليّة يتم خلالها تحديد وتعريف قطبيّة (إيجابيّة أو سلبيّة أو محايدة) لنص معين [18]. تبدأ هذه العمليّة بجمع التّعليقات ووسمها باستخدام طريقة crowdsourcing. ثم يقومون بالمعالجة القبليّة للنص لاختصار الكلمات العربيّة إلى جذورها. حيث سيتم استخدام هذه الكلمات لبناء متغيّرات

الإدخال باستخدام عدّة توليفات من عمليّات الاستخراج والتّرجيح.

العديد من الأعمال تُنشئ مجموعة بيانات ضخمة لاختبار نماذجها مثل Al-Obaidi and Samawi (2016) [15] ممّا أدّى إلى إنشاء مجموعة بيانات للتّنقيب عن الرّأي من اجل اللهجات المتنوّعة للّغة العربيّة (OMCCA). تتألّف مجموعة البيانات من 28576 مراجعة موسومة على أنها إيجابيّة أو سلبيّة أو محايدة. كانت اللهجات المدروسة هي الأردنيّة والسّعوديّة. OMCCA المتاحة للجميع. أبلغ المؤلّفون عن تجارب على OMCCA باستخدام ميزات وتقنيّات تصنيف مختلفة. في Al-Suwaidi et al. (2016) [16]، حجم مجموعة البيانات هو 1000 واللهجة المدروسة هي اللهجة الإماراتيّة، بينما في Alomari et al. (2017) [12]، تسمّى مجموعة البيانات بـ : مجموعة بيانات التّغريدات العامّة العربيّة الأردنيّة (AJGT) وتتكوّن من 1800 تغريدة. بالإضافة إلى ذلك، وسم عسيري وآخرون. (2016) مجموعة بيانات حجمها 4700 لتحليل المشاعر باللهجة السّعوديّة مع (K = 0.807) [17].

3. الإسهام

يتكوّن عملنا من 3 نقاط أساسيّة:

- استخدمنا اللهجة الجزائريّة مع 4 مصنّفات وهي كالتّالي: Support Vector Machines (SVM) Decision Tree (DT)، Random Forest (RF) وNave Bayes (NB).

- قمنا بوسم مجموعة البيانات الخاصّة باللهجة الجزائريّة يدويًّا والمحتوية على 2891 تعليقًا.

- قمنا بإنشاء قاموس مكوّن من 1328 كلمة موسومة باللهجة الجزائريّة.

1.3 الوسم

وهي عمليّة تحديد صنف كلّ تعليق عن طريق التّعليق عليه على أنّه إيجابي أو سلبي أو محايد. استخدمنا اثنين من المقيّمين لتحديد مشاعر التّعليقات يدويًّا. إمّا أن يتوافقوا فيكون لديهم درجة عاليّة من الاتّفاق في تصنيفهم للتّعليقات أوّلا يتوافقون؛ ففي هذه الحالة يتمّ استخدام مقيم ثالث لتحديد الصّنف النّهائي. حيث استخدمنا ثلاثة أقطاب 1 ، -1 ، 0 تتوافق على التّوالي مع المشاعر الإيجابيّة، السّلبيّة والحياديّة. يوضّح الجدول 1 الأقطاب المختلفة للتّعليقات الموسومة في مجموعة البيانات الخاصّة بنا.

الأقطاب (الأصناف)	إيجابي	سلبي	حيادي	**المجموع**
عدد التّعليقات	975	525	1391	**2891**

جدول 1: عدد التّعليقات لكلّ قطب (صنف)

في الصّورة 1، نقدم بعض الأمثلة التي تصف بعض التّعليقات بأقطابها.

التّعليقات	الأقطاب (الأصناف)
هذي جياحة نريسكي بعمري على حاجة تستاهل ماشي علاجال بوزوج رجلين يجري مور البالون	سلبي
شفت المواطن مش واعي خلاص داير المزبلة حذاه يكل منا او يرمي منا	سلبي
في بسكرة خويا لعزيز طريق سيدي عقبة ؟	محايد
هاذو روشيات تع رخام ماشي ثلج	محايد
ربي يكثر من امثالك	ايجابي
درس في حسن ضيفة كل الاحترام وتقدير ناس المديّة	ايجابي

صورة 1 : بعض التّعليقات مع أصنافها

2.3 القاموس

لم نجد معجمًا محدّدًا لكلمات اللهجة الجزائريّة. ولهذا، قمنا بإنشاء قاموسنا الخاصّ. حيث قمنا بتجميع قاموس للكلمات الإيجابيّة والسّلبيّة والحياديّة باللغة العربيّة من خلال مسح عدد كبير من الوثائق. بناءً على قاموس الاستقطاب هذا، قمنا بجمع مجموعة كبيرة من التّعليقات من الشّبكات الاجتماعيّة. استغرقنا حوالي شهر من العمل الشّاق لجمع 1328 كلمة باللهجة الجزائريّة من وسائل التّواصل الاجتماعي. كما طلبنا مساعدة أصدقاء من شمال الجزائر الجنوب، الشّرق والغرب لإنشاء قاموس للهجة الجزائريّة يغطي أكبر عدد ممكن من مناطق بلادنا.

يمثّل الجدول 2 تقسيم كلمات القاموس وفقًا للأقطاب.

الأقطاب (الأصناف)	إيجابي	سلبي	محايد
عدد التّعليقات	565	763	1328

جدول 2 : توزيع كلمات قاموسنا على مختلف الأقطاب

تمثّل الصّورة 2 مثالاً لبعض الكلمات الإيجابيّة في القاموس.

```
تـألـق
فـخر
تـهـبـل
نـاجـح
الـبـسـمـة
جـيـد
بـرافـو
احسنـتـي
تـحـيـاتـنـا
زيـنـة
مـزيـان
جـمـيـل
مـمـتـاز
روعه
```

صورة 2: الكلمات الإيجابيّة في القاموس

تمثّل الصّورة 3 مثالاً للكلمات السّلبيّة في القاموس.

```
عنف
فـاشل
ردئ
مـهـزلـة
خـامـجـين
دمـار
وسخ
تـبـلـعـيـط
بـلـعـاط
كـارثـة
الـنـحس
رهج
```

صورة 3: الكلمات السّلبيّة في القاموس

4. التّجارب والنّتائج

في هذا القسم، نستخدم Python كلغة برمجة في التّجارب، نظرًا لوجود العديد من المكتبات الجيدة (والسّريعة) للتّعامل مع نصوص اللغة الطبيعيّة، وتغطيّة لجميع المهام اللازمة لبناء أداتنا، مثل مكتبات CSV وGensim وScikit-learn. أيضًا، نستخدم أربع خوارزميّات: SVM، DT، RF وNB.

تـمّ إجـراء التّجـارب على جهـازين بمعالجـات Intel Core2 Quad Q9400 (4 أنوية لكل معالج) بسرعة 2.66 جيغا هرتز وذاكرة 4 جيغا بايت.

Feature	Abbreviation	signification
Has Positive Word	HPW	1 او 0
Has Negative Word	HNW	1 او 0
Positive Word Count	PWC	>= 0
Negative Word Count	NWC	>= 0
CommentLength	CL	صحيح > 0
SentimentLevel	SL	-1 <= V <= 1

جدول 3: الخصائص المستخدمة في عملنا

1.4 المميّزات

استخدمنا ستّ مميّزات رئيسيّة. يمثّل الجدول 3 المميّزات المختلفة المستخدمة في عملنا.

استخدمنا طريقة التّعلم الخاضعة للإشراف والنّهج القائم على المعجم. وبالتّالي، يتعين علينا تقسيم النّصوص إلى قسمين، 80٪ للتدريب و20٪ للاختبار. قمنا بإجراء العديد من الاختبارات، وتظهر نتائج الدّقة في الجدول 4.

الاختبار	المميزات	المصنف			
		SVM	DT	RF	NB
1	كل المميزات	85.14	83.07	**85.31**	84.28
2	HPW, HNW, PWC, NWC	84.11	84.45	84.28	82.38
3	PWC, NWC, CL, SL	84.62	83.24	85.31	83.76
4	PWC, NWC	84.11	84.28	84.45	67.87
5	HPW, HNW, CL, SL	84.11	84.11	84.11	48.35
6	HPW, HNW	84.11	84.11	84.11	72.30
7	CL, SL	47.49	47.66	48.18	48.18

جدول 4: نتائج التّجارب

نلاحظ أننا نحصل على أفضل النّتائج عندما نستخدم جميع المميزات مع المصنّف Random Forest (RF).

2.4 المقارنة والنّتائج

في هذا القسم، سنقارن عملنا بالأعمال الأخرى التي تستخدم لهجات مختلفة مثل التّونسيّة والمغربيّة والمصريّة والسّعوديّة والأردنيّة. يمكننا أن نستنتج أنّ المصنّفات الخاصّة بنا حقّقت نتائج جيّدة. يلخّص الجدول 5 المقارنة التي أجريناها للتّحقق من صحّة عملنا.

اللهجات	المصنّف	النّتائج	
		الدّقة	F-measure
عملنا	RF	85.31 %	84.93 %
الأردنيّة [12]	SVM	87.2 %	/
الجزائريّة [9]	/	79.13 %	/
التّونسيّة [10]	MLP	/	78 %
المغربيّة [14]	SVM	/	78 %
المصريّة [3]	SVM	/	72.5 %
السّعوديّة [13]	SVM	/	69 %

جدول 5: مقارنة النّتائج مع أعمال أخرى

5. الخاتمة والأعمال المستقبليّة

في هذه الورقة، قمنا بتحليل مشاعر النّصوص تحتوي على 2891 تعليقًا باللهجة الجزائريّة، مقسّمة إلى ثلاثة أصناف: 975 تعليقًا إيجابيًا، و525 تعليقًا سلبيًا و1391 تعليقًا محايدًا. لقد استخدمنا أربعة مصنّفات للتّعلم الآلي وهي Support Vector Machine (SVM) وDecision tree (DT) ، وRandom Forest (RF) وNaive Bayes (NB) ، حيث قمنا بتقييم هذه المصنّفات باستعمال 20٪ من النّصوص المستخدمة. استخدمنا ست مميّزات وهي HasPositiveWord وHasNegativeWord وPositiveWordCount وNegativeWordCount وCommentLength وSentimentLevel. ثمّ قمنا بإجراء سبعة اختبارات مختلفة، تمّ إجراء الاختبار الأوّل باستخدام جميع المميّزات أمّا الاختبارات الأخرى تمّ إجراؤها عن طريق التّبديل بين هذه المميّزات. وأخيرا قارنا نتائج الاختبار للمصنّفات الأربعة. النّتائج التي تمّ الحصول عليها مشجّعة للغايّة، حيث حقّقنا أفضل دقّة 85.31٪ عند استخدام مصنّف Random Forest (RF).

نذكر هنا بعض الأعمال المستقبليّة:

1. إثراء قاموسنا بمزيد من كلمات اللهجة الجزائريّة التي تغطي مناطق أخرى لأنّ الجزائر كبيرة جدًا وتحتوي على عشر لهجات.

2. إثراء مجموعة البيانات بتعليقات أخرى باللهجة الجزائريّة للحصول على نتائج دقيقة.

3. تطبيق مصنفات ومميّزات أخرى.

4. التّحليل باستخدام الصّنف المختلط بالإضافة إلى الأصناف السّابقة: الإيجابيّة والسّلبيّة والحياديّة.

5. استخدام إعدادات أخرى مثل: الثّنائي bigram الثّلاثي trigram والمختلط mixed.

مراجع (References)

[1] Harrat, S., Meftouh, K., Smali, K. (2017). Machine translation for Arabic dialects (survey). Information Processing Management.

[2]Syiam, M. M., Fayed, Z. T., Habib, M. B. (2006). An Intelligent System FOR Arabic Text Categorization. Information Sciences, 6.(1)

[3]Shoukry A. and Rafea A., sentence-level Arabic Sentiment Analysis, SoMNet, 2012, 2-5.

[4]Al-Subaihin, A. A., Al-Khalifa, H. S., Al-Salman, A. S. (2011). A proposed sentiment analysis tool for modern Arabic using human-based computing. Proceedings of the 13th international conference on information integration and web-based applications and services, 543546.

[5]Al-Subaihin, A. S. and Al-Khalifa, H. S. (2014). A system for sentiment analysis of colloquial Arabic using human computation. The Scientic World Journal, 2014, 18.

[6]Albraheem, L., Al-Khalifa, H. S. (2012). Exploring the problems of sentiment analysis in informal Arabic. Proceedings of the 14th international conference on information integration and web-based applications services415418.

[7] Nawaf A. Abdulla, Nizar A. Ahmed, Mohammed A. Shehab, Mahmoud Al-Ayyoub, Arabic Sentiment Analysis: Lexicon-based and Corpus-based, AEECT, 2013,1-6.

[8] Al-Kabi, M., Gigieh, A., Alsmadi, I., Wahsheh, H., Haidar, M., An opinion analysis tool for colloquial and standard Arabic, Proceedings of the 4th international conference on information and communication systems (ICICS)13.

[9] M. Mataoui, O. Zelmati, M. Boumechache (2016), A Proposed Lexicon-Based Sentiment Analysis Approach for the Vernacular Algerian Arabic, Research in Computing Science, 2016, 55-68.

[10] Medha_ar, S., Bougares, F., Esteve, Y., Hadrich-Belguith, L. (2017). Sentiment analysis of Tunisian dialects: Linguistic resources and experiments. Proceedings of the third Arabic natural language processing workshop, 5561.

[11] Baly, R., El-Khoury, G., Moukalled, R., Aoun, R., Hajj, H., Shaban, K. B. (2017). Comparative evaluation of sentiment analysis methods across Arabic dialects, Procedia Computer Science, 117, 266273.

[12] Alomari, K. M., ElSherif, H. M., Shaalan, K. (2017). Arabic tweets sentimental analysis using machine learning. Proceedings of the international conference on industrial, engineering and other applications of applied intelligent systems602610.

[13] N. Al-Twairesh, H. Al-Khalifa, A. AlSalman Y., Al-Ohali, Sentiment Analysis of Arabic Tweets: Feature Engineering and A Hybrid Approach, Computation and Language (cs.CL), 2018

[14] Elouardighi A , Maghfour M., Hammia H. Aazi F.,Z. , Analyse des sentiments partir des commentaires Facebook publis en Arabe standard ou dialectal marocain par une approche dapprentissage, Confrence Internationale sur l'Extraction et la Gestion des Connaissances, Paris, France,2018, 329-334.

[15] Al-Obaidi, A.,Samawi, V. (2016). Opinion mining: Analysis of comments written in Arabic colloquial. Proceedings of the world congress on engineering and computer science 2016 (wcecs 2016).

[16] Al Suwaidi, H., Soomro, T. R., Shaalan, K. (2016). Sentiment analysis for Emirati dialects on twitter. Sindh University Research Journal, 48(4), 707710.

[17] Assiri, A., Emam, A., Al-Dossari, H. (2016). Saudi twitter corpus for sentiment analysis. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 10(2), 272275.

[18] Al-Harbi, W. A.,Emam, A. (2015). E_ect of Saudi dialect preprocessing on Arabic sentiment analysis. International Journal of Advanced Computer Technology, 4(6), 9199.