الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جـــامـعـة حـمـه لـخـضـر الـــوادي

كلية العلوم الدقيقة

قسم الاعلام الالي

مذكرة نهاية الدراسة تدخل ضمن متطلبات الحصول على

شهادة الماستر الأكاديمي

الميدان:رياضيات واعلام الي

الشعبة:الاعلام الالي

التخصص:الأنظمة الموزعة والذكاء الاصطناعي

الموضوع:

نهج ذكي لتصنيف النصوص في الشبكات الاجتماعية حسب اللهجات العربية

من انـجاز الطالبات:

- ✓ بالمهدي حنين
- ✓ بالطاهر دنـيا

نوقشت يوم 2022/06/16 من طرف أعضاء لجنة المناقشة:

أعضاء لجنة المناقشة			
السيد عمار بوشريط	أستاذ محاضر	جامعة الوادي	رئيسا
السيد سمير عثماني	أستاذ محاضر	جامعة الوادي	مناقشا
السيد محمد شرف الدين مفتاح	أستاذ محاضر	جامعة الوادي	مشرفا

السنة الجامعية :

2022/2021

بسم الله الرحمن الرحيم

الاهداء

الحمد لله وكفى والصلاة على الحبيب المصطفى واهله ومن وفى اما بعد الحمد لله الذي وفقني لتثمين هذه الخطوة في مسيرتي الدراسية بهذه المذكرة ثمرة الجهد والنجاح على الرغم من جميع الصعوبات التي واجهتني مهداة الى والديا العزيزين سندي وقوتي حفظهما الله وادامهما نورا لدربي . لكل العائلة الكريمة من اخوة واخوات الى كل الاصدقاء الاعزاء رفقاء الدرب الى كل من احبهم قلبي ونسيهم قلمي..

حنين

الاهداء

من رياحين الكلام وياسمين الجمل وشقائق الأحرف أنظم أطواق الشكر والعرفان إلى التي حملتني تسعة اشهر وتحملت تربيتي حتى وصلت إلى ما أنا عليه إلى والديا الغاليان إلى من أضاء شموع دربي إلى من حملوا عبئ حياتي وإلى رفقائي في الحياة وإلى إخوتي الأعزاء.

دنيا

شكر وعرفان

{ فَوَجَدَا عَبْدًا مِنْ عِبَادِنَا آتَيْنَاهُ رَحْمَةً مِنْ عِنْدِنَا وَعَلَّمْنَاهُ مِنْ لَدُنَّا عِلْمًا (65) قَالَ لَهُ مُوسَىٰ هَلْ أَتَّبِعُكَ عَلَىٰ أَنْ تُعَلِّمَنِ مِمَّا عُلِّمْتَ رُشْدًا (66) قَالَ إِنَّكَ لَنْ تَسْتَطِيعَ مَعِيَ صَبْرًا (67) و كَيْفَ تَصْبِرُ عَلَىٰ مَا لَمْ تُحِطْ بِهِ خُبْرًا (68) قَالَ سَتَجِدُنِي إِنْ شَاءَ اللَّهُ صَابِرًا وَ لَا أَعْصِي لَكَ أَمْرًا (69) }

الحمد لله الذي أنار لنا درب العلم و المعرفة وأعاننا على أداء هذا الواجب ووفقنا على إتمام هذا العمل والصلاة والسلام على نبيه الحبيب أما بعد يقول الله سبحانه وتعالى : "يرفع الله الذين آمنوا منكم والذين أوتوا العلم درجات" الحمد لله الذي قدرنا وأنعم علينا باتمام هذه الرسالة العلمية ثمرة جهدنا فنرجو أن تكون إضافة علمية ينتفع بها ويستفيد منها غيرنا نتقدم بالشكر الجزيل الخالص للذي أشرف على مذكرتنا الأستاذ الفاضل محمد شرف الدين مفتاح الذي لم يبخل علينا بتوجيهاته القيمة وحرصه على متابعة إنجاز هذه المذكرة ولا يفوتنا أن نتقدم بالشكر إلى أساتذتنا الكرام الذين شاركونا في مشوارنا الدراسي خلال مسيرتنا الجامعية ، و لا ننسى جميع الزملاء و الأحبة دون استثناء.

FREE PRINTABLE BABY SHOWER INVITATION - BEESHOWER.COM

الملخص

يشهد العالم المعاصر مجموعة من التحولات المتسارعة في مجال الاتصال وتقنية المعلومات؛ خاصة مع الاستخدام الواسع لمواقع التواصل الاجتماعي إذ أنها ركيزة أساسية لمختلف التعاملات في كل مجالات. يعتبر تنقيب البيانات وتعدين النصوص من التقنيات و المجالات المهمة في العالم وخاصة الشبكة الافتراضية وتنوع اللغة العربية في لهجاتها في المغرب العربي والخليج. حيث تعبر كل لهجة عن هوية صاحبها وما يحاول إيصاله لطرف آخر أدى إلى خلق حجم بيانات هائل مليى بمعلومات مفيدة. خاصة في مجال التجارة الكترونية والتسويق . في عملنا هذا عملنا على تطبيق خوارزميات التعلم الآلي لتصنيف النصوص وتعليقات شبكات التواصل الاجتماعي حسب اللهجة : الجزائرية، التونسية ،المصرية، الشامية ،الخليجية .

الكلمات المفتاحية : شبكات التواصل الاجتماعي ، الذكاء الاصطناعي ، تعلم الآلة ،اللهجات العربية ، التصنيف ، تعدين النصوص.

Résumé

Le monde d'aujourd'hui est témoin d'une série de transformations rapides dans la communication et la technologie de l'information; En particulier avec l'utilisation généralisée des sites de médias sociaux car il est un pilier fondamental de différentes transactions dans tous les domaines. L'exploration de données et l'exploration de textes sont des technologies et des domaines importants dans le monde, en particulier le réseau virtuel et la diversité de la langue arabe dans ses dialectes au Maghreb et dans le Golfe. Chaque dialecte exprime l'identité du propriétaire et ce qu'il essaie de communiquer à une autre partie a créé un énorme volume de données remplies d'informations utiles. En particulier dans le domaine du commerce électronique et du marketing. Dans notre travail, nous avons travaillé sur l'application d'algorithmes d'apprentissage automatique pour classer les textes et les commentaires des réseaux sociaux en termes de dialecte : Algérie, Tunisie, Egypte, Shamia, Golfe.

Mots-clés : réseaux sociaux, intelligence artificielle, apprentissage automatique, dialectes arabes, classification, extraction de texte.

Abstract

Today's world is witnessing a series of rapid transformations in communication and information technology; Especially with the wide use of social media sites as it is a fundamental pillar of different dealings in every field. Data exploration and text mining are important areas in the world, especially the virtual network and the diversity of Arabic language in its dialects in the Maghreb and Gulf. Each dialect expresses the identity of its owner and what he tries to communicate to another party that has created an enormous volume of data, especially useful information in the field of electronic commerce and marketing In our work, we have worked on applying machine learning algorithms to classify texts and comments of social networks according to dialect: Algeria, Tunisia, Egypt, Shamia, Gulf .

Keywords: social media networks, artificial intelligence, machine learning , Arabic dialects, classification, text mining

فهرس المحتويات

العناوين **الصفحة**

اهداء.....
شكر وعرفان.....
ملخص.....
الفهرس.....
قائمة الجداول.....
قائمة الأشكال.....
قائمة الصور.....
مقدمة عامة..... أ

الفصل الاول الذكاء الاصطناعي وتعلم الآلة

1 الذكاء الاصطناعي وتعلم الآلة..... 1

1-1 مقدمة **1**

1-2تعريف الذكاء الاصطناعي..... **1**

1-3 تاريخ الذكاء الاصطناعي..... **1**

1-4 انواع الذكاء الاصطناعي **3**

1-5 مجالات الذكاء الاصطناعي..... **4**

1-6 أهمية الذكاء الاصطناعي **5**

1-7 فوائد استخدام الذكاء الاصطناعي في مختلف القطاعات **6**

2- تعلم الآلة **8**

2-1 تعريف تعلم الالة..... 8
2-2 مكونات الرئيسية لتعلم الالة..... 8
2-3 أنواع تعلم الآلة 9
2-4التعلم الآلي الخاضع للإشراف وغير الخاضع للإشراف 10
2-5 خوارزميات تعلم الآلة 10
2-6 التدريب والتحقق والاختبار (Training, validation, test) لخوارزميات ML
..... 13
2-7 قيود التعلم الالي..... 14
2-8 علاقات التعلم الآلي بمجالات مختلفة..... 15
2-9 أهمية تعلم الآلة 17
خاتمة..... 18

الفصل الثاني الشبكات الاجتماعية وتعدين النصوص:

مقدمة 20
1-اللغة العربية ولهجاتها 21
1-1مفهوم اللغة 21
1-2تعريف اللهجة 21
1-3 تعريف العامية 21
1-4 تعريف اللغة العربية..... 21
1- 5 الفرق بين اللغة و اللهجة: 22
1-6 العلاقة بين اللغة و اللهجة..... 22
1-7أسباب نشأة اللهجات: 22
1-8 تصنيف اللهجات..... 23
1-9 مقارنة بين الأسئلة في اللهجات العربية المختلفة..... 25

2- شبكات التواصل الاجتماعي 26
2-1 مفهوم التواصل الاجتماعي 26
2-2 مفهوم شبكات التواصل الاجتماعي 26
2-3 وسائل التواصل الاجتماعي Social Media 26
2-4 مواقع التواصل الاجتماعي..... 27
2-5 أنواع التواصل على مواقع التواصل الاجتماعي :2-5-1 التواصل الشفهي:.... 28
2-6 اللغة العربية في مواقع التواصل الاجتماعي:..... 28
2-7 تأثير الذكاء الاصطناعي على وسائل التواصل الاجتماعي:..... 31
2-8 دور الذكاء الاصطناعي في وسائل التواصل الاجتماعي: 32
3- تنقيب البيانات، تعدين النصوص وتحليل المشاعر 33
3-1 تعريف البيانات الكبيرة Big Data..... 33
3-2 تعريف تنقيب البيانات Data Mining 33
3-3 تعدين النصوص Text Mining 33
3-4 تعريف تحليل المشاعر 38
3-5 دراسات في تحليل البيانات الضخمة من الشبكات الاجتماعية 41
4- دراسات مشابهة لعملنا: 43
خاتمة 44

الفصل الثالث التصميم والنهج المقترح (مجموعة البيانات DataSet والقواميس):

مقدمة 46
1- جمع البيانات والتصميم..... 47
1-1 تعريف مجموعات البيانات Data Set 47

1-2 مراحل العمل 47
1-3 مميزات مجموعة البيانات dataset..... 47
1-4 القواميس: 48
1 -5انشاء قاعدة بيانات التدريب dataset..... 48
1-6انشاءالقواميس : 52
الخاتمة..... 58

الفصل الرابع التنفيذ و تحليل النتائج

مقدمة 60
1- بايثون Python : 61
2- أين تستخدم بايثون؟ 61
3-Anaconda :..... 62
4-Spyder : 62
5- مكاتب البايثون 62
6-الوسائل المادية: 63
7- التنفيذ: 64
الخاتمة..... 76
الخاتمة العامة..... 78
قائمة المراجع : 80

قائمة الجداول :

جدول (1.2) : يوضح الفرق بين التعلم الآلي الخاضع للإشراف وغير الخاضع للإشراف.10

جدول (1.1) : جدول يوضح الاختلاف بين الاسئلة في اللهجات العربية 25

جدول (1.4) : دراسات مشابهة..... 43

جدول (1.1) : عدد التعليقات المجمعة 48

جدول (2.1) : مجموعة البيانات (Dataset) للهجة الجزائرية 49

جدول (3.1) : مجموعةالبيانات (DataSet) للهجة المصرية..... 50

جدول (4.1) : مجموعةالبيانات (DataSet) للهجة التونسية..... 50

جدول (5.1) : مجموعة البيانات (DataSet) للهجة الشامية..... 51

جدول (6.1) : مجموعة البيانات (DataSet) للهجة الخليجية..... 52

جدول (7.1) : قاموس اللهجة الجزائرية 53

جدول (8.1) : قاموس اللهجة التونسية 54

جدول (9.1) : قاموس اللهجة المصرية 55

جدول (10.1) : قاموس اللهجة الشامية 56

جدول (11.1) : قاموس اللهجة الخليجية 57

قائمة الأشكال

الشكل (1.1) : الذكاء الاصطناعي الان تورنغ..... 3
الشكل (1.2) : منحنى الانحدار اللوجستي.....11
الشكل (2.2) : منحنى دعم آلات المتجهات (SVM) 12
الشكل (1.3) : نموذجي Skip-gram وCBOW 37

المقدمة العامة

مقدمة عامة:

لقد واكب عصرنا هذا تطورا هائلا في العديد من المجالات، ولعل أبرزها المجال التكنولوجي الذي بات يشهد حداثة لا مثيل لها، خصوصا تلك التي تستعمل للتواصل والاتصال أو ما يعرف عليها بمواقع التواصل الاجتماعي، وما هي في الحقيقة إلا مجتمعات تحاكي بعضها البعض بحيث نجدها تحوي العديد من التطبيقات التي تسهل هذه العملية من بينها الانستغرام، الواتس آب التويتر، الفايسبوك الذي يعد الأكثر رواجا بين الناس نظرا لسهولة استعماله وأيضا كثرة الناشطين به من مختلف الأعمار ولقد كان لظهور شبكات التواصل الاجتماعي أثر كبير على جميع مناحي الحياة الإنسانية؛ وبرزت آثارها على الحياة الثقافية في الوطن العربي وعلى استخدام اللغة العربية خاصّة، كتابة وقراءة، حيث ظهرت في الفترة الأخيرة تغيّرات كثيرة وسريعة في استخدام اللغة العربية في وسائل التواصل الاجتماعي ما بين الفصحى والعامية، وخليط من اللهجات المحلية واللغات الأجنبية الهدف من هذا العمل هو دراسة التعليقات على مستوى الشبكات الاجتماعية. على وجه التحديد فيسبوك (Facebook) من أجل اكتشاف المجتمعات العربية ولهجاتهم المختلفة أبرزها اللهجات التالية: الجزائرية التونسية المصرية الشامية والخليجية في الفصل الأول، حاولنا إعطاء تعريف للذكاء الاصطناعي ونذكر أهميته أنواعه بعد ذلك، نقدم تعريف. للتعلم الآلي، و مختلف أنواعه والخوارزميات المتعددة. أما بالنسبة للفصل الثاني، فإننا نخصصه لدراسة الموضوعية للهجات العربية ووسائل التواصل الاجتماعي والعلاقة بينهما كما نقدم لمحة عن مفهوم تنقيب البيانات وتحليل المشاعر في الفصل الثالث نقدم هيكلة لنظام وبعض التعاريف والشروحات ، من خلال جمع البيانات (مجموعة البيانات) وإنشاء القواميس. في الفصل الرابع سنتطرق فيه إلى وصف لتنفيذ النظام وتقديم النتائج المتحصل عليها.

الفصل الاول :

الذكاء الاصطناعي وتعلم الآلة

1 الذكاء الاصطناعي وتعلم الآلة

1-1 مقدمة

لا يخفى على أي متعلم لمجال علوم الحاسوب كثرة الاهتمام بمجال الذكاء الاصطناعي وتعلم الآلة، وكذلك الأمر بالنسبة لمستخدم التقنية العادي الذي بات يرى تطوراً كبيرا في الآلات والتقنيات التي تحيط به بدءا من المساعد الصوتي الآلي في جواله وحتى سيارته وبقية الأشياء الذكية المحيطة به. تتوالى الاختراعات والاكتشافات يوما بعد يوم وتتنافس كبرى الشركات حول من يحرز أكبر تقدم ليخطف الأضواء من غيره ولم يكن ليأتي هذا التطور الذي وصلنا إليه في مجال الذكاء الاصطناعي وتعلم الآلة عبثا وإنما نتيجة جهود وأعمال كبيرة على مر التاريخ. في هذا الفصل نهدف إلى تسليط الضوء على المفاهيم الأساسية للذكاء الاصطناعي من خلال تحديد المفهوم الدقيق للذكاء الإصطناعي ومعرفة خصائصه،أهدافه وإعطاء لمحة تعلم الالة وخوارزمياتها.

1-2 تعريف الذكاء الاصطناعي

تعريف1

الذكاء الاصطناعي هو لفظ يمثل مظلة تصف مجموعة التكنولوجيات التي تسعي لأداء المهام المرتبطة بالذكاء البشري عادة. ويرجع لجون ماكرثي(John McCarthy) الفضل في صياغة مصطلح "الذكاء الاصطناعي (ArtificialIntelligence) في عام 1955 حيث فسره كعلم وهندسة عمل الآلات الذكية [1].

تعريف 2

هو ذلك الفرع من علوم الحاسوب(Science Computer) الذي يمكن بواسطته خلق وتصميم برامج الحاسبات التي تحاكي أسلوب الذكاء الإنساني لكي يتمكن الحاسب من أداء بعض المهام بدل من الإنسان والتي تتطلب التفكير والتفهم والسمع والتكلم والحركة بأسلوب منطقي ومنظم .وترجع بدايته إلى التحول من نظم البرمجة التقليدية بعد الحرب العالمية الثانية إلى استحداث برامج للحاسبات تتسم بمحاكاة الذكاء الإنساني في إجراء الألعاب ووضع الحلول لبعض الألغاز والتي أدت بدورها إلى نظم اكبر للمحاكاة ، والتي تبلورت بعد ذلك و أصبحت نظم للذكاء الاصطناعي[2].

1-3 تاريخ الذكاء الاصطناعي

إن أول عمل جوهري في مجال الذكاء الاصطناعي قام به عالم الرياضيات ورائد الحاسوب البريطاني آلان تورينج، حيث أعلن تورينج في عام 1950 أنه في يوم من الأيام سيكون هناك آلة يمكنها مضاهاة الذكاء البشري بكل طريقة وإثبات ذلك من خلال اجتياز اختبار متخصص، وفي هذا الاختبار سيتم طرح

أسئلة متطابقة عشوائية على جهاز حاسوب وإنسان مخفي عن الأنظار، وإذا نجح الحاسوب فلن يتمكن السائل من تمييز الآلة عن الشخص بالإجابات. وبحلول أوائل القرن الحادي والعشرين، لم يقترب أي برنامَج للذكاء الاصطناعي من اجتياز اختبار تورينج، ومع ذلك فقد حققت بعض البرامج مستويات أداء الخبراء البشريين في أداء بعض المهام المحددة، ويمكن العثور على الذكاء الاصطناعي بهذا المعنى المحدود في تطبيقات متنوعة مثل التشخيص الطبي وترجمة اللغات وتصميم الحاسوب والتعرف على الصوت أو الكتابة اليدوية. ونظرًا لأن الهدف النهائي للذكاء الاصطناعي هو إنشاء أجهزة حاسوب يمكنها التفكير كما يفعل البشر فقد اقترح بعض مؤيدي الذكاء الاصطناعي أنه يجب تصميم أجهزة الحاسوب على غرار الدماغ البشري والذي يتكون أساسًا من شبكة من الخلايا العصبية، وتم تطوير أول شبكة عصبية اصطناعية في عام 1954 عندما كان هدف الذكاء الاصطناعي القوي نظام يقترب من الذكاء البشري يتقاسمه الكثيرون، وفي أوائل العقد الأول من القرن الحادي والعشرين كانت الشبكات العصبية الاصطناعية قادرة على مجموعة من المهام المعقدة بما في ذلك التعرف على الوجوه والأشياء الأخرى من البيانات المرئية لكن التفاؤل بشأن تحقيق ذكاء اصطناعي قوي أفسح المجال لتقدير الصعوبات الشديدة التي ينطوي عليها الأمر. وقد أكد بعض باحثي الذكاء الاصطناعي أن الذكاء الحقيقي ينطوي ببساطة على القدرة على العمل في بيئة حقيقية، فكان هذا النهج المعروف باسم الذكاء الاصطناعي الجديد رائدًا في مختبر الذكاء الاصطناعي التابع لمعهد ماساتشوستس للتكنولوجيا بواسطة العالم الأسترالي رودني بروكس، وأحد الأمثلة الشهيرة للذكاء الاصطناعي الجديد هو الروبوت المحمول هيربيرت من بروكس والذي صُمِمَ للتجول في مساحة مكتبية وجمع علب الصودا الفارغة والتخلص منها، ومنذ أن تم الكشف عن هربرت في أواخر الثمانينيات صمم بروكس وطلابه روبوتات أخرى لتطهير حقول الألغام واستكشاف المرِّيخ بالإضافة إلى روبوت بشري اسمه Cog والذي تزيد معرفته بشكل متزايد من خلال تفاعلاته مع البيئة.[3]

لم يكن آلان تورينغ الوحيد في المضمار (مضمار تطور الذكاء الاصطناعي) وإنما انضم لمرافقته بعض العلماء الأخرين مثل: عالم الفيزيولوجيا العصبية وارن ماكولوتش وعالم الرياضيات الشاب والتر بيتس عندما نشرا ورقة علمية في عام 1943تتحدث عن كيفية عمل الخلايا العصبية ونمذجا للمرة الأولى شبكة عصبية بسيطة باستخدام دوائر كهربائية ولهما تنسب الفكرة الأساسية للخلايا العصبية الاصطناعية التي نستخدمها في أيامنا الحالية42. وفي الخمسينيات بدأ علماء الحواسيب بتطبيق هذه الفكرة في عملهم. وكان من بينهم العالم الأمريكي أرثر ساموبل عندما أنشأ برنامج يعلم نفسه بنفسه لعبة الداما (checkers)كما أنه أول من أبتكر مصطلح تعلم الآلة وكان ذلك في عام 1952.[4]

الشكل (1.1) : الذكاء الاصطناعي الان تورنغ [5] .

1-4 انواع الذكاء الاصطناعي

1-4-1 الذكاء الاصطناعي الضيق(ANI) يشمل هذا الشكل من الذكاء الاصطناعي جميع أنواع الذكاء الاصطناعي المتاحة حاليًا ، بما في ذلك أكثر أنظمة الذكاء الاصطناعي تعقيدًا وقوة حتى الآن. يشير الذكاء الاصطناعي الضيق إلى أنظمة الذكاء الاصطناعي التي يمكنها فقط تنفيذ وظيفة واحدة بشكل مستقل مع محاكاة المهارات البشرية. تمتلك هذه الآلات نطاقًا محدودًا جدًا أو ضيقًا من الكفاءات حيث يمكنها فقط أداء ما صُممت من أجله. ترتبط هذه الأنظمة بجميع أنظمة الذكاء الاصطناعي التفاعلية والذاكرة المحدودة ، وفقًا لنهج التصنيف المذكور أعلاه. تشمل ANI حتى أكثر أنظمة الذكاء الاصطناعي تقدمًا والتي تستخدم التعلم الآلي والتعلم العميق لتدريب نفسها.[6]

1-4-2 الذكاء الاصطناعي العام(AGI): من حيث الوظائف ، يرتبط الذكاء الاصطناعي العام بنظرية العقل. هذا النوع من الذكاء الاصطناعي لا يزال قيد العمل. يدرس الباحثون الذكاء الاصطناعي عن طرق عملية لدمج المشاعر الإنسانية وغيرها من المهارات إلى آلات بحيث يمكن أن تتصرف مثل البشر. الهدف هو بناء أجهزة كمبيوتر يمكنها إجراء اتصالات عبر عدة مجالات بمفردها. هذا يقترب جدا من محاكاة كيفية عمل الدماغ البشري. ونتيجة لذلك، فإن الذكاء العام الاصطناعي ينمو باستمرار.[6]

1-4-3 الذكاء الاصطناعي الفائق(ASI): سيكون إنشاء أنظمة ASI في المستقبل إنجازًا كبيرًا في أبحاث الذكاء الاصطناعي. لتنفيذ الوظائف واختيار الخيارات أفضل من البشر ، تتطلب أنظمة الذكاء الاصطناعي كميات هائلة من الذاكرة. يمكن للآلات التي تحتوي على كمية كبيرة من الذاكرة وقدرات المعالجة السريعة ومعدل أسرع لاتخاذ القرارات الذكية أن تنجز المهام الصعبة بسهولة وفي وقت أقل. سيكونون قادرين على إصدار أحكام معقدة تتأثر بمجموعة متنوعة من الظروف بطرق لم يكونوا قادرين عليها من قبل .[6].

1-5 مجالات الذكاء الاصطناعي

- إنشاء اللغات الطبيعية (Natural Language Generation NLG)

تعمل الـ (NLG) على تكوين محتوى مفيد من خلال تفسير مجموعة البيانات المتوفرة، حيث أن هذه التكنولوجيا لها القدرة على معالجة كمياتٍ هائلةٍ من البيانات في ثوانٍ قليلةٍ،وتحويلها إلى لغةٍ مكتوبةٍ يسهل على الإنسان فهمها، إذ أظهر هذا العلم نفسه من خلال قدرته على عرض التقارير المالية وأوصاف المنتجات وملفات تعريف الشبكات وخطط التسويق وغيرها خلال وقتٍ قصيرٍ جدًا[7].

- التعرف على الكلام (Speech Recognition)

أبرز مجالات الذكاء الاصطناعي التطبيقية، وهو يعني قدرة البرنامج على تحديد الكلمات والعبارات المحكيّة وتحويلها إلى نمطٍ قابلٍ للقراءة آليًّا، إذ يحتوي نظام التعرف على الكلام البدائي على مفرداتٍ وعباراتٍ محددةٍ، كما يتطلب منك التحدث بشكلٍ واضحٍ جدًّا لفهم هذا الكلام، أمّا في الأنظمة الحديثة الأكثر تطورًا فإنّها قد أصبحت تمتلك القدرة على فهم ما يُقال في حالات التكلم بصورةٍ طبيعيةٍ[7].

- التعلم الآلي (Machine Learning)

أو تعلم الآلة، وهو واحدٌ من مجالات الذكاء الاصطناعي والتي تؤمن أنظمةً لديها القدرة تلقائيًّا على التعلم والتطور من خلال تجاربها دون الحاجة إلى أن تكون مبرمجةً فعليًّا على ذلك، حيث يركز التعلم الآلي على تطوير برامج الكمبيوتر بحيث تستطيع الوصول إلى البيانات واستخدامها لتعليم أنفسها بشكلٍ آليٍّ[7].

- إدارة القرار (Decision Management)

إذ إنّ هنالك أجهزةً ذكيةً لديها القدرة على وضع مجموعةٍ من القواعد لجعل أنظمة الذكاء الاصطناعي أكثر منطقيةً، وبذلك فإنّها سوف تتمكن من استخدامها في عمليات التدريب الأولي وعمليات الصيانة المستمرة وغيرها. لقد تم فعليًّا إدخال إدارة القرار في مجموعةٍ متنوعةٍ من تطبيقات الشركات لتتمكن من اتخاذ القرار الصحيح بشكلٍ آليٍّ، ومما لاشك أنّ هذا الإجراء سيجعل العمل مريحًا للغاية.[7]

- التعلم العميق (Deep Learning)

التعلم العميق أحد أكثر مجالات الذكاء الاصطناعي شيوعًا اليوم، وهو عبارةٌ عن وظيفةٍ من وظائف الذكاء الاصطناعي تحاكي طريقة عمل العقل البشري في معالجة البيانات وإنشاء أنماطٍ يمكن استخدامها في صنع القرار، وهو فعليًّا مجموعةٌ فرعيةٌ من Machine Learning ولكن أكثر عمقًا وتعقيدًا، و يطلق عليه أيضًا التعلم العصبوني العميق[7].

- أتمتة العمليات الآلية (Robotic Processes Automation PRA)

يشير الـ PRA إلى البرامج التي يمكن برمجتها بسهولةٍ لتقوم بمجموعةٍ من المهام الأساسية من خلال تطبيقاتٍ قادرةٍ على فعلِ ما يفعله الإنسان، حيث أنّه من الممكن تلقينها آلية سير العمل عن طريق مجموعةٍ من الخطوات المتتالية، إذ أنّ الهدف من إنشاء هذا المجال هو تقليل عبء أداء المهام البسيطة ذات العمليات التكرارية على الموظفين وما سيؤديه ذلك إلى استثمار الكثير من وقتهم وجهدهم[7].

- تحليل النص (Text Analysis)

من أبرز مجالات الذكاء الاصطناعي اليوم، وهي عبارةٌ عن عمليةٍ يستطيع فيها البرنامج تحليل النصوص بطريقةٍ تمكنه من فهم معناها بشكلٍ آليٍّ، إذ إنّ الغاية من تحليل النص هو الحصول على بياناتٍ منظمةٍ ذات مغزىً، وبذلك يمكن اعتبار هذه العملية نوعًا من التشريح للمستندات غير المنظمة وتحويلها إلى بياناتٍ سهلة الإدارة وممكنة التفسير[7].

- التعرف على ردات فعل الوجه (Emotion Recognition)

وهي عبارةٌ عن تقنيةٍ موجودةٍ في برنامجٍ ما بحيث يتمكن من قراءة ردات فعل وجه الإنسان باستخدام تقنيات معالجة الصورة المتقدمة، وهو توجّهٌ حديثٌ في مجالات الذكاء الاصطناعي اليوم، وقد كما قد حاولت الشركات الجمع بين الخوارزميات المعقدة وتقنيات معالجة الصورة التي ظهرت في السنوات العشر الأخيرة للحصول على فهمٍ أعمق لمشاعر الإنسان من خلال صورة وجهه[7].

- التعرف على الصور (Image Recognition)

حازت هذه التكنولوجيا اهتمامًا واسعًا دونًا عن كلّ مجالات الذكاء الاصطناعي الأخرى، إذ يمكن من خلالها كشف وتحديد عنصرٍ أو ميزةٍ ما في الفيديوهات والصور، لتتمكن من إيجاد صورٍ أخرى ذات صلةٍ بها، إذ أنّ الذكاء الاصطناعي اليوم قد أصبح قادرًا أثناء عمليات البحث عن الصور من القيام بمجموعةٍ من الأبحاث في مواقع التواصل الإجتماعي للحصول على صورٍ ومقارنتها مع مجموعةٍ واسعةٍ من البيانات لتحديد أيّها أكثر صلةً.[7]

1-6 أهمية الذكاء الاصطناعي

يؤثر الذكاء الاصطناعي على مستقبل كل قطاع صناعي وعلى كل إنسان على هذا الكوكب كما ويعد المحرك الأساسي لجميع التقنيات الناشئة مثل جمع البيانات الضخمة والروبوتات وإنترنت الأشياء، ومن المتوقع أن يلعب الذكاء الاصطناعي دورًا أكبر خلال السنوات القادمة. وبفضل الذكاء الاصطناعي يتطور مجال الرعاية الصحية بسرعة متزايدة ويرافق ذلك زيادة كبيرة في كَمّيَّة البيانات والتحديات في ما يخص التكلفة ونتائج المرضى لذلك تم استخدام تطبيقات الذكاء الاصطناعي للحد من هذه الصعوبات،

كما ويتم استخدام الذكاء الصناعي لتفادي إجراء الفحوصات المخبرية الروتينية غير الضرورية، وتضييق دائرة التحاليل المخبرية التي قد يحتاجها المريض، ولتحسين سير العمل السريري، والتنبؤ بالأمراض المُكتسبة من المستشفيات. كما يزيد الذكاء الاصطناعي من كفاءة الأعمال وسرعة تنفيذها ويزيد من قيمتها ويساهم في تطور الأعمال باستمرار، كما يزيد من عدد المتفاعلين مع هذه الأعمال بسبب التطور المستمر للأدوات والبرمجيات المتعلقة بها. وللذكاء الاصطناعي أهمية في حياتنا اليومية فقد أحدث استخدام تطبيقات الذكاء الاصطناعي ثورة كبيرة في مجال صناعة السيارات حيث يستخدم برنامج القيادة الذاتية من جوجل الذي يستخدم تقنيات الذكاء الاصطناعي لتقليل نسبة الحوادث وتخفيف الازدحام المروري، وتستخدم تطبيقات الذكاء الاصطناعي في مواقع التجارة الإلكترونية للحصول على صورة واضحة لسلوك العملاء في عمليات الشراء عبر الموقع وتقديم التوصيات، وتستخدم أيضًا شبكات التواصل الاجتماعي تطبيقات الذكاء الاصطناعي مثل فيسبوك للكشف عن وجود اختراق لصور المستخدم.[3]

1-7 فوائد استخدام الذكاء الاصطناعي في مختلف القطاعات

1-6-1 الذكاء الاصطناعي في الرعاية الصحية

الذكاء الاصطناعي هو كلمة تشير إلى نوع من التكنولوجيا التي تساعد ممارسي الرعاية الصحية في تحديد الأمراض. يتم توظيف المساعدين الافتراضيين المدعومين بالذكاء الاصطناعي لتتبع المشكلات الصحية للمريض. يستخدم الذكاء الاصطناعي التعلم الآلي (ML) والتعلم العميق لتوفير رؤى دقيقة لبيانات المريض. يمكن للتطبيقات التي تعمل بالذكاء الاصطناعي اكتشاف المشكلات الصحية الخطيرة. أدوات الرعاية الصحية بالذكاء الاصطناعي ، على وجه الخصوص ، تكتشف السرطان في مرحلة مبكرة. يوفر الذكاء الاصطناعي في الرعاية الصحية تنبؤات دقيقة بنسبة 99٪ ، مما يسمح للأطباء بتقديم العلاج في الوقت المناسب[8].

1-7-2 الذكاء الاصطناعي في التجارة الإلكترونية

يغير استخدام الذكاء الاصطناعي في التجارة الإلكترونية طريقة تسوق الناس. يساعد الذكاء الاصطناعي المتاجر عبر الإنترنت في توفير المزيد من التخصيص والسرور للمستهلكين. في عالم التجارة الإلكترونية الذكاء الاصطناعي والتعلم الآلي في الأعمال يساعد في دراسة سلوك العميل وتفضيلاته. تعد السرعة التشغيلية ، وقوة اتخاذ القرار ، والتقديم السريع للمنتجات كلها مزايا للذكاء الاصطناعي. استخدامه في البيع بالتجزئة له أيضًا مزايا مهمة أخرى ، مثل زيادة المستهلكين والمبيعات. في عام

2019 ، تم تنفيذ الذكاء الاصطناعي من قبل ما يقرب من نصف جميع متاجر الإنترنت العالمية. علاوة على ذلك ، يخطط عدد متزايد من التجار للاستثمار في الذكاء الاصطناعي[8].

1-7-3 الذكاء الاصطناعي في البنوك والتمويل

يشمل الذكاء الاصطناعي مجموعة واسعة من التقنيات الحديثة ، من التعلم الآلي والتعلم العميق إلى الشبكات العصبية. الشركات المصرفية والمالية تتبنى الذكاء الاصطناعي بأعداد كبيرة. تعد تجربة العملاء المحسّنة من أهم مزايا الذكاء الاصطناعي في الأعمال المصرفية والمالية. يتيح الذكاء الاصطناعي والتعلم الآلي في الأعمال للفرد أن يكون لديه فهم أعمق لعملائه ، وتعزز هذه الوظيفة تجربة العميل بأكملها. توفر تقنيات الذكاء الاصطناعي (AI) تجربة دفع أكثر تخصيصًا. يمكن للمستخدمين اختيار طريقة الدفع المفضلة لديهم باستخدام تطبيقات الهاتف المحمول التي تعمل بالذكاء الاصطناعي بأمان[8].

1-7-4الذكاء الاصطناعي في المبيعات

في المبيعات ، تساعد الحلول المستندة إلى الذكاء الاصطناعي في تحديد وفرز آفاق الجودة. علاوة على ذلك ، تتعقب تقنيات الذكاء الاصطناعي الطلبات وتتواصل مع العملاء المحتملين باستمرار. نتيجة لذلك ، تساعد تقنية الذكاء الاصطناعي فريق المبيعات على تحسين معدلات التحويل وهوامش الإيرادات. فيما يلي عدد قليل من مزايا مبيعات الذكاء الاصطناعي الأساسية:

- ✓ في المبيعات قد تتعامل برامج الذكاء الاصطناعي مع كمية كبيرة من البيانات من الأجهزة السحابية وأجهزة إنترنت الأشياء.
- ✓ نظرًا لأن الذكاء الاصطناعي يعمل على أتمتة عملية المبيعات النموذجية ، فإنه يساعد شركات المبيعات في تقليل تكاليف العمالة.
- ✓ يعزز عملية توليد العملاء المتوقعين[8]

2- تعلم الآلة

2-1 تعريف تعلم الالة

تعريف1:

هو أحد فروع الذكاء الاصطناعي الذي يمكن نظم الحاسب الآلي من أداء مهام تتسم بالذكاء. هذه النظم تؤدي عمليات معقدة بواسطة التعلم من البيانات، بدلا من اتباع القواعد المبرمجة مسبقا. وقد شهدت السنوات الحديثة تقدما كبيرا في قدرات تعلم الآلة كنتيجة زيادة توافر البيانات، الخوارزميات المتقدمة، وزيادة قوة الحوسبة. ويتفاعل كثير من الناس مع النظم الممكنة والمدفوعة بتعلم الآلة كتلك المستخدمة لوضع علامات علي الوسائط الاجتماعية في نظم التعرف علي الصوت، وتلك المستخدمة من خلال المساعدين الافتراضيين الشخصيين، وفي النظم الموصي بها مثل تلك المستخدمة بواسطة تجار التجزئة على الخط [1] .

2-2 مكونات الرئيسية لتعلم الالة

لو أردنا اختصار جميع الأهداف الكامنة وراء مجال تعلم الآلة فسيكون الهدف الوحيد هو توقع النتائج معينة بناء على البيانات المدخلة (أي التعلم من البيانات المدخلة) وهذا خلاصة الأمر ، إذ يمكن تمثيل جميع مهام تعلم الآلة بهذه الطريقة. كلما زاد تنوع البيانات (تسمى في بعض الأحيان بالعينات) المجمعة لديك ، كان مهمة العثور على الأنماط ذات الصلة و التنبؤ بالنتيجة أسهل نسبيا . لذلك، فإن أي نظام يستخدم تعلم الآلة سيحتاج لثلاثة مكونات رئيسية وهي:

2-2-1 البيانات (Data)

كلما كانت البيانات أكثر تنوعا، كانت النتيجة أفضل في بعض الأحيان تكون عشرات الآلاف من سجلات البيانات هي الحد الأدنى لاستنتاج معلومة معينة وفي البعض الآخر نحتاج إلى ملايين العينات[4].

2-2-2 الميزات

ويعرف أيضا باسم المعاملات (Parameters) أو المتغيرات (Variables) و التي يمكن أن تعبر عن المسافة المقطوعة بالسيارات،أو جنس المستخدم، أو سعر السهم، أو تكرار كلمة معينة في النص بعبارة أخرى، هذه هي الميزات التي يجب أن تنظر لها الآلة[4].

2-2-3 الخوارزميات (Algorithms)

وهو الجزء الأسهل والأكثر وضوحا، إذ يمكن حل أي مشكلة بطرق مختلفة ، أن الطريقة التي تختارها ستؤثر على دقة النموذج النهائي وأدائه وحجمه. هناك فارق بسيط واحد مهم : إذا كانت البيانات سيئة فلن تساعدك حتى أفضل خوارزمية موجودة، ففي بعض الأحيان يشار إليها بمصطلح "الدخل السيئ سيؤدي إلى نتائج سيئة" ، لذلك لا تهتم كثيرا لنسبة الدقة ، و حاول الحصول على المزيد من البيانات كبداية.[4]

- **الشبكات العصبية الاصطناعية (Artificial Neural Networks) :**

وهي من أحد أشهر الطرق الشعبية في مجال تعلم الآلة، ولكن هناك طرق أخرى جيدة أيضا . [4]

- **التعلم العميق (Deep Learning) :**

هو طريقة حديثة لبناء و تدريب واستخدام الشبكات العصبية . و هي بالأساس هيكلية جديدة للشبكات العصبية. وحاليا لا أحد يفصل التعلم العميق عن "الشبكات العصبية العادية ."حتى أننا نستخدم نفس المكتبات لهم. من الأفضل دوما تسمية نوع الشبكة وتجنب استخدام الكلمات الرنانة. [4]

2-3 أنواع تعلم الآلة

2-3-1 التعلم الخاضع الإشراف

في الأنظمة القائمة على التعلم الآلي ، يعمل النموذج وفقًا لخوارزمية. في التعلم تحت الإشراف ، يتم الإشراف على النموذج. أولاً ، مطلوب تدريب النموذج. من خلال المعرفة المكتسبة ، يمكنه التنبؤ بالإجابات للحالات المستقبلية. يتم تدريب النموذج باستخدام مجموعة بيانات مصنفة. عندما يتم إعطاء بيانات خارج العينة للنظام ، يمكنه التنبؤ بالنتيجة[10].

2-3-2 التعلم غير الخاضع للإشراف

في التعلم غير الخاضع للإشراف ، لا يخضع النموذج للإشراف. يعمل النموذج من تلقاء نفسه ، للتنبؤ بالنتائج. يستخدم خوارزميات التعلم الآلي للوصول إلى استنتاجات بشأن البيانات غير المسماة. بشكل عام ، تعد خوارزميات التعلم غير الخاضعة للإشراف أصعب من خوارزميات التعلم الخاضعة للإشراف بسبب قلة المعلومات. التجميع هو نوع من التعلم غير الخاضع للإشراف. يمكن استخدامه لتجميع البيانات غير المعروفة باستخدام الخوارزميات. التجميع القائم على الكثافة والكثافة هما خوارزميتان للتجميع. بشكل عام ، النتائج الناتجة عن خوارزميات التعلم غير الخاضعة للإشراف ليست دقيقة وموثوقة كثيرًا لأن الجهاز يجب أن يحدد بيانات الإدخال وتسميتها قبل تحديد الأنماط والوظائف المخفية[10] .

2-4التعلم الآلي الخاضع للإشراف وغير الخاضع للإشراف

	التعلم الخاضع للإشراف	التعلم غير الخاضع للإشراف
الوظيفة الرئيسية	في التعلم الخاضع للإشراف ، يتنبأ النموذج بالنتيجة بناءً على بيانات الإدخال المسمى.	في التعلم غير الخاضع للإشراف ، يتنبأ النموذج بالنتيجة بدون بيانات مصنفة من خلال تحديد الأنماط من تلقاء نفسه.
دقة النتائج	النتائج الناتجة عن طرق التعلم الخاضعة للإشراف أكثر دقة وموثوقية.	النتائج الناتجة عن طرق التعلم غير الخاضعة للإشراف ليست دقيقة وموثوقة كثيرًا.
الخوارزميات الرئيسية	خوارزميات للانحدار والتصنيف في التعلم تحت الإشراف.	خوارزميات للتجميع في التعلم غير الخاضع للإشراف.

جدول (1.2) : يوضح الفرق بين التعلم الآلي الخاضع للإشراف وغير الخاضع للإشراف [10]

2-5 خوارزميات تعلم الآلة

2-5-1 الانحدار الخطي Linear regression

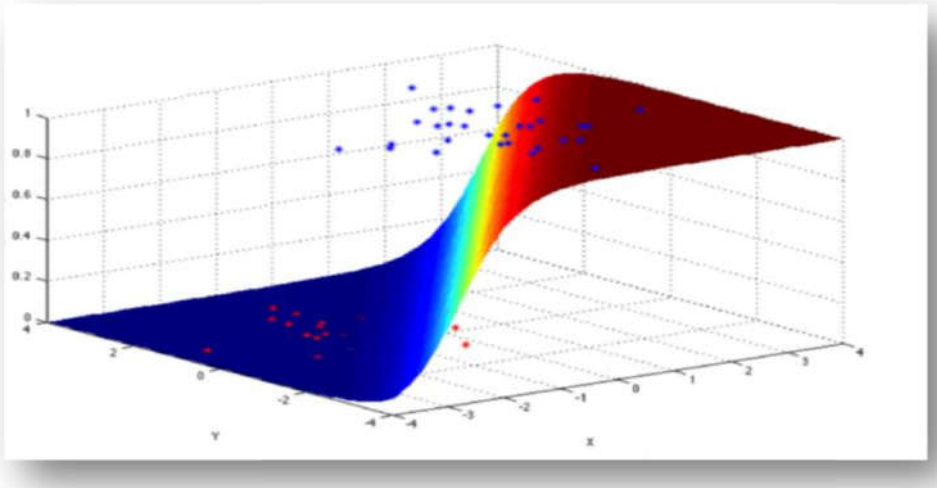
ربما تكون خوارزمية الانحدار الخطي أحد أكثر الخوارزميات شهرة وفهمًا في الإحصاء والتعلم الآلي. الانحدار الخطي هو نموذج خطي، على سبيل المثال، نموذج يفترض وجود علاقة خطية بين متغيرات الإدخال (x) ومتغير إخراج واحد (y) مما يعني أنه يمكن حساب (y) بسهولة باستخدام العلاقة الخطية، فعندما يكون هناك متغير إدخال واحد (x)، يشار إلى المعادلة باسم الانحدار الخطي البسيط، وعندما يكون هناك أكثر من متغير إدخال، يشار إلى المعادلة على أنها انحدار خطي متعدد[11].

المعادلة : "y = B0 + B1 * x" (حيث أن "x" متغير إدخال، "y" هو متغير الإخراج و "B0 و B1" معاملات)

2-5-2 الانحدار اللوجستي Logistic Regression

الانحدار اللوجستي هو خوارزمية تصنيف تعتمد على الدالة المستخدمة في الوظيفة اللوجستية أو الوظيفة السينية، وتكون منحنى على شكل حرف (S) يُستخدم للتنبؤ بنتيجة ثنائية مثل (0/1) أو (نعم/لا)، بالنظر إلى مجموعة من المتغيرات المستقلة. ويمكن اعتبارها أيضًا حالة خاصة من الانحدار الخطي عندما يكون متغير الإخراج محدد، حيث نستخدم سجل الاحتمالات كمتغير تابع. يتنبأ باحتمالية حدوث الحدث عن طريق ملاءمة البيانات لوظيفة تسجيل الدخول[11].

المعادلة ("p(X) = e^(b0 + b1*X) / (1 + e^(b0 + b1*X"))



الشكل (1.2) : منحنى الانحدار اللوجستي [12]

2-5-3"K" أقرب الجيران K-Nearest Neighbours

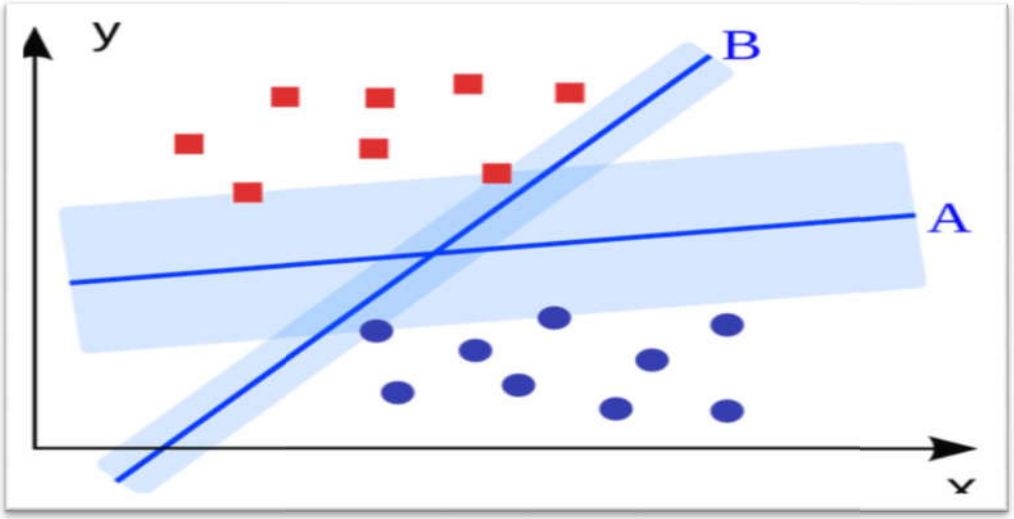
خوارزمية "K" الأقرب للجيران أو (KNN) للاختصار، هي خوارزمية تعلم آلي خاضعة للإشراف يمكن استخدامها لحل مشاكل التصنيف والانحدار. تعمل الخوارزمية من خلال إيجاد المسافات بين نقطة البيانات الجديدة المضافة والنقاط الموجودة بالفعل في الفئتين المنفصلتين[11].

المعادلة: "EuclideanDistance(x, xi) = sqrt(sum((xj — xij)²))"

2-5-4 دعم آلات المتجهات (SVM)

وهي خوارزمية تعلم آلي خاضعة للإشراف يمكن استخدامها أيضًا لكلتا المهمتين التصنيف والانحدار. مع ذلك، فإنها تُستخدم في الغالب في مشاكل التصنيف. تقوم الخوارزمية بإيجاد الحدّ بين فئتي البيانات اللذين يتم تكبيرها، وهناك مستويات أخرى أيضًا يمكنها فصل الفئتين، لكن الـ (hyperplane) في خوارزمية دعم آلات المتجهات يمكنها زيادة الهامش بين الفئات فقط[11].

المعادلة: "B0 + (B1 * X1) + (B2 * X2) = 0" حيث يحدد "B1 و B2" ميل الخط و "B0" (التقاطع) الذي وجدته خوارزمية التعلم. "X1 و X2" هما متغيرا الإدخال.



الشكل (2.2) : منحنى دعم آلات المتجهات (SVM) [12]

2-5-5 أشجار القرار (Decision Trees)

يشار إلى خوارزميات شجرة القرار باسم (CART) أو أشجار التصنيف والانحدار، وهو مخطط انسيابي مثل هيكل الشجرة، حيث تشير كل عقدة داخلية إلى اختبار السمة، ويمثل كل فرع نتيجة للاختبار، وكل عقدة طرفية (عقدة نهائية) تحمل تسمية فئة. تعطي "درجة جيني" فكرة عن مدى جودة التقسيم من خلال مدى اختلاط فئات الاستجابة في المجموعات التي تم إنشاؤها بواسطة التقسيم[11].

2-5-6 الغابة العشوائية (Random Forest)

الغابات العشوائية هي أسلوب تعلم جماعي يُبني من أشجار القرار، حيث تتضمن خوارزمية الغابات العشوائية إنشاء أشجار قرارات متعددة باستخدام مجموعات بيانات تمهيدية من البيانات الأصلية واختيار عشوائي لمجموعة فرعية من المتغيرات في كل خطوة من خطوات شجرة القرار، ثم يختار النموذج وضع جميع تنبؤات كل شجرة قرار[11].

2-5-7 خوارزمية التعزيز التكيفي (AdaBoost)

التعزيز التكيفي هي أيضًا خوارزمية تستفيد من طرق التعبئة والتعزيز لتطوير مؤشر المُحسَّن. تقوم الخوارزمية بإنشاء غابة من جذوع الأشجار بدلاً من الأشجار (الجذع هو شجرة مكونة من عقدة واحدة وعقدتين طرفيتين فقط)، تتخذ الخوارزمية نهجًا أكثر تكرارية بمعنى أنها تسعى إلى التحسين المتكرر من الأخطاء التي ارتكبها الجذع (الجذور) السابق[11].

2-5-8 دفعة التدرج Gradient Boost

تعد خوارزمية دفعة التدرج أيضًا خوارزمية مجموعة تستخدم طرق التعزيز لتطوير متنبئ محسّن. على عكس خوارزمية التعزيز التكيفي الذي يبني جذوعًا، تقوم الخوارزمية ببناء أشجارًا عادةً تتراوح ما بين (8 و 32) عقد طرفية، ترى الخوارزمية أن مشكلة التعزيز هي مشكلة تحسين، حيث تستخدم دالة الخسارة (وهي الدالة التي تحسب المسافة بين المخرجات الحالية للخوارزمية والمخرجات المتوقعة) وتحاول تقليل الخطأ، لهذا يطلق عليها تعزيز التدرج؛ لأنها مستوحى من نزول المنحدر[11].

2-5-9 خوارزمية Light GBM

إنه نوع آخر من خوارزمية التعزيز التي أظهرت أنها أسرع وأحيانًا أكثر دقة من "XGBoost". تستخدم هذه الخوارزمية تقنية فريدة تسمى أخذ العينات من جانب واحد على أساس التدرج (GOSS) لتصفية مثيلات البيانات للعثور على قيمة مقسمة[11].

2-5-10. خوارزمية Naive Bayes

وهي خوارزمية تصنيف تستخدم لمشاكل التصنيف الثنائية (من فئتين) ومتعددة الفئات. يتم استخدامها عندما يكون متغير المخرجات منفصلًا. وكما يوحي الاسم، تعتمد هذه الخوارزمية بالكامل على نظرية "Bayes"، حيث تقول النظرية أنه يمكننا حساب احتمال جزء من البيانات ينتمي إلى فئة معينة إذا تم إعطاء معرفة مسبقة[11].

الصيغة البايزية :

$$P(A/B) = \frac{P(B/A)\ P(A)}{P(B)}$$

2-6 التدريب والتحقق والاختبار (Training, validation, test) لخوارزميات ML

يتيح التعلم الآلي للشركات تحويل الكثير من البيانات إلى تنبؤات يمكن أن تساعد الأعمال. توفر خوارزميات التعلم الآلي التنبؤية هذه الكثير من إمكانات الربح. ومع ذلك، تتطلب خوارزميات التعلم الآلي الفعالة (ML) جودة التدريب واختبار البيانات – وغالبًا الكثير منها – لإجراء تنبؤات دقيقة. تخدم

مجموعات البيانات المختلفة أغراضًا مختلفة في إعداد خوارزمية لاتخاذ التنبؤات والقرارات بناءً على بيانات العالم الحقيقي.

- **بيانات التدريب(training data)**

هذا النوع من البيانات يبني خوارزمية التعلم الآلي. يغذي عالم البيانات بيانات مدخلات الخوارزمية، والتي تتوافق مع الناتج المتوقع. يقوم النموذج بتقييم البيانات بشكل متكرر لمعرفة المزيد عن سلوك البيانات ثم يعدل نفسه لخدمة الغرض المقصود منه[13].

- **بيانات التحقق(validation data)**

أثناء التدريب، تضخ بيانات التحقق بيانات جديدة في النموذج لم يتم تقييمها من قبل. توفر بيانات التحقق الاختبار الأول مقابل البيانات غير المرئية، مما يسمح لعلماء البيانات بتقييم مدى جودة عمل النموذج للتنبؤات بناءً على البيانات الجديدة. لا يستخدم جميع علماء البيانات بيانات التحقق من الصحة، ولكن يمكنهم توفير بعض المعلومات المفيدة لتحسين مقاييس فرط البارامامتر، مما يؤثر على كيفية تقييم النموذج للبيانات[13].

- **بيانات الاختبار(testing data)**

بعد بناء النموذج، تؤكد بيانات الاختبار مرة أخرى أنه يمكنه إجراء تنبؤات دقيقة. إذا كانت بيانات التدريب والتحقق تتضمن ملصقات لرصد مقاييس الأداء للنموذج، فيجب عدم تسمية بيانات الاختبار. توفر بيانات الاختبار فحصًا نهائيًا في العالم الحقيقي لمجموعة بيانات غير مرئية لتأكيد أن خوارزمية ML قد تم تدريبها بشكل فعال[13].

2-7 قيود التعلم الالي

على الرغم من أن التعلم الآلي قد تحول في بعض المجالات ، إلا أن برامج التعلم الآلي غالباً ما تفشل في تحقيق النتائج المتوقعة .أسباب ذلك عديدة: نقص البيانات (المناسبة) ، عدم الوصول إلى البيانات ، تحيز البيانات ، مشاكل الخصوصية ، المهام والخوارزميات التي تم اختيارها بشكل سيئ ، الأدوات الخاطئة والأشخاص ، نقص الموارد ، ومشاكل التقييم[14] .

2-8 علاقات التعلم الآلي بمجالات مختلفة

2-8-1 العلاقة مع الذكاء الاصطناعي

كان بعض الباحثين مهتمين بتعلم الآلات من البيانات. حاولوا معالجة المشكلة بأساليب رمزية مختلفة ، وكذلك ما أطلق عليه بعد ذلك "الشبكة العصبية" ؛ كانت هذه في الغالب (perceptrons) و نماذج أخرى (ADALINE) التي تم اكتشافها لاحقًا على أنها اختراعات لـ نموذج خطي معمم للإحصاءات . تم استخدام المنطق الاحتمالي أيضاً ، خاصة في التشخيص الطبي الآلي. ومع ذلك ، أدى التركيز المتزايد على النهج المنطقي القائم على المعرفة إلى حدوث شقاق بين الذكاء الاصطناعي والتعلم الآلي. تعاني الأنظمة الاحتمالية من المشكلات النظرية و العملية للحصول على البيانات وتمثيلها.بحلول عام 1980 ، أصبحت النظم الخبيرة تهيمن على الذكاء الاصطناعي ، وكانت الإحصاءات غير مواتية.استمر العمل في التعلم الرمزي القائم على المعرفة داخل AI ، مما أدى إلى البرمجة المنطقية الاستقرائية ، ولكن الخط الإحصائي للبحث أصبح الآن خارج مجال الذكاء الاصطناعي الصحيح ، في التعرف على الأنماط و استرجاع المعلومات. تم التخلي عن أبحاث الشبكات العصبية بواسطة الذكاء الاصطناعي و علوم الكمبيوتر في نفس الوقت تقريبًا. هذا الخط ، أيضًا ، استمر خارج مجال الذكاء الاصطناعي علوم الكمبيوتر ، باسم "اتصالية" ، من قبل باحثين من تخصصات أخرى بما في ذلك هوبفلد روملهارت و هنتون. جاء نجاحهم الرئيسي في منتصف الثمانينيات مع إعادة اختراع الانتشار العكسي. بدأ التعلم الآلي ، الذي أعيد تنظيمه كحقل منفصل ، في الازدهار في التسعينيات. غيّر المجال هدفه من تحقيق الذكاء الاصطناعي إلى معالجة المشاكل القابلة للحل ذات الطبيعة العملية. حولت التركيز بعيدًا عن المقاربات الرمزية التي ورثتها عن الذكاء الاصطناعي ، ونحو الأساليب والنماذج المستعارة من الإحصاءات و نظرية الاحتمالات.اعتبارًا من عام 2019 ، تواصل العديد من المصادر التأكيد على أن التعلم الآلي لا يزال مجالًا فرعيًا للذكاء الاصطناعي. ومع ذلك ، فإن بعض الممارسين ، على سبيل المثال الدكتور دانييل هولم ، الذي يقوم بتدريس الذكاء الاصطناعي ويدير شركة تعمل في هذا المجال ، و الذي يجادل بأن التعلم الآلي والذكاء الاصطناعي منفصلان[14].

2-8-2 علاقة بالاستكشاف والبحث عن البيانات

غالبًا ما يستخدم التعلم الآلي و استخراج البيانات نفس الأساليب و يتداخلان بشكل كبير ، ولكن بينما يركز التعليم الآلي على التنبؤ ، استنادًا إلى الخصائص "المعروفة" والتي تم الاستفادة منها من بيانات التدريب يركز استخراج البيانات على الاكتشاف للخصائص (غير المعروفة سابقًا) في البيانات (هذه هي خطوة التحليل لـ اكتشاف المعرفة في قواعد البيانات). تستخدم عملية اكتشاف و التنقيب عن البيانات العديد من أساليب التعلم الآلي ، لكن مع أهداف مختلفة ؛ من ناحية أخرى ، يستخدم التعلم الآلي أيضًا أساليب استخراج البيانات كـ "تعلم غير خاضع للإشراف" أو كخطوة قبل المعالجة لتحسين دقة المتعلم. الكثير من الالتباس بين هاتين المجموعتين من البحوث (التي غالباً ما يكون لها مؤتمرات منفصلة ومجلات منفصلة ، ECML PKDD كونها استثناءً رئيسيًا) يأتي من الافتراضات الأساسية التي تعمل بها: في التعلم الآلي ، يتم تقييم الأداء عادةً بالأخذ بعين الاعتبار "إعادة إنتاج إدراكاً أو علماً معروفاً" ، بينما في اكتشاف المعرفة واستخراج البيانات (KDD) المهمة الأساسية هي اكتشاف علماً أو إدراكاً "غير معروفاً" سابقًا. عند تقييمها فيما يتعلق بالعلم المعروف، سيتم بسهولة تفوق طريقة (غير خاضعة للإشراف أو غير مراقبة) و غير معروفة بطرق أخرى خاضعة للإشراف ، بينما في مهمة KDD النموذجية، لا يمكن استخدام الطرق الخاضعة للإشراف بسبب عدم توفر بيانات التدريب [14].

2-8-3 العلاقة بعملية التحسين

يرتبط التعليم الآلي أيضًا بعلاقة قوية بـ التحسين: حيث تم صياغة العديد من مشكلات التعلم كتقليل لبعض دالة الخسارة و الضياع على مجموعة من الأمثلة التدريبية. تعبر وظائف أو توابع الخسارة عن التناقضات بين تنبؤات النموذج الذي يتم تدريبه وحالات المشكلة الفعلية (على سبيل المثال ، في التصنيف ، يريد المرء تعيين تسمية للحالات ، ويتم تدريب النماذج على التنبؤ بشكل صحيح بالتسميات المعينة مسبقًا لمجموعة من أمثلة). ينشأ الاختلاف بين الحقلين من هدف التعميم: في حين أن خوارزميات التحسين يمكن أن تقلل من الخسارة في مجموعة التدريب ، بينما يهتم التعلم الآلي بتقليل الخسارة في العينات غير المرئية[14].

2-9 أهمية تعلم الآلة

تمكنّ فرع تعلم الآلة من اكتساب أهمية بالغة في العصر الحديث، وتحديداً بعد التطورات الهائلة التي طرأت على الذكاء الاصطناعي وخوارزمياته، وتتمثل أهميته فيما يلي:

- ✓ توفير كم أكبر من البيانات الضرورية لاتخاذ القرار.
- ✓ القدرة على تخزين البيانات بأكبر قدر ممكن.
- ✓ معالجة البيانات حاسوبياً يعتبر أقل تكلفة مادية من توظيف الأيدي البشرية.
- ✓ تحليل أكبر قدر من البيانات بمختلف مستوياتها سواءً أكانت بسيطة أو معقدة.
- ✓ ضمان الحصول على نتائج وقرارات أكثر دقةً وبأسرع وقت.
- ✓ تمكين المنظمات ومختلف المنشآت من رصد الفرص الأنسب لتحقيق الأرباح وتفاديها للأخطار المجهولة.
- ✓ المساعدة في اختيار القرار الأمثل من بين مجموعة من البدائل المتاحة[11] .

خاتمة

في هذا الفصل، درسنا و قدمنا تعريف شاملاً للذكاء الاصطناعي وكيفية عمله، و حددنا التعلم الآلي وذكرنا أنواعه بما في ذلك التعلم تحت الإشراف والتعلم بدون إشراف واهم خوارزمياته التي يستخدمها.

الفصل الثاني:

الشبكات الاجتماعية وتعدين النصوص

مقدمة

تحتل اللغة العربية الدرجة السابعة في قائمة اللغة الأكثر انتشار على الانترنت، ويتزايد عدد مستخدمي الانترنت باستمرار. حوالي 46 مليون شخص يبحثون في الانترنت باللغة العربية فقط. مع وجود 280 مليون شخص يتكلمون العربية في 28 بلدا، فإن إهمال هذه اللغة يعني فرص اقل لغزو السوق التجاري. يعتبر الشرق الأوسط الأكثر تطورا في العالم، مع وجود حجم تعامل تجاري كبير لديه. لذا الاستثمار في اللغة العربية الآن لتكسل حدا في مجال المنافسة، بينما أقل من 2% من معلومات شبكة الانترنت مكتوبة باللغة العربية. اللغة العربية هي لغة سامية وتتكون من العديد من اللهجات الإقليمية المختلفة. على الرغم من أن هذه اللهجات هي شكل من أشكال اللغة الأم الحقيقية ، يتم استخدامها بشكل عام فقط في الاتصالات اليومية غير الرسمية . إن احد نتائج استخدام وسائل التواصل الاجتماعي على نطاق واسع وفر قدر كبير جدا من البيانات فمع وجود مئات الملايين من الناس يقضون ساعات لا تحصى على هذه الوسائل ليتشارك ويترابطوا ويتواصل ويتفاعل بعده فان ذلك اوجد بيانات يولدها المستخدمون بمعدلات الغير مسبقة إلى الحد الذي يمكن أن نقوله معه إننا نعيش عصر البيانات .توفر قدرا ضخما من البيانات تعد مصدرا ثريا يتيح فرصا لا مثيل لها وإمكانات كبيره للبحث والتطوير. ومن هذا المنطلق نشا نوع جديد من أنواع الأعمال وهو تنقيب بيانات في وسائل التواصل الاجتماعي ويمثل تحليل المشاعر احد أنواع تنقيب البيانات الذي يستخدم المجموعة من التقنيات لتحديد المشاعر اتجاه موضوعات معينه يتم التعبير عنها داخل وسائل التواصل الاجتماعي وذلك في بهدف التوصل إلى الشعور العام تجاه هذه الموضوعات. وقد برز تحليل المشاعر أو التنقيب عن الرأي كمجال يدرس النصوص ويعد من أكثر المجالات البحثية نشاطا في الوقت الراهن لكن يبقى إشكال في كيفية تحليل هذه النصوص وتقييمها، وبما أن تحليلها يدويا يستغرق وقتا طويلا إن لم نقل مستحيلا فقد حاول الباحثون إيجاد حلول تساعد في تحليلها آليا حيث ظهرت أنظمة المستندة على القواعد وأنظمة المستندة على التعلم آلي. في هذا الفصل سنتطرق بشكل عام حول اللغة العربية ولهجاتها المختلفة وواقعها في الشبكات الاجتماعية كما نعطي فكرة حول تنقيب البيانات والتعلم الآلي في تصنيف النصوص وتحليلها وتقييمها.

1-اللغة العربية ولهجاتها

1-1مفهوم اللغة

اللغة من وجهة نظر أنيس فريحة هي: " ظاهرة سيكولوجية اجتماعية ثقافية مكتسبة، لا بيولوجية، ملازمة للفرد، تتألف من مجموعة رموز صوتية لغوية، اكتسبت عن طريق أخبار معاني مقررة في الذهن، وبهذا النظام الرمزي الصوتي تستطيع جماعة ما أن تتفاهم صفة وتتفاعل، وباللغة فقط صار الإنسان إنسانا". اللغة هي نظام اتصال منظم يتضمن مجموعات معقدة من المكونات، مثل الحروف والكلمات والجمل وما إلى ذلك. علم اللغة هو الدراسة المنهجية للغة[15].

1-2تعريف اللهجة

اللهجة في الاصطلاح العلمي الحديث هي: مجموعة من الصفات اللغوية تنتمي إلى بيئة خاصة، تشترك في هذه الصفات جميع أفراد هذه البيئة، وبيئة اللهجة هي جزء من بيئة أوسع وأشمل تضم عدة لهجات لكل منها خصائصها، ولكنها تشترك جميعها في مجموعة من الظواهر اللغوية التي تسير اتصال أفراد البيئة بعضهم ببعض، وفهم ما قد يدور بينهم من حديث فهما يتوقف على قدر الرابطة التي تربط بين هذه اللهجات[15].

1-3 تعريف العامية

تستخدم في الشؤون العادية، والتي يجرى بها الحديث اليومي ويتخذ مصطلح العامية أسماء عدة عند بعض اللغويين المحدثين: كاللغة العامية، والشكل اللغوي الدارج، واللهجة الشائعة، واللهجة العربية المحكية، واللهجة الدارجة، لغة الشعب والعامية لا تخضع لقوانين تضبطها وتحكم عباراتها؛ لأنها تلقائية متغيرة تتغير تبعا بتغير الأجيال، وتغير الظروف المحيطة بها[15].

1-4 تعريف اللغة العربية

تعد اللغة العربية من اللغات السامية يتحدث بها سكان المناطق في شمال أفريقيا وشبه الجزيرة العربية والشرق الأوسط، ومن المنظور الديني فإن هذه اللغة تعد في الإسلام لغة القرآن واللغة المعتمدة في الدين الإسلامي بصيغتها الفصحى الثابتة وغير المتغيرة في مختلف بقاع العالم العربي على عكس النمط العامي منها، والذي يتضمن العديد من اللهجات التي تبدو في بعض الأحيان غير مفهومة للأطراف الأخرى من العرب أنفسهم، حيث تنحصر تلك اللهجات في مجموعات لهجية رئيسية مثل المصرية، والعراقية، والمغربية، والخليجية، والسورية، كلها نشأت تحت تأثير اللغة الفصحى باستثناء تلك اللهجة التي يتكلم بها الجزائريون[16].

1- 5 الفرق بين اللغة و اللهجة:

على ضوء علم اللغة لا فرق بين لغة و لهجة، فكل لهجة هي لغة قائمة بذاتها بنظامها الصوتي، و بصرفها، و بنحوها، و بتركيبها، و بمقدرتها على التعبير، و يرى البعض أن اللغة هي التى تغاير لغة أخرى بأصواتها و بمفرداتها، وبتراكيبها مغايرة لا يستطيع معها أن يتفاهم زيد، و عمرو. أما إذا كانت الفروق في الأصوات، و المفردات، و التركيب من النوع الذي يمكن فيه التفاهم بين الجماعات، فإن هذه تُحسب لهجات، و هذا الرأي يجعل التفاهم مقياساً للفرق بين لهجة و لغة[17].

1-6 العلاقة بين اللغة و اللهجة

هي العلاقة بين العام و الخاص، فاللغة تشمل عادة على عدة لهجات، لكل منها ما يميزها، و جميع هذه اللهجات تشترك في مجموعةٍ من الصفات اللغوية، و العادات الكلامية التي تؤلف لغةً مستقلة عن غيرها من اللغات فليس في العالم لغة غير متشعبة إلى هذا المستوى من العامية و الفصحى، و اللغة العربية كغيرها من اللغات تعرضت عبر الزمن إلى عوامل أدت إلى انشعابها إلى عدة لهجات، و ظاهرة العامية ليست ظاهرة تنفرد بها اللغة العربية بل عرفتها كل اللغات–تقريباً–وهي ليست أكثر من لهجات الفصحى المتشعبة منها[17].

1-7أسباب نشأة اللهجات:

لنشأة اللهجات أسباب قررها علماء اللغة المحدثون، وفصلوا القول فيها، وليس لنا أن نعدها مجتمعة سببا في تكون اللهجات، وأنها تسهم متساوية في ذلك التكون؛ إذ إن اللهجات تتفاوت فيما بينها في أثر هذا السبب، أو ذاك في تكونها، وأهم الأسباب التي تسهم في نشأة اللهجات هي:

1-7-1اختلاف البيئات الجغرافية:

فالأرض التي يعيش عليها البشر مختلفة، فيها الجبال، والسهول والوديان، فيها الأراضي الزراعية القاحلة، ومتى اختلفت البيئة الجغرافية فإن ذلك يؤدي إلى اختلاف اللغة، فإذا انتشرت جماعة لغوية تعيش في مكان معين على أرض واسعة تختلف طبيعتها فإن ذلك يؤدي مع تطاول الزمن إلى استيعاب اللغة الواحدة إلى لهجات[15].

1-7-2عوامل اجتماعية سياسية:

تتعلق باستقلال المناطق التي انتشرت فيها اللغة بعضها عن بعض، وضعف السلطان المركزي الذي كان يجمعها، ويوفق ما بينها من علاقات؛ وذلك أن اتساع الدولة وكثرة المناطق التابعة لها، واختلاف الشعوب الخاضعة لنفوذها يؤدي غالبا إلى ضعف سلطانها، وتفككها من الناحية السياسية، وانقسامها إلى دويلات، وهذا الانقسام السياسي يؤدي إلى انقسام الوحدة الفكرية[15] .

1-7-3 الاحتكاك اللغوي:

ويحدث ذلك نتيجة التجاور مع اللغات الأخرى، أو نتيجة الغزو والاستعمار والهجرات ، ويعّد هذا العامل من العوامل الرئيسية التي تسهم في تكون اللهجات، وهو ما حدث في العصر الحديث ؛ إذ أن الاستعمار ساعد على تعميق الاختلافات بين اللهجات العربية المعاصرة، والى تطور مستقل في تلك البيئات التي استحوذ عليها، ويلاحظ أن تأثير الطبقات التحتية في مصر والشام والمغرب كان واضحا في اللهجات[15].

1-8 تصنيف اللهجات

الفرق الأكبر بين اللهجات العربية يكون بين لهجات البدو ولهجات أهل القرى والمدن وثم بين لهجات اهل الحضر في المشرق ولهجات اهل الحضر في المغرب،. تختلف لهجات العربية العامية كثيراً الآن في المفردات وفي الأصوات والنحو والصرف (في اللهجات الدارجة وليس في أصل اللغة الفصحى)؛ فمثلاً، في لهجات الشام العامية يبدأ الفعل المضارع بالسابقة "ب"، والنفي يكون باستعمال "ما" (أنا ما بعرف، أنت ما بتعرف، ما درست، ما لعبت، إلخ.)، أما دول الشمال الأفريقي فتظهر اللاحقة "ش" وتكون اللهجتين المصرية الأردنية وسط بين الطرفين إذ تستخدم السابقة "ب" وتنفي باستخدام اللاحقة "ش" (بعرفش، مرحتش، معملتش، مكلتش). من الظواهر الهامة: تشابه لغات العرب البدو في المشرق والمغرب بعيدا عن لهجات الحضر في كلتا المنطقتين، ويدخل في ذلك أيضا تشابه أغراض اللغة وفنونها ويبرز ذلك في الشعر الشعبي عند البدو في كلتا المنطقتين وتشابه تراثهم البدوي العام. اللهجات العربية لا يزال الفهم سهلا ممكنا بين معظمها لتشابه المفردات في الأغلب. أدى الإنتاج التلفزيوني المصري والسوري واللبناني- إلى انتشار لهجات تلك الدول وإلى حد ما أصبحت تلك اللهجات مفهومة لدى غالبية الجيل العربي الحديث. تختلف اللهجات في نطق القاف كثيرا، فينطق "ق"، أو "گ" عند البدو في ليبيا والخليج، أو همزة في مصر وسوريا، أو "ك" في أرياف فلسطين خاصة بمنطقة المثلث وأرياف الضفة الغربية، بالإضافة إلى اختلافات طفيفة في لفظ الضاد[18] .

1-8-1 اللهجات واصنافها

- **اللهجة المصرية:**

اللهجة العربية المصرية، المعروفة محليًا باسم العامية المصرية أو المصري، يتحدث بها معظم المصريين المعاصرين. المصرية هي لهجة شمال أفريقية للغة العربية وهي فرع سامي من عائلة اللغات الأفروآسيوية .نشأت في دلتا النيل في مصر السفلى حول العاصمة القاهرة[19]

- **اللهجات الشامية**

✓ **لبنانية:** تقسم إلى لهجات بيروتية (حسب الأحياء مثل لهجة الأشرفية، لهجة البسطة، الخ.)، لهجات شمالية (لهجة طرابلسية، لهجة زغرتاوية، لهجة بشراوية، لهجة الكورة، لهجة عكارية وغيرها)، لهجة جنوبية (صور، بنت جبيل، الخ)، لهجة بقاعية (لهجة زحلية، بعلبك–الهرمل،)، لهجة أهل جبل لبنان (اللهجة الكسروانية واللهجة الدرزية الخ)

✓ **سورية:** تقسم إلى :لهجة دمشقية، لهجة حمصية، لهجة حلبية، لهجة ساحلية، لهجة ماردلية ولهجة ديرية ومارونية قبرصية و لهجة يبرودية.

✓ **فلسطينية:** تقسم إلى :لهجة نابلسية، لهجة خليلية، لهجة مقدسية وهي تشبه أو تضم أيضاً لهجة بيت لحم ورام الله .لهجة الشمال: والتي تضم كل مناطق الجليل وحيفا، وتتشابه لهجة منطقة الجليل مع لهجة أهل الناصرة .ويوجد أيضاً لهجة مدن المركز يافا واللد والرملة .وبشكل عام أغلب مناطق الداخل الفلسطيني يتحدث أهلها باللهجة الفلسطينية المدنية في ما عدا منطقة المثلث الذين يتحدثون باللهجة الريفية ذات الطابع الخاص. ويوجد أيضاً لهجة غزاوية، ولهجة بدوية يتحدث بها أهل النقب.

✓ **أردنية:** تقسم إلى لهجات شمالية في مدن (إربد، عجلون، جرش، الخ)، لهجات جنوبية في مدن (مادبا، الكرك، الطفيلة، معان، الخ)،لهجات مدن الوسط(عمان، السلط، الخ) ولهجة العقبة، ويوجد أيضاً اللهجة البدوية في بعض المناطق وفي شرق الطريق الصحراوي.[18]

- **اللهجة الخليجية**

اللهجة الخليجية هي لهجة مُعيَّنة يتحدث بها سكان ساحلي الخليج العربي، ويبلغ عدد مُتحدِّثيها حوالي السبعة ملايين نسمة. تتفرع اللهجات الخليجية من عائلة لهجات شبه الجزيرة العربية، إلا أن مصطلح «خليجي» و«لهجة خليجية» في الاستخدام الشائع قد يشير إلى اللهجات التي يُتحدث بها في دول مجلس التعاون عمومًا.[19]

- **لهجة المغرب العربي:**

✓ **اللهجة الجزائرية** :اللهجة الجزائريّة أو اللهجة العربية الجزائرية هي لغة التواصل المشترك الأساسية في الجزائر، اللغة الأم لـ 75–80٪ من السكان وتستعمل من طرف 95–100٪ من سكان الجزائر .ويسميها المتكلمون بها الدَّارْجة أو العَرْبِية[19] .

✓ **اللهجة الليبية**

✓ **اللهجة التونسية**

✓ **اللهجة المغربية**

1-9 مقارنة بين الأسئلة في اللهجات العربية المختلفة

العربية الفصحى	السعودية النجدية	ليبيا	المغرب	الجزائر	مصر	الاردن و فلسطين	الشامية
لماذا	ليش/شوله	وعلاش/عليش	علاش/لاش	علاش	ليه	ليش/لويش	ليش/
كيف	كيف/وشلون	كيف	كيفاش	كيفاش	ازاي	كيف/شلون	كيف/ شلون
متى	متى	امته	وقتاش/ فوقاش/ ايمتى	وقتاش	امته	وينتا/ ايمتى/ متى	ايمتى/ اميت
ماذا	وشهو/وشو/ايش	شني/شنو	اش/شنو	وشنو/ شوالا/ وشي/شتاهو	ايه	شو/ ايش	شو/ ايش
هل	هل/يعني	هل	واش/وشنو	واش/وشنو	الضمير +[+]		شي
كم	كم	قداش /ابكم	شحال/مشحال اش حال	شحال/ قداش	كام	قديش/ كم	قديش

جدول (1.1): جدول يوضح الاختلاف بين الاسئلة في اللهجات العربية [18]

2- شبكات التواصل الاجتماعي

2-1 مفهوم التواصل الاجتماعي

مفهوم التواصل الاجتماعي هو نقل الأفكار والتجارب وتبادل الخبرات والمعارف بين الذوات والأفراد والجامعات بتفاعل ايجابي وبواسطة رسائل تتم بين مرسل ومتلقي كل حسب قدرته وطاقته واستطاعته وهو جوهر العلاقات الإنسانية ومحقق تطورها[20].

2-2 مفهوم شبكات التواصل الاجتماعي

الشبكات الاجتماعية مصطلح ظهر نهاية التسعينات وبداية الالفيه الثانية هي تركيبه اجتماعيه إلكترونيه تم صناعتها من قبل أفراد وجماعات أو مؤسسات تقدم مجموعه من الخدمات التي من شانها تدعيم التواصل والتفاعل بين أعضاء الشبكة الاجتماعية من خلال من خلال الخدمات والوسائل المقدمة مثل التعارف الصداقة المراسلة المحادثة الفورية مشاركه الوسائط مع الآخرين كالصور والفيديوهات والبرمجيات إنها مجموعة من المواقع على شبكة الانترنيت العالمية تتيح التواصل بين الأفراد في بيئة مجتمع افتراضي يجمعهم الانضمام أو الانتماء لبلد أو مدرسة أو فئة معينة في نظام عالمي لنقل المعلومات. فهي مواقع الكترونية تسمح للأفراد بالتعريف بأنفسهم والمشاركة في شبكات اجتماعية من خلالها يقومون بإنشاء علاقات اجتماعية[20].

2-3 وسائل التواصل الاجتماعي Social Media

على الرغم من أن التعريف لا يزال نوعا ما في حاله التغير إلا أن مصطلح social media أو وسائل التواصل الاجتماعي يعرف على نطاق واسع بأنه منصات الكترونيةplatformesElectronic أو قنوات تتيح للناس التواصل وتبادل المحتوى كما وإنها تسهل أقامه مجموعه متنوعة من العلاقات عبر الانترنت يصف ذلك المصطلح أنواع تطبيقات شبكه الويب www مثل المدوناتblogs والمدونات الصغيرة microblogs مثل تويترTwitter ومواقع الشبكات الاجتماعية Social Network sites أو منصات تشارك الفيديو والصور والملفات أو الويكي Wikis إن وسائل التواصل الاجتماعي وبحسب ما جاء في تقرير الإعلام الاجتماعي العربي هي تطبيقات قائمه على الانترنت تسهل إنشاء شبكات منظمه من الأفراد أو الكيانات أو أشياء الافتراضية المرتبطة على أساسي الخصائص الاجتماعية وهي تسهل التفاعلات المتعددة بما في ذلك إنشاء والمحتوى ومشاركته من قبل المستخدمين[21].

2-4 مواقع التواصل الاجتماعي

هي منظومة من الشبكات الالكترونية عبر الانترنت تتيح للمشترك فيها إنشاء موقع خاص به ومن ثم ربطه بنظام اجتماعي الكتروني مع أعضاء آخرين لديهم الاهتمامات والهويات نفسه وتشمل مواقع التواصل الاجتماعي الفيسبوك تويتر انستغرام وغيرها[22].

2-4-1 فيسبوك (Facebook)

تم إنشاء Facebook في عام 2006 ، وهو عبارة عن شبكة اجتماعية تسمح بمشاركة المعلومات والصور ، مقاطع الفيديو والحالات المزاجية والموسيقى مع الأصدقاء و المجتمع. من الممكن استخدام Facebook بشكل خاص أو مهني ستظهر كل مشاركة على صفحة الأصدقاء. يقدم Facebook أيضًا خدمة البريد الإلكتروني والدردشة (الدردشة الفورية). يستمر Facebook في الريادة باعتباره الشبكة الاجتماعية الأكثر استخدامًا (71٪).[23]

2-4-2 تويتر Twitter

التويتر (Twitter) هو أحد مواقع الشبكات الاجتماعية، وهو يسمح لمستخدميه بنشر رسائل قصيرة التي يمكن لمستخدمي التويتر الآخرين رؤيتها، وتُعرف هذه الرسائل بالتغريدات(tweets) ، ويمكن لهذه الرسائل أن تشمل حوالي 140 حرف، أو أقل، وتأسس موقع تويتر عام 2006، وكان عدد المشتركين بين مستخدمي تويتر اعتباراً من عام 2008 من 4 إلى 5 ملايين مستخدماً، وكان ثالث أكثر مواقع الشبكات الاجتماعية انتشاراً بعد موقعي فيس بوك(Facebook) ، وماي سبيس(MySpace) ، وقد استخدم مستخدمي التويتر التطبيق في التواصل الأساسي بين الأصدقاء والعائلة، أو كطريقة للتعريف بحدث معين، أو كأداة علاقات العملاء للشركات من أجل التواصل مع المستهلكين[24]

2-4-3 انستغرام

يُمكن استخدام تطبيق الإنستغرام بشكل مجاني عبر الهواتف المحمولة التي تعمل بأيّ من نظامي تشغيل أندرويد أو(iOS) ، ويُتيح التطبيق لمُستخدِميه إمكانية مشاركة الصور ومقاطع الفيديو مع المُستخدِمين الآخرين للتطبيق، كما يُمكن من خلاله التعليق على المُشاركات المُختلفة للمُستخدِمين وإبداء الإعجاب بها، ويُسمح للأشخاص الذين تزيد أعمارهم عن 13 عام استخدام هذا التطبيق عبر هواتفهم من خلال تحميله والتسجيل كمستخدم جديد. يسمح تطبيق الإنستغرام لمُستخدِميه إضافة علامات مرجعية على المنشورات التي يتمّ نشرها عبر الحسابات الخاصة بهم؛ كعلامات الهاشتاق والعلامات الجغرافية المُستندة إلى الموقع، حيث يسمح هذا الأمر بفهرسة تلك المنشورات وجعلها قابلةً للبحث من قبل المُستخدِمين الآخرين للتطبيق، وتظهر المشاركات التي ينشرها المُستخدِم عبر الإنستغرام من خلال ما يُعرف بـ

(Feeds)الخاصة بمُتابعي حساب هذا المُستخدِم، ويوفّر الإنستغرام ميزة جعل الملف الشخصي الخاص بالمستخدِم خاصّاً (Private) بحيث لا يتمكّن الأشخاص غير المُتابعين لحساب المُستخدِم من مُشاهدة ما ينشره[25]

2-5 أنواع التواصل على مواقع التواصل الاجتماعي:

2-5-1 التواصل الشفهي:

وتستخدم فيه اللغة الشفهية والأصوات المعبرة عن الأفكار والمعارف التي يراد نقلها إلى المستقبِل باستخدام آليات كالهاتف ومكبر الصوت ،أو الحاسوب واللوحات الذكية؛ وتكون صوتية فقط أو صوتية ومرئية، وهذه الأخيرة تلعب فيها الإيماءات وحركات الوجه والجسم دورا فعالا لا تتوافر في التواصل الكتابي.[20]

2-5-2 التواصل الكتابي:

وتستخدم فيه الكتابة، ويكون أكثر رسمية من التواصل بالكلام الشفهي، حيث إن الكتابة تعتمد على استخدام اللغة والمهارة في عرض المكتوب، لذلك يجب أن تكون الرسالة المكتوبة كاملة في ذاتها لكي تتجنب الفهم الخاطئ لها. ولدراسة هذا الموضوع المعنون أعلاه فقد قمنا بدراسة وتتبع اللغة المكتوبة في هذه المواقع، فوجدنا أن اللغة العربية المكتوبة في هذه المواقع أشكال ومستويات[20].

2-6 اللغة العربية في مواقع التواصل الاجتماعي :

إن ظاهرة استخدام مواقع التواصل الاجتماعي أصبحت تحتل موقع الصدارة في عمليات الاتصال الإنساني علي مستوي كل المجتمعات بل باتت هي الوسيلة الأهم التي يعتمدها الناس في اتصالاتهم مع بعضهم البعض خاصة فئات الشباب وقد صحب ذلك الاستخدام أحد الظواهر التي يراها كثير من المختصين جانبا سالبا ومهددا يزيد من ضعف اللغة العربية كلغة يفترض أن تكون هي اللغة الأصل المستخدمة في وسائل الاتصال في المجتمعات العربية والظاهرة التي نشير اليها هي ظاهرة كتابة الكلمات العربية بحروف اللاتينية فيما يسمي بلغة الجات، حيث يتجه أغلب مستخدمي هذه المواقع إلى كتابة اللغة العربية بحروف لاتينية، خصوصا في التواصل عبر موقعي التواصل الاجتماعي فيس بوك وتويتر، والدردشة على البريد الالكتروني، ويعزو البعض السبب في ذلك إلى سهولة الكتابة بالحروف اللاتينية على لوحة مفاتيح الأجهزة الالكترونية خاصة حال استخدامهم لهذه المواقع [25] .

2-6-1 اللغة العربية المكتوبة في وسائل التواصل الاجتماعي:

ظهرت في السنوات الأخيرة أنماط جديدة من الاستعمال اللغوي، وهي أنماط مرتبطة بوسائل الاتصال الاجتماعي الحديثة، وتتميز هذه الأنماط بتنوع الكتاب ما بين متعلمين ومثقفين وطلاب وعامة، ومن ثم نشأت طريقة جديدة للتعبير، وألفاظ جديدة ومصطلحات جديدة تختلف في طريقتها عن الطريقة السائدة المستعملة المعروفة لدى أهل اللغة العربية والمتخصصين فيها. ومن خلال ترصدنا وتتبعنا اليومي للغة

المكتوبة في وسائل التواصل الاجتماعي، فقد وجدنا اللغة التي يكتب بها المتراسلون قد تنوعت طبقا للمستوى الثقافي والعلمي للمتواصلين وهي ما بين فصحى بمفردها وعامية بمفردها وخلط بين المستويين، والمستويين معا مع لغة أجنبية فرنسية أو انجليزية (في أغلب الأحيان)، أي: لغة هجين، مع اختصارات كثيرة للكلمات والجمل، واستعمال الأرقام بدل الحروف...

أ)**الفصحى** : فهي نمط من الكتابة والنطق بالعربية الفصحى كما وضع قواعدها النحاة، ويتميز هذا النمط باللغة العالية النموذجية من الناحية الصوتية والصرفية والتركيبية، ويستعمله قلة من المتخصصين في الدراسات العربية، وأساتذة التعليم العالي خاصة بين الأساتذة الذين ينتمون إلى بلدان عربية متباينة، وبينهم وبين متخصصين من أهل الغرب والعجم الذين تعلّموا العربية وأجادوها من أجل أن يحصل الفهم الجيّد بينهم لمحتوى الرسائل، وتتم عملية التواصل بشكل جيّد. وقد حاولوا المحافظة عليها بكتابتها كتابة سليمة صحيحة؛ ومع هذا الحرص لا تخلوا هذه التعبيرات والصفحات من بعض الأخطاء الإملائية التي تنال من أربابها، لكن لو روعيت لكان أفضل وأكمل، وهي أخطاء ناشئة في الغالب من سرعة الكتابة أو عدم مساعدة الأجهزة الحديثة وإسعافها أحيانا للمتعامل معها، أو عدم الأخذ ببعض الأمور التي تبدو مهمة في اللغة العربية، كالهمزة وعلامات التعجب والاستفهام، أو النقاط والفواصل والأقواس، وعلامات التخصيص، مع العلم أن هؤلاء لا يكتبون لمتخصصين فقط. إلاّ أن هذه الملاحظات لا يمكن تعميمها على كل المتواصلين مع هذه الوسيلة من أهل العربية وآدا¨ا، فجلّهم قلّما تصدر منه هذه الملاحظات أو نلاحظها في لغته وأسلوبه الذي يتعامل معه. مثال على ذلك لتعليق من موقع فيسبوك:[20]

– مرحبا اخي لقد خرجنا اليوم مع الزملاء كان يوما رائعا!!

– ساذهب غدا للجامعة باكرا لدي بعض المحاضرات واعمال .

ب)**الخلط بين الفصحى والعامية** : وبه بعض الملاحظات النحوية والإملائية كذلك التي تدل على أن هؤلاء المستخدمين غير متخصصين، وأن أكثرهم من المتعلمين الذين تلقوا تعليما باللغة العربية، وتأثّروا بعد ذلك بالبيئة الاجتماعية لهم، وهي فصحى مختلطة بأنواع من العاميات، كلّ حسب العامية التي نشأ عليها أو تأثّر بها. مثال على ذلك لتعليق من موقع فيسبوك:[20]

– اليوم تقرر عليكم مايلي:تكملو وش بقا من بحث تعكم وتبعثوه فايميل

* نلاحظ ان اول الجملة (اليوم تقرر عليكم مايلي) جملة كتبت بالفصحى

* واخر الجملة (تكملو وش بقا من بحث تعكم وتبعثوه فايميل)كتبت بالعامية

ج) **العامية** : هو مستوى أدنى من المستويين السابقين، ولكنّه أكثر استعمالا منه على وسائل التواصل الاجتماعي خاصة أن كثيرا من مستخدمي هذه المواقع لا يقعون في دائرة المتخصصين، وقد زاد هذا المستوى استعمالا على صفحات التواصل الاجتماعي، وهذا من شأنه أن يفتح الباب أمام العامية،

ويعمل على شيوعها وانتشارها بكثرة، والابتعاد عن الفصحى كثيرا؛ ومستخدمي هذه اللغة ينتمون إلى فئات اجتماعية وثقافية متنوعة، وهي عامية تقع في إطارها عدد من العاميات منهاعامية الشباب، وعامية الأميّين، عامية الطبقة الاجتماعية الدنيا و...، ويذكر أنّه لكل مستوى من هذه العاميات ميزات وسيمات لغوية تختلف من مستوى لآخر. مثال على ذلك لتعليق من موقع فيسبوك:[20]

– ليوما روحت عند جداتي بعثني ليها بابا

– ضرك حناكلو فريسطو ومباعد نروحو للكلاسة نقرو تيدي عزوج

د) اللغة الهجين : وهو النمط أو المستوى اللغوي الأكثر خطورة على اللغة العربية، ويطلق عليه في العصر الحاضر عند مجموعة من الباحثين واللسانيين مصطلح "**العربيزي**"(**ARAB–EZ**) ؛ لا توجد دراسات كثيرة تتناول هذه الظاهرة لحداثتها، فهي ظاهرة جديدة تجتاح اللغة العربية في نظام كتابتها؛ تكتب باللغة العامية مع الفصحى أحيانا لكن بالحروف اللاتينية والأرقام مع الخلط بينها و بين اللغات الأجنبية، و قد "اشتهرت بأسماء كثيرة منها (الفرانكو، والفرانكو آراب، والعربيزي، والإنجلو عربي، والأرابيش ...) " ،انتشرت مع التوسع في استعمال الهواتف الذكية ومواقع التواصل الاجتماعي خاصة عند فئة الشباب من الجنسين، مما حدا بكثير من المتخصصين في الدراسات اللغوية إلى انتقاد هذه الظاهرة؛ وما برحوا يحذرون من خطورتها على الهوية اللغوية لدى الفئات المستخدمة لها، ويرون أنها باتت تهدد حروف اللغة العربية بالانقراض وتمحوا خصوصيتها". وقد تطورت هذه الظاهرة بشكل سريع جدا في وقتنا الحالي مع دخول التقنية والانترنيت، والهواتف الذكية، فأضيفت إليها الأرقام لتعبر عن بعض الحروف، وأصبح لها عدة أشكال وأسماء. وقد فرق الدكتور سعد بن طفلة العجمي بين هذا المصطلح الشائع (العربيزي) ومصطلح(العربتيني) (ARABATIN) حيث قال: "وهناك فرق بين (ARABATIN)،والعربيزي(ARAB–EZ) وتكتب بالإنجليزية أحيانا العربتيني (ARABIZI)أو(3ARABIZI) فظاهرة العربيزي تعني: الخلط في الكلام أثناء الحديث بين العربية والانجليزية تحديدا. [20]

كأن يقول أحدهم:

– "أنا رايح هناك، .. سي يو"(SEE YOO) ؛

– "وبعدين جلسنا في بيتهم .. أوكي "OK ؛

– "وكنت أنا وماي فريندز"(MY FRIENDS) ؛

– وعبارات "باي"(BYE) و"تيك كير"(TAKE CARE) .. وهكذا .

ولكن العربتيني مصطلح يستعمل الأحرف اللاتينية بدلا من العربية في الوسائل الرقمية، و في الحوارات

أو الدردشة الالكترونية؛ أي كتابة العربية بالأحرف اللاتينية، والكلمة (**العربتيني–ARABATIN**) منحوتة من كلمتي (العربي) و (اللاتيني) ، وهي ظاهرة يمارسها في الغالب الجيل الرقمي الجديد". و سواء أكانت هذه الظاهرة تسمى العربتيني التي هي كتابة نصوص عربية من حيث المحتوى والنطق، ولكن باستخدام حروف لاتينية، أم تسمى العربيزي التي هي الخلط في الكلام أثناء الكتابة و الحديث بين العربية و الانجليزية مع استخدام بعض الأرقام عوض الحروف فإن الظاهرة هجين لغوي عصري رقمي إلكتروني سابقة في عصرها تنذر بخطر يهدد نظام كتابة العربية ونطقها على السواء، وللتعرف أكثر على هذه الظاهرة فقد اخذنا بعض هذه الكتابات من موقع التواصل الاجتماعي فيسبوك أمثلة من ظاهرة الهجين اللغوي الرقمي (العربيزي) :

1)كتابة العربية بالحروف اللاتينية والأرقام نحو الكلمات والتراكيب الآتية:

- (Na3am) نعم؛
- (Aham shay) أهم شيء؛
- (Inshallah)إن شاء الله؛
- (3eidek mebarek) عيدك مبارك؛
- (Amine) آمين؛
- (A7lleme sa3ida) أحلام سعيدة؛
- (Nalta9i gadan) نلتقي غدا؛
- (Ramazan mubarak) رمضان مبارك...

2) كتابة الفرنسية أو الانجليزية –في أغلب الأحيان– المختصرة وسط نصوص عربية نحو:

- نلتقي في سوق السيارات OK.أو (أوكي)
- NP سآتي غدا.(NP= No problem) يعني: لا مشكل سآتي غدا[20].

2-7 تأثير الذكاء الاصطناعي على وسائل التواصل الاجتماعي:

يجب على منصات التواصل الاجتماعي مثل Facebook و LinkedIn استخدام الذكاء الاصطناعي لفهم حجم البيانات البشرية الضخمة القادمة إليهم. مع ما يقدر بنحو 2.77 مليار مستخدم للشبكات الاجتماعية حول العالم، تمتلك منصات الربط الشبكي قدرًا لا يمكن تصوره من البيانات. ان مدى انتشار الخدمات الاجتماعية تتوسع فيه الشبكات بمعدل ينذر بالخطر في جميع أنحاء العالم في عام 2019. هذه الأرقام ستستمر في النمو. توفر وسائل التواصل الاجتماعي كمية هائلة من بيانات المستخدمين. هذا هو المكان الذي يلعب فيه الذكاء الاصطناعي دور مهما . الذكاء الاصطناعي هو منطقة علوم

الكمبيوتر التي تهدف إلى فك شفرة البيانات من العالم الطبيعي غالبًا باستخدام التقنيات المعرفية المصممة لفهم وإكمال المهام التي يمتلكها البشر[23].

2-8 دور الذكاء الاصطناعي في وسائل التواصل الاجتماعي:

إنه عصر وسائل التواصل الاجتماعي. سواء كان الأمر يتعلق بالتواصل مع العملاء الجدد أو رعاية العلاقات التجارية القائمة، أصبحت وسائل التواصل الاجتماعي جزءًا لا يتجزأ من التفاعل مع الشركات أو المؤثرين أو العملاء. تساعد الإدارة الفعالة لوسائل التواصل الاجتماعي تعزيز قوة العلامة التجارية وزيادة محادثات وسائل التواصل الاجتماعي. إنه فعال بطريقة إشراك الملايين من مستخدمي وسائل التواصل الاجتماعي. يساعد الذكاء الاصطناعي في تحليل البيانات الضخمة إلى تحديد الموضوعات الشائعة وعلامات التصنيف والأنماط لفهم سلوك المستخدم. يمكن لخوارزميات الذكاء الاصطناعي مراقبة ملايين تعليقات المستخدم غير المنظمة لفهمها لتقديم تجربة شخصية. مع فعالية التجزئة، يمكن أن يساعد الذكاء الاصطناعي في توفير محتوى يعتمد على النشاط عبر الإنترنت والتركيبة السكانية. استحوذت العديد من مواقع الشبكات الاجتماعية على أعمال الذكاء الاصطناعي للانتقال إلى المستوى التالي. تستكشف مجموعة متنوعة من الشركات التي تقدم خدمات التسويق عبر الإنترنت طرقًا جديدة. استخدم وسائل التواصل الاجتماعي مع الذكاء الاصطناعي أيضًا. لقد بدأوا في استخدام الذكاء الاصطناعي لتحديد الجديد على أساس التحويلات السابقة. تعتمد أدوات الذكاء الاصطناعي هذه على خوارزميات التحليلات التنبؤية، والتي يمكنها إطالة المعلومات عن جميع المستخدمين المعروفين في شبكة اجتماعية معينة[23].

3- تنقيب البيانات، تعدين النصوص وتحليل المشاعر

3-1 تعريف البيانات الكبيرةBig Data

إن البيانات الكبيرة هي مصادر البيانات ذات الحجم والسرعة الكبيرة وهي تمثل مجموعه متنوعة من البيانات التي تتطلب أدوات وطرق جديدة لاستخراجها وتنظيمها وإدارتها ومعالجتها بطريقه فعاله ويعرفها الاتحاد الدولي للاتصالات بأنها نموذج لتمكين وجمع وتخزين و أدارة وعرض مجموعات واسعة من البيانات ذات الخصائص غير المتجانسة[21].

3-2 تعريف تنقيب البيانات Data Mining

استخدم مصطلح تنقيب البيانات أو اكتشاف المعرفة من البيانات هو مصطلح شامل يصف العديد من الأنشطة والتقنيات لاستخراج المعلومات من البيانات واقتراح أنماط في قواعد بيانات كبيره للغاية والتي تشتمل على تجمعات ضخمه ومستودعات للبيانات والتحليل الإحصائي وتقنيات التعلم الإستنتاجي. ويشير جيمس واتسون إلى أن تنقيب البيانات هو مجموعة من التقنيات الآلية لاستخراج المعلومات المخفية والتي لم تكن معروفه مسبقا من قواعد البيانات الكبيرة إن عمليه إن عمليات تنقيب البيانات الناجحة تجعل من الممكن استخراج المعلومات ذات قيمه أو الكشف عن الأنماط والعلاقات ومن قبل استخدام هذه المعلومات قرارات الأعمال القائمة على المعرفة الإستباقية [21].

3-3 تعدين النصوص Text Mining

تعدين النصوص، المعروف أيضًا باسم تعدين البيانات النصية، هو عملية تحويل النص غير المنظم إلى تنسيق منظم لتحديد الأنماط الهادفة والرؤى الجديدة. من خلال تطبيق التقنيات التحليلية المتقدمة، مثل Naieve Bayes و Support Vector Machines (SVM) وخوارزميات التعلم العميق الأخرى، يمكن للشركات استكشاف واكتشاف العلاقات المخفية داخل بياناتها غير المنظمة[26].

يُستخدم على نطاق واسع في المنظمات القائمة على المعرفة، تعدين النصوص هو عملية فحص مجموعات كبيرة من الوثائق لاكتشاف معلومات جديدة أو المساعدة في الإجابة أسئلة بحثية محددة. يحدد تعدين النصوص الحقائق والعلاقات والتأكيدات التي كانت ستبقى لولا ذلك مدفونة في كتلة البيانات النصية الضخمة. بمجرد استخراج هذه المعلومات يتم تحويلها في شكل منظم يمكن مواصلة تحليله أو عرضه مباشرة باستخدام جداول HTML المجمعة وخرائط العقل والمخططات وما إلى ذلك. يستخدم تعدين النصوص مجموعة متنوعة من منهجيات لمعالجة النص، و من أهمها منهجيات طبيعية تجهيز اللغات (NLP) [23] .

3-3-1 هيكلة وتنظيم البيانات (النصوص)

يعد النص أحد أكثر أنواع البيانات شيوعًا داخل قواعد البيانات. وتبعا لقاعدة البيانات، يمكن تنظيم هذه البيانات على النحو التالي:

- **البيانات المنظمة:** تم توحيد هذه البيانات في شكل جدولي مع العديد من الصفوف والأعمدة، مما يسهل تخزين ومعالجة خوارزميات التحليل والتعلم الآلي. يمكن أن تتضمن البيانات المنظمة مدخلات مثل الأسماء والعناوين وأرقام الهواتف[26].
- **البيانات غير المنظمة:** لا تحتوي هذه البيانات على تنسيق بيانات محدد سلفا. يمكن أن يتضمن نصوصًا من مصادر، مثل مراجعات وسائل التواصل الاجتماعي أو المنتجات، أو تنسيقات وسائط غنية مثل ملفات الفيديو والصوت[26].
- **البيانات شبه المنظمة:** كما يوحي الاسم، هذه البيانات عبارة عن مزيج بين تنسيقات البيانات المنظمة وغير المنظمة. في حين أن لديها بعض التنظيم، إلا أنها لا تملك هيكلًا كافيًا لتلبية متطلبات قاعدة البيانات العلائقية. تشمل أمثلة البيانات شبه المنظمة ملفات XML و JSON و HTML.

- نظرًا لأن 80٪ من البيانات في العالم توجد في شكل غير منظم، فإن تعدين النصوص هو ممارسة قيمة للغاية داخل المنظمات. أدوات تعدين النصوص وتقنيات معالجة اللغة الطبيعية (NLP)، مثل استخراج المعلومات تسمح لنا بتحويل المستندات غير المنظمة إلى تنسيق منظم لتمكين التحليل وتوليد رؤى عالية الجودة. وهذا بدوره يحسن عملية صنع القرار في المنظمات، مما يؤدي إلى نتائج أفضل للأعمال التجارية[26].

3-3-2 تقنيات تعدين النصوص

تتضمن عملية تعدين النصوص العديد من الأنشطة التي تمكنك من استنتاج المعلومات من بيانات نصية غير منظمة. قبل أن تتمكن من تطبيق تقنيات مختلفة لتعدين النصوص، يجب أن تبدأ بمعالجة النصوص مسبقًا، وهي ممارسة تنظيف وتحويل البيانات النصية إلى تنسيق قابل للاستخدام. هذه الممارسة هي جانب أساسي من معالجة اللغة الطبيعية (NLP) وعادة ما تتضمن استخدام تقنيات مثل تحديد اللغة، والترميز، ووضع العلامات على جزء من الكلام، والقطع، وتحليل التركيب لتنسيق البيانات بشكل مناسب للتحليل. عند اكتمال المعالجة المسبقة للنص، يمكنك تطبيق خوارزميات تعدين النصوص لاستخلاص رؤى من البيانات. وتشمل بعض هذه التقنيات الشائعة لتعدين النصوص ما يلي:

3-3-2-1 استرجاع المعلومات

استرجاع المعلومات (IR) يعيد المعلومات أو المستندات ذات الصلة بناءً على مجموعة محددة مسبقًا من الاستفسارات أو العبارات. تستخدم أنظمة الأشعة تحت الحمراء الخوارزميات لتتبع سلوكيات المستخدم وتحديد البيانات ذات الصلة. يُستخدم استرجاع المعلومات بشكل شائع في أنظمة كتالوج المكتبة ومحركات البحث الشائعة، مثل Google. وتشمل بعض المهام الفرعية المشتركة في مجال الموارد البشرية ما يلي:

- ✓ **الترميز (Segmentation)**: هذه هي عملية تقسيم النص الطويل إلى جمل وكلمات تسمى «الرموز». تستخدم في النماذج، مثل حقيبة الكلمات، لمهام تجميع النصوص ومطابقة الوثائق[26].
- ✓ **التوقف(Emboutissage)** : يشير هذا إلى عملية فصل البادئات واللواحق عن الكلمات لاشتقاق الكلمة الجذرية الشكل و المعنى. تعمل هذه التقنية على تحسين استرجاع المعلومات عن طريق تقليل حجم ملفات الفهرسة[26].

3-3-2-2 استخراج المعلومات

استخراج المعلومات (IE) يظهر أجزاء البيانات ذات الصلة عند البحث في مختلف المستندات. كما يركز على استخراج المعلومات المنظمة من النص الحر وتخزين هذه الكيانات والسمات ومعلومات العلاقة في قاعدة بيانات. وتشمل المهام الفرعية المشتركة لاستخراج المعلومات ما يلي:

- ✓ **اختيار الميزة:** أو اختيار السمة، هو عملية اختيار الميزات المهمة (الأبعاد) للمساهمة بشكل أكبر في إخراج نموذج التحليلات التنبؤية[26].
- ✓ **استخراج الميزات:** هو عملية اختيار مجموعة فرعية من الميزات لتحسين دقة مهمة التصنيف. هذا مهم بشكل خاص للحد من الأبعاد[26].
- ✓ **التعرف على الكيانات المسماة(La reconnaissance des entités nommées 'NER')** المعروف أيضًا باسم تحديد الكيان أو استخراج الكيان، إلى العثور على كيانات معينة وتصنيفها في النص، مثل الأسماء أو المواقع. على سبيل المثال، تعرف NER «كاليفورنيا» على أنها موقع و «ماري» على أنها اسم امرأة[26].

3-3-2-3 استخراج البيانات

استخراج البيانات هو عملية تحديد الأنماط واستخراج رؤى مفيدة من مجموعات البيانات الضخمة. تقوم هذه الممارسة بتقييم كل من البيانات المنظمة وغير المنظمة لتحديد المعلومات الجديدة، ويتم استخدامها بشكل شائع لتحليل سلوكيات المستهلك في التسويق والمبيعات. يعد تعدين النصوص في الأساس مجالًا فرعيًا لتعدين البيانات حيث يركز على جلب الهيكل إلى البيانات غير المنظمة وتحليلها لتوليد رؤى جديدة

والتقنيات المذكورة أعلاه هي أشكال لاستخراج البيانات ولكنها تندرج في نطاق تحليل البيانات النصية[26].

3-3-2-4 معالجة اللغة الطبيعية (NLP)

تستخدم معالجة اللغة الطبيعية، التي تطورت من اللغويات الحسابية، أساليب من مختلف التخصصات، مثل علوم الكمبيوتر والذكاء الاصطناعي واللغويات وعلوم البيانات، لتمكين أجهزة الكمبيوتر من فهم اللغة البشرية في كل من الأشكال المكتوبة واللفظية. من خلال تحليل بنية الجملة والقواعد، تسمح مهام NLP الفرعية لأجهزة الكمبيوتر «بالقراءة». وتشمل المهام الفرعية المشتركة ما يلي:

- ✓ **التلخيص(Résumé)**: توفر هذه التقنية ملخصًا لقطع نصية طويلة لإنشاء ملخص موجز ومتماسك للنقاط الرئيسية للوثيقة[26].
- ✓ **علامة جزء من الكلام (PoS)**: تخصص هذه التقنية علامة لكل رمز في مستند بناءً على الجزء الخاص به من الكلام – أي الإشارة إلى الأسماء والأفعال والصفات وما إلى ذلك. تتيح هذه الخطوة التحليل الدلالي للنص غير المنظم[26].
- ✓ **تصنيف النصوص(Catégorisation du texte)**: هذه المهمة، المعروفة أيضًا باسم تصنيف النصوص، مسؤولة عن تحليل وثائق النصوص وتصنيفها بناءً على مواضيع أو فئات محددة مسبقًا. وهذه المهمة الفرعية مفيدة بوجه خاص عند تصنيف المرادفات والمختصرات[26].
- ✓ **تحليل المشاعر(Catégorisation du texte)**: تكتشف هذه المهمة المشاعر الإيجابية أو السلبية من مصادر البيانات الداخلية أو الخارجية، مما يسمح لك بتتبع التغييرات في مواقف العملاء بمرور الوقت . يتم استخدامه بشكل شائع لتوفير معلومات حول تصورات العلامات التجارية و المنتجات والخدمات. يمكن لهذه الأفكار أن تدفع الشركات إلى التواصل مع العملاء وتحسين العمليات وتجارب المستخدمين[26].

3-3-3 طرق تمثيل البيانات

يمكن القول أن المشكلة في العمل على النصوص هو صعوبة الحصول على المعلومات من النص الذي هو عبارة عن سلاسل طويلة و إدخالها في واجهة برمجة تطبيقات API .لذلك هناك طرق لجعل هذه النصوص مدخلات يمكن استعمالها.

- **حقيبة الكلمات (Bag Of Words)**

و هي طريقة شائعة لازالت تستخدم حاليا. تعمل طريقة التمثيل هذه بأخذ كل كلمة و حساب عدد المرات التي تظهر فيها في كل وثيقة أو نص، أي أننا نحول سلسلة الكلمات إلى شعاع (vector) حيث يمثل عدد الكلمات طول الشعاع. مساوئ هذه الطريقة هي غياب ترتيب الكلمات و هو أمر مهم في أي لغة . [27]

-ترميز من1الىn (One Hot Encoding)

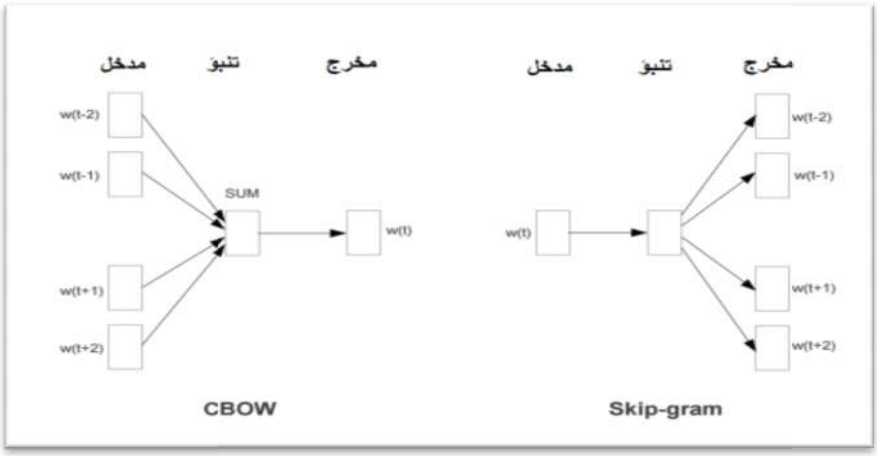
يستخدم هذا التمثيل الأحرف بشكل فردي أي يأخذ ترميز واحد لكل حرف في النص، لكن في أي لغة مفهوم الكلمة مهم و مفيد جدا لذلك تمرير المدخلات كحروف يجعل النموذج يتعلم عن اللغة أكثر ، ما يجعله خام أكثر مالم يكن لدينا كميات هائلة من البيانات، هذا ما أدى إلى ظهور تحويل آخر بين هذا التحويل و حقيبة الكلمات و هو ترميز الكلمة(Word Encoding/Word Embeding) [27].

- ترميز الكلمة

يقوم بتحويل كل كلمة من الوثيقة أو النص إلى شعاع من الأرقام و تكون جميع الأشعة بنفس الطول، على سبيل المثال في الجملة "أنا أحب السفر"، اذا افترضنا أن طول الأشعة هو 4 ، سيكون لكلمة "أنا" شعاع بنفس الأربع أرقام دوما و نفس الشيء لباقي الكلمات أي أن تمثيل كل كلمة يبقى ثابت. بدل قيامنا بحساب الترميزات يمكننا استخدام ترميزات محسوبة مسبقا، و من أشهر هذه الترميزاتWord2Vec [27]

- خوارزمية (Word2 Vec)

هي خوارزمية لتمثيل الكلمات بأشعة تأخذ بعين الاعتبار سياق الكلمة بحيث يتم تمثيل الكلمات المتشابهة في السياق بأشعة متقاربة .تحتوي خوارزمية Word2Vecعلى نموذجين CBOW و Skip-gram (الشكل1.3) حيث يقوم الأول بالتنبؤ بمعنى الكلمة بناء على سياقها لكي يتمكن من تمثيلها . أما الثاني فيقوم بالتنبؤ بالكلمات المجاورة للكلمة.[27]



الشكل (1.3) : نموذجي Skip-gram وCBOW [27]

3-4 تعريف تحليل المشاعر

تحليل المشاعر والمعروف أيضا باسم استخراج الرأي هو المهمة الأساسية لتصنيف نص أساسي أليا أو تلقائيا وتحديد ما إذا كان اتجاهه أو ميله ايجابيا أو سلبيا أو محايدة وتحليل المشاعر وتحليل النص يهدف إلى تحديد اتجاه أو موقف متكلم أو الكاتب في ما يتعلق ببعض الموضوعات أو كتابة السياق الكلي جزء من النص ويهتم تحليل المشاعر أو استخراج الرأي بتطوير البرامج لاستخراج آراء المستخدمين تلقائيا حول المنتجات أو الكيانات الأخرى من النص وعاده ما تكون من مصادر عبر الانترنت إن كان تحليل المشاعر يمكن إن يختلف عن استخراج الرأي إذا يشتملوا على البرامج التي تستخرج المشاعر من النص لأغراض مثل تقدير رحال العاطفية لمؤلف النص ذاته.[21]

3-4-1 طرق وتقنيات تحليل المشاعر

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

3-4-1-1 نهج تعلم الالة

يعتمد على مجموعة من البيانات لإنشاء أو تصميم نماذج تستخدم في تصنيف البيانات الجديدة أو غير المعروفة بدقه ويعد هذا المدخل هو الأسلوب الأكثر استخداما في الدراسات تحليل المشاعر نظرا لمعدلات الدقة العالية التي يحققها ومع ذلك فانه يعتمد على المجال ويحتاج إلى مجموعه بيانات معلمه ويوم طويل ذلك على كثير من الجهد في البيئة مثل الشبكات الاجتماعية التي تحتوي على الكثير من المجالات وهذا النهج ينقسم إلى ثلاثة تصنيفات فرعية فقد يكون خاضع للإشراف ويعتمد على خوارزمية وتسمية لكل مثال أو قد يكون غير خاضع للإشراف ولا يعتمد على شيء وأخيرا قد يكون خاضعا لشبه الإشراف تطلب تقنيه تعلم الآلة في حاله التعلم تحت الإشراف القيام بعمليه التصنيف اليدوي لعينات من قاعدة البيانات تكوين مجموعه تدريب بهدف بناء نماذج تستند إلى النص تكون بمثابة مقاييس في تحليل البيانات الآلية[21]

3-4-1-2 النهج القائم على المعجم

ويتم بأسلوبين الأول هو استخدام القواميس التي تحتوي على مجموعه واسعة من المصطلحات أو العبارات الاستنتاج المشاعر المنقولة في النص والثاني يعتمد على موارد التحويل مجموعه من الكلمات ويمكن أن نسمي ذلك النهج أسلوب التوجيه الدلالي الذي يعتمد على تطبيق أو استخدام مجموعه كلمات تم إنشائها سابقا تحتوي على مصطلحات موجبه وسالبه وبالتالي يتم تحديد تلقائي للقطبية باستخدام تكرار هذا المصطلحات في النص دون الحاجة إلى نموذج وقاعدة بيانات مصنفه سابقا إن استخدام قاموس الكلمات أو الذي قد يكون موجبا أو سالبا يحدد وزن الكلمات الرئيسية وفقا لتكرارها أو ترددها وتنبأ بقطبيتها ايجابية أم سلبية غير إن هذا النهج المعتمد على المعجم وحده لا يكون فعالا دائما في تحديد

المشاعر بدقة نظرا لان النصوص تحوي بعض الاختصارات وبعض الكلمات العامية والرموز والعناصر غير التقليدية الأخرى مما يجعل من الصعب وجود قاموس محدث باستمرار من اجل تحديث تحديد المشاعر ولهذا السبب يلجا الباحثين في دراستهم إلى نهج تعلم الآلة.[21]

3-4-1-3 النهج المختلط

وهو يجمع بين أساليب النهج القائم على المعجم وبين تقنيات التعلم الآلي من اجل الاستفادة من خصائص كل منهما.[21]

3-4-2 مهام تحليل المشاعر

تتمثل المهام الأساسية لتحليل المشاعر في الكشف عن الذاتية في النصوص حيث يتم تصنيف النصوص على أنها تحتوي على تعبيرات ذاتيه غير موضوعيه عن المشاعر والكشف عن قطبيه النصوص وتصنف النصوص على أنها إما ايجابية أم سلبية بشكل عام وكشف قوه المشاعر حيث تصنف النصوص بحسب القوه الكلية للمشاعر الايجابية أو السلبية وتحديث قطبيتها وأخيرا كشف العاطفة حيث تصنف النصوص بحسب العاطفة السائدة على سبيل المثال حزين غاضب وباختصار يهدي التحليل المشاعر إلى فهم المعلومات الذاتية مثل الآراء والمواقف والمشاعر المعبر عنها في النص وتشمل مهام تحليل المشاعر على سبيل المثال لا الحصر ما يلي

- ✓ **تصنيف المشاعر:**

الذي يصنف جزءا معينا من النص على انه ايجابي أو سلبي أو محاور محايد[21]

- ✓ **استرجاع الرأي:**

الذي يسترد الآراء ذات الصلة بموضوع الاستعلام معين [21]

- ✓ **تلخيص الرأي:**

الذي يلخص الآراء على مصادر نصيه متعددة نحو موضوع معين [21]

- ✓ **تحديد هوية صاحب الرأي:**

الذي يحدد من يعبر عن رأي محدد.[21]

- ✓ **تتبع ديناميكية الموضوع أو المشاعر:**

تهدف إلى تتبع المشاعر وتغيرات الموضوع مع مرور الوقت الكشف عن البريد المزعج الذي يحدد الآراء المزيفة أو الكاذبة[21]

- ✓ **التنبؤ:**

الذي يتنبأ سلوكيات الناس واتجاهات السوق ونتائج الانتخابات السياسية وما إلى ذلك بناء على الآراء أو المشاعر المعبر عنها في المحتويات على الانترنت[21]

3-4-3 لماذا تحليل المشاعر

تعتبر الآراء مهمة جدا لاتخاذ القرارات عند البشر، ففي القديم إذا أراد الشخص اتخاذ قرار ما يقوم بسؤال العائلة أو الأصدقاء أما الآن و مع ظهور وسائل التواصل الاجتماعي و المدونات أصبحت هذه الآراء و المعلومات متاحة للجميع، و لكن مع هذا الكم الهائل من الآراء في الانترنت قد يجد هذا الشخص صعوبة في تحليلها، هنا نجد الحاجة لأنظمة تحليل المشاعر و التي اعتمدتها العديد من الشركات من أجل تطوير منتجاتها بمراقبة آراء عملائها، إضافة إلى استعمال هذه الأنظمة أيضا في مجالات أوسع مثل القضايا السياسية، مدى ملائمة مقاطع الفيديو للأطفال بناء على التعليقات، مراقبة العلامات التجارية.[27]

3-4-4 معايير تحليل المشاعر

لدراسة رأي ما يجب أن يتوفر على الأقل على بعض المعايير الآتية:

- **لكيان (Object)**: و هو الفرد أو المنظمة أو المنتج المعني بالرأي مثل: حاسوب.[27]
- **الجوانب (Attributes)**:و هي خصائص أو مكونات الكيان مثل: الذاكرة العشوائية (RAM) البطارية، مساحة التخزين، حجم الشاشة[27] ...
- **صاحب الرأي (Opinion holder)**: و هو الشخص أو المنظمة التي تعبر عن الرأي مثل: الشخص الذي اشترى الحاسوب[27] .
- **قطبية الرأي (Opinion orientation polarity)**:أو الاتجاه و هو الموقف العام للرأي، و عادة ما يكون إيجابيا أو سلبيا أو محايدا[27] .
- **شدة الرأي (Opinion strength)**:هو مستوى أو درجة المشاعر المعبر عنها مثل: أنا أكرهه جدا جدا جدا، فهي تعبر عن مشاعر سلبية قوية[27]

3-4-5 مستويات تحليل المشاعر

- **مستوى الوثيقة Document Level**

يتم تصنيف الوثيقة حسب الشعور العام إذا ما كان إيجابي أو سلبي أو حيادي، و يفترضن تحتوي الوثيقة على رأي حول موضوع أو كيان واحد[27]

- **مستوى الجملة Sentence Level**

في مستوى الجملة يكون التصنيف بتحديد ما إذا كان الرأي في هذه الجملة إيجابيا او سلبيا أو محايدا [27]

- **مستوى الجانب Aspect Level**

يطلق عليه أيضا تحليل المشاعر القائم على الجانب، يدرس هذا المستوى الجمل متعددة الآراء فهو يسعى إلى إكتشاف المشاعر على الكيانات و/أو جوانبها بالنظر مباشرة إلى الرأي و هدفه (الكيان أو أحد

جوانبه) تعطي الجملة "هذا العطر غال جدا لكنني أحبه" إنطباع إيجابي لكن لايمكن القول عنها أنها إيجابية تماما إذ أنها تحتوي على مشاعر سلبية حول سعر العطر ويعتبر هذا المستوى الأكثر إستخداما في أنظمة تحليل المشاعر الواقعية في الصناعة[27] .

3-5 دراسات في تحليل البيانات الضخمة من الشبكات الاجتماعية

اضطلع Kauffmann et al (2020). إلى إنشاء إطار لتحليل البيانات الضخمة في الوسائل الاجتماعية التجارية؛ لمساعدة مديري التسويق والمستهلكين في الحصول على إجابات للأسئلة المهمة التي تطرأ عليهم؛ لمساعدتهم في عملية صنع القرار. واعتمدت الدراسة على منهج دراسة الحالة لموقع أمازون Amazon. وتم استخدام تحليل المشاعر لتحليل المراجعات على الموقع. وتوصلت الدراسة إلى إطار معياري يستخدم تقنيات البرمجة اللغوية العصبية، بما في ذلك تحليل المشاعر، وأدوات كشف المراجعات المزيفة، والتنقيب عن المعلومات النصية المدرجة في مراجعات المستخدمين، وذلك للكشف عن المراجعات الزائفة وإزالتها؛ لأنها تُؤثِر سلبًا على النتائج التي يتم الوصول إليها. كما أظهرت النتائج أنَّ تلك الأداة مفيدة لاتخاذ قرارات أكثر شمولاً فيما يتعلق بالمنتجات. وهدفت دراسة Liu et al (2019). إلى فحص أدوار التَّرفيه، والتَّفاعل، وأبعاد التَّخصيص، للتسويق عبر الشبكات الاجتماعية للعلامات التجارية الفاخرة على تفاعل العملاء مع محتوى الشبكات الاجتماعية، وذلك باستخدام مجموعة بيانات twitter كمصدر للبيانات الضخمة التي تسمح بالتقاط، وقياس، وتحليل أنشطة مشاركة كل من الشركة والعملاء في الشبكات الاجتماعية. وأظهرت النتائج إلى أنَّ التَّركيز على مجموعة من الأبعاد عبر الشبكات الاجتماعية للعلامة التجارية الفاخرة يزيد بشكل كبير من مشاركة العملاء، وهذه النتيجة لها آثار مهمة على تصميم، وتسليم، وإدارة التسويق عبر الشبكات الاجتماعية للعلامات التجارية الفاخرة؛ لإشراك العملاء في محتوى الشبكات الاجتماعية. وأوضحت الدراسة أهمية الاستثمار في جوانب الترفيه، والتفاعل، لأنشطة الشبكات الاجتماعية للشركة محل الدراسة. وأنَّ استخدامهم للشبكات الاجتماعية لتعزيز التَّرفيه والتَّفاعل يُؤدِّي إلى زيادة مشاركة العملاء مع محتوى الشبكات الاجتماعية المرتبط بالعلامة التجارية[28]

3-6 تحليل البيانات الضخمة من الشبكات الاجتماعية

تزايدت في السنوات الأخيرة الدراسات التي تعرَّضت للبيانات الضخمة. وبشكل أساسيٍ يُشار إلى البيانات الضخمة، أنها مصطلح يصف مجموعات البيانات الكبيرة جدًا من تيرابايت (TB) إلى إكسابايت (EB)، وغير المنظمة والمعقدة في الشبكات الاجتماعية، أو تطبيقات الهواتف الذكية، أو بيانات الأدوات المستندة إلى الإنترنت. على سبيل المثال، يستضيف الفيس بوك Facebook أكثر من 500 تيرابايت (TB) من البيانات يوميًا، بما في ذلك الصور التي يتم تحميلها، والإعجابات، ومنشورات المستخدمين (Xu et al., 2016). وبالمقارنة بالبيانات الصغيرة، فإن البيانات الضخمة تتطلَّب تِقنِيَّات وإجراءات متطورة ونظام

لاستخلاص الأهمية، للوصول من ذلك إلى مخرجات جديدة، مكونة نماذج تنبُّئيّة مُستنتجة من الرؤى والأفكار التحليلية (وانج والصبحي، 2019: 313). وفي ظل ما سبق، أصبحت تَّحلِيلَات البيانات الضخمة big data analytics ممارسة شائعة تتبناها العديد من المنظمات؛ بهدف الحصول على معلومات قيِّمة. على الرَّغم من ذلك، هناك معرفة محدودة حول كيفيَّة قيام الشركات والمؤسسات بتحويل الإمكانات التي توفرها البيانات الضخمة إلى قيمة اجتماعية واقتصادية حقيقية (Kauffmann et al., 2020). حيث تُشير تلك التَّحلِيلَات إلى مجموعة متنوعة من البيانات التي يتم جمعها بواسطة التِقنِيَّات من مصادر مختلفة غير متجانسة، وتوحيدها، وفحصها، واستغلالها، ومن ثم تحسن أداء الأعمال واتخاذ القرارات الإدارية المحسنة، وبذلك تُمكِن الجِهات التجارية من اكتساب ميزة على منافسيها (Shabbir & Gardezi, 2020).

وبالنظر إلى تحليل البيانات الضخمة big data analysis من منظور تسويقي، فهي باختصار تُعبِر عن الاستجابة لمُتطلبات العملاء الفعليَّة. وبالرَّغم من ذلك، لا يزال البحث حول تطبيقات البيانات الضخمة للتسويق في مرحلة مُبكرة، مما يجعل من الضروريّ زيادة الجهود المبذولة للاعتراف بالبيانات الضخمة كأداة رئيسية في مجال التسويق (Kauffmann et al., 2020). وكَاستِقراء مستقبلي، سنكون أمام عدد مُتزايد من مجموعات البيانات التي تُكَون بيانات ضخمة. سيؤثِّر ذلك بشكل إيجابي على عمليات التسويق والترويج؛ نظرًا لأن حجم مجموعة البيانات يتجاوز سعة قاعدة البيانات التقليديَّة، مما يعني ضمناً أن مجموعات البيانات الضخمة يتم تجميعها في منصات وموارد متعددة مثل: الشبكات الاجتماعية social networks، والمواقع الإلكترونية websites، وسجلات مبيعات العملاء customer sales records، حيث يمكن أنْ يكون الحجم كبيرًا مثل بعض التيرابايت (TB) للعديد من البيتابايت (Wang) (Wang, 2020 & PB). وهذا ما يتفق مع جزء من الرؤية التي تسعى الدراسة إلى تأطيرها. وتجدر الملاحظة بأن تَّحلِيلَات البيانات الضخمة تساعد في فهم واستخراج المعرفة القَيِّمة من الأحجام الهائلة من البيانات المتاحة على الشبكات الاجتماعية، وبعد ذلك يمكن استخدام تلك المعرفة لتجويد أداء العمليات في الجهات التسويقية. وفي ضوء ما سبق، تطرح الباحثتان مفهوم إجرائي لتحليل البيانات الضخمة من الشبكات الاجتماعية، بأنها كمية البيانات الهائلة والعشوائية التي يتم استخلاصها من الشبكات الاجتماعية بهدف تحليلها، وتنظيمها، والوصول من خلالها إلى معرفة تصب في مصلحة اتخاذ القرارات التَّسويقيَّة، حتى تصل المنظمات بذلك إلى إدارة المعرفة التَّسويقيَّة المستدامة، وهو ما سيتم تناوله في السِّياق اللَّاحق[28].

4- دراسات مشابهة لعملنا:

قد وجدنا بعض اعمال ومشاريع في تصنيف البيانات ونقصد بيانات الشبكات الاجتماعية ومن هذه الأعمال :

تاريخ النشر	العنوان	الكاتب	المجلة الناشرة	التقنية المستخدمة	النتائج	
2014	Automatic identification of arabic dialects in social media	1-Sadat Fetiha 2-Kazemi Farnazeh 3-Farzindar Atefeh	Proceedings of the first international workshop on Social media retrieval and analysis	Naive Bayes classifier with bi-gram model	Identify 18 AD accuracy = 98%	
2015	Improved Arabic dialect classification with social media data	Huang Fei	Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing			
2019	Classifying Arabic dialect text in the social media Arabic dialect corpus (SMADC)	1-Alshutayri Areej 2-Atwell Eric	Proceedings of the 3rd Workshop on Arabic Corpus Linguistics	Weighted voting and SMADC Weight average metric and SMADC	accuracy	
					before clean	after clean
					74%	77.6 %
					88%	90%

جدول (1.4) : دراسات مشابهة

خاتمة

في هذا الفصل قدمنا مفهوم للهجات العربية وتنوعها وتعريف بسيط لشبكات التواصل الاجتماعي وأمثلة عن اللغات واللهجات العربية المتعامل بها في مواقع التواصل ، بعد ذلك قدمنا تعريفًا لتنقيب البيانات ولتعدين النصوص و معالجة اللغة الطبيعية (NLP) كما تطرقنا إلى إعطاء فكرة حول تحليل البيانات الضخمة في الشبكات الاجتماعية وأهميتها .

الفصل الثالث:

التصميم والنهج المقترح (مجموعة البيانات DataSet والقواميس)

مقدمة

في هذا الفصل سنقدم فيه مشروعنا بدءا من جمع التعليقات ونصوص للهجات العربية المختلفة من مواقع التواصل الاجتماعي فيسبوك بشكل خاص نظرا لتوافد البشري عليه ثم سنقدم مجموعة البيانات والقواميس المهيكلة ومنسقة بشكل منظم قابل للمعالجة.

1- جمع البيانات والتصميم

1-1 تعريف مجموعات البيانات Data Set

مجموعات البيانات Data Set هي مجموعات من البيانات تتطابق بشكل عام مع محتويات جدول قاعدة بيانات مفرد ، أو مصفوفة بيانات إحصائية واحدة ، حيث يمثل كل عمود في الجدول متغيرًا معينًا، ويقابل كل صف عضوًا معينًا من مجموعات البيانات المعنية، وتسرد القيم لكل من المتغيرات، مثل ارتفاع ووزن كائن ، لكل عضو في مجموعة البيانات، تعرف كل قيمة في مجموعة البيانات باسم المسند.

1-2 مراحل العمل

1- انشاء مجموعة بيانات Data Set :جمع تعليقات من خمس لهجات الجزائرية التونسية المصرية الشامية و الخليجية .

2 - إنشاء القواميس: (قاموس لكل لهجة: الجزائرية التونسية المصرية الخليجية والشامية) .

3 - تطبيق خوارزميات التعلم الالي الخاضعة للإشراف : آلات المتجهات الداعمة (SVM) وشجرة القرار (DT) والغابة العشوائية (RF) و(Naïve Bayes (NB

1-3 مميزات مجموعة البيانات dataset

أهم ما يميز تعليقات اللهجات التي تم جمعها حيث حددنا خصائص كل لهجة والتي سنشرحها كالتالي

لهجة دول المغرب العربي (اللهجة الجزائرية والتونسية كمثال) :

- ✓ العامية غالبة كتابة اكثر من العربية حيث تستخدم الاخيرة في محادثات الرسمية.
- ✓ استخدام اللغة الهجين (العربتيني) اللغة الفرنسية واللغة الانجليزية
- ✓ الخلط بين العامية والعربية واللغات الاجنبية (الفرنسية اكثر)
- ✓ استخدام بعض الملصقات واختصارات المعبرة في كتابة
- ✓ استخدام اللغة امازيغية (الشاوية القابائلية...)

لهجة دول المشرق العربي (اللهجة الشامية والخليجية كمثال) :

- ✓ العامية غالبة في الكتابة اكثر من العربية حيث تستخدم الاخيرة في محادثات الرسمية.
- ✓ الخلط بين العامية والعربية واللغات الاجنبية (الانجليزية اكثر)
- ✓ استخدام اللغة الهجين (العربتيني) اللغة الفرنسية واللغة الانجليزية
- ✓ استخدام بعض الملصقات واختصارات المعبرة في كتابة

1-4 القواميس:

مجموعة من الكلمات تستخدم لتصنيف تعليقات كل لهجة حيث كل لهجة لديها قاموس خاص بكلمات مميزة لها وغير مكررة في اللهجات الاخرى وذلك من أجل الحصول على تصنيف أكثر دقة وتقليل نسبة الخطأ.

1 -5انشاء قاعدة بيانات التدريب dataset

العدد الكلي لتعليقات المجمعة من كل اللهجات العربية (الجزائرية التونسية المصرية الخليجية والشامية) هو7185. حيث جمع من مواقع التواصل الاجتماعي فيسبوك بدرجة اولى حيث تم جمع تعليقات ومنشورات من المجموعات الفيسبوكية في كل المجالات (الاجتماعية السياسية الثقافية ...) نظرا لتوافد البشري عليهم والكل حر في ابداء رايه و التعبير عنه بلغته حيث جمعنا من كل لهجة حسب الاحصاء التالي :

اللغة العربية	الجزائرية	المصرية	التونسية	الخليجية	الشامية
عدد التعليقات	**2000**	**1000**	**1000**	**1119**	**2066**

جدول (1.1) : عدد التعليقات المجمعة

- وقد تم العمل والتجريب على755 تعليقا واعطى نتائج جيدة و ونحن في اطار تحسين النتائج وذلك بتكبير عدد التعليقات وتحسين القواميس للحصول على دقة ونتائج افضل

➢ **تم تنسيق هذه البيانات وهذه بعض النماذج**

اللهجة	التعليقات
الجزائرية	رانا جينا خويا البشير وربحناكم ب أربعة
الجزائرية	أن شاء الله الكاس جزائرية
الجزائرية	منتخبنا الوطني بقوته الضاربة ماشاء الله هذه البداية و مازال الخير
الجزائرية	ماشاء الله شاف .وهل يخفى القمر . بالتوفيق
الجزائرية	يمسيكم بالخير كي مسيتو
الجزائرية	اناقهي ازلط ارهيغ ايلا تسرفين سجنجغ ايلى امنكاض نسان اكشقن ارشاك يلى
الجزائرية	كي راكم دايرين صفا
الجزائرية	سيد ضربوه في راسي هو راه يخرج في دم من فمو هههههه
الجزائرية	نروح نجومك شوي ونجي
الجزائرية	صباح الخير. متا لحوال نون ربي اشوني يسنجح واشومي يحفظ
الجزائرية	ينغاي لاز خسغ اذتش
الجزائرية	الديب يوكل وحدو كي موت واش يدي معاه
الجزائرية	وينا قناة رح يفوتوه

جدول (2.1) : مجموعة البيانات (dataset) للهجة الجزائرية

اللهجة	التعليقات
المصرية	حاجة تكسف مش كفاية روبى كمان
المصرية	اللهم صبرنا على ما ابتليتنا به مناظركم دى هى اللى زودت القرف الى بنشوف في الشارع
المصرية	ارحمونه الفن خرب من الاقارب والعالم
المصرية	ياراجل فن ايه هوه ده فن دي قلة

المصرية	ارحمونه الفن خرب من الاقارب والعالم
المصرية	الاخت روبى مش مكفيها اللى عاملاه جايبه لنا اختها هى المشرحة ناقصة
المصرية	مجتش عليها يعني إشمعنا هيفا طلعت أختها جت على الغلبانة إللي كل
المصرية	ماتقوله آه يقولها
المصرية	، نتفرج علي ولادك يا مصر ، . ، شرم الشيخ صيف
المصرية	ديما نسمع عن ناس بتهرب وبردو بنسمع عن ناس بتموت وبردو بنسمع ان
المصرية	ما تخليك واضح يا استاذ ما الكل عارف ان ده بيقرب للحرب ما تتكلم
المصرية	بصراحه واللى عاجبه يوافقك و اللى مش عاجبه

جدول (3.1) : مجموعةالبيانات (DataSet) للهجة المصرية

اللهجة	التعليقات
التونسية	تحددها انت، وتنجم زادا تسكروا في أي وقت حتى لو أنت خارج المنزل. طفلك قبل ما يصب أي برنامج أو لعبة لازم موافقتك أولاً عن طريق إشعار
التونسية	يجيك على تليفونك وتنجم زادا تسكر برنامج يخدم على تلفونه. يعني نظام أمان وحماية للأطفال من أضرار التكنولوجيا عشاق الكوجينة التركية و
التونسية	الفخامة والقعدات المزيانة عشاق الماعون التركي الملون و المذخم. أعمل طلة على الرابط هذ انا الجمعة الجايةباش نكتب الصداق على شكون ما
التونسية	نحبوش وما حاجتيش بيه انا لي نحبوو ما وافقوش عليه هاذي حاجة ربي .. وهذا عطوني
التونسية	وارتحيت بعدها الحالة النفسية الخايبة لي صحيح لكنت فبها اللحظات لي عشتهم وقتها صعاب ومنجمش نتفكرهم
التونسية	برشا أما الحمد لله رحمة ربي واسعة وطيبة زادا تعدا كل شي وبقى كان الحديث راهو صحيح لوغتنا هيّا إلّي
التونسية	تجمع سكان لبلاد ودنيا كل هيّا إلي تحكي على تاريخو ، بيها نقيمو نفوسنا توصل قيمة ثقافة لبلاد. في تونس
التونسية	الحرة وزينة لوغتنا نحكيو بيها أثناش مليون مافاما حتى حاجة تفرقنا تونسي ، أمّا اليوم هيّا لهجة ، موش بش
التونسية	تكون فتونس بركا معترف بها لوغة،هذاك علاش احنا منجموش نحكو بيها تكونت جمعية "دارجه" عام 2016
التونسية	بش ندافعو على لوغتنا باش نوريو قيمة لوغتنا ونرجعولها وبلاصتها عند ناس كل بلاصته معروفة زادة شبيك
التونسية	عليّ غضبانة ... محمد صالح بن عوانة السّحابات ملبّدة ، السّماء مكشّرة ، المطر هابطة غاضبةمكدّرة ، الوجوه
التونسية	بنات خاص بالبروبلمات أدخل أعمل طلة جربتش تعيش حادث
التونسية	لحضة لكن أنا تحديت ضعفي ورجع فيا الأمل وهاكا شخصية قوية وكبشت في الدنيا بسنيا وقاومت مشاكل كل

جدول (4.1) : مجموعةالبيانات (DataSet) للهجة التونسية

اللهجة	التعليقات
الشامية	عنجد عنجد اشي . .
الشامية	يا جماعة الصيام اجرو عند الله كبير لسا ما بلش
الشامية	شو رتب المعتمرين3 ليش ما حدا يعرف بهيك رحلات الا بعد السفر
الشامية	اذا كان هاد جائز دينيا ياريت نعمل زيهم لانه حرارة بلدنا صارت زي مصر . انا بوافق تعليق 1 انه بنضف الجسم الصيام بس حتى لو غيرنا الساعه رح اضل نفس الساعات لانه حتى الامساك بكون ابكر بساعه بس بنخفف ساعه من
الشامية	حرام يا جماعه والله اللي بصير حرام وذنب كبير الله يلطف
الشامية	شو الموضوع يا عالم شو اللي
الشامية	شو بدنا نحكي شو بدنا نعلق خلص تعبنا ملينا الامر لله بيده الخير وهو على كل شي قدير هو قادر على كل
الشامية	طيب ين دور الجهات المسؤولة عن المقولةالمواطن اغلى ما نملك
الشامية	الله أكبر شو
الشامية	يعني الآخرة رح يستمتع فيها
الشامية	والله لو بعرفوا شو نتائج هادي الحركات الغبية متلهم رح يبقوا طول العمر يكفروا عن
الشامية	أما غباء وأما ناس عم تقتل حالها وفكرها رح
الشامية	اللهم اشفيهم يا رب ديرو بالكو على خزاناتكو نظفوها بعدين اي شي في الصيف قابل لان يحدث هالامور الجو الحار يفسد

جدول (5.1) : مجموعة البيانات (DataSet) للهجة الشامية

اللهجة	التعليقات
الخليجية	ليش ما يكون هناك مثل معهد الدراسات الفنية للقوات الجوية حيث اثبت الخريجون كفائتهم في اصلاح طائرات القوات
الخليجية	ليش مايكون معهد الدراسات الأمنية للأمن العام دراسة 2 عامين يتخرج وكيل
الخليجية	الحل الوحيد ان الهيئة تنلغي و يقوم بمهامها قسم من أقسام الشرطة يسمى شرطة
الخليجية	صحيح فأن الحارس الكبير في سن لا يقدر على حماية البنات من اصحاب النفوس الضعيفه ولكن فيهم خير
الخليجية	فعلاً من تعاملي مع الاهل احس مخ الحرمة فيه نقص .
الخليجية	طيب ما قلتوا لنا السر :

جدول (6.1) : مجموعة البيانات (DataSet) للهجة الخليجية

6-1انشاء القواميس :

بعد انشاء dataset عملنا على انشاء القواميس حيث يحتوي كل قاموس على كلمات مميزة للهجة ولا تتكرر في لهجات اخرى وهذه بعض النماذج من قواميس

اللهجة	الكلمة
الجزائرية	خويا
الجزائرية	البشير
الجزائرية	ربحناكم
الجزائرية	يمسيكم
الجزائرية	بالخير
الجزائرية	كي
الجزائرية	راكم

الجزائرية	دايرين
الجزائرية	صفا
الجزائرية	شوي
الجزائرية	الديب
الجزائرية	يوكل
الجزائرية	واش
الجزائرية	وينا
الجزائرية	رح
الجزائرية	جبتلي
الجزائرية	مالقيتش

جدول (7.1) : قاموس اللهجة الجزائرية

اللهجة	الكلمة
التونسية	تاليفون
التونسية	تابلات
التونسية	تبّع
التونسية	هاذي
التونسية	متنساش
التونسية	تحط
التونسية	مافماشي
التونسية	شنيا
التونسية	شنوة
التونسية	متاع

التونسية	تابلات
التونسية	مينجمش
التونسية	تحدو
التونسية	تنجم
التونسية	تسكروا
التونسية	يصب
التونسية	يجيك
التونسية	الكوجينة
التونسية	المزيانة
التونسية	حاجتيش
التونسية	متاعنا

جدول (8.1) : قاموس اللهجة التونسية

اللهجة	الكلمة	
المصرية		ازاي
المصرية		معملتش
المصرية		بردو
المصرية		اوي
المصرية		مينفعش
المصرية		زي
المصرية		شباب
المصرية		كتير
المصرية		حلو
المصرية		بلاد
المصرية		بحري
المصرية		وقبلي
المصرية		القاهرة
المصرية		المدن
المصرية		مختلفة
المصرية		بشكل
المصرية		واضح
المصرية		للدرجة
المصرية		اللي
المصرية		بيها
المصرية		ما كنتش
المصرية		بعرف
المصرية		زعلت

جدول (9.1) : قاموس اللهجة المصرية

اللهجة	الكلمة
الشامية	ساعدني
الشامية	وقويني
الشامية	وساعدني
الشامية	اخد
الشامية	الغريبه
الشامية	اه
الشامية	يا
الشامية	بطني
الشامية	بس
الشامية	ولا
الشامية	مستحيل
الشامية	يطلع
الشامية	واحد
الشامية	يفشل
الشامية	نفسه
الشامية	كذا
الشامية	ولاه
الشامية	امرك
الشامية	الوهابيين
الشامية	يكرعون
الشامية	ولك
الشامية	تفيبيه
الشامية	عليك
الشامية	ولك

جدول (10.1) : قاموس اللهجة الشامية

اللهجة	الكلمة
الخليجية	احنا
الخليجية	الحين
الخليجية	بعصر
الخليجية	تكنولوجيا
الخليجية	ما
الخليجية	ينخاش
الخليجية	شي
الخليجية	الحين
الخليجية	صار
الخليجية	كلش
الخليجية	واضح
الخليجية	للكل
الخليجية	ياخذك
الخليجية	ترى
الخليجية	كلش
الخليجية	وارد
الخليجية	داحين
الخليجية	ناس
الخليجية	كلها
الخليجية	لسوته
الخليجية	قوته
الخليجية	گام
الخليجية	وشال

جدول (11.1) : قاموس اللهجة الخليجية

الخاتمة

في هذا الفصل عملنا بكثافة ومجهود كبير لجمع مجموعة البيانات(dataset) ،حيث قدمنا جزء من مجموعة البيانات التي تم جمعها وانشاءها وحددنا المميزات العامة التي تخص كل لهجة . ،كما أنشأنا قواميس لكل لهجة حيث كل قاموس يميز لهجة معينة من خلال الكلمات الخاصة بها دون غيرها . كل هذا عبارة عن خطوة اولية وسباقة للفصل القادم.

الفصل الرابع :

التنفيذ و تحليل النتائج

مقدمة

في هذا الفصل ، سنتحدث عن Python ومكتبات التعلم الآلي ، ثم خطوات البرنامج سنقوم بتحليل ومناقشة نتائج التصنيف المختلفة التي تم الحصول عليها بعد تطبيق الخوارزميات الأربعة وهي آلات ناقلات الدعم machines à vecteurs de support (SVM) ، وشجرة القرار arbre de décision (DT) والغابات العشوائية forêt aléatoire (RF) ،نايف باييز naïf Bayes (NB).

1- بايثون Python :

بايثون لغة سهلة القراءة للغاية ومتنوعة ومتعددة الاستخدامات، واسمها مستوحى من مجموعة كوميدية بريطانية باسم «Monty Python » ، وكان أُحد الأهداف الأساسية لفريق تطوير بايثون هو جعل اللغة مرحة وسهلة الاستخدام، وإعدادها بسيط، وطريقة كتابتها مباشرة وتعطيك تقريرًا مباشرًا عند حدوث أُخطاء، وهي خيار ممتاز للمبتدئين والوافدين الجدد على البرمجة لغة بايثون هي لغة متعددة الاستعمالات،وتدعم مختلف أُنماط البرمجة مثل كتابة السكربتات والبرمجة كائنية التوجه (object-oriented) وهي مناسبة للأغراض العامة، واستعمالها يتزايد في سوق العمل إذ تعتمدها منظمات مثل «United Space Alliance» (شركة في مجال إرسال مركبات فضائية وتتعاقد معها ناسا) و «Industrial Light & Magic»(أُستوديو للتأثيرات السينمائية وللرسوم المتحركة)، وتوفر بايثون قدراتٍ كثيرةٍ لمن يريد تعلم لغة برمجة جديدة. طوَرت اللغة في نهاية الثمانينات من القرن الماضي، ونِشرت أُول مرة في عام ، 1991طِورت بايثون من قبل Guido van Rossumوهو عضوٌ نشط للغاية في المجتمع وتعد بايثون على أُنها بديل عن لغة ،ABCوأُول إصدار منه كان يتضمن التعامل مع الاستثناءات exception handlingوالدوال والأصناف (classes)مع إمكانية الوراثة فيه اوعندما أُنشئ منتدى محادثة في Usenetباسم comp.lang.python في ،1994 فبدأُت قاعدة مستخدمي بايثون بالنمو، مما مهد الطريق لها لتصبح واحدة من أُكثر لغات البرمجة شيوعا وخصوصً ا لتطوير البرمجيات مفتوحةالمصدر[29]

2- أين تستخدم بايثون؟

تستخدم لغة بايثون في كل المجالات، فهي لغة برمجة متعددة الأغراض ، و من مجالات استخدامها تحليل البيانات، والروبوتات، وتعلم الآلة، وتطبيقات ،(REST) وتطوير المواقع والألعاب، و الرسوم ثلًاثية الأبعاد، والأتمتة وبرمجة الأنظمة المدمجة ، والكثير من المجالات الأخرى التي لا يسعنا حصرها هنا. تستخدم الكثير من المواقع والشركات العملًاقة لغة بايثون ، و منه (Spotify) و (Google) و ، (Amazon)إضافة إلى (Facebook) التي تستخدم بايثون لمعالجة الصور وفي كل يوم تتحول شركات جديدة إلى استخدام بايثون، مثل (Instagram) التي قررت مؤخرًا استخدامها وفضلتها على (PHP).تستخدم بايثون أُيضا من قبل بعض الجهات العلمية والبحثية، مثل وكالة الفضاء الأمريكية ناسا ، و التي لها مستودع خاص بالمشاريع المطورة ببايثون[29]

3-Anaconda :

هو توزيع مجاني ومفتوح المصدر للغات البرمجة Python و R المستخدمة على نطاق واسع في الحوسبة العلمية (علوم البيانات ، وعلوم البيانات ، والتعلم الآلي ، والعلوم ، والهندسة ، والتحليلات التنبؤية ، والبيانات الضخمة ، إلخ) يقوم بتثبيت عدد كبير من التطبيقات المستخدمة على نطاق واسع في هذه التخصصات دفعة واحدة ، بدلاً من الاضطرار إلى تثبيتها واحدة تلو الأخرى. أكثر من 1400 وهي الأكثر استخدامًا في هذه التخصصات.[30]

4-Spyder :

سبايدر (بيئة تطوير بايثون العلمية) هو بيئة تطوير تفاعلية قوية للغة بايثون .لديه ميزات تحرير متقدمة ، واختبار تفاعلي ، وتصحيح أخطاء واستبطان ، وبيئة حوسبة رقمية. بفضل دعم (IPython) (تحسين مترجم بايثون التفاعلي) ومكتبات (Python) الشهيرة مثل (NumPy) أو (SciPy) أو (matplotlib)

(2D / 3Dالتآمر التفاعلي) يمكن أيضًا استخدام (Spyder) كملف مكتبة توفر عناصر واجهة مستخدم قوية ذات صلة بوحدة التحكم لتطبيقاتنا المستندة إلى .PyQt يمكن استخدامه لدمج وحدة تصحيح الأخطاء مباشرة في تصميم واجهة المستخدم الرسومية[31]

5- مكاتب البايثون

5-1مكتبة Pandas

هي مكتبة Python أخرى هي الأنسب لمعالجة البيانات ودمجها. تُستخدم Pandas بشكل أساسي لسهولة وسرعة التلاعب بالبيانات، وتجميع البيانات ، وتصور البيانات. يستخدم Pandas لإنشاء إطارات البيانات (Python Objects) من ملف CSV[32]

5-2مكتبةMatplotlib

هي مكتبة Python مفيدة أخرى لتصور البيانات. يعد التحليل الوصفي وتصور البيانات مهم جدًا لأي منظمة. يوفر Matplotlib طرقًا مختلفة لتصور البيانات بشكل فعال. يتيح لك Matplotlib إنشاء رسوم بيانية خطية ومخططات دائرية ومخططات بيانية وأرقام احترافية أخرى بسرعة. باستخدام Matplotlib ،

يمكن للمرء تخصيص كل جانب من جوانب الشكل. يحتوي Matplotlib على ميزات تفاعلية مثل التكبير والتخطيط وحفظ الرسم البياني بتنسيق رسومات[32] .

5-3 مكتبةScikit-Learn

هي واحدة من مكتبات التعلم الآلي الأكثر ديناميكية وانتشارًا لخوارزميات ML الكلاسيكية. إنه مبني على أعلى مكتبتين أساسيتين في Python ، وهما NumPy و .SciPy يوفر Scikit-Learn الدعم لمعظم خوارزميات التعلم الخاضعة للإشراف وغير الخاضعة للإشراف. يمكن أيضًا استخدام هذه المكتبة لاستخراج البيانات ، وجمع البيانات، وتحليل البيانات، مما يجعلها أداة رائعة لتعلم ML للمبتدئين[32] .

5-4 مكتبة Seaborn

هي في الغالب مكتبة لتصور البيانات مبنية على قمة .Matplotlib تزودك هذه المكتبة بالقدرة على تنظيم المرئيات الإعلامية والإحصائية جنبًا إلى جنب مع الرسوم البيانية التوضيحية. يجعل Seaborn تصور البيانات ، وهو جزء لا غنى عنه في استكشاف البيانات وتحليلها. المكتبة هي الأنسب لفحص العلاقات بين المتغيرات المتعددة[32] .

5-5مكتبة CSV

تعد صيغة CSVأكثر صيغ الملفات شيوعًا (Comma Separated Values القيم المفصولة بفاصلة) والتي تستخدم لتصدير واستيراد البيانات من جداول البيانات spreadsheets وقواعد البيانات. وقد استخدمت هذه الصيغة لفترة طويلة قبل أي محاولة لوصف هذه الصيغة بطريقة معيارية عبر المعيار RFC .4180 إن غياب معيار محدّد ومعروف يعني وجود اختلافات طفيفة بين البيانات التي الناتجة أو المستخدمة من قبل التطبيقات المختلفة، ما يجعل من التعامل مع ملفات CSV القادمة من مصادر مختلفة أمرًا مزعجًا في بعض الأحيان. وعلى الرغم من أنّ الفواصل delimiters ومحارف علامات الاقتباس quoting characters مختلفة بين تطبيق وآخر، ولكنّ النسق العام لهذه الملفات متشابه بما يكفي لإنشاء وحدة يمكن استخدامها للتعامل مع مثل هذه البيانات بكفاءة عالية، وإخفاء تفاصيل كتابة البيانات وقرائتها عن المبرمج[33] .

6-الوسائل المادية:

جهاز كمبيوتر LENOVO وخصائصه كالتالي:

- RAM_4GO
- .PROCESSEUR_intel(R)_celeron(R)_CPU N3060
- نوع النظام : نظام تشغيل Windows

7- التنفيذ:

7-1خطوات التنفيذ:

قمنا بتطبيق الخوارزميات على مجموعة بيانات (dataset) التعليقات من اللهجات الخمسة (الجزائرية التونسية المصرية الخليجية والشامية)

اولا :

1- التصريح بالمكتبات اللازمة:

```
ML_Dialects.py

# In[1]:
import csv
from   operator import add
import pandas as pd
from   sklearn.svm import SVC
#from   sklearn.externals import joblib
from   sklearn.model_selection import train_test_split
from   sklearn.tree import DecisionTreeClassifier
from   sklearn.ensemble import RandomForestClassifier
from   sklearn.naive_bayes import MultinomialNB
from   sklearn.metrics import accuracy_score
from   sklearn.metrics import precision_recall_fscore_support

#########################################################################
#                                                                       #
#                          L'ensemble des                               #
import seaborn as sns
import matplotlib.pyplot as plt
#########################################################################
```

2 - قراءة مجموعة البيانات dataset للمعالجة:

```
# All Data set
Corpus = pd.read_csv(r"C:/Data/Arabic_dialects/Dict/12.csv" ,encoding = 'UTF-8' , engine='python' )
```

3 - قراءة القواميس:

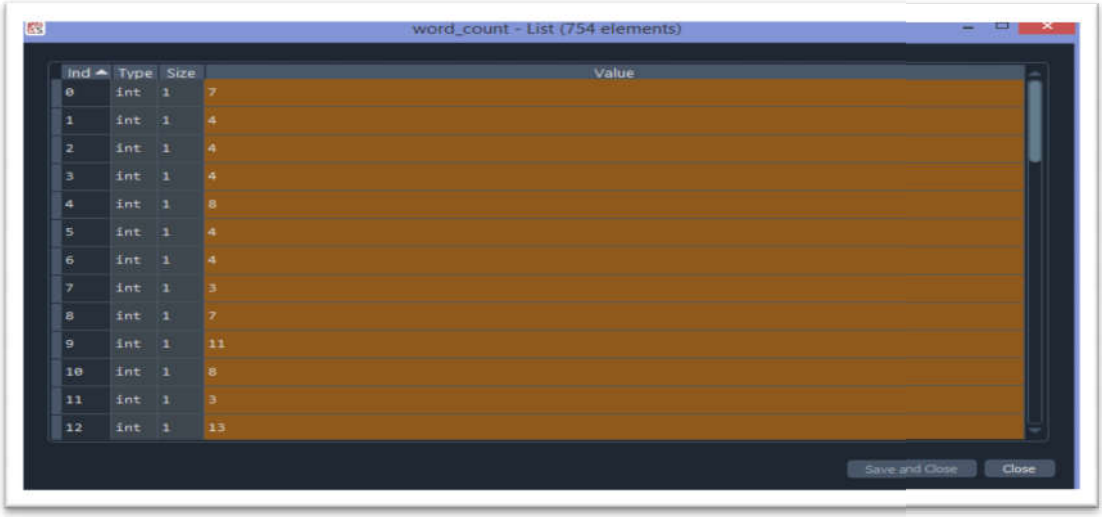
```
# In[3]:
# Features of Algerian Dialect
Dect_Dziri      = pd.read_csv(r"C:/Data/Arabic_dialects/Dict/Dziri.csv"     , header=0, sep='delimiter')
# Features of Tunizian Dialect
Dect_Tunsi      = pd.read_csv(r"C:/Data/Arabic_dialects/Dict/Tunsi.csv" , header=0, sep='delimiter')
# Features of Egytsian Dialect
Dect_Masri      = pd.read_csv(r"C:/Data/Arabic_dialects/Dict/Masri.csv"  , header=0, sep='delimiter')
# Features of Shamist Dialect
Dect_Shami      = pd.read_csv(r"C:/Data/Arabic_dialects/Dict/Shami.csv"   , header=0, sep='delimiter')
# Features of Khalijist Dialect
Dect_Khalidji   = pd.read_csv(r"C:/Data/Arabic_dialects/Dict/Khalidji.csv"  , header=0, sep='delimiter')
```

ثانيا :

1- حساب عدد الكلمات في كل تعليق من مجموعة البيانات

لحساب عدد الكلمات في كل تعليق من مجموعة البيانات dataset تقوم الدالة () Split بالفصل بين الكلمة و الكلمة في التعليق اعتمادا على الفراغ و تعرض النتيجة في جدول Word_count حيث يعطينا عدد الكلمات الموجودة في كل تعليق بعد تقسيمه.

```
80  # In[4]
81  # Calculate the number of words in each comment from data set
82  word_count = []
83  for i in Corpus.Comment:
84      word_count.append(len(i.split()))
85
86  Corpus["word_count"] = word_count
87
```



- هنا تتم اضافة العمود Word_count في جدول Corpus.

2- إنشاء جدول Corpus_words

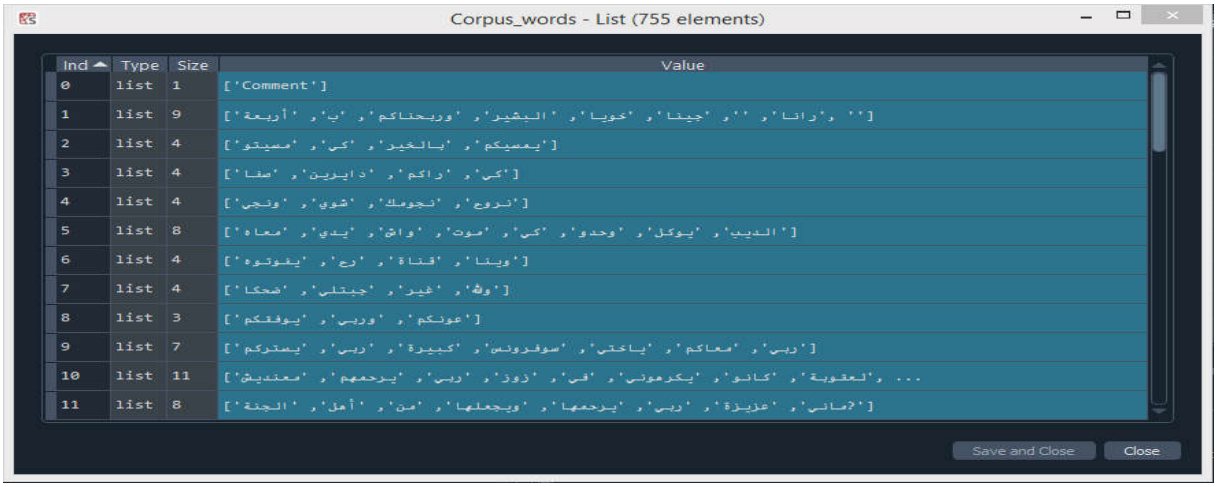
بعد تنفيذ يقوم البرنامج بإنشاء جدول Corpus_words بحيث يفصل كل تعليق كلمة بكلمة.

```
# In[ 5]:

Corpus_words       =[]

csv_reader = csv.reader(open(r'C:/Data/Arabic_dialects/Dict/12.csv' , encoding="utf-8" ), delimiter=',')
arr = list(csv_reader)
for line in arr:
        Corpus_words.append(line[0].split(" "))
```

- النتيجة



- تقسيم كلمات قواميس الى وحدات منفصلة

```
# In[6]:
Dect_Dziri_words      =[]
Dect_Tunsi_words      =[]
Dect_Masri_words      =[]
Dect_Shami_words      =[]
Dect_Khalidji_words   =[]

csv_reader = csv.reader(open(r"C:/Data/Arabic_dialects/Dict/Dziri.csv" , encoding="utf-8" ), delimiter=',')
arr = list(csv_reader)
for line in arr:
    Dect_Dziri_words.append(line[0].split(" "))
```

- النتيجة

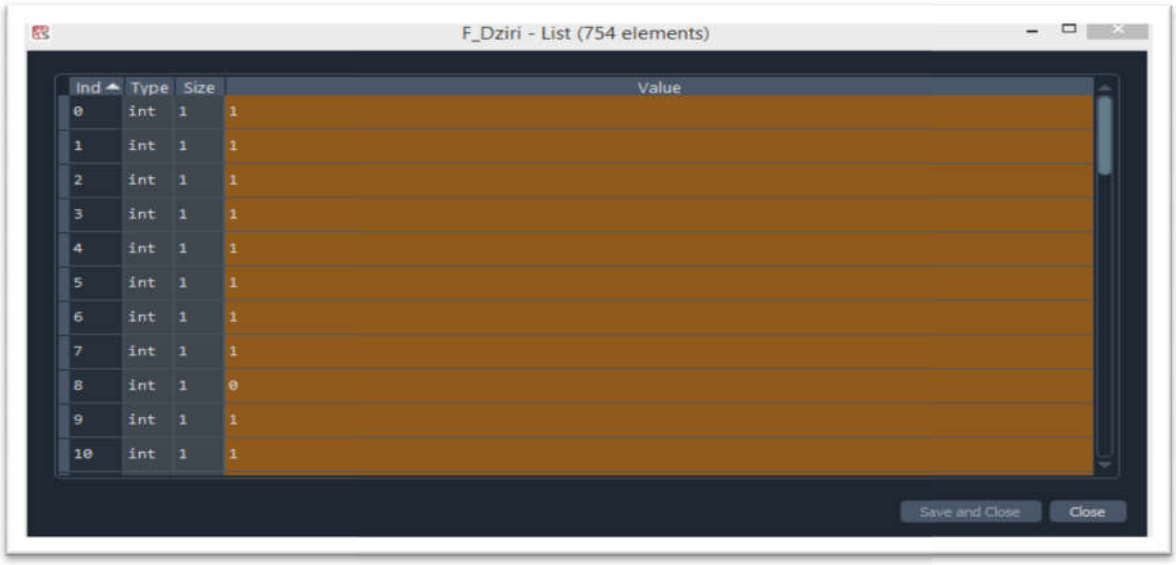
3- يتم مسح جدول تعليقات Corpus على جدول القواميس حيث تتم مقارنة بين بيانات قاموس (كلمات) وكل تعليق حيث تعطينا 1 اذا تطابقت مقارنة و 0 اذا لم تتطابق ثم يتم ادخال هذه قيم في عمود في جدول Corpus

```
# In[14]:

F_Dziri     = []
F_Tunsi     = []
F_Masri     = []
F_Shami     = []
F_Khalidji  = []
#.....................................................................................
for t in Corpus.Comment:
    if any(w in t.split() for w in Dect_Dziri.text):
        F_Dziri.append(1)
    else :
        F_Dziri.append(0)

Corpus["One_F_Dziri"] = F_Dziri
```

- النتيجة



4- بعد إيجاد التطابقات نقوم بمعالجة على مستوى قاموس كل لهجة ثم إدخال النتيجة في أعمدة جدول مجموعة البيانات Corpus حيث لدينا هنا اعمدة Dziri_count و Tunsim_countو Masr_count و P_Shamim_count و Khalidji_count

```
# In[22]:

CW_Dziri      = []
CW_Tunsi      = []
CW_Masri      = []
CW_Shami      = []
CW_Khalidji   = []


#.....................................................................................

count = 0
for t in Corpus.Comment:
    for word in t.split():
        for word1 in Dect_Dziri.text:
            if word == word1:
                count +=1
    CW_Dziri.append(count)
    count = 0
fn = []
fn= list( map(add, CW_Dziri, F_Dziri) )

Corpus["Dziri_count"] = fn

#.....................................................................................
```

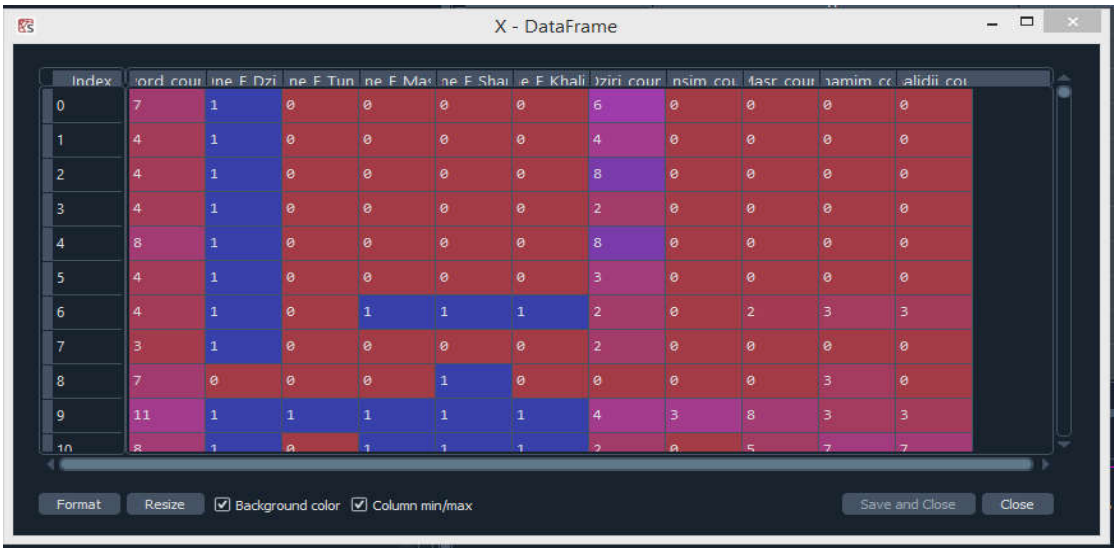
- النتيجة

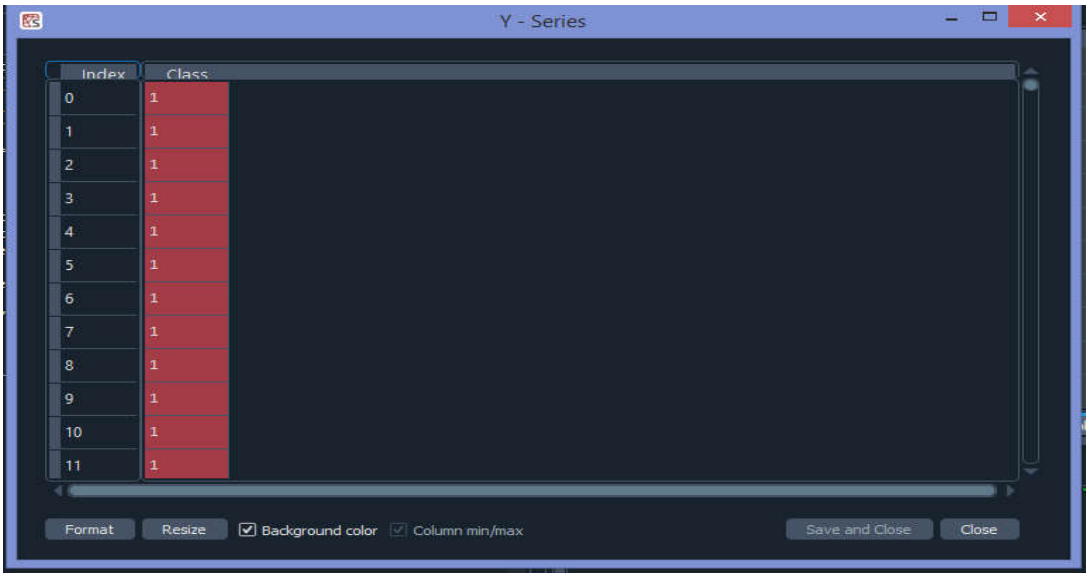
5- التدريب والتجريب

تدريب الآلة والتجريب لاخذ نتائج: يتم تدرب النظام على القيم والبيانات ثم تأتي مرحلة التجريب test

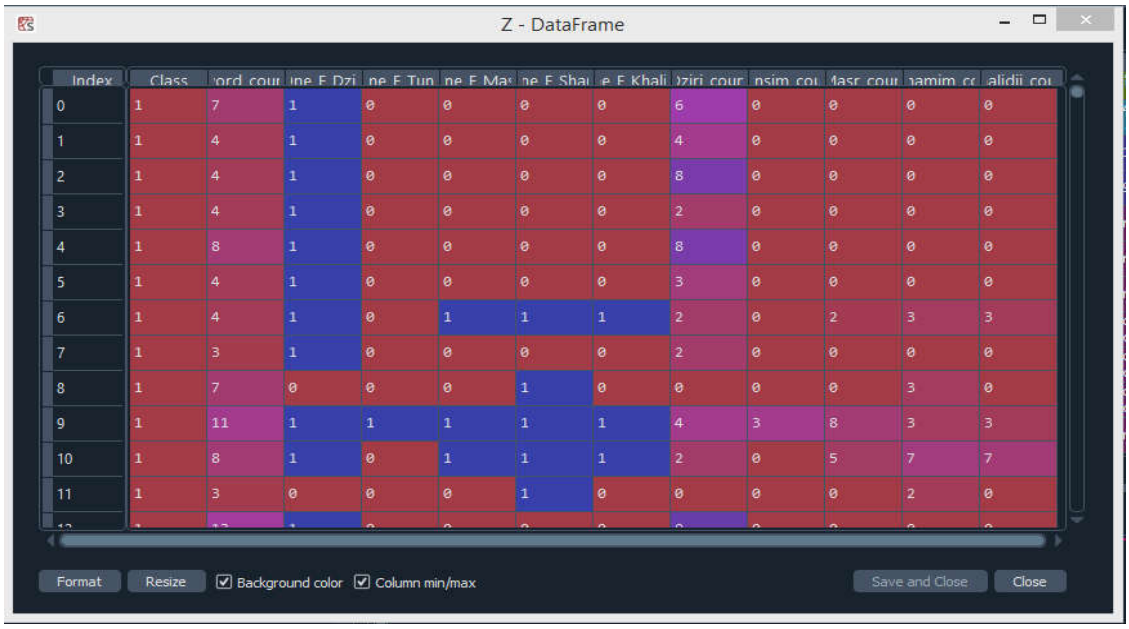
```
# In[]:

Z = Corpus.get(Corpus.columns[2:])
Y = Corpus.get(Z.columns[0])
X = Corpus.get(Z.columns[1:])
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=4)
```





– يتم اخذ قيم عمود Class من Corpus واضافتها في مصفوفة Z حيث ان الدالة () get تعطي القيم العددية ل copus، وهي قيم X و والقيم Y هي قيم Class



– ثم نقوم بتمرير البيانات عبر الخوارزميات الأربعة على النحو التالي

```
# In[ ]:

svm = SVC()
svm.fit(x_train, y_train)
acc1_svm = svm.score(x_test, y_test)
pr1_svm = svm.predict(x_test)
ac = accuracy_score(y_test, pr1_svm)
ww1 = precision_recall_fscore_support(y_test, pr1_svm, average='macro')

# In[ ]:

dt = DecisionTreeClassifier()
dt.fit(x_train, y_train)
acc2_dt = dt.score(x_test, y_test)
pr2_dt = dt.predict(x_test)
ww2 = precision_recall_fscore_support(y_test, pr2_dt, average='macro')
ac2 = accuracy_score(y_test, pr2_dt)

# In[ ]:

rf = RandomForestClassifier(n_estimators=40)
rf.fit(x_train, y_train)
acc3_rf = rf.score(x_test, y_test)
pr3_rf = rf.predict(x_test)
ww3 = precision_recall_fscore_support(y_test, pr3_rf, average='macro')
ac3 = accuracy_score(y_test, pr3_rf)

# In[ ]:

NB = MultinomialNB()
NB.fit(x_train, y_train)
acc4_NB = NB.score(x_test, y_test)
pr4_NB = NB.predict(x_test)
ac4 = accuracy_score(y_test, pr4_NB)
ww4 = precision_recall_fscore_support(y_test, pr4_NB, average='macro')
```

النتيجة

- نتحصل على دقة كل خوارزمية كالتالي

Name	Type	Size	Value
ac	float64	1	0.7019867549668874
ac2	float64	1	0.6423841059602649
ac3	float64	1	0.6490066225165563
ac4	float64	1	0.5960264900662252

- نلاحظ أن دقة SVMهي الافضل 0.7019867549668874 لذلك SVM هي أفضل خوارزمية

8-مقاييس تصنيف الدقة

في عملنا ، بعد تنظيف مجموعة البيانات ، وتمرير مجموعة البيانات وفقًا لتقنية البرمجة اللغوية العصبية المقترحة ، مررنا البيانات الرقمية التي تم الحصول عليها مع ميزاتها إلى مجموعة الخوارزميات الذكية الخطية. للوصول إلى تقييم جيد لخوارزمية التعلم الآلي لعملنا ، ومنحنا دقة تصنيف نتائج جيدة عند التقييم ، نستخدم المقاييس الثلاثة التالية:

8-1 الدقة: مقياس يعطي للخوارزمية دقة التدريب أو الاختبار على مجموعة البيانات. وفقًا لـ ISO 5725-1 [34] إنه مصطلح عام لوصف مدى قرب القياس من القيمة الحقيقية. يتضمن تطبيق المصطلح على مجموعة بيانات ذات طبيعة مماثلة مكون خطأ عشوائي ومكون خطأ منهجي. وتحسب على النحو التالي:

$$\text{Accuracy} = \frac{(\text{truepositives} + \text{truenegative})}{(\text{truepositives} + \text{truenegative} + \text{falsenegative} + \text{truenegative})}$$

8-2 F1_score:

يستخدم لتقييم المعلومات وأنواع عديدة من نماذج التعلم الآلي ، ولا سيما في معالجة اللغة الطبيعية. تعني درجة F1 أن النموذج مثالي. ويحسب كالتالي: هو المتوسط التوافقي للدقة والاسترجاع[35].

$$F1 = 21recall \times 1precision = 2 \times precision \times recallprecision + recall$$

حيث الدقة = truepositives/ (truepositives+falsenegative)

و $Recall$ = truepositives + falsenegative

```
335
336 # In[]
337
338 text_classifier = SVC(probability=True)
339 text_classifier.fit(x_train, y_train)
340
341 # In[]
342 predictions = text_classifier.predict(x_test)
343 # In[]
344 print(confusion_matrix(y_test,predictions))
345 print(classification_report(y_test,predictions))
346 print(accuracy_score(y_test, predictions))
347
348
```

DecisionTree ✓

```
              precision    recall  f1-score   support

           1       0.67      0.76      0.71        41
           2       0.71      0.57      0.63        42
           3       0.65      0.71      0.68        28
           4       0.45      0.50      0.48        20
           5       0.67      0.60      0.63        20

    accuracy                           0.64       151
   macro avg       0.63      0.63      0.63       151
weighted avg       0.65      0.64      0.64       151

0.6423841059602649
```

SVM ✓

```
              precision    recall  f1-score   support

           1       0.86      0.73      0.79        41
           2       0.76      0.69      0.72        42
           3       0.79      0.68      0.73        28
           4       0.41      0.75      0.53        20
           5       0.76      0.65      0.70        20

    accuracy                           0.70       151
   macro avg       0.72      0.70      0.69       151
weighted avg       0.75      0.70      0.71       151

0.7019867549668874
```

✓ **RandomForest**

	precision	recall	f1-score	support
1	0.70	0.80	0.75	41
2	0.71	0.52	0.60	42
3	0.56	0.68	0.61	28
4	0.48	0.50	0.49	20
5	0.67	0.60	0.63	20
accuracy			0.64	151
macro avg	0.62	0.62	0.62	151
weighted avg	0.64	0.64	0.63	151

0.6357615894039735

✓ **Naïve Bayes NB**

	precision	recall	f1-score	support
1	0.67	0.85	0.75	41
2	0.80	0.67	0.73	42
3	0.43	0.21	0.29	28
4	0.23	0.30	0.26	20
5	0.62	0.75	0.68	20
accuracy			0.60	151
macro avg	0.55	0.56	0.54	151
weighted avg	0.60	0.60	0.58	151

0.5960264900662252

استنتاج

من خلال هذه النتائج المتحصل عليها نجد نسب pricision في خوارزميات 4 تتراوح بين 0.86% و0.23 وافضل نسب تكون في DT وSVM حيث نتائج DT فلهجة الجزائرية(1) 0.67 و والتونسية(2) =0.71 و بالنسبة F1score كانت ايضا نتائجه عالية في DT بنسب 0.71 للجزائرية و0.63 لتونسية ودقة تصل الى 0.64

و SVM تحصلناعلى نسب pricision في الجزائرية 0.86 و التونسية 0.76 و بالنسبة ل f1score كانت نسب 0.79 للجزائرية و0.72 لتونسية ودقة قدرت ب0.70 ومنه نجد ان SVM هو احسن نموذج.

الخاتمة

في هذا الفصل قدمنا لمحة صغيرة عن لغة Python ومكتباتها الأكثر أهمية ، و قمنا بتطبيق 4 خوارزميات تعلم تحت الإشراف في مجموعة البيانات وعرض نتائجها التي اوضحت ان خوارزمية SVM هي افضل في نتائج والدقة.

الخاتمة العامة

الخاتمة العامة

الغرض من هذا العمل هو تحقيق نهج ذكي لتصنيف التعليقات حسب اللهجات العربية في الشبكات الاجتماعية. لتطبيق هذا النهج ، استخدمنا برنامج Python لأنه يحتوي على مكتبات تعلم الآلة ويقوم بمعالجة ملفات مختلفة. في الفصل الأول ، تعرفنا على الذكاء الاصطناعي وواهميته ودوره و أهم مجالاته ، وتطرقنا الى التعلم الآلي و أنواعه(التعلم الآلي الخاضع للإشراف وغير الخاضع للإشراف) ، وخوارزمياته في الفصل الثاني قدمنا لمحة عن اللهجات العربية وانواعها واستخدامتها في مواقع التواصل الاجتماعي كما قمنا بتعريف مصطلح الشبكات الاجتماعية وتحدثنا عن التنقيب في البيانات و طرق تمثيلها وكما اشرنا الى تعدين النصوص وهيكلتها وتقنياتها، و كذلك دور معالجة اللغة الطبيعية NLP في الفصل الثالث ، بدأنا في تنفيذ النهج ذكي ، حيث بدأنا في جمع البيانات في ملفات csv ،، مما سمح لنا بإنشاء القواميس ، قاموس لكل لهجة(الجزائرية التونسية المصرية الخليجية والشامية) في الفصل الرابع : بدانا بالتنفيذ باستخدام Python ، حيث قمنا بتنزيل المكتبات المطلوبة ، ، والمطبقة باربعة خوارزميات آلات ناقلات الدعم machines à vecteurs de support (SVM) ، وشجرة القرار arbre de décision (DT) و الغابات العشوائية forêt aléatoire (RF) ، باييز الساذج (Bayes naïf NB)،، ونفذنا البرنامج بعدة مراحل ووجدنا أن SVM هو الذي يعطينا أفضل نتيجة .

قائمة المراجع

قائمة المراجع :

[1] أ.د. محمد محمد الهادي . تاثير الذكاء الاصطناعي واثاره على العمل والوظائف

[2] أحمد كاظم .الذكاء الاصطناعي .سنة 2012 .

[3] الموقع الالكتروني الذكاء الاصطناعي https://www.twinkl.com

[4] محمد لحلح . مدخل إلى الذكاء الاصطناعي وتعلم الآلة .

[5] الموقع الالكتروني مقدمة في الذكاء الاصطناعي https://www.t2.sa

[6] الموقع الالكتروني انواع الذكاء الاصطناعي https://www.marketin8.com

[7] الموقع الالكتروني مجالات الذكاء الاصطناعي https://www.arageek.com

[8] الموقع اهمية الذكاء الاصطناعي وتعلم الالة في الاعمال https://www.affdu.com

[9] الموقع الالكتروني اهمية الذكاء الاصطناعي . جواهر الخالدي https://mufahras.com

[10]الموقع التعلم الآلي الخاضع للإشراف وغير الخاضع للإشراف https://ar.strephonsays.com

[11] الموقع خوارزميات التعلم الآلي التي عليك معرفتها. 2021/12/21 https://e3arabi.com

[12]الموقع الالكتروني أفضل عشر خوارزميات لتعلم الآلة https://arabicprogrammer.com

[13] https://www.applause.com Training Data vs.Validation Data vs.Test Data

[14] الموقع الالكتروني التعلم الالي https://www.marefa.org

[15] مذكرة اللهجات العربية الحديثة وعلاقتها باللغة العربية الفصحى لهجة قسنطينة أنموذجا.عبير دروش.مريم بـــــوصيــد.2019/2018

[16] الموقع الالكتروني التعريف باللغة العربية. https://www.noor-book.com

[17] الموقع الالكتروني الفرق بين اللغة واللهجة https://arbtech.ahlamontada.com

[18] الموقع الالكتروني لهجات عربية . https://areq.net

[19] الموقع الالكتروني https://ar.wikipedia.org

[20] الاستعمال اللغوي في وسائل التواصل الاجتماعي عند الشباب العربي الواقع والأسباب والآثار. د،صافية كساس.2019/07/15

[21] تنقيب بيانات وسائل التواصل الإجتماعى واستخداماته فى البحوث الإعلامية ... تحليل المشاعر نموذجا.د طارق الخليفي

[22] الاخطاء اللغوية واسلوبية في مواقع التواصل الاجتماعي وتاثيرها في سلامة اللغة العربية (الفيسبوك انموذجا).د.رياض محمد كاظم

[23] مذكرة machine learning for religious communities detection on social network.adaika khaoula,dou youmna ,2020/2019

[24] الموقع الالكتروني ما هو تويتر https://teb21.com

[25] تهجين اللغة العربية في مواقع التواصل الاجتماعي . سالمة شداني. جامعة الجزائر 3.الجزائر

[26] الموقع الالكتروني https://www.ibm.com text mining

[27] مذكرة : تحليل المشاعر: نهج التعلم العميق. اروى رزاق .أميمة رزاق.2019/2020

[28] تحليل البيانات الضخمة من الشبكات الاجتماعية كعامل تمكين لادارة المعرفة المستدامة التسويقية للاتجاهات. شهد عبد الحليم العشي. ضحى هزاع بادي

[29] البرمجة بلغة البايثون تعلم البرمجة وكتابة البرامج وتنقيحها بلغة بايثون. ليزا تاغليفيري.2020

[30] الموقع الالكتروني. anaconda. https://www.ikkaro.com

[31]الموقع الالكتروني spyder, https://ubunlog.com

[32] الموقع الالكتروني أفضل 10 مكتبات لعلوم البيانات في بايثون. https://coursee.org

[33] الموقع الالكتروني https://wiki.hsoub.com .python/csv

[34] BS ISO 5725-1: «Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions», p.1 (1994)

[35] Sasaki, "The truth of the F-measure", Y, 2007