الجمهورية العربية السورية
جامعة دمشق
كلية الهندسة الميكانيكية والكهربائية
قسم هندسة الحواسيب والأتمتة

تطوير شبكة أكاديمية اجتماعية معتمدة على الحوسبة السحابية

رسالة أُعدّت لنيل درجة الماجستير في هندسة الحواسيب وشبكاتها

إعداد: م. أمجد علي مخلوف

المشـــــرف: أ.د.م. سمير كرمان

2023_2024

شكر وتقدير

أتوجه بجزيل الشكر والامتنان والتقدير:

إلى أستاذي ومشرفي الذي منحني الوقت والجهد والإرشاد والتشجيع وقدم لي الدعم الكامل والملاحظات القيمة خلال هذه الفترة الدراسية لأتمكن من إنهاء هذا العمل بالشكل السليم **الأستاذ الدكتور سمير كرمان.**

إلى من كان له فضل لا ينسى وأمدني بالعون والملاحظات القيمة في جميع الخطوات **الدكتور رائوف حمدان.**

إلى من قدم التوجيهات المناسبة لإتمام العمل وكان خير معين **الدكتور مأمون يونس.**

إلى الصرح العظيم **جامعة دمشق.**

خالص الشكر و الحب و الامتنان إلى جميع **أعضاء الهيئة التدريسية في قسم الحواسيب والأتمتة** - كلية الهندسة الميكانيكية والكهربائية.

وآمل أن يرقى هذا العمل ليكون بداية لما فيه الخير والمنفعة.

فهرس المحتويات

ملخص: ..... - 11 -

1- الفصل الأول: الإطار العام للدراسة ..... - 13 -

1-1- مقدمة: ..... - 14 -

1-2- المشكلة العلمية للبحث: ..... - 14 -

1-3- تساؤلات البحث: ..... - 14 -

1-4 - الهدف من البحث: ..... - 15 -

1-5- أهمية البحث: ..... - 15 -

1-6- مبررات البحث: ..... - 15 -

1-7- محدوديات البحث: ..... - 15 -

1-8- الدراسات المرجعية: ..... - 16 -

2- الفصل الثاني: الشبكات الاجتماعية الأكاديميةACADEMIC SOCIAL NETWORKS ASNS..... - 18 -

2-1 - مقدمة[1] : ..... - 19 -

2-2- تفاصيل وظائف بعض ASNs: ..... - 19 -

2-2-1- ResearchGate (RG): ..... - 19 -

2-2-2- Academia: ..... - 19 -

2-2-3- Mendeley : ..... - 20 -

2-2-4- Zotero : ..... - 20 -

2-3- مقارنة بين ACADEMIC SOCIAL NETWORKS ASNs الحالية: ..... - 20 -

2-4- لمحة تاريخية عن ACADEMIC SOCIAL NETWORKS ASNs: ..... - 23 -

2-5- علاقة المنصات بالنشر المفتوح الوصول: ..... - 24 -

2-6- METRICS ..... - 26 -

2-7- التركيبة السكانية والمنصة الاجتماعية: ..... - 27 -

3- الفصل الثالث: الحوسبة السحابية CLOUD COMPUTING: ..... - 29 -

3-1- مقدمة ..... - 30 -

3-2- أنواع نشر الحوسبة السحابية: ..... - 30 -

3-2-1- السحابة العامة public cloud: ..... - 30 -

3-2-2- السحابة الخاصة private cloud: ..... - 30 -
3-2-3- السحابة الهجينة hybrid cloud: ..... - 31 -
3-2-4- سحابة المجتمع Community Cloud: ..... - 31 -
3-2-5- السحابة البينية Inter Cloud: ..... - 31 -
3-2-6- المزايا والعيوب لنماذج التطوير السحابي: ..... - 31 -
3-3- خدمات الحوسبة السحابية: ..... - 33 -
3-3-1- البنية التحتية كخدمة (IaaS): ..... - 33 -
3-3-2- النظام الأساسي كخدمة (PaaS): ..... - 34 -
3-3-3- البرمجيات كخدمة (SaaS): ..... - 34 -
3-4- مكونات السحابة: ..... - 35 -
3-4-1- مستهلك السحابة: ..... - 35 -
3-4-2- مزود السحابة: ..... - 35 -
3-4-3- مدقق السحابة: ..... - 36 -
3-4-4- وسيط السحابة: ..... - 36 -
3-4-5- الناقل السحابي: ..... - 37 -
3-5- المفاهيم المرتبطة بالحوسبة السحابية: ..... - 37 -
3-5-1- الحوسبة السحابية الموزعة: ..... - 37 -
3-5-2- مقاييس موازنة تحميل السحابة: ..... - 39 -
3-6- فوائد الحوسبة السحابية للتعليم عن بعد: ..... - 41 -
3-6-1- تشجيع التعاون والتواصل: ..... - 41 -
3-6-2- التكلفة المادية: ..... - 41 -
3-6-3- حفظ السجلات: ..... - 41 -
3-6-4- وسائل الراحة: ..... - 41 -
3-6-5- سهولة التنفيذ: ..... - 42 -
3-6-6- الإمكانية: ..... - 42 -
3-6-7- الوصول إلى السعة القصوى لتكنولوجيا المعلومات: ..... - 42 -
3-6-8- إعادة توزيع المعلمين: ..... - 42 -
3-6-9- القابلية للقياس: ..... - 42 -

4- الفصل الرابع: تعليم الآلة MACHINE LEARNING:..... - 43 -

4-1- مقدمة:..... - 44 -

4-2- تصنيفات تعليم الآلة: ..... - 44 -

4-2-1- التعلم الخاضع للإشراف Supervised Learning: ..... - 44 -

4-2-2- التعلم غير الخاضع للإشراف Unsupervised Learning: ..... - 44 -

4-2-3- التعلم شبه الخاضع للإشراف Semi-supervised Learning:..... - 45 -

4-2-4- التعلم المعزز Reinforcement Learning : ..... - 45 -

4-3- المصطلحات المستخدمة في مفاهيم تعلم الآلة: ..... - 46 -

4-4- أنواع خوارزميات تعليم الآلة: ..... - 47 -

4-4-1- خوارزميات الانحدار: ..... - 47 -

4-4-2- خوارزميات التصنيف..... - 47 -

4-4-3- خوارزميات التجميع(Clustering): ..... - 49 -

4-5- تقييم خوارزميات تعلم الآلة:..... - 49 -

4-5-1- الدقة Accuracy: ..... - 50 -

4-5-2- مصفوفة الارتباك Confusion Matrix: ..... - 50 -

4-5-3- الدقة والاسترجاع Precision and Recall: ..... - 50 -

4-5-4- مقياسF (F Measure):..... - 51 -

4-5-5- مقاييس الانحدار Regression Metrics: ..... - 51 -

4-5-6- التحقق المتقاطع k-fold (k-Fold Cross-validation):..... - 52 -

4-5-7- التحقق المتقاطع k-fold المقسم إلى طبقات(Stratified k-Fold Cross-validation):. - 52 -

4-5-8- متوسط الخطأ التربيعيMean Squared Error ..... - 52 -

4-5-9- متوسط الخطأ المطلق MAE (Mean Absolute Error): ..... - 52 -

4-6- طرق لزيادة دقة خوارزميات تعليم الآلة (النماذج): ..... - 53 -

4-6-1- زيادة حجم البيانات:..... - 53 -

4-6-2- معالجة القيم المفقودة والشاذة:..... - 53 -

4-6-3- هندسة الخصائص:..... - 53 -

4-6-4- اختيار الميزة:..... - 54 -

4-6-5- اختبار الخوارزميات المتعددة:..... - 55 -

4-6-6- ضبط الخوارزمية: ..... - 55 -
4-6-7- طرق التجميع: ..... - 55 -
4-6-8- التحقق من الصحة المتقاطعة: ..... - 55 -
-5 الفصل الخامس: تضمين الكلماتWORD EMBEDDING ..... - 56 -
5-1- المقدمــة INTRODUCTION: ..... - 57 -
5-2- تمثيل الكلمات باستخدام ترميز :ONE HOT ENCODING ..... - 57 -
5-3 - مساوئ تمثيل الكلمات باستخدام ترميز ONE HOT ENCODING..... - 58 -
5-4- تضمين الكلماتWORD EMBEDDING ..... - 58 -
5-5- نماذج الكلمة إلى متّجه WORD2VEC : ..... - 60 -
5-6 - نموذجا تخطّي الكلمات SKIP-GRAM وحقيبة الكلمات المستمرّةCBOW ..... - 61 -
5-7- تفاصيل نموذج تخطّي الكلماتSKIP-GRAM ..... - 62 -
5-8 - الطبقة الخفيّة لنموذج تخطّي الكلماتSKIP-GRAM ..... - 63 -
5-9- طبقة الخرج لنموذج تخطّي الكلماتSKIP-GRAM ..... - 65 -
5-10- تدريب نموذج تخطّي الكلماتSKIP-GRAM ..... - 66 -
5-11- نموذج حقيبة الكلمات المستمرّةCBOW ..... - 66 -
6- الفصل السادس: - 69 - بنية المنظومة المقترحة: ..... - 69 -
6-1- مقدمة: ..... - 70 -
6-2- البنية العامة للمنظومة..... - 70 -
6-3- المواثقة: ..... - 71 -
6 -3-1- آلية عمل بروتوكول الـ oAuth؟..... - 72 -
6-3-2- طرق التسجيل في المنظومة:..... - 73 -
6-4- أنواع الحسابات من حيث الصلاحيات: ..... - 73 -
7- الفصل السابع: التطبيقات المقترحة ضمن المنظومة المراد تنفيذها ..... - 75 -
7-1- مقدمة:..... - 76 -
7-2- برنامج المحادثة CHATTING APPLICATION:..... - 76 -
7-2-1 بنية قاعدة البيانات: ..... - 77 -
7-3- برنامج المؤتمرات CONFERENCE APPLICATION:..... - 79 -

7-3-1- بنية قاعدة البيانات: ..... - 79 -
7-4- برنامج الاستفسارات QUESTIONS APPLICATION: ..... - 80 -
7-4-1- بنية قاعدة البيانات: ..... - 81 -
7-5- تطبيق مراكز التسجيل REGISTERATION PLACES MANAGER APPLICATION: ..... - 83 -
7-5-1- بنية قاعدة البيانات: ..... - 84 -
7-6- تطبيق الصلاحيات:PERMISSIONS PROGRAM : ..... - 84 -
7-6-1- بنية قاعدة البيانات: ..... - 84 -
7-7- تطبيق المقالاتDOCUMENTS PROGRAM : ..... - 85 -
7-7-1- بنية قاعدة البيانات: ..... - 86 -
7-8- تطبيق التحكيم JUDGING PROGRAM: ..... - 87 -
7-8-1- بنية قاعدة البيانات: ..... - 88 -
8- الفصل الثامن: تصميم نموذج تعلم آلي لتصنيف المقالات العلمية..... - 90 -
8-1- المقدمة: ..... - 91 -
8 -2- أهمية النموذج المقترح وأسباب القيام به: ..... - 91 -
8-3- النموذج المقترح: ..... - 92 -
8-3-1- بنية النموذج المقترح: ..... - 92 -
8-3-2- تجميع العينات: ..... - 93 -
8-3-3- استخراج السمات عبر تقنية Word2Vec :[25] ..... - 93 -
8-3-4- تطبيق خوارزمية MLP في التصنيف[27] ..... - 95 -
8-4- العمل المنجز: ..... - 95 -
8-4-1- دراسة الأطر والدراسات المرجعية ذات الصلة: ..... - 95 -
8-4-2- تجميع وتحضير العينة: ..... - 96 -
8-4-3- تجهيز وبرمجة تقنية Word2Vec: ..... - 97 -
8-4-4- تجهيز وبرمجة خوارزمية MLP: ..... - 97 -
8-4-5- تحضير وتجهيز مجموعة التدريب والاختبار: ..... - 98 -
8-4-6- ضبط بارامترات النموذج: ..... - 98 -
8-4-7- المقياس: ..... - 98 -
8-5- النتائج والمناقشة: ..... - 99 -

8-5-1- الآفاق المستقبلية لتطوير البحث: ..... - 102 -

9- المراجع REFERENCES :..... - 103 -

فهرس الأشكال

الشكل 1: مقارنة بين ASNs الحالية من حيث نسب الاستخدام حسب الدول ..... - 21 -
الشكل 2: مقارنة بين ASNs حسب عدد المستخدمين والمشاهدات وتحميل البيانات ..... - 22 -
الشكل 3: لمحة تاريخية عن ASNs Academic Social Networks:..... - 23 -
الشكل 4: مزايا وعيوب السحب العامة والخاصة والمختلطة ..... - 32 -
الشكل 5: بنية السحابة حسب الخدمات المقدمة..... - 32 -
الشكل 6: تكوين الحوسبة السحابية NIST ..... - 37 -
الشكل 7: ترميز الكلمات باستخدام One Hot Encoding ..... - 57 -
الشكل 8: جدول يوضح مميّزات كلّ كلمة..... - 59 -
الشكل 9: العلاقة بين الكلمات ..... - 60 -
الشكل 10: بنية نموذجي تخطّي الكلمة و حقيبة الكلمات المستمرّة..... - 61 -
الشكل 11: بعض الأمثلة عن عيّنات تدريب نموذج skip-Gram..... - 62 -
الشكل 12: الشبكة العصبونيّة لنموذج تخطي الكلمات ..... - 63 -
الشكل 13: معماريّة الشبكة العصبونيّة لنموذج تخطّي الكلمات ..... - 64 -
الشكل 14: أوزان الطبقة الخفيّة لشبكة النموذج تخطّي الكلمات..... - 64 -
الشكل 15: تحديد صفّ الفهرس فقط بقيمة البعد 1 في المتّجه المقابل في المصفوفة..... - 65 -
الشكل 16: خرج الشبكة العصبونيّة في نموذج تخطّي الكلمة..... - 66 -
الشكل 17: بعض الأمثلة عن عيّنات تدريب نموذج حقيبة الكلمات المستمرّة..... - 67 -
الشكل 18:بنية الشبكة العصبونيّة في نموذج حقيبة الكلمات المستمرّة ..... - 67 -
الشكل 19: البنية العامة للمنظومة المقترحة ..... - 70 -
الشكل 20: بنية الجداول المخصصة للمواثقة..... - 72 -
الشكل 21: طرق التسجيل في المنظومة:..... - 73 -
الشكل 22: أنواع الحسابات المتوفرة في المنظومة من حيث الصلاحيات..... - 74 -
الشكل 23: بنية قاعدة البيانات لبرنامج المحادثة ..... - 77 -
الشكل 24 :بنية قاعدة البيانات لبرنامج المؤتمرات ..... - 80 -
الشكل 25: بنية قاعدة البيانات لبرنامج الاستفسارات ..... - 81 -
الشكل 26: بنية قاعدة البيانات لتطبيق مراكز التسجيل ..... - 83 -

الشكل 27: بنية قاعدة البيانات لتطبيق الصلاحيات ..... - 85 -
الشكل 28: بنية قاعدة البيانات لتطبيق المقالات ..... - 86 -
الشكل 29: بنية قاعدة البيانات لتطبيق التحكيم..... - 88 -
الشكل 30: : مخطط يوضح بنية النموذج المقترح..... - 93 -
الشكل 31: :آلية عمل تقنية Word2Vec ..... - 94 -
الشكل 32: آلية عمل خوارزمية MLP في النموذج المقترح..... - 95 -
الشكل 33: العينة ضمن جدول Pandas بعد التجميع..... - 96 -
الشكل 34::مخطط يوضح عدد التقطيع مع عدد المقالات ..... - 97 -
الشكل 35: النتائج النهائية لتدريب النموذج واختباره..... - 99 -
الشكل 36: دقة النتائج لبيانات التدريب و الاختبار ..... - 99 -
الشكل 37: الخسارة في النتائج لبيانات التدريب و الاختبار ..... - 100 -
الشكل 38: التداخل الحاصل بين فئة الفيزياء وبقية الفئات بالنسبة لأول سمتين ..... - 101 -

ملخص:

مع تطور العلم ظهرت مفاهيم جديدة غيرت المجتمع التقليدي وغيرت طبيعة العلاقات الاجتماعية ومن هذه المفاهيم امكانية التواصل الاجتماعي بدون الحاجة الى التواصل المباشر،حيث يمكن تعريف مواقع التواصل الاجتماعي كمجموعة خدمات شبكية تسمح للأشخاص بإنشاء ملف شخصي عام وتحديد من يمكن أن يشترك في الاتصال والسماح أو المنع لهذه الجهات بالوصول إلى مايتم مشاركته، ولكن مع ظهور هذه المفاهيم ظهرت الحاجة لتطوير مستوى وثوقية العلاقات عبر الإنترنت في الشبكات الاجتماعية التي تستند إلى علاقات العالم الحقيقي لتشكيل سحابة اجتماعية ديناميكية ولتمكين المستخدمين من مشاركة الموارد في سياق شبكة اجتماعية، لذلك ظهر مفهوم الشبكة الاجتماعية الاكاديمية Acadimic Social Networks (ASNs) والتي تشمل تنوعاً من المنصات الفعالة تهدف إلى الاستفادة من الشبكات الاجتماعية لأغراض أكاديمية خاصة.

في هذا البحث تم تصميم شبكة اجتماعية أكاديمية محلية خاصة تضمن مشاركة أكاديميين فعليين، كما تم وضع بنية عامة للتطبيقات وطرق المواثقة وأنواع المستخدمين ضمن الشبكة، كما تم تصميم نموذج تعلم آلي لتصنيف المقالات العلمية.

جدول المصطلحات والاختصارات

المصطلح العربي	الاختصار	المصطلح الأجنبي
الشبكات الاجتماعية الأكاديمية	ASNs	Acadimic Social Networks

1- الفصل الأول:

الإطار العام للدراسة

1-1- مقدمة:

مع تطور العلم ظهرت مفاهيم جديدة غيرت المجتمع التقليدي وغيرت طبيعة العلاقات الاجتماعية ومن هذه المفاهيم امكانية التواصل الاجتماعي بدون الحاجة الى التواصل المباشر،يمكن تعريف مواقع التواصل الاجتماعي كمجموعة خدمات شبكية تسمح للأشخاص بإنشاء ملف شخصي عام وتحديد من يمكن أن يشترك في الاتصال والسماح أو المنع لهذه الجهات بالوصول إلى مايتم مشاركته،[3] ولكن مع ظهور هذه المفاهيم ظهرت الحاجة لتطوير مستوى وثوقية العلاقات عبر الإنترنت في الشبكات الاجتماعية التي تستند إلى علاقات العالم الحقيقي لتشكيل سحابة اجتماعية ديناميكية ولتمكين المستخدمين من مشاركة الموارد في سياق شبكة اجتماعية، لذلك ظهر مفهوم الشبكة الاجتماعية الاكاديمية Acadimic Social Network(ASN) والتي تشمل تنوعاً من المنصات الفعالة تهدف لى الاستفادة من الشبكات الاجتماعية لأغراض أكاديمية خاصة[3].

2-1- المشكلة العلمية للبحث:

في الفترة الأخيرة ظهر الاعتماد بشكل مباشر على وسائط التواصل الاجتماعية المعروفة مثل facebook,whatsapp وغيرها كما أن تبادل الملفات وتخزينها على الشبكة العامة ازداد بشكل كبير نتيجة الحاجة إلى الوصول إلى هذه الملفات من أماكن مختلفة، ولكن هذا التوجه إدى إلى فقد خصوصية العديد من المواضيع وأماكن العمل، لذلك ازدادت الحاجة إلى الشبكات الاجتماعية الأكاديمية وهي الشبكات الاجتماعية التي تشكلت من خلال الأنشطة والمعلومات الأكاديمية، يمكن لهذا التعبير دراسة ASNs من مختلف المقاييس الجغرافية والزمنية لتوصيف أنماط المجالات العلمية الجديدة وتسريع إمكانات العلم ، حيث يكون المؤلفون المشاركون هم الشكل الأكثر رسمية للأنشطة الأكاديمية[1] ، كما أن تنفيذ منظومات واقع افتراضي خاصة بأهداف معينة تحقق متطلبات العمل وتحافظ على العنصر البشري، وهذه الحاجة إلى تصميم منظومة خاصة يمكن اعتمادها في المراكز العلمية لتؤمن مستلزمات العمل والمصادر اللازمة لتنفيذ الأعمال، حيث تعتمد الحوسبة السحابية وتكون مرجعا يمكن استخدامه في المجمعات العلمية والأكاديمية التي تتطلب امكانيات خاصة من وسائط تواصل وتخزين ومعالجة وأمان.

3-1- تساؤلات البحث:

1- هل يمكن تصميم شبكة اجتماعية أكاديمية محلية خاصة تضمن مشاركة أكاديميين فعليين، حيث أن الشبكات الاجتماعية الحالية تفتقد إلى الخصوصية والموثوقية.

2- هل يمكن لهذه الشبكة تأمين كل مايلزم كبيئة العمل بحيث تحل كبديل عن الحاجة إلى التقيد بمكان العمل وتحقق الثقة والخصوصية التي يحتاجها المدير والموظف والعميل.

3- هل يمكن الاستفادة من هذه الشبكة للمصادقة في المواقع الأخرى بحيث تكون مرجعا معتمدا للشهادات الرقمية للأكاديميين، وتكون بوابة للعبور إلى المواقع العلمية والمنتديات الأكاديمية العالمية.

1-4- الهدف من البحث:

الشبكات الاجتماعية الأكاديمية (ASNs) عبارة عن شبكات غير متجانسة معقدة تتكون من عدد كبير من الكيانات (المنشورات والعلماء وما إلى ذلك) وعلاقاتهم (الاستشهادات والتأليف المشترك وما إلى ذلك) [1]، غالبًا ما يشارك الأكاديميون بياناتهم البحثية ومقالاتهم والمحتويات الفكرية على حد سواء جنبًا إلى جنب مع معلوماتهم الشخصية. إذا لم يتم الحفاظ على الخصوصية المناسبة من قبل ASNs ، فقد يتم إساءة استخدام هذه البيانات المشتركة من قبل أي شخص في أي لحظة .[2] تبين من خلال الأبحاث أن 70٪ لم يعرفوا الشركة التي يمكن إرسال معلوماتهم الشخصية إليها للمعالجة أو التخزين ، و 17٪ لم يكونوا متأكدين من معلومات الشركة و 13٪ يعرفون الشركة التي يمكن إرسال معلوماتهم الشخصية إليها للمعالجة أو التخزين،[2] لذلك تهدف هذه الدراسة إلى تقديم دراسة لتطوير شبكة اجتماعية أكاديمية خاصة محلية أو عربية تعتمد مفهوم الحوسبة السحابية وتطويعها لتلبية حاجات الباحثين في العالم العربي وتسريع العملية البحثيّة.

1-5 - أهمية البحث:

1- إمكانية اعتماد البنية المقترحة لتكون نواة لبيئة عمل متكاملة ضمن المجتمعات الأكاديمية الخاصة كمراكز البحوث والجامعات بحيث تضمن الموثوقية والخصوصية.
2- إمكانية اعتماد البنية المقترحة لتكون مرجعا يضم الأكاديميين وذلك للاستفادة منه عند الحاجة إلى اختصاصيين في مجال ما.
3- تلبية حاجات الباحثين في العالم العربي وتسريع العملية البحثيّة.
4- القدرة على إيجاد العلماء الآخرين الباحثين في نفس المجال.
5- خلق بيئة حوار آمنة وموثوقة خاصة يكون لها فعالية في تطوير العملية البحثية.
6- القدرة على رفع الأوراق، والمقالات، والكتب وغيرها وتخزينها ضمن بيئة موثوقة خاصة.

1-6- مبررات البحث:

ظهرت الحاجة للبحث مع ظهور أزمة كورونا والحاجة إلى التزام أماكن الإقامة، ظهور هذه الأزمة أوقف مختلف الفعاليات ونتج عنها ظهور أزمات على مختلف المجالات، فكان لابد من إيجاد طرق بديلة تحقق متطلبات العمل والإنتاج، وتضمن استمرارية العمل في مجالات لاتحتاج إلى التواصل المباشر مثل البحث العلمي والتعليم، وعدم اعتماد الطرق العمومية وشبكات التواصل العامة لما لها من تأثير على الخصوصية والأمان.

1-7- محدوديات البحث:

استخدام التخزين السحابي بشكل حقيقي.

1-8- الدراسات المرجعية:

1- University Applications Based on Cloud Computing: North Technical University Case [4]

تصف هذه الورقة مشروعاً يقدم تطبيقات جامعية تستند إلى تنفيذ بنية تحتية للحوسبة السحابية لمجتمع جامعة الشمال التقنية، حيث يتم تقديم الخدمات الافتراضية سواء من التخزين أو المعالجة أو أداء الأجهزة الافتراضية المخصصة للمختبرات ذات الاختلافات في مجالات الهندسة في الإلكترونيات وشبكات الاتصالات.

تم تصميم تطبيقات مختلفة كمراكز اتصال هاتفي عبر بروتوكول الإنترنت واتصالات موحدة وترحيل خدمات مدير المسح والكواشف وبيئات التعلم الافتراضية التي ستسمح خصائص تنفيذها بالحصول على خدمة قوية تلبي متطلبات التطبيق المصمم للأغراض الجامعية، كما تتيح النتائج للمعلمين والطلاب تحديد نماذج جديدة للتعاون والتواصل والتعلم في الفصل الدراسي.

2- Privacy Awareness Among The Academic Social Network Users [2]

في هذه الدراسة تم استطلاع عدة تساؤلات تتعلق بالخصوصية في مجال الشبكات الاجتماعية الاكاديمية ، ومن هذه التساؤلات:هل يهتم المستخدمون لهذه الشبكات بموضوع أمان البيانات والخصوصية بالنسبة للبيانات التي توضع على المشاركة، وهل جميع مستخدمي شبكات ASN مهتمين بإعدادات الأمان التي تؤمنها شبكات ASNs ، وهل هناك بنية متاحة تسمح لمزودي خدمان ASNs بتصميم سياسات خصوصية قادرة أن تتعامل مع متطلبات الخصوصية للمستخدمين كما تم استطلاع إذا كانت كل الشبكات الاجتماعية الاكاديمية تلزم المستخدم بإدخال بياناته الشخصية وغيرها من التساؤلات التي أغنت البحث وأوضحت الآراء المختلفة للمتعاملين مع الشبكات اللأكاديمية بما يساعد في تطوير متطلبات وأمكانيات الشبكات الأكاديمية

من سيئات هذه الدراسة ان استطلاع الرأي حول الأسئلة المطروحة شمل 457 أكاديميا فقط لذلك فإن النتائج التي تمك التوصل إليها في هذه الدراسة قد لا تكون شاملة

3- Estimating educational outcomes from students' short texts on social media [5]

أحد مجالات الاهتمام الخاصة هو تقدير خصائص مختلف المستخدمين من نصوصهم على وسائل التواصل الاجتماعي.

في هذا البحث تم استخدام بيانات من لجنة تمثيلية وطنية من الطلاب للتنبؤ بنتائجهم التعليمية المقاسة باختبارات معيارية من نصوص قصيرة على موقع شبكات اجتماعية روسي شهير.

أخيراً، استكشفوا تنبؤات الأداء الأكاديمي المرتفع والمنخفض للحصول على رؤى حول العوامل المرتبطة بالنتائج التعليمية المختلفة.

يمكن الاستفادة من هذا البحث لوضع نظام تقييم ضمن الشبكة الأكاديمية لاكتشاف وجود حالات دخول غير مشروع من خلال التبادلات النصية بين المستخدمين.

4- Smart Monitoring and Controlling of Government Policies Using Social Media and Cloud Computing [6]

تقدم هذه الورقة نهجًا عملياً يجمع بين قدرات كل من الحوسبة السحابية وتحليلات الوسائط الاجتماعية نحو المراقبة الفعالة والتحكم في السياسات الحكومية من خلال المشاركة العامة.

لقد قدم النظام المقترح بعض النتائج المشجعة عندما تم اختباره لتطبيق ضريبة السلع GST والخدمات من قبل الحكومة الهندية وتثبت أنه يمكن تنفيذه بنجاح من أجل صنع السياسات وتنفيذها بكفاءة.

2- الفصل الثاني:

الشبكات الاجتماعية الأكاديمية

Academic Social Networks ASNs

2-1- مقدمة[1] :

إن الشبكات الاجتماعية الأكاديمية هي خدمات عبر الإنترنت تركز على دعم الأنشطة الموجهة نحو البحث عبر الإنترنت بالإضافة إلى بناء شبكات اجتماعية للعلماء والباحثين. إن أحد عوامل استخدام الطالب لمواقع التواصل الاجتماعي هو القيام بنوع من الدراسة التعاونية أو العمل الأكاديمي أو البحثي. اكتسبت مجموعة متنوعة من منصات الشبكات الاجتماعية الأكاديمية، بما في ذلك ResearchGate و Academia.edu وGoogle Scholar وMendeley وZotero شهرة على مدار العقد الحالي. [1]

2-2- تفاصيل وظائف بعض ASNs:

في الوقت الحاضر، يستخدم الأشخاص مواقع ASN لأغراض عديدة. ويرجع ذلك إلى القدرة التشغيلية العالية لمواقع ASN على توفير خدمة مجتمع عبر الإنترنت وتمركزها جماعياً عبر مشاركة المعلومات والتفاعل مع الآخرين والبحث عن الخبرة في مجال اهتمامهم. فيما يلي تفاصيل وظائف بعض ASNs وفوائدها النسبية للتطوير المهني. [1]

2-2-1- ResearchGate :

ResearchGate هو ASN يسمح للباحثين بتحميل أوراقهم البحثية والانضمام إلى المناقشات ومتابعة أنشطة الباحثين الأخرى. الغرض الرئيسي من RG هو التواصل مع العلماء في جميع أنحاء العالم لمشاركة اهتماماتهم. يوفر حساب تسجيل مجاني للمستخدمين لإنشاء ملف تعريف عبر الإنترنت ومشاركة أفكارهم والبحث في قوائم المؤتمرات. بالإضافة إلى ذلك، يمكن أن يكون هذا الموقع ملزماً للغاية في تحديد موقع مواد المؤتمر مثل المقالات والملصقات والشرائح التي لم يتم أرشفتها في ASNs الأخرى. بخلاف ذلك، يمكن لمستخدمي RG الحصول على درجات على منشوراتهم بما في ذلك عدد المرات التي تم فيها عرض أوراقهم واستشهد بها مستخدمون آخرون في RG. توفر RG منصة لأغراض الاتصال التي تميزها عن ASNs الأخرى.

من خلال RG، يمكن للباحثين:

1- طرح الأسئلة والرد على جميع المستخدمين فيما يتعلق بموضوعهم.
2- الحصول على إخطارات كلما تم تحميل عمل جديد. رسالة مباشرة لطلب التعليقات من خلال نظام المراسلة الخاص بـ RG .
3- طلب مشاركة النسخ الكاملة للمقالات.
4- يسمح بإنشاء سجلات المشروع على مشاريعهم الحالية.
5- طلب تقديم أوراق بحثية كأعداد خاصة في هيئة المجلة. [1]

2-2-2- Academia:

Academia.edu هو شبكة اجتماعية أكاديمية ASN الذي يوفر الأوراق الأكاديمية التي يمكن مشاركتها بين الباحثين. من خلال Academia، يمكن للمستخدمين إنشاء ملفاتهم الشخصية وتحميل قائمة بمخرجات أبحاثهم.

يتضمن حساب المستخدم هذا إحصائيات حول المشاهدات والتفاعلات الاجتماعية مثل المتابعين. لتحقيق هدفهم، تم تطوير Academia.edu كوسيلة لتزويد الأكاديميين بمشاركة أبحاثهم أو مقالاتهم مع أكثر من مليار باحث حول العالم. يمكن للمستخدمين التسجيل مجاناً ومن السهل إنشاء ملف تعريف. الميزة الرئيسية لـ Academia.edu هي أنه يمكن للباحثين تحميل مستندات متنوعة ومتابعة مجموعة متنوعة من الموضوعات التي تهمهم. يمكن للمستخدمين على Academia.edu استيراد جهات الاتصال من حسابات الشبكات الاجتماعية الأخرى للعثور على الأعضاء الذين لديهم ملفات تعريف Academia.edu. ومع ذلك، لتجنب تلقي إشعار إزالة من المحرر قبل استخدام هذه الخدمات، يجب على المستخدمين اتباع إرشادات السياسة والتحقق منها لكل مجلة قبل مشاركة مقال. [1]

2-2-3- Mendeley:

Mendeley.com هو أداة مرجعية مجانية وASN يساعد الأكاديميين على تنظيم أوراقهم، والتعاون مع الآخرين افتراضياً، واكتشاف أحدث الأبحاث حول العالم. يساعد Mendeley المستخدمين على تحميل ومشاركة الملفات التي تشجع على التعاون مع المجموعات الأخرى من خلال مشاركة الأوراق ومتابعة التحديثات وإبداء التعليقات وتتبع التقدم داخل المجموعات. وظيفة أخرى في Mendeley هي السماح للمستخدمين بالبحث وتحميل الأوراق إلى مكتبة ملفاتهم الشخصية وترك تعليق مباشر على أوراق زملائهم. Mendeley ليس فقط مجانياً للتسجيل، ولكن يمكن للمستخدمين تنزيله على سطح المكتب الشخصي. [1]

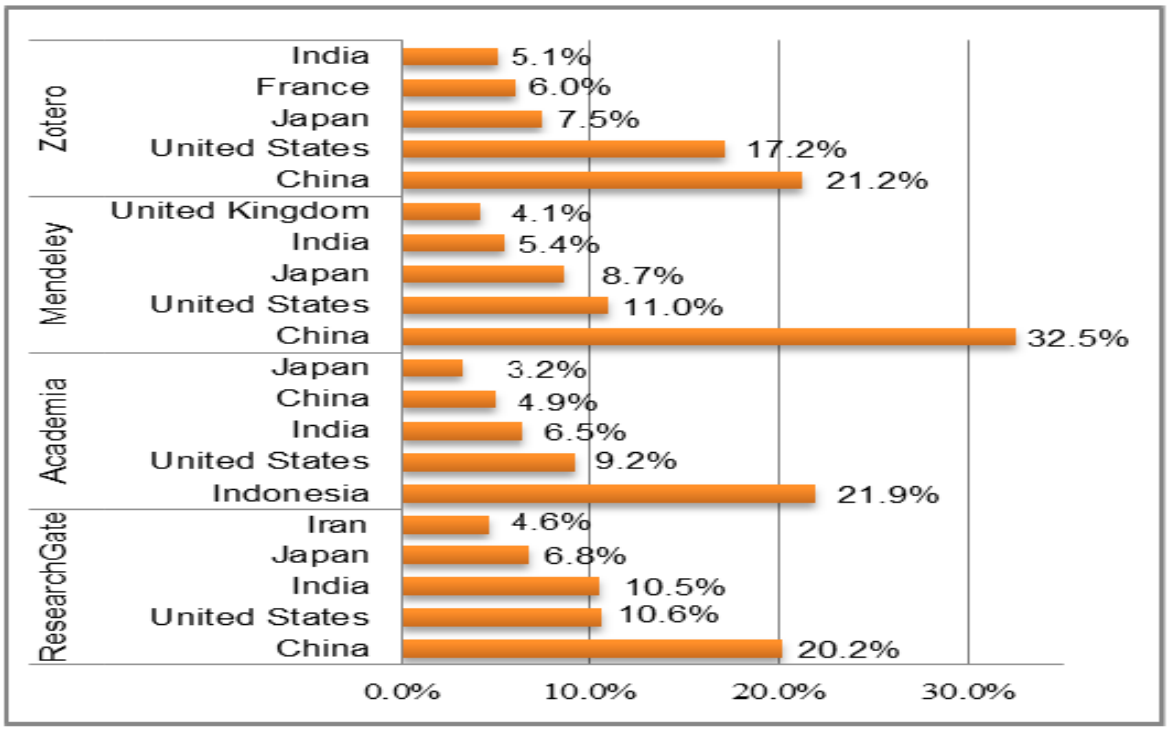
2-2-4- Zotero:

بخلاف RG و Academia.edu وMendeley، يساعد Zotero المستخدمين على الاستشهاد بمصادر دراستهم وجمعها ومشاركتها وتنظيمها بسهولة. إنه موقع معروف ومفيد يضم مجتمعات أو مجموعات مهمة. تستضيف مجموعات Zotero التي تسمح للمستخدمين بالاتصال والعمل مع خبراء آخرين واكتشاف أعمال أخرى من مختلف المجالات بحيث يمكن إبقاء المستخدمين على اطلاع دائم بالمشكلة والتواصل مع الأشخاص المحيطين بهم. التسجيل مجاني وقابل للتنزيل على جهاز الكمبيوتر.

2-3- مقارنة بين Academic Social Networks ASNs الحالية:

من خلال الاطلاع توجد سمات مشتركة لهذه ASNs على النحو التالي:

(مستودع الملفات- تسجيل مجاني- لوحات المناقشة-البريد الإلكتروني - المقاييس البديلة-عدد الاقتباسات-مساحة المجموعة-الملف الشخصي العام / شبه العام -إدارة المراجع -رؤية الشبكة-معالجة المستندات بشكل تعاوني - تحميل المنشورات- رابط إلى مواقع التواصل الاجتماعي). [1]



الشكل 1: مقارنة بين ASNs الحالية من حيث نسب الاستخدام حسب الدول

تم العثور على Academia.edu ليكون أكثر ASNs شعبية بين الشبكات الأخرى مع أكثر من 58 مليون مستخدم مسجل. تم غزو مواقع Academia.edu من قبل إندونيسيا (21.9٪)، بينما كانت الولايات المتحدة ثاني أعلى دولة تستخدم Academia.edu، تليها الهند (6.5٪) والصين (4.9٪) واليابان (3.2٪). سيطرت الصين على أعلى استخدام لموقع Mendeley.com، والذي كان 32.5٪. بناءً على الإحصائيات، كانت الدول الآسيوية مثل الصين وإندونيسيا واليابان والهند هي الأكثر استخداماً لشبكات ASNs.

حاز موقع ResearchGate.net على المرتبة الأولى في تحميل البيانات، يليه Academia.edu، و Mendeley.com، و Zotero.org. من وجهة النظر هذه، يعد Zotero هو الأقدم والأدنى في تصنيف تحميل البيانات مقارنةً بـ Mendeley الذي يتميز بواجهة جذابة ومساحة تخزين مجانية تبلغ 2 غيغابايت والكثير من الموارد التي يمكن الوصول إليها. لذلك، فإن هذه الاختلافات في ترتيب تحميل البيانات مهمة حقاً عند استهداف مستخدمين محددين. بشكل عام، يمكن استنتاج أن الباحثون من المجالات الاجتماعية مثل اللغويين وعلماء الاجتماع والعلوم الإنسانية أكثر نشاطاً باستخدام Academia.edu، بينما يجذب ResearchGate انتباه الباحثين من المجالات البيولوجية. وبالتالي، يشير هذا إلى أن باحثي الدراسات الاجتماعية هم مستخدمون نشطون لمواقع ASN

لأن لديهم معدلات عالية من الاتصال الاجتماعي والاستخدام مقارنة بمجالات البحث الأخرى. كشفت النتائج من الدراسات السابقة أن الموقف الأكاديمي له تأثير على استخدام ASNs.

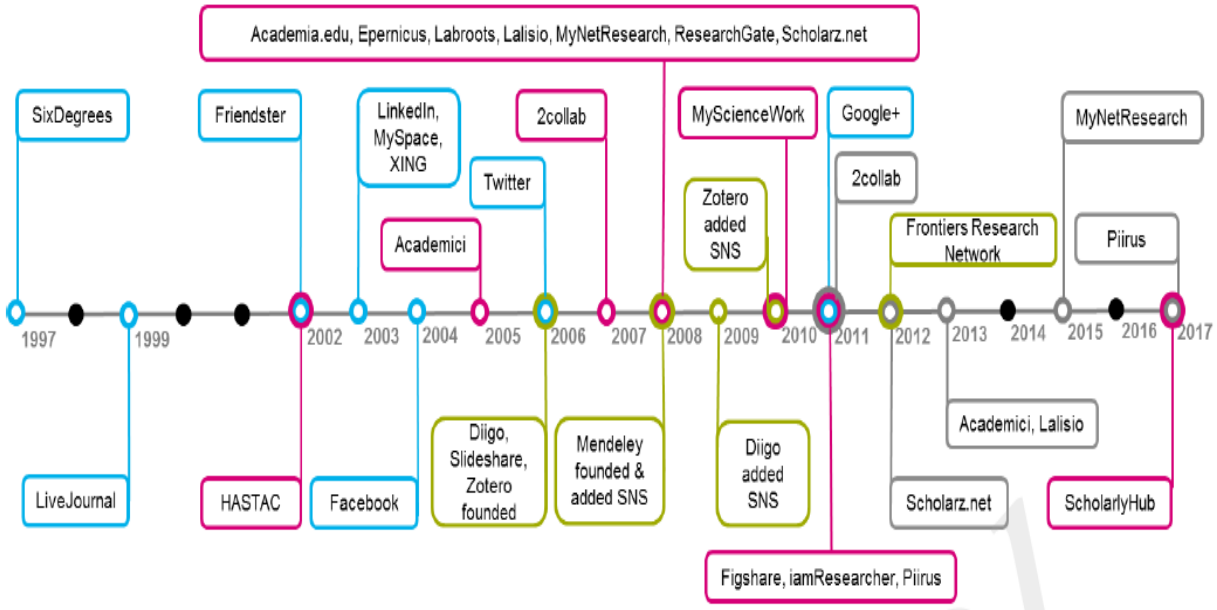
كبار الأكاديميين مثل الأساتذة والمحاضرين وطلاب الدراسات العليا هم أعلى مستخدمي ASNs مقارنة بالمناصب الأكاديمية الأخرى.

ASNs	Traffic Rank	Number of Users	Page Views per User	Average Time on Site	Creation Year
Academia.edu	591	>58 million	1.71	2:25	Sept, 2008
Mendeley.com	4,930	>6 million	2.77	4:27	Aug, 2008
ResearchGate.net	227	>11 million	2.12	2:36	May, 2008
Zotero.org	13,987	>2 million	3.21	4:04	Oct, 2006

الشكل 2: مقارنة بين ASNs حسب عدد المستخدمين والمشاهدات وتحميل البيانات

هذا يرجع إلى المحاضرين والأساتذة الذين يحتاجون إلى اتصال مع ASNs الأخرى في نشر دراساتهم. ومع ذلك، لا يتطلب مساعدي البحث والزملاء الباحثين سوى عدد قليل من المراجعات الأدبية في أبحاثهم لأنهم مدعومون ببعض الأعمال البحثية. [1] يوضح الشكل 1 مقارنة بين ASNs حسب عدد المستخدمين والمشاهدات وتحميل البيانات.

4-2- لمحة تاريخية عن Academic Social Networks ASNs:



الشكل 3: لمحة تاريخية عن ASNs Academic Social Networks:

يوضح الشكل 3 لمحة تاريخية عن ASNs، حيث يمكن تقسيم ASNs إلى فئتين:

1- ASNs تم تطويرها في المقام الأول لتسهيل إنشاء الملف الشخصي والاتصال (على غرار Facebook؛ تشمل الأمثلة Academia.edu وResearchGate). [2]

2- ASNs تركز بشكل أساسي على نشر ومشاركة المحتوى المرتبط بالأكاديمية وبعد ذلك تم إضافة إمكانات الشبكات الاجتماعية (مثل Mendeley). [2]

عادةً ما يكون ASNs مجانياً للاستخدام، على الرغم من أن هذه ليست سمة مميزة، وقد تم إيقاف العديد من الخدمات في السنوات الأخيرة. يوضح الشكل 3 إطلاق ASNs (وإغلاقه في بعض الحالات)، وتواريخ إطلاق ASNs السائدة للمقارنة. من بين الخدمات التي تم تأطيرها منذ البداية باسم ASNs ، تهيمن على السوق الآن منصتان: Academia.edu وResearchGate. تم إطلاق كلتا المنصتين في عام 2008 وشهدتا في البداية نمواً ثابتاً، حيث وصل كل منهما إلى مليوني مستخدم مسجل في عام 2012. في الوقت نفسه، تجاوز عدد مستخدمي Mendeley مليوني مستخدم. منذ عام 2012، تسارع النمو، حاليا يدعو موقع Academia.edu زوار الموقع للانضمام إلى أكثر من 64 مليون مستخدم حالي، بينما تنص ResearchGate على أن لديها أكثر من 15 مليون عضو. تقود هذه المنصات المجال من حيث عدد المستخدمين والوصول الدولي، على الرغم من أن المنصات هي في الأساس ناطق باللغة الإنجليزية وأن المنصات المحلية الأخرى موجودة (مثل منصة Lattes المستخدمة في البرازيل). [2]

Academia.edu وResearchGate شركتان تعتمدان رأس المال الاستثماري الربحي. بينما تلقت Academia.edu استثمارات بقيمة 17.8 مليون دولار حتى عام 2018، حصلت ResearchGate على 87.6 مليون دولار، بما في ذلك مستثمرون بارزون مثل Bill Gates وCrunchbase. في عام 2013، اشترت شركة Elsevier موقع Mendeley كجزء من استراتيجيتها لاكتساب مجموعة واسعة من الأدوات العلمية عبر الإنترنت وإثبات نفسها كمنصة. ومع ذلك، فإن نموذج عمل ASNs ونواياها لتحقيق الدخل من الشبكات لا يزال غامضاً. تختلف الوظائف التي توفرها الأنظمة الأساسية وتغيرت بمرور الوقت. في البداية، تم تصنيف Academia.edu على أنه موقع للباحثين، [2] واستخدم واجهة مرئية لتعيين العلاقات الأكاديمية بهذه الطريقة، والتي تم إيقافها منذ ذلك الحين. في المقابل، يعتمد كل من Academia.edu وResearchGate بشكل أساسي على الملفات الشخصية، والتي يمكن تحميل المنشورات إليها بالإضافة إلى المعلومات الشخصية، والقدرة على متابعة الآخرين. تشمل الوظائف الأخرى نشر الأسئلة، ومراجعة التعليقات، والمقاييس، وإعلانات الوظائف. على الرغم من أن الشبكة هي خاصية مميزة، إلا أن المنصات تضع نفسها في منافسة مع صناعة النشر العلمي بدلاً من وسائل التواصل الاجتماعي. ومع ذلك، فإن استضافة المواد المحمية بحقوق الطبع والنشر قد جعلت المنصات في صراع مباشر مع الناشرين الأكاديميين. في عام 2013، بدأت Elsevier في إصدار إشعارات الإزالة لمستخدمي Academia.edu الذين قاموا بتحميل مقالات تنتهك حقوق الطبع والنشر الخاصة بهم. وبالمثل، قامت مجموعة من خمسة دور نشر بدأت رسمياً بإثارة مخاوف ودعت إلى اتخاذ إجراء مع ResearchGate في عام 2017. بينما كان يُنظر إلى الصراع الأولي بين Academia.edu وElsevier إلى حد ما على أنه عمل إيجابي للمقاومة ضد الناشرين التقليديين، فقد تغيرت هذه المشاعر بحلول عام 2016. ظهر هاشتاج #DeleteAcademiaEdu كرد فعل عنيف استجابةً للاقتراحات بأن المنصة ستقدم المزيد من رؤية أوراق المستخدمين مقابل رسوم. كان الهاشتاج بمثابة تذكير بأن ASNs هي أعمال هادفة للربح، وتتطلع في النهاية إلى إيجاد طرق لاستثمار الشبكة. بعد عشر سنوات من إطلاق اثنين من ASNs الرئيسيين. [2]

2-5- علاقة المنصات بالنشر المفتوح الوصول:

يتضمن هذا الموضوع المنشورات التي تدرس الجوانب العملية لـ ASNs كمنصات نشر مفتوحة الوصول. يعد توفير مساحة استضافة عبر الإنترنت للأكاديميين لتحميل أوراقهم كشكل من أشكال الأرشفة الذاتية جزءاً أساسياً من الخدمات التي تقدمها Academia.edu وResearchGate. يرتبط دور الوصول المفتوح للمنصات بقضية المقاييس metrics، والتي حظيت بتركيز بحثي مكثف درست العديد من الدراسات إلى أي مدى يختار الأكاديميون تحميل الأوراق إلى ASNs. في عام (2017) استطلعت آراء أعضاء هيئة التدريس في جامعة رود آيلاند، حيث أفادت أن 72.5٪ من مستخدمي ResearchGate قاموا بتحميل نصوص كاملة، على الرغم من أن ما يقرب من نصف (47.0٪) من إجمالي أعضاء هيئة التدريس كانوا من مستخدمي ResearchGate. تشير هذه النتائج إلى أنه بينما يستخدم العلماء الأفراد ASNs كمنصة مفتوحة الوصول، هناك تباين كبير في مدى الاستيعاب عبر قطاع التعليم العالي. هناك أيضاً سؤال حول مكان وجود ASNs فيما يتعلق بالنظام البيئي الأوسع للمستودعات على الإنترنت والموارد الببليوغرافية، ولا سيما المستودعات المؤسسية. أظهرت دراسة في عام 2017 أن

ResearchGate تفوقت على المستودعات المؤسسية من حيث توافر الأوراق، حيث توفرت 54.8٪ من الأوراق المنشورة خلال عام 2014 كنصوص كاملة من خلال ResearchGate مقارنة بـ 11.1٪ فقط متاحة للتنزيل من خلال المستودعات المؤسسية، على الرغم من وجود 84.5٪ من الأوراق البحثية تم نشرها في المنافذ التي من شأنها دعم الأرشفة. كان هناك مقارنة أخرى عام 2017 بين ResearchGate والمستودع المؤسسي، حيث تفوقت ResearchGate مرة أخرى على المستودعات المؤسسية (20.3٪ من أعضاء هيئة التدريس قاموا بتحميل النصوص الكاملة إلى ResearchGate، مقارنة بـ 15.4٪ من خلال مستودع مؤسسي). عدد قليل نسبياً من أعضاء هيئة التدريس (6.3٪) استخدم كلا الأمرين، بينما لم يستخدم الغالبية (70.6٪) أي منهما. يعزو المؤلفون انخفاض الإقبال على الوصول المفتوح المؤسسي إلى تفضيل حقوق الملكية للإصدارات المقبولة من المؤلفين من المخطوطات، بينما يفضل الأكاديميون توزيع النسخة النهائية (على الرغم من أنه من المرجح أن يكون هناك مشكلات متعلقة بحقوق النشر). وبالمثل أظهرت مقالة لدراسة حالة لأرشفة المقالات عبر الإنترنت من قبل 125 أكاديمياً مرتبطين بكلية هانكين للاقتصاد في فنلندا. بالتركيز على 587 منشوراً مدرجاً في قاعدة البيانات المؤسسية للأعوام 2012 و2013 و2014، فإن منصات ASNs الأكثر استخداماً لاستضافة الأوراق، مقارنة بالمستودع المؤسسي. تم العثور على نسخ نصية كاملة من 15.8٪ من 587 ورقة على ASNs، مقارنة بـ 9.9٪ في المستودع المؤسسي. [2] مرة أخرى، هذا يعكس الإقبال المنخفض على الوصول المفتوح بشكل عام، ولكن استخداماً أكبر لـ ASNs مقارنة بالمستودعات المؤسسية.

أظهرت دراسة في عام (2017) أخذ عينات من وثائق ResearchGate المرتبطة بأربع جامعات عالمية رائدة، والتي أظهرت تبايناً كبيراً في الإصدارات النهائية وحالة الوصول المفتوح للمستندات التي تم تحميلها وفقاً للمؤسسة. [2] علاوة على ذلك، يوفر ResearchGate للمستخدمين خيار تقديم منشورات معينة على أنها "بحث مميز". حيث ظهر من خلال دراسة في عام (2018) بفحص العوامل المرتبطة بالمنشورات التي يختار الأكاديميون وضعها كأبحاث مميزة. واعتمدت عينتهم على 2708 عضو ResearchGate من مجموعة مختارة من مؤسسات التعليم العالي الأمريكية، والتي تضمنت 95424 منشوراً في المجموع، تم تصنيف 11821 منها على أنها أبحاث مميزة، وُجد أن موقع تأليف الأكاديميين، وعدد الاستشهادات، وسمعة الناشر، كلها مرتبطة بأبحاث مميزة (المرجع نفسه). تجدر الإشارة هنا إلى أنه تم دراسة ميزة الاقتباس المرتبطة باستضافة المنشورات على المنصات من خلال بحث مرتبط بـ Academia.edu حيث تم فحص عينة من 31216 بحثاً. تم الإبلاغ عن تلقي الأوراق التي تم تحميلها على Academia.edu زيادة في الاقتباسات عند مقارنتها بالأوراق غير المتاحة عبر الإنترنت. عند مقارنة الأوراق التي تم تحميلها على Academia.edu بتلك المستضافة في مكان آخر عبر الإنترنت، تلقت أوراق Academia.edu زيادة في الاقتباس بمعدل 58٪ بعد خمس سنوات. قام بحث في عام (2017) بفحص ميزة الاقتباس في ResearchGate من خلال أخذ عينات من المستندات المستضافة فيما يتعلق بأربع جامعات تعكس مواقع في جميع أنحاء العالم. [2] على الرغم من أن التحليل كان محدود النطاق، تشير الدراسة إلى زيادة معدل الاقتباس المرتبط بتوافر أوراق الوصول المفتوح من خلال المنصة. في عام (2017) تم تقديم تطبيق جديد لـ ASNs داخل النظام البيئي الأكاديمي عبر الإنترنت، مما يدل على أنه يمكن استخدام ملفات تعريف

ResearchGate كطريقة لتأكيد هويات المؤلفين والتعامل مع توضيح الغموض بين المؤلفين الذين لديهم أسماء متشابهة.

في النهاية، يُظهر البحث في مدى توفر المنشورات الأكاديمية من خلال ASNs أن الأكاديميين من المرجح أن يقوموا بتحميل أوراقهم إلى ASNs بدلاً من المستودعات المؤسسية. تشير الدراسات هنا أيضاً إلى أن الأكاديميين يفضلون مشاركة المخطوطة النهائية بدلاً من المخطوطة المقبولة للمؤلف (والتي تفضلها المؤسسات عادةً). في حين أن ASNs لا يمنع تحميل المخطوطات النهائية، فإن نشرها عبر الإنترنت يحمل مخاطر أكبر من حيث حقوق النشر. قد ترتبط استضافة أوراق الوصول المفتوح من خلال ASNs بمعدل اقتباس أعلى. [2]

Metrics -6-2

Metrics هو تسجيل الآثار الرقمية لكيفية وصول المستخدمين إلى الأنواع المختلفة من المواد المستضافة على المواقع، هناك الكثير من الطرق التي يمكن من خلالها ASNs تسجيل الآثار الرقمية لكيفية وصول المستخدمين إلى المواد المستضافة. يعكس الاهتمام بالمقاييس المستمدة من المنصات مفهوم المقاييس الببليومترية باعتباره انعكاساً للمكانة الأكاديمية. قد تكون المنصات مصدراً مفيداً للمؤشرات المبكرة للأبحاث عالية التأثير، قد تعمل مقاييس المنصة كطريقة لإنشاء المقارنات، أو تسمح بقياس أداء الأفراد بالنسبة إلى المعايير الأخرى. ومع ذلك، نظراً لأن وظائف الأنظمة الأساسية توفر أكثر من مجرد الوصول إلى الأوراق البحثية، تُظهر ASNs أيضاً إمكانية وجود طرق مختلفة لقياس المشاركة بما يتجاوز أوجه التشابه مع عدد الاقتباسات وعوامل التأثير. يوفر Mendeley الفرصة لتقدير التفاعل مع الأوراق الفردية من خلال أعداد القراء، والتي ثبت أنها وكيل جيد لأعداد الاقتباسات. حيث أظهرت الدراسات العلاقة بين القراء وعدد الاقتباسات. تلقى المقالات قدراً أكبر من الاهتمام من خلال أعداد قراء Mendeley مقارنةً بالاقتباسات. يوفر ResearchGate أوسع نطاق من المقاييس، على عدد من المستويات المختلفة، بما في ذلك الأوراق الفردية والأكاديميين الفرديين والمؤسسات. واحدة من الخصائص الرئيسية لـ ResearchGate هي مقياسها الخاص، RG Score. لا يذكر موقع ResearchGate صراحة كيفية حساب الدرجة، على الرغم من أنه يبدو أنه يعتمد بشكل أساسي على عامل التأثير للمجلات التي نشرها الأكاديمي، والنشاط في المناقشات التي يتوسط فيها الموقع. درست العديد من الدراسات الارتباطات بين درجات RG ومقاييس أخرى للمكانة الأكاديمية والتأثير، حيث أبلغ عن الارتباطات بين نقاط RG والقراءات، وجهات النظر الشخصية، وعدد النصوص الكاملة، والمتابعين. يعرض ResearchGate أيضاً الأرقام الإجمالية لـ "مجموع نقاط RG" للصفحات المؤسسية، على الرغم من أنه تم تحديد مشكلات الموثوقية مع درجة RG المؤسسية. إن ASNs ليست مجرد مستودعات للأوراق بل تقدم طرقاً جديدة محتملة لقياس التأثير. على سبيل المثال، [2] يمكن للآثار الرقمية للتفاعلات من خلال ASNs أن تخبر الأكاديميين بالضبط عن جمهورهم من خلال الموقع. تم استخدام ملفات تعريف Mendeley كطريقة لقياس مدى القراء الدوليين أو فحص القراء من حيث الفئات المتعلقة بالوظائف الأكاديمية.

بشكل عام، تُظهر المقاييس من خلال ASNs درجة جيدة من الارتباط مع المقاييس الببليومترية التقليدية ومقاييس المكانة، مثل عوامل التأثير، وعدد الاقتباسات والتصنيفات. تعتبر درجات RG الإجمالية، مثل تلك الموجودة على

المستوى المؤسسي، استثناءً وأقل موثوقية، ويبدو أنها تتأثر بحجم المؤسسة. ومع ذلك، قد تكون مقاييس ASNs طريقة فعالة لقياس الأداء الفردي، تُظهر مقاييس الشبكات الاجتماعية كطريقة جديدة لقياس التأثير العلمي من خلال التفاعل مع الآخرين من خلال المنصة يتكون هذا الموضوع من نوعين رئيسيين من الأوراق: الدراسات التي تركز على طرح الأسئلة والإجابة عليها من خلال المنصات، وتلك التي تنظر في تكوين المجموعات. كل نوع خاص بمنصة معينة. في حين أنه من الجدير بالذكر أن هذه الوظائف لا تمثل النطاق الكامل للطرق التي يمكن أن يحدث بها الاتصال غير الرسمي بين المستخدمين عبر المنصات، إلا أنها تمثل الطرق الوحيدة التي خضعت للبحث حتى الآن. في البداية، عرض كل من Academia.edu وResearch Gate التسهيلات للمستخدمين ليتمكنوا من إرسال الأسئلة إلى المجتمع. ومع ذلك، أوقفت Academia.edu هذه الميزة في عام 2015 ،بينما تظل القدرة على طرح الأسئلة والإجابة عليها نشطة في ResearchGate ومن المرجح أن تستمر، [2] لأن نشاط السؤال هو مساهمة رئيسية في نقاط RG، وهي "نقطة فريدة" لمنصتهم.

تركز الدراسات في موضوع التفاعلات مع الآخرين من خلال المنصات على وظيفتين خاصتين للغاية بالمنصة. يوفر Mendeley القدرة على تكوين مجموعات، [2] بينما يسمح ResearchGate للمستخدمين بنشر الأسئلة والإجابة عليها (كان لدى Academia.edu أيضاً هذه الوظيفة، ولكن تم إيقافها). شوهدت توزيعات منحرفة للمشاركة لكليهما؛ الغالبية العظمى من مستخدمي ResearGate لا يطرحون أسئلة، ومعظم مجموعات Mendeley لديها عضو واحد. يدعم كل منAcademia.edu وResearchGate الآن أشكال مراجعة النظراء المفتوحة والتعليق على منشورات محددة، على الرغم من أن هذا لم يتم فحصه بعد.

2-7- التركيبة السكانية والمنصة الاجتماعية:

ركزت الدراسات في هذا الموضوع على توصيف مجتمع الأكاديميين الذين يختارون أن يصبحوا مستخدمين لـ ASNs، ونوع الشبكة الاجتماعية التي تعززها المنصات. من خلال معالجة هذه القضايا، يمكن إجراء استنتاجات حول العلاقة بين المنصات والأوساط الأكاديمية الرسمية، مثل الأسئلة المتعلقة بمدى تمثيل المنصات لمساحة جديدة في الأوساط الأكاديمية. نظراً لأن ملفات تعريف ASNs غنية بالمعلومات الشخصية ذات الصلة بالحالة المهنية للأكاديميين، فقد تمت معالجة فهم التركيبة السكانية لمجتمع المستخدمين من خلال تجميع خصائص الملف الشخصي عبر الويب. أظهرت دراسة في عام 2017 تحليلاً لـ 29133 ملفاً شخصياً من Academia.edu مستمدة من أربعة مجالات تخصصية (الأنثروبولوجيا، والكيمياء، وعلوم الكمبيوتر، والفلسفة)، وأربعة مستويات من الأقدمية الأكاديمية (أعضاء هيئة التدريس، وباحثو ما بعد الدكتوراه، وطلاب الدراسات العليا، والباحثون المستقلون). [2] تم تحديد جوانب الملفات الشخصية والتعبير عنها عددياً. وشمل ذلك مدى اكتمال الملف الشخصي، والاهتمامات البحثية، والعلاقات (عدد المتابعين وعدد الأشخاص الذين يتابعهم المستخدم)، والمتابعة (عدد الذين يتابعونهم) وتكرار النشاط. تم العثور على أكاديميي الأنثروبولوجيا والفلسفة ليكونوا مستخدمين أكثر نشاطاً من الكيمياء أو علوم الكمبيوتر. عبر التخصصات، كان أعضاء هيئة التدريس والباحثون ما بعد الدكتوراه أكثر نشاطاً، لا سيما فيما يتعلق بتحميل المواد. يعزز باحثو ما بعد الدكتوراه أكبر عدد من العلاقات (يتبعون الآخرين)، بينما يظهر

طلاب الدراسات العليا أقل مستويات الاستخدام. مع التركيز أيضاً على Academia.edu، كذلك تم جمع وتحليل البيانات من 428،30 ملفاً شخصياً، وتحديد خصائص الملف الشخصي واختبار الاختلافات بناءً على العوامل الفئوية بما في ذلك الأقدمية الأكاديمية وفئة تطوير الدولة وفئة التصنيف الجامعي، [2] لم يختلف عدد الأسئلة المطروحة وعدد الأسئلة التي يتابعها المستخدمون إحصائياً وفقاً للوضع الأكاديمي؛ أظهرت جميع العناصر الأخرى فروقاً ذات دلالة إحصائية بناءً على المنصب، حيث كان عدد أكبر من الأكاديميين الكبار أكثر انتشاراً في كل جانب من الباحثين المبتدئين. كذلك تم فحص ما إذا كان Academia.edu يعكس المعايير المرتبطة بالأوساط الأكاديمية أو وسائل التواصل الاجتماعي، من خلال تحليل ملفات تعريف جميع الأكاديميين البالغ عددهم 30167 المرتبطين باهتمام البحث "الفلسفة". تعكس النتائج نشر الطلاب عدداً أقل من العناصر في ملفاتهم الشخصية ويكتسبون عدداً أقل من المشاهدات مقارنة بأعضاء هيئة التدريس. بالإضافة إلى ذلك، قامت دراسة بفحص الاختلافات من حيث الجنس، على أساس أنه ثبت أن الإناث يتمتعن بميزة في وسائل التواصل الاجتماعي بشكل عام، على الرغم من أن الفلاسفة لديهم آراء شخصية أقل من الذكور. امتد هذا النهج ليشمل القانون والتاريخ وعلوم الكمبيوتر، بالنسبة لفئة "العلوم الإنسانية والاجتماعية"، فإن Academia.edu هي المنصة الأكثر شعبية، في حين أن ResearchGate هو الأكثر شعبية في العلوم الطبيعية.

بالإضافة إلى ذلك، كان الباحث العلمي من Google أكثر شهرة بشكل ملحوظ في "العلوم والتكنولوجيا المادية" و"الموارد الطبيعية"، بينما كانت مستويات Mendeley منخفضة نسبياً في جميع المجالات الدراسية. حيث أظهر Academia.edu زيادة في عدد كبار الأكاديميين، بينما يُظهر ResearchGate نمواً من حيث عدد الأكاديميين المبتدئين وطلاب الدراسات العليا.

يقترح المؤلفون أن هذه البيانات يمكن أن تكون طريقة محتملة لقياس مكانة المؤسسات. ومع ذلك، نظراً للاختلافات وفقاً للأقدمية التي وجدتها دراسات أخرى، فإن الاختلافات وفقاً للمؤسسة قد تعكس النسب المختلفة بين الطلاب والموظفين داخل الأقسام بدلاً من المكانة على هذا النحو. تشير الأبحاث في التركيبة السكانية للمنصات إلى أنها تعكس التسلسل الهرمي للأوساط الأكاديمية الرسمية. قد توفر الأشكال الأخرى من وسائل التواصل الاجتماعي، مثل Twitter، فرصة أكبر للأكاديميين أو الطلاب الأقل مرتبة ليكونوا جزءاً نشطاً من شبكة مهنية. هناك أيضاً اختلافات انضباطية ملحوظة في استخدام النظام الأساسي، مع تفضيل ResearchGate من قبل العلوم الطبيعية، و Academia.edu المفضل من قبل العلوم الإنسانية. [2]

3- الفصل الثالث:

الحوسبة السحابية Cloud Computing:

3-1- مقدمة

تعرف الحوسبة السحابية ببساطة بأنها تأجير وحدة تخزين البيانات والتطبيقات وخدمات الحوسبة الأخرى عبر الإنترنت. [3]ببساطة، الحوسبة السحابية هي تقديم خدمات الحوسبة - بما في ذلك الخوادم والتخزين وقواعد البيانات والشبكات والبرمجيات والتحليلات والذكاء - عبر الإنترنت ("السحابة") لتقديم ابتكار أسرع وموارد مرنة واقتصاديات الحجم. عادةً ما تدفع فقط مقابل الخدمات السحابية التي تستخدمها، مما يساعدك على خفض تكاليف التشغيل، وتشغيل البنية التحتية الخاصة بك بشكل أكثر كفاءة، وتوسيع نطاق عملك مع تغير احتياجات عملك. [4]الحوسبة السحابية هي مصطلح عام لأي شيء يتضمن تقديم خدمات مستضافة عبر الإنترنت. تنقسم هذه الخدمات إلى ثلاث فئات أو أنواع رئيسية من الحوسبة السحابية: البنية التحتية كخدمة (IaaS) والنظام الأساسي كخدمة (PaaS)والبرمجيات كخدمة [5] (SaaS).

3-2- أنواع نشر الحوسبة السحابية:

تتميز الحوسبة السحابية بالعديد من الجوانب بما في ذلك نموذج التطوير السحابي الذي يوفر نوعاً معيناً من البيئة السحابية، والتي تتميز بشكل أساسي بحجمها وملكيتها وإمكانية الوصول إليها. في الواقع، يتم تمكين الحوسبة السحابية من خلال مشاركة الموارد بين الأجهزة الفردية أو الخوادم المحلية. يرتبط الغرض من السحابة وطبيعتها بنموذج النشر. يشتمل نموذج النشر على ثلاثة أنواع أساسية، وهي السحابة العامة والسحابة الخاصة والسحابة المختلطة، [6]يضاف إليها في بعض التصنيفات سحابة المجتمع والسحابة البينية.

3-2-1- السحابة العامة Public Cloud:

يتم تشغيلها من قبل CSP، الذي يمتلك البنية التحتية ومراكز البيانات. البنية التحتية قائمة ويمكن للمؤسسات الوصول إلى الخدمات عند الطلب وعلى أساس الدفع أولاً بأول. يتم توفير الخدمات للمؤسسات والمستخدمين عبر شبكة عامة من خلال متصفح. السحب العامة هي مواقع مستقلة وموثوقة وقابلة للتوسع بدرجة كبيرة، ولكنها أقل أماناً وغير قابلة للتخصيص. [7]

تمتلك الحكومة أو الأعمال أو الأكاديميات سحابة عامة وتقوم بتشغيلها. أيضاً، يمكن للمؤسسات توفير وصول مفتوح عبر الإنترنت أو بوابة أخرى في هذا النموذج. في هذا النوع من السحابة، تظهر العديد من التحديات في مجال تحديد موقع الموارد واكتشاف الملكية. علاوة على ذلك، من الصعب للغاية حماية الموارد من أنواع مختلفة من عمليات الاقتحام والهجمات. [6]

3-2-2- السحابة الخاصة Private Cloud:

يتم إعداد السحابة الخاصة وتشغيلها بشكل حصري لمؤسسة معينة، ولكن يتم منح المؤسسات الخارجية حق الوصول لإدارتها نيابة عن مالك السحابة. يمكن تشغيل السحابة الخاصة داخل المقر أو خارجه. السحابة الخاصة تتمتع بالخصوصية والأمان والتحكم. التكلفة وكفاءة الطاقة جيدة أيضاً. السحب الخاصة لها قابلية محدودة للتوسع وهي مقيدة بمنطقة معينة. [7]

تتعامل الحوسبة السحابية مع مركز البيانات في المؤسسة. في هذا النموذج، يتم تخصيص الموارد لمنظمة واحدة أو عدة مؤسسات والعمل بين الوظائف. لهذا السبب، فإن البنية التحتية مملوكة ومدارة من قبل نفس المؤسسة في هذا النوع من السحابة. كما أن تحديد العلاقة بين العملاء والبائعين والكشف عن مخاطر الأمان أسهل بكثير. [6]

3-2-3- السحابة الهجينة Hybrid Cloud:

تمثل Hybrid Cloud مزيجاً من مجموعة مختارة أو جميع أنواع نشر السحابة، أي السحابة الخاصة أو العامة أو المجتمعية. تتم استضافة الأنشطة الأساسية على سحابة خاصة، بينما يتم إسناد الخدمات الأقل أهمية إلى السحابة العامة. تظل كل سحابة كياناً فريداً، ولكنها مرتبطة ببعضها البعض بواسطة تقنية موحدة. [7]

هذا النموذج هو الأفضل من حيث المزايا، حيث يقدم سحابة خاصة مرتبطة بواحدة أو أكثر من الخدمات السحابية الخارجية، في حين أن البيانات والتطبيق مرتبطان معاً ويتم إدارتهما مركزياً. تجدر الإشارة إلى أن أمان السحابة المختلطة أكثر موثوقية من السحابة العامة إذا كانت الكيانات تصل عبر الإنترنت. ومع ذلك، لكل من نماذج النشر مزاياها وعيوبها. في الواقع، لكل نموذج نشر ميزة وعيوب محددة فيما يتعلق بتجربة المستخدم.

3-2-4- سحابة المجتمع Community Cloud:

يتم استضافة سحابة المجتمع من قبل العديد من المنظمات أو المؤسسات التي تشترك في الاهتمام المشترك. الأمثلة النموذجية هي الجامعات التي تستخدمها للتعلم والبحث. قد تقرر المنظمات إدارة النظام السحابي بنفسها في الموقع أو خارجه وقد تقرر أيضاً الاستعانة بمصادر خارجية للتشغيل اليومي للنظام إلى مؤسسة طرف ثالث. [7]

3-2-5- السحابة البينية Inter Cloud:

يمكن الإشارة ببساطة إلى السحابة البينية على أنها ربط بيني للسحب، مثل شبكة الشبكات [11]. يتيح ذلك الاتصال بالبنية التحتية السحابية المتعددة لتوفير المزيد من موارد الحوسبة.

هناك نوعان من السحب البينية هما: السحابة الاتحاد والسحابة المتعددة.

أ. سحابة الاتحاد: سحابة الاتحاد عبارة عن إعداد بين السحب حيث تقوم مجموعة من موفري السحابة بدمج بنيتهم التحتية السحابية عمداً، مما يتيح سهولة مشاركة الموارد بين المنظمات المشاركة. يتعاون كل مزود خدمة سحابية طواعية لتبادل الموارد.

ب. السحابة المتعددة: تتضمن السحابة المتعددة استخدام بنية سحابية مستقلة متعددة تعمل كبنية واحدة. على سبيل المثال، يمكن للمؤسسات تشغيل نشاط سحابي معين يتطلب موارد هائلة على السحابة الخاصة وتشغيل أنشطة سحابية أخرى تتطلب مستوى أقل من موارد الشبكة أو قدرة التخزين على السحابة العامة. [8]

3-2-6- المزايا والعيوب لنماذج التطوير السحابي:

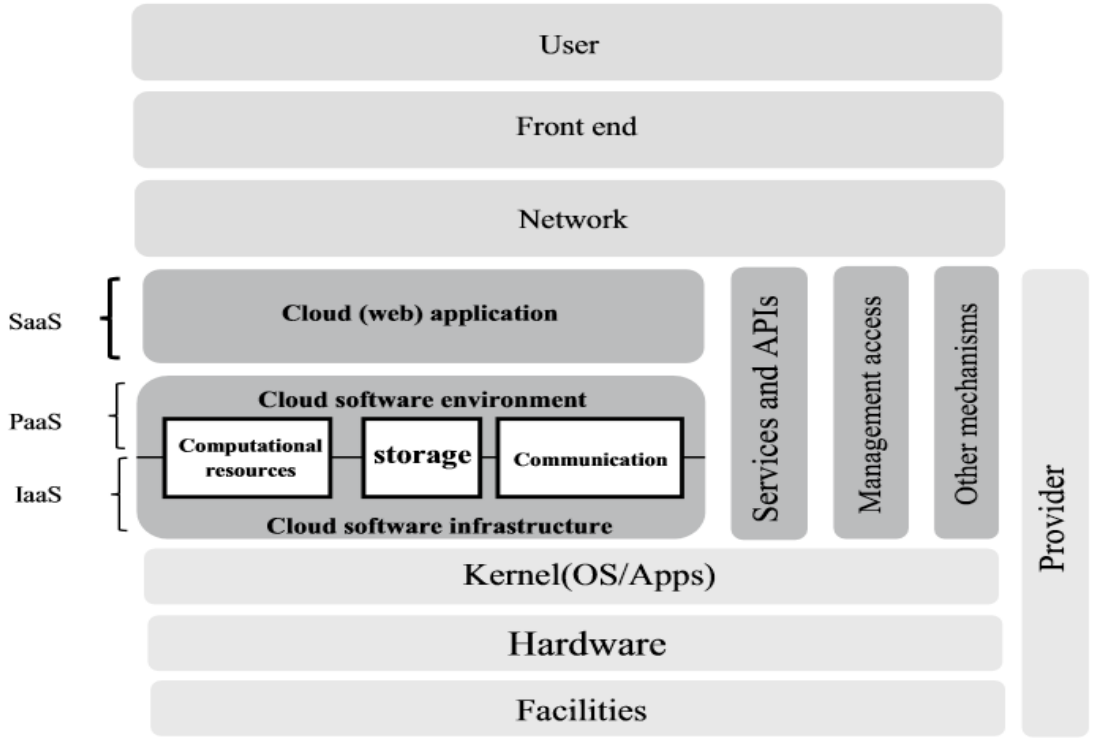
توفر السحابة الخاصة تحكماً كاملاً في تجربة المستخدم. ومع ذلك، في بعض الحالات، لا تتحكم السحابة العامة في تجربة المستخدم وتعترف السحابة المختلطة بأن التحكم في تجربة المستخدم يعتمد على الاتفاق المبرم مع

المستهلك. بالإضافة إلى ذلك، [6] يوضح الشكل 4 مزايا وعيوب السحب العامة والسحابة الخاصة والسحابة المختلطة.

Cloud development models	Advantages	Disadvantages
Public cloud	Scalability and reliability with on-demand resources	Can be unreliable
	Easy to use	Less secure
Private cloud	Organization-specific	More costly
	Customizable	Requires IT expertise
Hybrid cloud	Flexible infrastructure	Lack of visibility
	Cost controls	Potential challenges in application and data integration
	Faster speeds	

الشكل 4: مزايا وعيوب السحب العامة والخاصة والمختلطة

ومع ذلك، هناك نموذج أمان مشترك للمسؤولية وراء جميع الخدمات التي تقدمها الحوسبة السحابية. يلعب كل من الموفر ومستهلك السحابة دوراً في أمان البنية التحتية المقيمة في السحابة والتطبيقات المقدمة عبر السحابة. تختلف مسؤولية الأمن في كل نموذج تسليم. في بعض الخدمات، يكون العميل مسؤولاً عن أمان البيانات مثل وصول المستخدم وإدارة الهوية في أي حالة من نماذج التسليم (IaaS وPaaS وSaaS) كما هو موضح في الشكل5:



الشكل 5: بنية السحابة حسب الخدمات المقدمة

يتحمل العملاء بعض المسؤوليات مثل التعاون مع المزود فيما يتعلق بصيانة النظام. في هذا الصدد، [6] يجب عليهم تثبيت التصحيحات على نظام التشغيل (OS) ومكدس التطبيقات، وأن يقوموا بدور مسؤول النظام في السحابة من خلال تصحيح أنظمة التشغيل والتطبيقات المختلفة وتحديثها. مسؤوليات العملاء الأخرى هي إدارة إنشاء حساب المستخدم والتزويد والحذف وآليات مصادقة الحساب على مستوى الخادم وسياسات كلمة المرور وما إلى ذلك. يمكن مراقبة نشاط بيانات العملاء باستخدام السجلات التي يتم تسليمها تلقائياً إلى حساباتهم. ومع ذلك، يتحمل كل من موفر السحابة وعميل السحابة المسؤولية عن جوانب مختلفة من النظام ويجب على كليهما اتخاذ بعض الإجراءات لتأمين الخدمة بشكل صحيح. وفقاً لتقرير Oracle وKPMG Cloud Threat Report عام 2018، كان 43٪ فقط من المستجيبين قادرين على التعرف بشكل صحيح على نموذج أمان المسؤولية المشتركة IaaS الأكثر شيوعاً. [6] يجب أن تستخدم المؤسسات العديد من ضوابط أمان الشبكة مثل جدران الحماية المادية والقائمة على الأجهزة الافتراضية، وأنظمة الكشف عن التسلل والوقاية، والبوابات بالإضافة إلى عبء العمل الهادف وعناصر التحكم في التطبيقات السحابية. أيضاً، قال 66٪ من المشاركين أنهم واجهوا حادثة أمنية إلكترونية أثرت على عملياتهم التجارية على مدار العامين الماضيين. كانت هذه الآثار هي تعطيل العمليات التجارية القياسية، وتعطيل تقديم الخدمة، وفقدان إنتاجية الموظفين، والتأخير في مشاريع تكنولوجيا المعلومات. في الواقع، قدمت الشركات السحابية الكثير من السمات مثل التوفر العالمي للخدمات عالية الأداء، ودعم عدد كبير من الخدمات، وتخزين كمية كبيرة من البيانات. [6]

3-3- خدمات الحوسبة السحابية:

يمكن تقسيم بنية السحابة إلى النهاية الخلفية والواجهة الأمامية. تكون الواجهة الأمامية مرئية للمستخدم من خلال الاتصالات بالإنترنت، مما يسمح بتفاعل المستخدم مع النظام [11]. تتكون النهاية الخلفية من نماذج الخدمات السحابية المختلفة. [8]

3-3-1- البنية التحتية كخدمة (IaaS):

تقدم IaaS إمكانات الحوسبة والتخزين الأساسي كخدمات موحدة عبر الشبكة. تتمثل الإستراتيجية الكامنة في المحاكاة الافتراضية في ترتيب أجهزة افتراضية مستقلة (VM) وعزلها عن كل من الأجهزة الأساسية والأجهزة الافتراضية الأخرى. يركز IaaS أيضاً على مجالات الأمان مثل جدار الحماية، واكتشاف التسلل، والوقاية (IDS / IPS)، ومراقبة الجهاز الافتراضي. تشهد الحوسبة السحابية نمواً كبيراً وسريعاً، واعتماد المزيد والمزيد من المنظمات التكنولوجيا السحابية كل يوم. ومع ذلك، هناك العديد من القضايا الأمنية الموازية التي يجب الاهتمام بها. تختار المؤسسات بعض البنى التحتية الآمنة عند نقل بياناتها إلى وجهات بعيدة. بهذه الطريقة، يستخدم كل عميل عادةً برامجه الخاصة على البنية التحتية. تعد Google Compute Engine (GCE) و Amazon Web Services (AWS) وCisco Metapod وMicrosoft Azure) من بين أهم أمثلة IaaS. [6]

هذا هو تسليم الخوادم والتخزين والشبكة ونظام التشغيل، كخدمة يوفر IaaS جهازاً تجريدياً بنظام تشغيل تم تثبيته وتكوينه بالفعل. تتيح IaaS تخزين البيانات في مواقع جغرافية مختلفة. يتحكم موفرو IaaS في الأنشطة في مراكز

البيانات السحابية مع السماح للمستخدمين بالمرونة لنشر وإدارة خدمات البرامج بأنفسهم]. يمكن للمستخدم الوصول إلى جهاز كمبيوتر افتراضي والتخزين والبنية التحتية للشبكة وموارد الحوسبة لنشر البرامج وتشغيلها. يقوم موفر السحابة بإدارة البرامج والأجهزة فقط، مثل الخوادم وأجهزة التخزين ونظام التشغيل المضيف وHypervisor للافتراضية. يتم عرض خدمات الرسم التخطيطي للبنية السحابية النموذجية المتاحة لمستخدمي السحابة في. أمثلة على IaaS: أ. شبكات توصيل المحتوى (CDNs): تسجل شبكات توصيل المحتوى محتوى المستخدم وملفاته لتحسين أداء النظام مثل السرعة والتكلفة المرتبطة بمحتوى التسليم للأنظمة القائمة على الويب. هذا مفيد للتعامل مع أنواع مختلفة من المحتوى لتسليمه إلى أي موقع ويب أو تطبيق جوال. ب. النسخ الاحتياطي والاسترداد: يوفر هذا القدرة على النسخ الاحتياطي واستعادة الملفات بسلاسة. ج. الحوسبة: يتضمن ذلك متطلبات الخادم للحفاظ على الأنظمة السحابية التي يمكن تكوينها وتوفيرها ديناميكياً. د. التخزين: تتوفر أيضاً قدرة تخزين قابلة للتطوير بدرجة عالية ومفيدة لتسجيل أنشطة التطبيقات والنسخ الاحتياطية للملفات واستعادة الملفات وتخزينها. [8]

3-3-2- النظام الأساسي كخدمة (PaaS):

في PaaS، تُستخدم طبقة من بيئة التطوير كخدمة ويمكن بناء مستوى أعلى من الخدمة. في نموذج PaaS، يقوم العميل بإنشاء تطبيقاته الخاصة التي تعمل على البنية التحتية للموفر. في الواقع، تقدم PaaS مجموعة من خوادم أنظمة التشغيل والتطبيقات مثل Microsoft Azure وGoogle App Engine ومنصة LAMP (Linux وApache وMySQL وPHP) وما إلى ذلك. وتجدر الإشارة إلى أن أحد المجالات الرئيسية لنموذج PaaS هو حماية البيانات. يصبح هذا مهماً بشكل خاص في حالة التخزين كخدمة. يجب أن يؤخذ في الاعتبار أن هذا النموذج يجب أن يكون قادراً على تشفير البيانات أثناء تخزينها على نظام أساسي تابع لجهة خارجية ويجب أن يكون على دراية بالمسائل التنظيمية مثل توفر البيانات في مناطق جغرافية مختلفة. [6]

PaaS هي خدمة تطوير تقدم للمستخدم عبر الإنترنت. لا يحتاج المستخدم إلى تثبيت أي برامج أو متطلبات للأجهزة، مما يوفر التكلفة. إنها برمجيات وسيطة تُبنى عليها التطبيقات. تحتوي PaaS على أدوات مضمنة وواجهات خدمة ويب وأمن مضمنة للتطبيقات المنشورة. يمكن دمج التطبيق المنشور مع تطبيقات أخرى على نفس النظام الأساسي وربطه بتطبيقات أخرى خارج النظام الأساسي. تحتوي PaaS على برامج تشتمل على قاعدة بيانات وبرامج وسيطة وأدوات تطوير. أمثلة على PaaS:

أ. ذكاء الأعمال. ب. قاعدة البيانات. ج. التطوير والاختبار. د. الاندماج. هـ. نشر التطبيق. [8]

3-3-3- البرمجيات كخدمة (SaaS):

يركز نموذج SaaS على مهام إدارة الوصول في تطبيقات مثل ضوابط السياسة. على سبيل المثال، يُسمح لأي شخص فقط بتنزيل معلومات معينة من التطبيقات. بهذه الطريقة، يستفيد العديد من المستخدمين النهائيين من مثيل واحد للخدمة.

اليوم، تقدم شركات مثل Google وMicrosoft Office 365 وDropbox وما إلى ذلك SaaS. [6]

يُعرض على المستخدم مجموعة مستضافة من البرامج تعمل على منصة وبنية تحتية مملوكة لمزود السحابة]. تم تصميم التطبيقات وتطويرها ليتم الوصول إليها في نفس الوقت من قبل مستهلكين مختلفين في السحابة عبر الإنترنت. تتم إدارة التطبيق المستضاف بواسطة CSP، الذي يحافظ على تشغيل النظام ويضمنه. يدعم التطبيق المستضاف متعدد الشركات، وهو متاح عند الطلب ويمكن زيادته إلى أسفل. يعمل بعض موفري SaaS على عروض PaaS أو IaaS الخاصة بمزود السحابة الآخرين. أمثلة على SaaS:

أ. إنتاجية البريد الإلكتروني والمكتب: تطبيقات البريد الإلكتروني ومحررات ومعالجات الكلمات وتطبيقات جداول البيانات وتطبيقات العروض التقديمية هي أمثلة نموذجية في هذه الفئة.

ب. الفواتير: هناك تطبيقات مصممة لمراقبة وإدارة فواتير العملاء. يتم تحديد ذلك من خلال استخدام نظام المستخدمين والاشتراكات في المنتجات والخدمات.

ج. إدارة علاقات العملاء CRM: هي تطبيقات مركز اتصال نموذجية.

د. المالية: هذه تطبيقات مفيدة لتتبع الأنشطة المالية والإبلاغ عنها بما في ذلك معالجة النفقات وإنشاء الفواتير وكشوف المرتبات وإدارة الضرائب. [8]

3-4- مكونات السحابة:

تتكون الشركة السحابية من الموارد المخصصة للطلبات. وفقاً لـ NITS، يحتوي تكوين الحوسبة السحابية على خمسة جهات فاعلة رئيسية مدرجة.

3-4-1- مستهلك السحابة:

مستهلك السحابة هو شخص أو مؤسسة تحصل على بعض الخدمات من موفر السحابة. يمكن للمستهلك السحابي اختيار الخدمات الأكثر ملاءمة عن طريق التدقيق في الخدمات التي يقدمها موفرو السحابة وإغلاق العقد. لإغلاق هذا العقد، يتعين على مستهلك السحابة تحديد الأداء الفني من خلال توقيع اتفاقية مستوى الخدمة (SLA) مع موفر السحابة. تغطي اتفاقيات مستوى الخدمة (كاتفاقية) عدة نقاط مثل اتساق جودة الخدمات والأمان والوقاية وفشل الأداء. ومع ذلك، يمكن للمستهلك السحابي اختيار مقدمي الخدمات الذين يقدمون خدمات أكثر ملاءمة وأسعاراً أفضل. [6]

3-4-2- مزود السحابة:

مزود السحابة هو شخص أو مؤسسة توفر خدمة لمستهلك السحابة. يقوم موفر السحابة بتنظيم البرامج السحابية وترتيبها من خلال الحصول على البنية التحتية السحابية وإدارتها. في SaaS، يوفر موفر السحابة خدمات بمستويات الخدمة المتوقعة من خلال نشر تطبيقات البرامج وتكوينها وصيانتها وتحديثها. وفقاً للتطبيقات الإدارية المحدودة للسحابة، فإن معظم مسؤوليات الإدارة والتحكم في البنية التحتية والتطبيق تقع على عاتق مزود SaaS. في PaaS، يدير موفر السحابة البنية الأساسية للحوسبة للنظام الأساسي ويتم توفير مكونات النظام الأساسي (مثل برنامج وقت

التشغيل أو قاعدة البيانات أو مكونات البرامج الوسيطة الأخرى) بواسطة البرامج السحابية. في IaaS، يكتسب موفر السحابة موارد الحوسبة المادية بما في ذلك الشبكات والمخازن والخوادم والبنية التحتية للاستضافة. [6]

3-4-3- مدقق السحابة:

يكون مدقق السحابة مسؤولاً عن فحص الخدمات السحابية بشكل مستقل. يتحقق المدقق من التوافق مع المعايير من خلال مراجعة الأدلة الموضوعية المختلفة. يمكن تقييم الخدمات التي يقدمها موفرو السحابة بواسطة مدقق السحابة فيما يتعلق بتأثيرات الخصوصية، وضوابط الأمان، والأداء، وما إلى ذلك. [6]

3-4-4- وسيط السحابة:

نظراً لأن تكامل الخدمات السحابية معقد للغاية بحيث لا يمكن إدارتها بواسطة مستهلكي السحابة، فإنهم يحصلون على الخدمات السحابية عبر وسيط سحابي بدلاً من الاتصال بمزود السحابة مباشرةً. في الواقع، الوسيط السحابي مسؤول عن إدارة استخدام الخدمات السحابية وأدائها وتقديمها بالإضافة إلى العلاقات بين مستهلكي السحابة وموفري الخدمات السحابية. يمكن تصنيف الخدمات التي يقدمها الوسطاء السحابيون إلى ثلاث فئات:

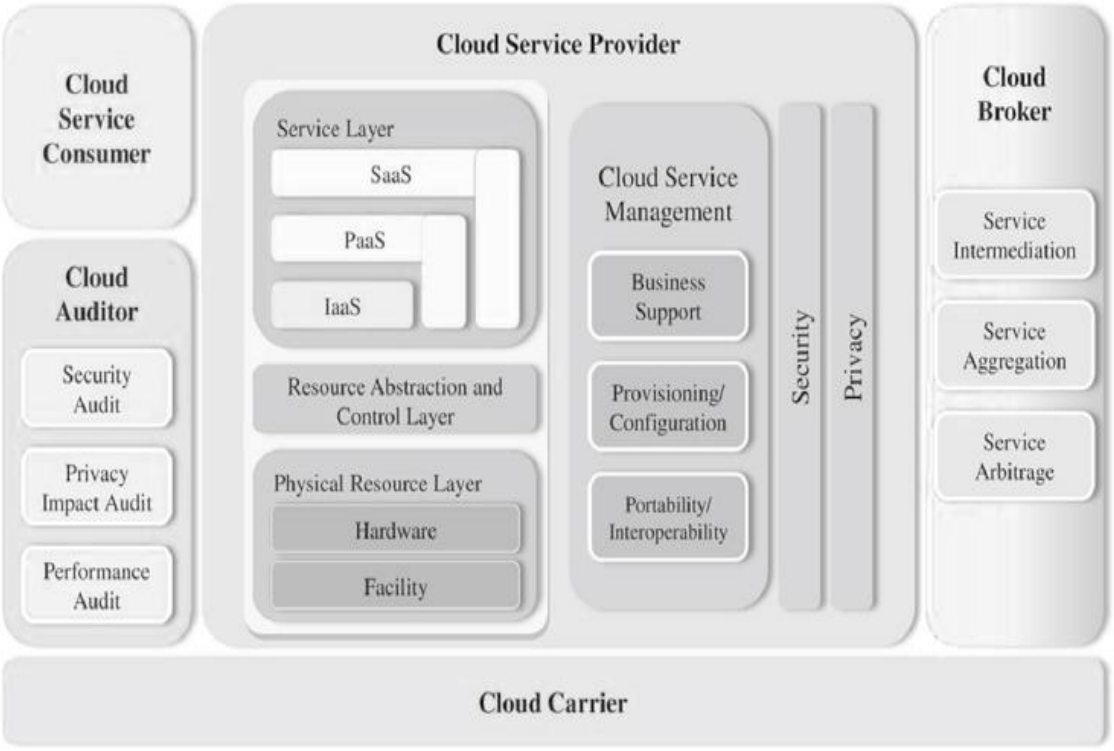
- وساطة الخدمة: يمكن للوسيط السحابي تعزيز خدمة معينة من خلال تحسين بعض القدرات المحددة وتقديم خدمات ذات قيمة مضافة لعملاء السحابة. تشمل هذه التحسينات إدارة الوصول إلى الخدمات السحابية، وتقارير الأداء، وإدارة الهوية، والأمن المحسّن، وما إلى ذلك.

- تجميع الخدمة: يشير إلى الإجراءات التي يتخذها وسيط السحابة لدمج أو دمج خدمات متعددة في واحدة أو أكثر من الخدمات الجديدة. في الواقع، يقوم الوسيط السحابي بدمج البيانات ويلعب دور الجسر بين مستهلك السحابة وموفري السحابة المتعددين.

- موازنة الخدمة: تعمل مثل تجميع الخدمات، ولكن لا يتم تجميع الخدمات لتكون ثابتة. في الواقع، تمنح موازنة الخدمة الوسيط المرونة لاختيار الخدمات من عدة وكالات. على سبيل المثال، يمكن للوسيط السحابي استخدام خدمة الائتمان لتقييم واختيار أفضل وكالة. [6]

3-4-5- الناقل السحابي:

يتوسط الناقل السحابي بين مستهلك السحابة ومزود السحابة لتقديم الخدمات السحابية. تحصل شركات النقل السحابية على إمكانية الوصول إلى المستهلكين من خلال الشبكة وأجهزة الوصول الأخرى. من خلال إنشاء اتفاقيات مستوى الخدمة مع مشغل شبكة سحابية، يمكن لمزود الخدمة السحابية تقديم خدمات متوافقة مع اتفاقيات مستوى الخدمة إلى مستهلكي السحابة. علاوة على ذلك، فإن الناقل السحابي مسؤول عن تخصيص الاتصالات الآمنة للمستهلك السحابي ومزود السحابة. في الشكل 6.



الشكل 6: تكوين الحوسبة السحابية NIST

يشير الرسم البياني إلى نظرة عامة على تكوين الحوسبة السحابية NIST بما في ذلك الجهات الفاعلة الرئيسية وأنشطتها ووظائفها في الحوسبة السحابية. وفقاً للمحتويات المذكورة، يصف الرسم البياني متطلبات وخصائص ومعايير الحوسبة السحابية. [6]

3-5- المفاهيم المرتبطة بالحوسبة السحابية:

3-5-1- الحوسبة السحابية الموزعة:

هناك حاجة إلى المعالجة المتوازية للتنفيذ الفعال للوظائف من خلال التطبيقات المعقدة. في العملية المتوازية، يتيح وجود الاتصال والمزامنة استخداماً أكثر فعالية لموارد وحدة المعالجة المركزية. وبالتالي بشكل عام، فإن الحفاظ

على مستوى استجابة الوظائف الموازية مع تحقيق الاستخدام الفعال للعقد أمر إلزامي لمركز البيانات [9]. من خلال الحوسبة السحابية، يمكن للعميل طلب المعلومات والموارد المشتركة والبرامج والخدمات الأخرى في أي وقت، وفقاً لمواصفاته. إنها خدمة عند الطلب، ويشيع المصطلح عبر الإنترنت. يمكن عرض الإنترنت بالكامل كسحابة. ناهيك عن أن استخدام السحابة يقلل من تكاليف رأس المال والتشغيل. ومع ذلك، فإن التحدي الرئيسي في الحوسبة السحابية هو موازنة التحميل، مطلوب دائماً حل موزع لهذه المشكلة. بسبب تعقيد السحابة والتوزيع الواسع لمكوناتها، من الصعب الحصول على موازنة تحميل فعالة من خلال تعيين وظائف للخوادم والعملاء المناسبين بشكل فردي، كما أنها ليست فعالة من حيث التكلفة أو عملية للوفاء بالمطالب المطلوبة من خلال الحفاظ على واحدة أو أكثر من الخدمات الخاملة[9] . تم اقتراح وتصميم بروتوكول، والغرض منه هو تعزيز إنتاجية الخادم وأدائه، وتعزيز استخدام الموارد ووقت التبديل. في هذا البروتوكول، يتم جدولة الوظائف في السحابة وفي البروتوكولات الحالية يتم حل العيوب. لتقليل وقت الانتظار والتبديل، يتم إعطاء الأولوية للمهمة التي تقدم أداءً أفضل للكمبيوتر. يتطلب حل عيوب البروتوكولات الحالية من خلال إدارة جدولة العمل جهداً كبيراً، إلى جانب تحسين الإنتاجية وكفاءة الخادم [9]. تسمح التطبيقات العلمية التي تستخدم الأنظمة القائمة على الحوسبة السحابية بالحوسبة عالية الإنتاجية (HTC). تحتوي التطبيقات في فيزياء الجسيمات على أنظمة ذات بنية تحتية موحدة تستخدم عدداً من سحب IaaS المتميزة.

هناك عدد من المعايير التي يعتمد عليها نظام الحوسبة السحابية لدينا. يجب تشغيل تطبيقات HEP الفردية المتوازية بشكل محرج في بيئة دُفعات في النظام. لا يلزم إجراء اتصالات داخلية أو اتصالات بين العمليات ؛ ومع ذلك، يتم تقليل بصمة الذاكرة التي تنتجها ذاكرة العملية المشتركة باستخدام مهام متعددة العمليات [9]. يتم تحقيق تكوين جزر الطوبولوجيا ومعالجة طوبولوجيا الشبكة السريعة من خلال المعالجة الطوبولوجية المتوازية CIM، والتي تعتمد عليها شبكة الطاقة وتمت مناقشتها. تم تصميم الإنتاجية العالية والموثوقية العالية والتوافر العالي للتخزين لمنصة التخزين السحابية من قبلهم. أيضاً، تم تقديم فكرة تصميم نموذج MySQL-CIM من قبلهم أيضاً وتم تحقيق نموذج أنبوب MySQL-CIM الفعال في البيانات من خلال تطوير Ogma. تم تطوير تطبيق معالجة طوبولوجيا شبكة الطاقة (NTP) [9]. الحوسبة السحابية مصطلح واسع يشير بشكل عام إلى الخدمات المستضافة. إنها المحاكاة الافتراضية للأجهزة المادية حيث يتم تنظيم البيانات في مراكز محددة. ومع ذلك، باعتبارها تقنية جديدة لها مزايا عديدة، فهي لا تخلو من مشاكلها، والتحدي الرئيسي هو موازنة الأحمال، وموضوع مهم في الحوسبة السحابية يتطلب العديد من الأبحاث والدراسات. نظراً لأن العديد من الأنظمة تشارك في هيكل مركز البيانات، فمن الصعب إجراء موازنة الحمل، خاصة في حالة الحوسبة السحابية. تم إجراء معظم الأبحاث حول هذا الموضوع في بيئة موزعة، ومع ذلك كان استخدام موازنة الحمل شبه الموزعة هدفاً لمجرد القليل من البحث. موازنة الحمل بطريقة شبه موزعة تسمح بتصميم خوارزمية جديدة للحوسبة السحابية [9].

تمت مناقشة مزايا الحوسبة اللامركزية بعيداً عن مراكز البيانات، جنباً إلى جنب مع النظر في استخدام البنية التحتية التي تم الحصول عليها من عدة مزودين وتغيير البنية التحتية للسحابة. تتطلب هذه الاتجاهات الجديدة بنية جديدة

للحوسبة ويجب أن تتحقق من خلال البنية التحتية السحابية في المستقبل. أنظمة التعلم الذاتي، ومساحة الخدمة، والحوسبة كثيفة البيانات، وتوصيلات الأشخاص - الأجهزة هي المجالات التي يُتوقع أن تتأثر أكثر بهذه البنى الجديدة. أخيراً، من أجل تحقيق إمكانات الجيل التالي من أنظمة السحابة، تم اقتراح خارطة طريق للتحديات التي يجب أخذها في الاعتبار. [9] تمت مناقشة إنشاء السحب المخصصة وتسخير الحوسبة للتطبيقات عبر الإنترنت والجوال على حد سواء على حافة الشبكة، باستخدام نماذج الحوسبة القائمة على توفير الموارد طواعية. إن فكرة الدفع مقابل جهاز افتراضي سحابي حتى لو كان الخادم المنفذ على الجهاز الافتراضي خاملاً هي فكرة تقليدية، وقد قدمنا نموذجاً للحوسبة لاستبداله. تم ذكر نموذج حوسبة للسحابة يدمج المرونة ويتم تعريفه بالبرمجيات. ستتأثر بعض المناطق ببنية الحوسبة حديثة التكوين والبنية التحتية السحابية المتغيرة. سيتم تسهيل نموذج إنترنت الأشياء لتعزيز الاتصال بين الأشخاص والأجهزة، وستلعب البنى الجديدة دوراً حيوياً. يوفر حجم البيانات تحدياً في الحوسبة كثيفة البيانات وهناك حاجة إلى تقنيات جديدة لمعالجة هذا الأمر. سيكون هناك اهتمام متزايد بالخدمات الجديدة مثل التسريع والحاويات والوظيفة. سيتم تحقيق نظام التعلم الذاتي عندما يكون هناك تقارب بين مناطق البحث والأنظمة السحابية. الأوساط الأكاديمية والصناعة هي القوى الرائدة في هذه التغييرات، ولكن هناك العديد من التحديات التي يجب حلها في المستقبل. تمت مناقشة تطوير الجيل التالي من الحوسبة السحابية بأنظمة مستدامة وإدارة فعالة والتعبير عن التطبيقات وتحسين الأمان ، جنباً إلى جنب مع نهجها واتجاهها [9]. تعد الحوسبة الموزعة والحوسبة المتوازية جزءاً من الحوسبة السحابية. أولاً، تم تقديم نموذجين تقليديين للبرمجة المتوازية، جنباً إلى جنب مع شرح مفهوم الحوسبة السحابية. ثانياً، يقوم بتحليل ودراسة مبادئ ومزايا وعيوب OpenMP وMPI وMap Reduce، على التوالي. أخيراً، يناقش ويقارن نماذج MPI وOpenMP وMap Reduce من زاوية الحوسبة السحابية. تهدف نتائج هذه الورقة إلى توفير مرجع لتطوير الحوسبة المتوازية. تم التركيز بشكل كبير على الحوسبة الموزعة. تمت دراسة الفروق بين الحوسبة الموزعة والمتوازية أيضاً، جنباً إلى جنب مع المصطلحات، وتخصيص المهام، ومعلمات الأداء، ومزايا ونطاق الحوسبة الموزعة، بالإضافة إلى نماذج الخوارزمية الموزعة المتوازية. [9]

3-5-2- مقاييس موازنة تحميل السحابة:

تتم مناقشة المقاييس المختلفة التي تم أخذها في الاعتبار في تقنيات موازنة الحمل الحالية في الحوسبة السحابية أدناه: [9]

1- يتم قياس عدد المهام المنفذة باستخدام معدل النقل، ويشير الرقم المرتفع إلى أداء جيد للنظام.

2- عند تطبيق خوارزمية لموازنة الحمل، يتم قياس مشاركة النفقات العامة من قبل الشركات الزميلة. تتكون عملية تعبئة العمليات بين العمليات والمعالجات وتعبئة المهام من النفقات العامة وكلما كانت تقنية موازنة الحمل أكثر كفاءة، قل مقدار الحمل. [9]

3- يحتاج موازنة الحمل إلى تقنية جيدة للتسامح مع الخطأ. وهي القدرة على القيام بموازنة تحميل موحدة بواسطة خوارزمية، على الرغم من فشل الارتباط أو العقدة التعسفية.

4- نظم الأداء الجيد قللت من أوقات الترحيل، وهو الوقت اللازم لترحيل الموارد أو الوظائف بين الإيماءات الفردية، فكلما قل ذلك كان أفضل. [9]

5- وقت الاستجابة هو عامل آخر، إذا تم تقليله يؤدي إلى أداء أفضل للنظام وهو الوقت اللازم لخوارزمية موازنة حمل معينة للاستجابة في نظام موزع.

6- يعد الاستخدام الفعال للموارد أمراً إلزامياً لتحقيق موازنة فعالة للأحمال ويجب إجراء التحسين.

7- يتم تحديد قابلية توسيع الخوارزمية من خلال قدرتها على أداء موازنة الحمل لأي عدد محدود من العقد في النظام. قابلية التوسع المحسن هو المطلوب. [9]

8- تقاس فعالية النظام من خلال أدائه، ومع ذلك فإن فعالية التكلفة تحتاج إلى النظر فيها وجعلها معقولة. مثال على ذلك هو الاحتفاظ بالتأخيرات المقبولة مع تقليل أوقات استجابة المهام.

3-6- فوائد الحوسبة السحابية للتعليم عن بعد:

3-6-1- تشجيع التعاون والتواصل:

يمكن للحوسبة السحابية خلق روح التعاون بين المتعلمين والمعلمين وغيرهم في الأوساط الأكاديمية. نظراً لسهولة الوصول إلى الملفات، يمكن للمستخدمين المتنوعين إجراء تعديلات على أي مخطوطة، مثل إدارة الخطط أو مشاريع الدراسة. سيكون من السهل على المعلم اقتراح تعديلات على مقال أو مهمة مقدمة من الطالب. كل ما عليك فعله هو الوصول إلى ملف الطالب في السحابة وحفظ ملاحظاته وإخطار الطالب من خلال النظام. لا مزيد من المواد الورقية، ولا مراسلات متعددة عبر البريد الإلكتروني، كل شيء مدمج في نظام واحد لسهولة الوصول. [10]

3-6-2- التكلفة المادية:

توفر السحابة الإلكترونية العديد من الموارد والخدمات المشتركة الضرورية، مما يمكّن الأساتذة والطلاب من تقليل الإنفاق على مواد ومنهجية الدراسة. يستطيع المحاضر تحميل المحاضرات على السحابة الإلكترونية. وبذلك تزول الحاجة إلى الكتب والمحاضرات الورقية وتقل تكاليف الطباعة والاستنساخ. ما يحتاجه الطالب بالكامل هو جهاز كمبيوتر يمكّنه من الوصول إلى مواد دراسية متنوعة على السحابة عبر الإنترنت. كما يمكن للأقسام العلمية تقليل النفقات المالية وتقليل نفقات التشغيل الجارية من خلال الحوسبة السحابية والاستفادة من خبرات الموظفين المتخصصين دون الحاجة إلى اللجوء إلى البحث عن موظفين وتوظيفهم وتدريبهم ودفع رواتبهم. [10]

3-6-3- حفظ السجلات:

تُمكِّن الحوسبة السحابية المعلمين والطلاب من فحص السجلات السحابية عند حدوث أي خلاف أو مشكلة، حيث من الممكن ببساطة التحقق من سجلات الإدخال، والنظر في الملفات التي تم تحميلها، والتحقق مما إذا تم تقديم الدورات الدراسية في وقت متأخر.

3-6-4- وسائل الراحة:

تقدم الحوسبة السحابية لمعلمي الجامعات والطلاب على حد سواء بمعرفة تعليمية أكثر ملاءمة وكفاءة. تحتفظ السحابة بجميع البيانات في مكان واحد، بدءاً من سجلات حضور الفصل وحتى الواجبات المنزلية وتصفح المناهج والمزيد. يمكن للجميع الوصول إلى النظام والوصول إلى مواد المناهج المختلفة، وهذه العملية ذات فائدة أكبر للطلاب الذين يدرسون عبر الإنترنت ويحتاجون إلى المرونة للنجاح في إكمال برنامجهم التعليمي. ربما، في المستقبل، سيكون للحوسبة السحابية تأثير كبير على البيئة التعليمية لأنها يمكن أن توفر البنية التحتية للمتعلم والموارد اللازمة لتنفيذ أي عدد من المهام على السحابة مع تقليل التكلفة وتزويدهم أيضاً بسهولة الوصول إلى النطاق الضخم. كمية المعلومات المتوفرة على الإنترنت. [10]

3-6-5- سهولة التنفيذ:

يمكن للمؤسسة التعليمية التكيف والنشر دون شراء الأجهزة والبرامج، وتثبيت تصاريح البرامج، والنشاط، وخدمات الصيانة، وتطبيقات الحوسبة السحابية. شراء الأجهزة والبرامج، وتراخيص البرامج، والتثبيت، والنشاط، والصيانة. [10]

3-6-6- الإمكانية:

لا تحتاج المؤسسات التي تستخدم الحوسبة السحابية إلى إضافة أجهزة وبرامج ذات معايير وكفاءات أعلى عند زيادة أعداد المعلمين. ليسوا ملزمين بشراء موسع جديد لشراء المزيد من أجهزة الكمبيوتر وأنظمة التخزين والمحولات وأجهزة التوجيه. تضمن خدمة الحوسبة السحابية السرعة في الانضمام والتعامل مع التقنيات الحديثة على الإنترنت. [10]

3-6-7- الوصول إلى السعة القصوى لتكنولوجيا المعلومات:

يمكن أن تسمح الحوسبة السحابية للمعلمين بالوصول إلى أعلى كفاءة من البرامج التي قد تجذب الطلاب لمتابعة دروسهم اليومية، وخاصة للمؤسسات التعليمية الصغيرة.

3-6-8- إعادة توزيع المعلمين:

من خلال حذف أو إزالة التحديثات التي تتطلبها الخوادم الثابتة ومشاكل الحوسبة الأخرى التي تقلل التكاليف أو الوقت أو المال أو تطوير التطبيقات، يمكن لأقسام تكنولوجيا المعلومات التركيز على المهام ذات القيمة الأعلى والتركيز على الكفاءات الأساسية. [10]

3-6-9- القابلية للقياس:

يمكن قياس استخدام موارد وخصائص الحوسبة السحابية. يجب إكمال هذا لكل طالب وتطبيقه يومياً وأسبوعياً وشهرياً وسنوياً [24].

4- الفصل الرابع:

تعليم الآلة Machine Learning:

4-1- مقدمة:

تعليم الآلة هو حقل فرعي من الذكاء الاصطناعي(AI) ، ويستخدم لاستكشاف بنية البيانات وتناسبها مع النماذج التي يمكن فهمها واستخدامها من قبل المستخدمين، ويجيب على السؤال حول كيفية إنشاء برنامج كمبيوتر باستخدام البيانات التاريخية، لحل مشكلة معينة وتحسين كفاءة البرنامج تلقائيًا. في السنوات الأخيرة، تم تطوير العديد من تطبيقات تعليم الآلة، مثل اكتشاف المعاملات المصرفية الاحتيالية، وأنظمة تصفية المعلومات التي تتعلم من تفضيلات القراءة للمستخدم، والدراسات البيولوجية العصبية، والمركبات المستقلة التي تتعلم القيادة على الطرق السريعة. وفي الوقت نفسه أيضًا، كان هناك تقدم مهم في المفاهيم والخوارزميات التي تشكل أساس التعلم الآلي. حالياً يتمتع كل مستخدم للتكنولوجيا بميزة من التعلم الآلي [18]. مثل:

-تقنية التعرف على الوجه لوسائل التواصل الاجتماعي.

-اقتراحات محركات التوصية للأفلام أو البرامج التلفزيونية لمشاهدتها بناءً على تفضيلات المستخدم.

-تقنية التعرف الضوئي على الأحرف (OCR) لتحويل صور النص إلى كتابة قابلة للتحرير.

-تعليم الآلة للسيارات ذاتية القيادة للتنقل في طريقها. [18]

4-2- تصنيفات تعليم الآلة:

يتم تصنيف تعليم الآلة على نطاق واسع على أنه تعلم خاضع للإشراف وغير خاضع للإشراف وشبه إشراف وتعلم معزز.

4-2-1- التعلم الخاضع للإشراف Supervised Learning:

هو نموذج تعليمي تم إنشاؤه لإجراء التنبؤ، بالنظر إلى مثيل الإدخال غير المتوقع. تأخذ خوارزمية التعلم الخاضع للإشراف مجموعة معروفة من بيانات الإدخال واستجاباتها المعروفة للبيانات (المخرجات) لتعلم نموذج الانحدار / التصنيف(regression/classification). ثم تقوم خوارزمية التعلم بتدريب نموذج لتوليد تنبؤ للاستجابة للبيانات الجديدة أو مجموعة بيانات الاختبار. تتضمن الخوارزميات الانحدار الخطي linear regression، والانحدار اللوجستي logistic regression ، الشبكات العصبية neural networks، شجرة القرار decision tree، Support Vector Machine (SVM)، الغابة العشوائية random forest، Naive Bayes، k-nearest neighbor [18] . تصنف نماذج التصنيف بيانات الإدخال إلى فئات. تشمل تطبيقات التصنيف الشائعة أو الرئيسية التصنيف الائتماني للبنك والتصوير الطبي والتعرف على الكلام. أيضًا، يستخدم التعرف على خط اليد التصنيف للتعرف على الأحرف والأرقام. يحاول الانحدار الخطي نمذجة العلاقة بين متغيرين من خلال ملاءمة المعادلة الخطية للبيانات المرصودة.

4-2-2- التعلم غير الخاضع للإشراف Unsupervised Learning:

في التعلم غير الخاضع للإشراف، يتمثل الهدف في العثور على الانتظام في المدخلات بحيث تحدث أنماط معينة في كثير من الأحيان أكثر من غيرها وتعلم رؤية ما يحدث وما لا يحدث بشكل عام، على سبيل المثال: تجميع المستندات وضغط الصور هي ميادين عملية لاستخدام التعلم غير الخاضع للإشراف. [18] يوجد في الواقع

طريقتان للتعلم غير الخاضع للإشراف. النوع الأول هو تعليم الوكيل باستخدام نوع من نظام المكافآت للإشارة إلى النجاح. ليس الهدف إنتاج تصنيف ولكن اتخاذ قرارات تزيد من المكافآت. يعمم هذا النهج بشكل جيد على العالم الحقيقي، هذا النوع من التعلم يمكن أن يكون قويًا لأنه يفترض عدم وجود تصنيف مكتشف مسبقًا للأمثلة. في بعض الحالات ، على سبيل المثال ، لعبة الشطرنج التي تعلمت من خلال التعلم غير الخاضع للإشراف أقوى من أفضل لاعبي الشطرنج البشريين فقط من خلال اللعب بأنفسهم مرارا و تكرارا. [19] اكتشفت هذه البرامج بعض المبادئ التي فاجأت خبراء لعبة الشطرنج وأداؤها أفضل من برامج لعبة الشطرنج المدربة على أمثلة مصنفة مسبقًا. النوع الثاني من التعلم غير الخاضع للإشراف يسمى التجميع(clustering). في هذا النوع من التعلم، لا يتمثل الهدف في تعظيم وظيفة المنفعة، ولكن ببساطة لإيجاد أوجه التشابه في بيانات التدريب. غالبًا ما يكون الافتراض أن المجموعات التي تم اكتشافها ستتطابق بشكل معقول مع تصنيف بديهي. على سبيل المثال، قد يؤدي تجميع الأفراد على أساس التركيبة السكانية إلى تجمع الأثرياء في مجموعة والفقراء في مجموعة أخرى. على الرغم من أن الخوارزمية لن تحتوي على أسماء لتعيينها لهذه المجموعات، إلا أنها يمكن أن تنتجها ثم تستخدم تلك المجموعات لتعيين أمثلة جديدة في واحدة أو أخرى من المجموعات. [19] هذا نهج يعتمد على البيانات ويمكن أن يعمل بشكل جيد عندما تكون هناك بيانات كافية؛ على سبيل المثال، تعتمد خوارزميات تصفية المعلومات الاجتماعية، مثل تلك التي يستخدمها موقع Amazon.com للتوصية بالكتب، على مبدأ العثور على مجموعات مماثلة من الأشخاص ثم تعيين مستخدمين جدد لمجموعات. في بعض الحالات، يمكن أن تكون المعلومات حول الأعضاء الآخرين في الكتلة (مثل الكتب التي يقرؤونها) كافية للخوارزمية لإنتاج نتائج ذات مغزى. في حالات أخرى، قد تكون المجموعات مجرد أداة مفيدة لمحلل بشري، هنا لابد لنا من معرفة أن التعلم غير الخاضع للإشراف يعاني أيضاً من مشكلة فرط تجهيز بيانات التدريب. [19]

4-2-3- التعلم شبه الخاضع للإشراف Semi-supervised Learning:

هو مزيج من التعلم الخاضع للإشراف والتعلم غير الخاضع للإشراف، ويتم استخدامه عندما يتم تحديد عدد أقل من البيانات المسماة (labeled) لتطبيق معين. يقوم بإنشاء وظيفة للتحويل من مدخلات كل من البيانات المسماة والبيانات غير المسماة. الهدف من التعلم شبه الخاضع للإشراف هو تصنيف بعض البيانات غير المسماة باستخدام مجموعة المعلومات المسماة، ومن الأمثلة الحقيقية لاستخدام هذا النوع: تصنيف تسلسل البروتين وتصنيف محتوى الويب وتحليل الكلام. [18]

4-2-4- التعلم المعزز Reinforcement Learning :

يتضمن التعلم المعزز التفاعل مع البيئة المحيطة، حيث يعالج التعلم المعزز مسألة كيف يمكن للعامل المستقل الذي يستشعر ويتصرف في بيئته أن يتعلم كيفية اختيار الإجراءات المثلى لتحقيق أهدافه. يُكافأ سلوك الوكيل بناءً على الإجراءات التي يتخذها في البيئة، فهي تنظر في عواقب أفعالها وتتخذ الخطوات المثلى إلى أبعد من ذلك، على سبيل المثال: الكمبيوتر الذي يلعب الشطرنج مع الإنسان، التعرف على الكلمات المنطوقة، وتصنيف الهياكل الفلكية

الجديدة. [18] هنا تتعلم الخوارزمية كيفية التصرف مع وجود تغذية عكسية، حيث يكون لكل إجراء بعض التأثير في البيئة، وتوفر البيئة تغذية عكسية لخوارزمية التعلم.

4-3- المصطلحات المستخدمة في مفاهيم تعلم الآلة:

إن مفهوم تعلم الآلة يؤطر البيانات في سياق الوظيفة الافتراضية (f) التي تهدف خوارزمية التعلم الآلي إلى تعلمها. تقوم الوظيفة f بحساب متغير الإخراج المتوقع (output) بالاعتماد على متغير الإدخال (input).

Output=f(Input)

من حيث قاعدة البيانات، يصف السطر كيانًا أو ملاحظة حول كيان. يُشار إلى عمود الصف على أنه سمة للملاحظة. المتغيرات هي خصائص أو نوع خصائص حدث أو كائن معين في مجموعة البيانات. يمكن أن تكون متغيرات تابعة أو مستقلة. المتغيرات المستقلة هي المتغيرات التي يمكن للباحثين أو محلل البيانات معالجتها أو تغييرها ومن ثم يتم قياس تأثيراتها ومقارنتها بها. تُعرف المتغيرات المستقلة أيضًا باسم المتنبئ، [18] حيث تسمى كذلك لأنها تتنبأ بقيم المتغير التابع في النموذج. تشير المتغيرات التابعة إلى نوع المتغير الذي يقيس تأثير المتغير (المتغيرات) المستقلة على وحدات الاختبار. يُعرف المتغير التابع أيضًا باسم المتغير المتوقَع لأنها القيم التي يتوقعها أو يفترضها المتنبئ أو المتغيرات المستقلة. لنفترض سيناريو تكون فيه درجة الطالب في الاختبار متغيرًا تابعاً، حيث قد يتغير اعتمادًا على عدة عوامل، مثل مقدار دراسته لموضوع معين، أو مقدار نومه في الليلة السابقة للامتحان، أو مدى جوعه عندما أخذ الامتحان. بشكل عام ، عندما يبحث المحلل عن علاقة بين شيئين ، فإنه يحاول معرفة ما الذي يجعل المتغير التابع يغير الطريقة التي يتغير بها .[18]

الهدف الرئيسي من التعلم الآلي هو بناء نموذج يعمل بشكل جيد في كل من مجموعة التدريب ومجموعة الاختبار، يوجد عدة مفاهيم يجب الإضاءة عليها منها:

أ- **التعميم**: هو مفهوم في التعلم الآلي يخبرنا عن مدى جودة أداء النموذج في البيانات الجديدة أو البيانات التي لم تكن مرئية من قبل. يمكن للنموذج ذو القدرة التعميمية القوية أن يشكل مساحة العينة بأكملها بشكل جيد للغاية. [18]

ب-**مجموعة التدريب**: هي البيانات المستخدمة لتدريب النموذج.

ج-**مجموعة التحقق من الصحة**: يتم تقسيم مجموعة البيانات المتاحة بشكل عشوائي إلى مجموعة تدريب ومجموعة التحقق من الصحة. يتناسب النموذج مع مجموعة التدريب، ويتم استخدام النموذج المناسب للتنبؤ بالاستجابة في مجموعة التحقق من الصحة. يتم تقييم معدل خطأ مجموعة التحقق الناتج باستخدام متوسط الخطأ التربيعي (Mean Squared Error-MSE)لتقدير معدل خطأ الاختبار. [18]

د- **مجموعة الاختبار**: هي البيانات المستخدمة لاختبار النموذج المبني بالفعل، وقد تكون هذه البيانات مشابهة لقيم مجموعة بيانات التدريب وقد لا تكون كذلك. يشير الخطأ الموجود في بيانات الاختبار إلى مدى جودة أداء المصنف في البيانات الجديدة. هذا هو التعميم ، ويعرف خطأ الاختبار باسم خطأ التعميم. [18]

ه- **الملاءمة الزائدة (overfitting):** هو سيناريو عندما يكون لدى المصنف خطأ تدريب منخفض وخطأ اختبار مرتفع. في حالة الملائمة الزائدة، يحاول المصنف أو النموذج التقاط كل نقطة عينة بدلاً من عينة أصلية عند النقطة.

و- **التباين Variance:** عندما يكون أداء النموذج جيدًا حقًا في مجموعة بيانات التدريب وضعيفًا عند المرور على كل البيانات كبيانات تحقق cross-validation data، يُقال إن للنموذج تباينًا. هذا يعني أن النموذج غير ملائم.

ز- **الملاءمة المنخفضة (underfitting):** يحدث عندما يكون النموذج بسيطًا جدًا ويتم إعلامه بميزات قليلة جداً أو أنه منظم بشكل كبير، مما يجعل النموذج جامدًا تجاه التعلم من مجموعة البيانات. على سبيل المثال، إذا كان وقت التعلم كبيرًا جدًا، أوتم إنهاء مرحلة التعلم قبل الأوان، أو إذا لم يستخدم المتعلم عددًا كافيًا من التكرارات، أو إذا حاول المتعلم وضع خط مستقيم في مجموعة التدريب التي تظهر أمثلتها طبيعة تربيعية، وهذا سيؤثر سلبًا على أدائه في عمل تنبؤات دقيقة حول البيانات الجديدة التي لم يرها أثناء التدريب. [18]

ح-**التحيز**: يُقال إن نموذج التعلم لديه تحيز عندما يكون أداؤه ضعيفًا على بيانات التدريب وكذلك على بيانات التحقق من الصحة. يمكن أن يكون الحل المحتمل للتغلب على التحيز هو: إنشاء نموذج معقد (أكبر).

4-4- أنواع خوارزميات تعليم الآلة:

4-4-1- خوارزميات الانحدار:

مثل الانحدار الخطي (Linear regression) والانحدار اللوجستي (Logistic Regression). تتعامل خوارزميات الانحدار مع نمذجة العلاقة بين المتغيرات التي يتم تنقيحها بشكل متكرر باستخدام مقياس الخطأ في التنبؤات التي قدمها النموذج. **الانحدار الخطي:** هو طريقة شائعة لتحليل البيانات الموصوفة في نموذج خطي بطبيعته، وهو محاولة لنمذجة العلاقة بين متغيرين من خلال ملاءمة معادلة خطية للبيانات المرصودة. [18]

الانحدار اللوجستي: يُستخدم الانحدار اللوجستي لتقدير احتمالية أن يكون الحدث "نجاحًا" وأن يكون الحدث "فشلًا". يجب على المرء أن يذهب إلى الانحدار اللوجستي عندما يكون متغير الخرج تابعاً ثنائيًا بطبيعته. يوجد أنواع أخرى من الانحدارات المتاحة هي الانحدار متعدد الحدود polynomial regression، الانحدار التدريجي regression stepwise، انحدار التلال ridge regression، انحدارlasso ، وانحدار الشبكة المرنة elastic net regression. [18]

4-4-2- خوارزميات التصنيف

التصنيف هو من أنواع التعلم الخاضع للإشراف. ويعبر عن خوارزميات تعليم الآلة الغير بارامترية أنها الخوارزميات التي لا تضع أي فرضيات قوية حول شكل وظيفة المقابلة(mapping). عادة ما تكون الطرق الغير بارامترية أكثر مرونة وتحقق دقة أفضل ولكنها تتطلب بيانات ضخمة ووقت تدريب. تعد SVMs والشبكات العصبية وأشجار القرار أمثلة على الخوارزميات الغير بارامترية. [18]

4-4-2-1- خوارزمية شجرة القرار:

الغرض من بناء شجرة قرار هو التنبؤ بالمتغير المستهدف بناءً على متغير الإدخال من خلال تعلم قواعد القرار، حيث تُستخدم أشجار القرار بشكل شائع لبيانات التصنيف والانحدار. تنتج نماذج شجرة التصنيف مجموعة منفصلة من القيم للمتغيرات المستهدفة، بينما تأخذ نماذج شجرة الانحدار مجموعة مستمرة من القيم للمتغيرات المستهدفة، فعند بناء نموذج لسمة التصنيف بناءً على السمات الأخرى، تأخذ شجرة القرار كائنًا كمدخل يتم وصفه بواسطة مجموعة من الخصائص. يعتمد القرار في كل مرحلة من مراحل الشجرة على عمليات التفريع السابقة، حيث يبدأ بناء الشجرة بالعقدة الجذرية، ثم يقوم بإجراء الاختبار الذي يتبع الحافة، ويكرر الاختبار حتى يصل إلى عقدة النهاية (الورقة)، وهنا تتوقع الشجرة النتيجة المرتبطة بها، أي تسمية الفئة. [18]

4-4-2-2- تصنيف Naive Bayes:

هي خوارزمية بسيطة وفعالة للنمذجة التنبؤية. يشتمل النموذج على نوعين من الاحتمالات التي يمكن حسابها مباشرة من بيانات التدريب:

1- احتمال كل فئة.

2- الاحتمال الشرطي لكل فئة مع إعطاء كل قيمة لـ x. بمجرد اكتمال عملية الحساب.

يمكن استخدام نموذج الاحتمالية لعمل تنبؤات للبيانات الجديدة باستخدام نظرية بايز. عندما تكون البيانات ذات قيمة حقيقية، فمن الشائع افتراض توزيع غاوسي (منحنى الجرس) بحيث يمكن للمرء بسهولة تقدير هذه الاحتمالات. يُطلق علىNaive Bayes اسم ساذج لأنه يفترض أن كل متغير إدخال مستقل. هذا افتراض قوي وغير واقعي للبيانات الحقيقية. ومع ذلك، فإن هذه التقنية فعالة للغاية في مجموعة كبيرة من المشاكل المعقدة. [18]

4-4-2-3- خوارزمية Support Vector Machine SVM

هي واحدة من أشهر الخوارزميات التي تفصل بين فئتين. تتميز بإمكانية إجراء التصنيف الخطي، وإدارة البيانات غير الخطية بكفاءة باستخدام طريقة تُعرف باسم وظيفة kernel والتي بدورها تقوم بتعيين مدخلات المتجه إلى مساحة ميزة عالية الأبعاد. تعتبر نظرية التعلم الإحصائي أساساً لـSVM . توجد عدة نماذج من هذه الخوارزمية: [18]

1-Linear SVM

2-Hard Margin Classifier: شكل صارم من SVMحيث يجب أن تفي جميع العينات بالشرط الهامشي. يعمل هذا النموذج على النحو المطلوب عندما تكون مجموعة البيانات خالية من الأخطاء (الضوضاء أو القيم المتطرفة)، أما في حالة وجود أخطاء، سيكون الهامش الذي يفصل بين الفئات صغيرًا أو قد يفشل هذا النموذج. [18]

3-Soft Margin Classifier: تحتوي البيانات الواقعية التي يتم تطبيق خوارزميات تعليم الآلة عليها على ضوضاء أو قيم متطرفة، ومن المستحيل عمليا فصل البيانات المفيدة عن الضوضاء. تتمثل إحدى طرق التعامل

مع هذا الحالة في تخفيف القيود المفروضة على وجود الحد الأقصى للهامش للخط الذي يفصل بين الفئات. يسمح هذا التراخي لبعض النقاط بانتهاك الخط الفاصل وتكون على الجانب الآخر.

Nonlinear SVM-4: النماذج السابقة تفترض حقيقة كون البيانات خطية. ولكن من الناحية العملية، لا يمكن فصل العديد من مجموعات البيانات خطيًا. لذا فإن افتراض أن البيانات خطية غير واقعي. هنا يمكن تحويل SVM من خطية إلى غير الخطية من خلال تطبيق وظائف النواة المعروفة أو حيل النواة. [18]

4-4-3- خوارزميات التجميع(Clustering):

التجميع هو وضع تعلم غير خاضع للإشراف، يهدف لاستخلاص استنتاجات من مجموعات البيانات التي تتكون من بيانات الإدخال بدون تسمية أو قيم مستهدفة، أو لتحليل البيانات الاستكشافية للعثور على أنماط مخفية. تقوم خوارزميات التجميع بتجميع البيانات المتشابهة مع بعضها البعض في مجموعة واحدة وتوزيع البيانات التي تختلف كثيرًا عن بعضها البعض في مجموعات مختلفة. يتم تشكيل المجموعات أو تنظيم البيانات بحيث تكون المسافة داخل المجوعة أقل والمسافة بين المجموعات أكبر. مثلا تجميع النصوص، دراسة تسلسل الجينات، التعرف على الكائن في صورة، وغيرها، [18] توجد عدة انواع لخوارزميات التجميع: k-means , k-medoids, hierarchical clustering, hidden Markov models , self-organizing maps , and fuzzy c-means clustering

4-4-3-1- k-Means Clustering:

k-meansعبارة عن خوارزمية تجميع ، يتم فيها تقسيم نقاط البيانات إلى عدد k من المجموعات، وكل مجموعة لديها مركز مجموعة يسمى النقطه الوسطى، وتتميز هذه الخوارزمية بأنها بسيطة وسهلة الفهم والتنفيذ، ولكن لا يمكن استخدام خوارزمية k-mean إلا إذا كانت قيمة "k" محددة جيدًا في البداية، أيضاً قد لا تعمل الخوارزمية بشكل جيد مع القيم المتطرفة، حيث يمكن أن تكون القيم المتطرفة خطأ في مجموعة البيانات أو نقطة بيانات خاصة ذات قيمة غير عادية.

4-4-3-2- التجميع الهرمي (Hierarchical Clustering):

تعمل خوارزمية التجميع الهرمي على مفهوم تجميع كائنات البيانات في تسلسل هرمي من "شجرة" المجموعات. ينقسم التجميع الهرمي إلى مجموعات تكتلية أو تقسيمية، وذلك اعتمادًا على ما إذا كان التحلل الهرمي قد تم تشكيله في نهج من أسفل إلى أعلى (دمج) أو من أعلى إلى أسفل (تقسيم). [18]

4-5- تقييم خوارزميات تعلم الآلة:

توجد اساليب متنوعة لتقييم خوارزميات تعليم الآلة وتحديد مدى جودتها وتحقيقها للمتطلبات منها:

4-5-1- الدقة Accuracy:

هي مقدار الحالات المصنفة بشكل صحيح من إجمالي الحالات. يتم تعريفه على أنه نسبة عدد التنبؤات الصحيحة إلى العدد الإجمالي للتنبؤات. ومع ذلك، إذا كانت بيانات الاختبار غير متوازنة (أي عندما تنتمي معظم الحالات أو السجلات إلى أحد الفئات)، فإن الدقة لا تمثل فعالية المصنف. [18]

4-5-2- مصفوفة الارتباك Confusion Matrix:

هو جدول يسجل عدد المثيلات في مجموعة البيانات التي تقع في فئة معينة، حيث يمكن أن تأخذ تسمية الصنف في مجموعة التدريب الثنائية قيمتين محتملتين، والتي تسمى فئة إيجابية وفئة سلبية. كما هو موضح في الجدول،

		Predicted class	
		Positive	Negative
Actual Class	True	TP	FN
	False	FP	TN

الجدول 1: مصفوفة الارتباك

فإن عدد الحالات الإيجابية والسلبية التي يتوقعها المصنف بشكل صحيح يسمى الإيجابيات الحقيقية (TP) والسلبيات الحقيقية(TN) ، على التوالي. تُعرف الحالات التي تم تصنيفها بشكل خاطئ باسم الإيجابيات الكاذبة (FP) والسلبيات الكاذبة .(FN)يقرر نموذج التعلم بشكل لا إرادي أي فئتين في مجموعة البيانات هي الفئة الإيجابية. إذا كانت تسميات فئة مجموعة البيانات عبارة عن سلاسل، يتم أولاً فرز الملصق أبجديًا وفي المستوى الأول يتم اختياره كفئة سلبية، بينما يتم اختياره في المستوى الثاني كفئة موجبة. إذا كانت تسميات الفئة منطقية أو عددًا صحيحًا في طبيعتها ، فسيتم تعيين مثيلات معنونة "1" أو (true) كفئة موجبة [18].

4-5-3- الدقة والاسترجاع Precision and Recall:

-الدقة precision: هي نسبة عدد الحالات ذات الصلة على عدد الحالات المسترجعة(الكلية).

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

-الاسترجاع Recall ويعرف أحيانا باسم الحساسية، يمكن تعريفه على أنه حاصل قسمة عدد الحالات المستردة على عدد الحالات ذات الصلة

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

True Positive: هو عدد الحالات ذات الصلة والتي حددها النموذج بشكل صحيح على أنها ذات صلة.

False Positive: هو عدد الحالات غير ذات الصلة ولكن النموذج حددها بشكل غير صحيح على أنها ذات صلة.

False Negative: هو عدد الحالات ذات الصلة والتي حددها النموذج بشكل غير صحيح على أنها غير ذات صلة.

يتم عادة قياس الدقة والاسترجاع معاً ويمكن الاستعاضة عنهم بـ(قياس F Measure – F)

4-5-4- مقياس F (F Measure):

في التحليل الإحصائي للتصنيف الثنائي، تعتبر الدرجة F أو (مقياس F) مقياساً لدقة الاختبار. يأخذ في الاعتبار دقة الاختبار واسترجاعه لحساب درجاته. قياس F هو المتوسط التوافقي للدقة والاسترجاع.

$$F = 2/\left(\frac{1}{\text{recall}}\right) + \left(\frac{1}{\text{precision}}\right)$$

يصل مقياس F إلى أفضل قيمته عند 1 (دقة واسترجاع مثاليين) وأسوأ قيمة عند 0 .[18]

4-5-5- مقاييس الانحدار Regression Metrics:

في التعلم الخاضع للإشراف، قد يتعلم النموذج التنبؤ بقيم رقمية. مثلاً التنبؤ بسعر السهم في الأيام القادمة، بالنظر إلى التاريخ السابق للشركة والسوق، يمكن التعامل مع هذه المهمة على أنها مهمة انحدار. يعتبر RMSE وخطأ الكميات هما مقياس التقييم الرئيسي للانحدار.

يتم تعريف RMSE على أنه الجذر التربيعي لمتوسط المسافة التربيعية بين الدرجة الفعلية والنتيجة المتوقعة. يقيس RMSE الانحراف المعياري للتنبؤات عن الحقيقة، حيث يعد RMSE إحدى طرق قياس أداء المصنف. يوجد طرق اخرى كمعدل الخطأ (أو عدد سوء التصنيف). لا يوصى باستخدام RMSE كوسيلة وحيدة لفهم مدى جودة أداء المصنف. عادةً ما يعطينا RMSE مدى بعد نموذجنا عن إعطاء الإجابة الصحيحة. [18]

كميات الأخطاء: تأتي RMSE مع عيب، لأنها تأخذ متوسط جميع نقاط البيانات. بافتراض أن مجموعة البيانات تتكون من مجموعة خارجية، فسيكون لها تأثير كبير على متوسط القيمة. يمكن تقليل تأثير القيم المتطرفة الكبيرة أثناء التقييم باستخدام مقياس قوي يسمى كميات الأخطاء. وهي تعتبر النسبة المئوية للخطأ المطلق الوسيط. [18]

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

هنا، تشير yi إلى الدرجة الحقيقية لنقطة البيانات i، وتشير y^i إلى القيمة المتوقعة. إحدى الطرق البديهية لفهم هذه الصيغة هي أنها المسافة الإقليدية بين متجه الدرجات الحقيقية ومتجه الدرجات المتوقعة، بمتوسط n، حيث n هو عدد نقاط البيانات. [20]

4-5-6- التحقق المتقاطع k-fold (k-Fold Cross-validation):

نود أن يكون لدينا طريقة يمكنها زيادة البيانات المستخدمة للتحقق دون تقليل بيانات مجموعة التدريب. يمكن تحقيق ذلك باستخدام التحقق المتقاطع k-fold ، حيث يتم تقسيم البيانات إلى k قسم، يتعين على المرء تنفيذ طريقة الانتظار بشكل متكرر ، حيث يتم استخدام واحدة من الأقسام k كبيانات التحقق ويتم استخدام البيانات المتبقية كمجموعة تدريب. يتم الحصول على تقريب الخطأ عن طريق حساب المتوسط عبر تجاربk . [18]

4-5-7- التحقق المتقاطع k-fold المقسم إلى طبقات(Stratified k-Fold Cross-validation):

هناك بعض الحالات التي قد تنحرف فيها مجموعة البيانات نحو قيمة معينة. على سبيل المثال، في مجموعة بيانات لمرض نادر، قد يكون هناك عدد كبير من التوقعات السلبية أكثر من التوقعات الإيجابية. في مثل هذه السيناريوهات، لا يعمل التحقق المتقاطع k-fold بشكل جيد، ويتم استخدام شكل مختلف من التحقق من صحة k-fold ، حيث يوجد تمثيل متساوٍ لكل نوع من تصنيفات الفئة، ونتأكد أيضًا من أن كل قسم له نفس النسبة المئوية للتمثيل . [18]

4-5-8- متوسط الخطأ التربيعيMean Squared Error

متوسط الخطأ التربيعي هو الفرق بين القيم المقدرة وما تحصل عليه نتيجة لذلك. تستند القيمة المتوقعة إلى معادلة ما وتخبرنا بما تتوقعه كمتوسط ولكن النتيجة التي تحصل عليها قد تختلف عن هذا التوقع وهو خطأ بسيط عن القيمة المقدرة. هذا الاختلاف يسمى MSE، يحدد هذا مدى جودة التقدير بناءً على المعادلة الخاصة بك. [20]

$$\mathrm{MSE}\left(y_{true}, y_{pred}\right) = \frac{1}{n_{samples}} \sum \left(y_{true} - y_{pred}\right)^2$$

4-5-9- متوسط الخطأ المطلق MAE (Mean Absolute Error):

هو قياس الفرق بين المتغيرين المستمرين. هو متوسط المسافة العمودية بين كل قيمة فعلية والخط الذي يطابق البيانات بشكل أفضل. MAE هو أيضًا متوسط المسافة الأفقية بين كل نقطة بيانات وأفضل خط مطابقة. [20]

$$\mathrm{MAE} = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_j|$$

4-6- طرق لزيادة دقة خوارزميات تعليم الآلة (النماذج):

تمر دورة تطوير النموذج بمراحل مختلفة، بدءاً من جمع البيانات إلى بناء النموذج، ولكن قبل استكشاف البيانات لفهم العلاقات (في المتغيرات)، يوصى دائمًا بإجراء إنشاء الفرضيات.

من المهم قضاء وقت في التفكير في المشكلة المعينة واكتساب المعرفة بالمجال. لذا، تساعد هذه الممارسة عادةً في بناء ميزات أفضل لاحقاً، والتي لا تكون واضحة بواسطة البيانات المتاحة في مجموعة البيانات. هذه خطوة حاسمة تعمل عادة على تحسين دقة النموذج. [21]

في هذه المرحلة، يتم تطبيق التفكير المنظم على المشكلة، أي عملية تفكير تأخذ في الاعتبار جميع الجوانب المحتملة لمشكلة معينة، ولكن هناك خطوات لزيادة دقة النموذج هي:

4-6-1- زيادة حجم البيانات:

الحصول على المزيد من البيانات فكرة جيدة دائمًا. فهو يسمح "للبيانات بالحديث عن نفسها"، بدلاً من الاعتماد على الافتراضات والعلاقات الضعيفة. يؤدي وجود المزيد من البيانات إلى نماذج أفضل ودقيقة.

4-6-2- معالجة القيم المفقودة والشاذة:

غالبًا ما يقلل الوجود غير المرغوب فيه للقيم المفقودة والمتقطعة في بيانات التدريب من دقة النموذج أو يؤدي إلى نموذج متحيز، ويؤدي إلى تنبؤات غير دقيقة، لذلك من المهم التعامل مع القيم الناقصة والمفقودة جيدًا. [21]

توجد طرق مختلفة للتعامل مع القيم الناقصة والناشئة:

القيم المفقودة: في حالة المتغيرات المستمرة، يمكن افتراض القيم المفقودة بحساب المتوسط. بالنسبة للمتغيرات الفئوية، يمكن التعامل مع المتغيرات كفئة منفصلة. حيث يمكن أيضًا بناء نموذج للتنبؤ بالقيم المفقودة. يوفر افتراض KNN خيارًا جيداً للتعامل مع القيم المفقودة.

القيم الشاذة: يمكن حذف القيم الشاذة وإجراء التحويل والتجميع والعرض (مثل القيم المفقودة) أو يمكن أيضًا معالجة القيم الخارجية بشكل منفصل. [21]

4-6-3- هندسة الخصائص:

تساعد هذه الخطوة على استخراج المزيد من المعلومات من البيانات الموجودة. يتم استخراج معلومات جديدة من حيث الميزات الجديدة. قد يكون لهذه الميزات قدرة أعلى على شرح التباين في بيانات التدريب. وهذا بالتالي يعطي دقة نموذجية محسنة.

تتأثر هندسة الميزات بشكل كبير بتوليد الفرضيات. تؤدي الفرضية الجيدة إلى ميزات جيدة. يمكن تقسيم عملية هندسة الميزات إلى خطوتين: [21]

- تحويل الميزة: هناك العديد من السيناريوهات التي تتطلب تحويل الميزة:

أ- تغيير مقياس متغير من المقياس الأصلي إلى مقياس بين صفر وواحد. على سبيل المثال: إذا كانت مجموعة البيانات تحتوي على متغير أول بالمتر، وثاني في سنتيمتر، وثالث بالكيلو متر، في مثل هذه الحالة، قبل تطبيق أي خوارزمية، يجب تسوية هذه المتغيرات بنفس المقياس.

ب- تعمل بعض الخوارزميات بشكل جيد مع البيانات الموزعة بشكل طبيعي. لذلك، يجب إزالة انحراف المتغير (المتغيرات). هناك طرق مثل الجذر التربيعي أو معكوس القيم لإزالة الانحراف.

ج- في بعض الأحيان، يعمل إنشاء كتل البيانات الرقمية بشكل جيد، لأنه يتعامل مع القيم الخارجية أيضًا. يمكن جعل البيانات الرقمية منفصلة عن طريق تجميع القيم في كتل. يُعرف هذا باسم تفتيت البيانات.

- إنشاء الميزة: يُعرف اشتقاق متغير (متغيرات) جديدة من المتغيرات الحالية بإنشاء الميزة. يساعد على إطلاق العنان للعلاقة المخفية لمجموعة البيانات. لنفترض أننا نريد توقع عدد المعاملات في متجر بناءً على تواريخ المعاملات. قد لا يكون لتواريخ المعاملة هنا علاقة مباشرة مع عدد المعاملات، ولكن إذا نظرنا إلى يوم من أيام الأسبوع، فقد يكون لها ارتباط أعلى. في هذه الحالة، يتم إخفاء المعلومات الخاصة بيوم الأسبوع. نحن بحاجة إلى استخراجه لجعل النموذج أفضل. [21]

4-6-4- اختيار الميزة:

اختيار الميزة هو عملية لاكتشاف أفضل مجموعة فرعية من السمات التي تشرح بشكل أفضل علاقة المتغيرات المستقلة مع المتغير المستهدف، حيث يمكن تحديد الميزات المفيدة بناءً على مقاييس مختلفة مثل:

معرفة المجال: بناءً على تجربة المجال، نختار الميزة (الميزات) التي قد يكون لها تأثير أكبر على المتغير المستهدف. [21]

التصور: كما يوحي الاسم، فإنه يساعد على تصور العلاقة بين المتغيرات، مما يجعل عملية اختيار المتغيرات أسهل.

البارامترات الإحصائية: تأخذ في الاعتبار قيم المعلومات والمقاييس الإحصائية لتحديد الميزات الصحيحة.

PCA: يساعد في تمثيل بيانات التدريب في مساحات ذات أبعاد أقل ، ولكن لا يزال يميز العلاقات المتأصلة في البيانات، وهو نوع من تقنيات تقليل الأبعاد. هناك طرق مختلفة لتقليل أبعاد (ميزات) بيانات التدريب مثل تحليل العوامل، والتباين المنخفض، والارتباط العالي، واختيار الميزات الخلفية / الأمامية وغيرها.

4-6-5- اختبار الخوارزميات المتعددة:

يعد الوصول إلى خوارزمية التعلم الآلي الصحيحة هو الأسلوب المثالي لتحقيق دقة أعلى. [21]
يأتي هذا الحدس مع الخبرة والممارسة المستمرة. تتناسب بعض الخوارزميات بشكل أفضل مع نوع معين من مجموعات البيانات أكثر من غيرها. ومن ثم، يجب تطبيق جميع النماذج ذات الصلة والتحقق من الأداء.

4-6-6- ضبط الخوارزمية:

إن خوارزميات التعلم الآلي لها بارامترات تحدد طريقة عملها. تؤثر هذه البارمترات بشكل كبير على نتائج عملية التعلم.

الهدف من ضبط البارامترات هو العثور على القيمة المثلى لكل بارامتر لتحسين دقة النموذج. لضبط هذه البارامترات، يجب أن يكون لدينا فهم جيد لهذه المعنى وتأثيرها الفردي على النموذج، ويمكن تكرار هذه العملية مع عدد من النماذج ذات الأداء الجيد. [21]

على سبيل المثال: في Random Forest، يوجد بارامترات مختلفة مثل max_features وnumber_trees وrandom_state وغيرها. سيؤدي التحسين البديهي لقيم البارامترات هذه إلى نماذج أفضل وأكثر دقة.

4-6-7- طرق التجميع:

هو النهج الأكثر شيوعًا. تجمع هذه التقنية ببساطة بين نتيجة عدة نماذج ضعيفة وتنتج نتائج أفضل. يمكن تحقيق ذلك بعدة طرق منها: التعبئة والتعزيز.[21]

4-6-8- التحقق من الصحة المتقاطعة:

يجب أن نستخدم تقنية التحقق المتقاطع. يعد التحقق من الصحة أحد أهم المفاهيم في نمذجة البيانات. يكون ذلك من خلال ترك عينة لا تدرب عليها النموذج واختبار النموذج على هذه العينة قبل إنهاء النموذج. [21]

5- الفصل الخامس:

تضمين الكلمات Word Embedding

5-1- المقدمــة Introduction:

عند القيام ببناء نماذج تعلم الآلة مهمّتها فهم وتفسير اللغات الطبيعيّة المتدفّقة مثل لغات البشر من غير الممكن لهذه النّماذج التّعامل مع البيانات النصيّة بشكل مباشر، فهي ليست ذكيّة بما فيه الكفاية لبدء معالجة النّص في شكلها الأصليّ، نماذج تعلّم الآلة تعتمد بشكل رئيسيّ على مبادئ الإحصاء والرّياضيات والتّحسين، ولا تفهم سوى لغة الأرقام الأمر الذي يستدعي إجراء معالجة مسبقة للبيانات النصيّة وتحويلها إلى تمثيل عدديّ مقابل لها، إنّ تحويل البيانات النصيّة إلى بيانات عدديّة مقابلة لها سيمكّن حتماً خوارزميّة التعلم الآليّ من فهمها والتعامل معها؛ حيث تعمل خوارزميّات التضمين Embedding على إيجاد تمثيل عدديّ مقابل للكلمات النصيّة بشكل فعّال.

5-2- تمثيل الكلمات باستخدام ترميز :One Hot Encoding

يتمّ تمثيل الكلمة بواسطة شعاع من الأعداد الصّحيحة {0,1}، وتتطلّب هذه الطّريقة أن تكون جميع المفردات محدّدة مسبقاً قبل إجراء عمليّة التّمثيل .

تتمثّل خطوات تمثيل الكلمة باستخدام ترميز ون هوت كالتالي:

تحديد جميع المفردات الموجودة في مجموعة البيانات.dataset

إجراء ترتيب أبجديّ لهذه المفردات.

إنشاء قاموس للمفردات وإضافة جميع المفردات إليه بعد ترتيبها أبجديّاً.

إسناد رقم فريد إلى كلّ كلمة موجودة في قاموس المفردات.

تمثيل كلّ كلمة بشعاع طوله يساوي عدد المفردات، حيث تكون جميع عناصر هذا الشعاع أصفاراً ماعدا عنصر واحد يحمل القيمة واحد، وتمثل القيمة واحد في الشعاع موضع الكلمة في قاموس المفردات.

مثال :

بفرض لدينا قاموس يحتوي على المفردات التّالية:

{قطة، سيارة، طائرة، قطار، كلب }

يكون تمثيل هذه المفردات باستخدام ترميز ون هوت كما في الشكل 7:

Dictionary D= {قطة , سيارة, طائرة, قطار, كلب}

1	0	0	0	0
0	**1**	0	0	0
0	0	**1**	0	0
0	0	0	**1**	0
0	0	0	0	**1**

الشكل 7: ترميز الكلمات باستخدام One Hot Encoding

عند ترميز الكلمة "قطار" نلاحظ أنّ الرّقم 1 أخذ الموضع الثّاني في الشّعاع والقيمة 2 تمثّل موضع الكلمة في قاموس المفردات.

5-3- مساوئ تمثيل الكلمات باستخدام ترميز One Hot Encoding

تتمثّل مساوئ تّرميز ون هوت في التكلفة الحسابيّة و الزّمنية اللازمة لمعالجة الأشعة الناتجة عن هذا التّرميز، فعلى سبيل المثال في حال كان لدينا قاموس يحتوي 10000 مفردة بالتالي نحن بحاجة إلى شعاع بحجم 1×10000 لتمثيل الكلمة الواحدة، وهذا رقم كبير بالنسبة لكلمة واحدة فلو أردنا على سبيل المثال تمثيل جملة تحتوي 100 كلمة، بالتالي سنحتاج إلى مليون قيمة من أجل إجراء تمثيل للجملة، من المساوئ الأخرى أيضاً أنّ الشعاع المعبّر عن الكلمة ليس له علاقة بمعناها، بمعنًى آخر إنّ كلمتين متشابهتين بالمعنى قد يتمّ تمثيلهم بأشعة بعيدة عن بعضهما، وبالنّظر إلى المثال التالي يتّضح الأمر:

بفرض لدينا مجموعة من المفردات تمّ ترتيبها أبجديّاً بغرض إيجاد تمثيل لها باستخدام ترميز ون هوت، سنأخذ من مجموعة المفردات عيّنة منها وهي كالتالي:

امرأة	ملكة	برتقال	رجل	ملك	تفاحة
(9853)	(7157)	(6257)	(5391)	(4914)	(456)

نلاحظ أنّ الكلمتين "ملك" و "ملكة" متشابهتان في المعنى ولكنّ أشعتهما بعيدة عن بعضهما جدّاً، ونلاحظ أيضاً أنّ الفرق بالأرقام بين الكلمتين "برتقال" و "ملكة" أقلّ من الفرق بين الكلمتين "برتقال" و "تفاحة" أي أنّ التشابه في المعنى بين الكلمتين "برتقال" و "ملكة" أكبر بكثير من الكلمتين "برتقال" و "تفاحة" وهذا خاطئ حتماً.

المشكلتان السّابقتان دفعت الباحثين إلى إيجاد أساليب أخرى لترميز الكلمات بشكل أكثر فعاليّة من الطريقة السّابقة، بحيث تقلّل من التكلفة الحسابيّة وأيضاً تمثّل الكلمات اعتماداً على سياقها؛ أي يتمّ تمثيل كلّ كلمتين متشابهتين بالمعنى بشعاعين قريبين من بعضهما (تكون المسافة الإقليديّة بينهما صغيرة) وبالمقابل أيضا الكلمات البعيدة تكون المسافة بينها بعيدة، وهذا ما يدعى بأسلوب التّضمين word embedding.

5-4- تضمين الكلماتWord Embedding

في هذه الطّريقة يتمّ أوّلاً تجميع الكلمات التي نرغب في إيجاد ترميز لها، ومن ثمّ اختيار بُعد الشعاع للكلمة، هنا البُعد يرمز إلى عدد الميّزاتfeatures ، هذه الميّزات ستصف الكلمة (سنوضح هذا لاحقاً)، في هذه الحالة نستطيع التحكّم ببُعد شعاع الكلمة (عدد الميّزات) ونستطيع جعله أقلّ بكثير من العدد الإجماليّ للمفردات، قيمة كلّ ميّزة في الشعاع تكون عدداً حقيقيّاً ضمن المجال [1،-1]، وكلّما كانت القيمة قريبة من الواحد كلّما كانت الميّزة تعبر عن الكلمة والعكس في حال اقتربت من الصّفر. على سبيل المثال بإمكاننا اختيار 300 ميّزة في حال كان لدينا قاموس

يحتوي على 10000 مفردة، وهذا ما يؤدّي إلى تخفيض التكلفة الحسابيّة والزمنيّة بشكل كبير مقارنة مع ترميز ون هوت .

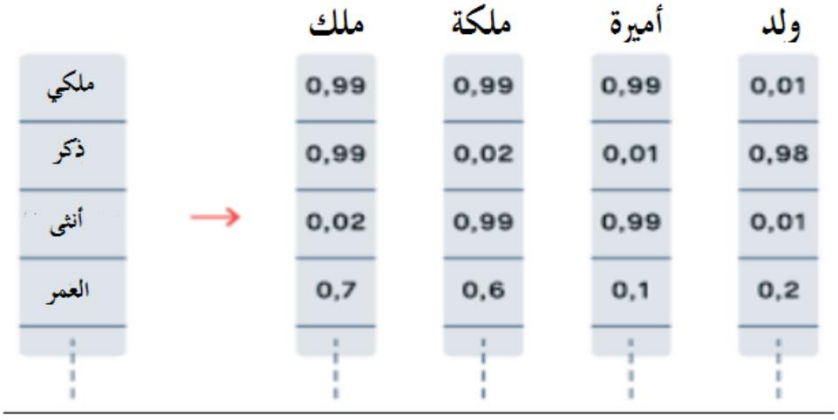
مثال توضيحيّ:

بفرض لدينا المفردات التالية

{ ولد، أميرة، ملكة، ملك }

والميّزات المستخدمة لهذه المفردات هي:(العمر، أنثى، ذكر، ملكي)

لدينا الجدول التالي يعبّر عن شعاع كلّ كلمة.



الشكل 8: جدول يوضح مميّزات كلّ كلمة.

نلاحظ من الجدول السّابق أنّ الميّزة "ذكر" تخصّ الكلمتين "ملك" و "ولد" فقط لذلك تكون قيمتها عند هاتين الكلمتين قريبة من الواحد بينما هذه الميّزة لا تخصّ الكلمتين "ملكة" و "أميرة" لذلك تكون قيمتها عند هاتين الكلمتين قريبة من الصفر. باستخدام هذه الطّريقة استطعنا تمثيل الكلمات التي تتشابه في المعنى بأشعة قريبة جدّاً من بعضها؛ أي أنّ المسافة الإقليديّة بينهما تكون صغيرة وهذا مفيد جدّاً في حالات كثيرة منها حالة تعميم الشبكة العصبونيّة.

فعلى سبيل المثال في حال أردنا إجراء تحليل مشاعر للجمل النصيّة وتصنيفها إلى ثلاثة أصناف هي إيجابيّ Positive وسلبيّ Negative وحياديّ Neutral. في حال كان لدينا الجملتين التاليتين:

بعد استخدام الجهاز لمدة شهر أدركت أنه جهاز جميل.

بعد استخدام الجهاز لمدة شهر أدركت أنه جهاز رائع.

نذكر أنّ الكلمتين "رائع" و "جميل" متشابهتان جدّاً بالمعنى ويتمّ استخدامهما في سياق متشابه، أثناء تدريب الشبكة العصبونيّة كانت الجملة الأولى موجودة في مجموعة بيانات التّدريب، Training Dataset وتمّ تدريب الشبكة العصبونيّة على أن يكون خرج هذه الجملة إيجابيّاً. الجملة الثانية موجودة في مجموعة بيانات الاختبار، Test dataset ولم تتدرّب عليه الشبكة أثناء عمليّة التّدريب، في حال أدخلنا الجملة الثانية إلى الشبكة العصبونيّة لمعرفة خرج هذه الجملة مع العلم أنّ كلمة "رائع" غير موجودة أبداً في بيانات التّدريب. وباستخدام أسلوب التّضمين سيتمّ تمثيل كلمة "رائع" بشعاع قريب جداً من "جميل" وبالتالي ستتمكن

الشبكة العصبونيّة من إعطاء خرج إيجابي، فنستنتج ميّزة كبرى لدى تضمين الكلمات Word Embedding وهي أنّها قادرة على إجراء تعميم بشكل كبير جداً للشبكة العصبونيّة.

هو بالضّبط ما يسعى نموذج الكلمة إلى متّجه Word2Vec إلى القيام به، إنّه يحاول تحديد معنى الكلمة من خلال تحليل الكلمات المجاورة لها (سياق الكلام).

توضيح :يقصد بمفهوم تعميم الشبكة العصبونية هو قدرة الشبكة العصبونية على التكيف واتخاذ القرارات الصحيحة من أجل البيانات الجديدة غير المرئية سابقًا والمستمدة من نفس التوزيع المستخدم في تدريب الشبكة العصبونية.

5-5- نماذج الكلمة إلى متّجه Word2Vec :

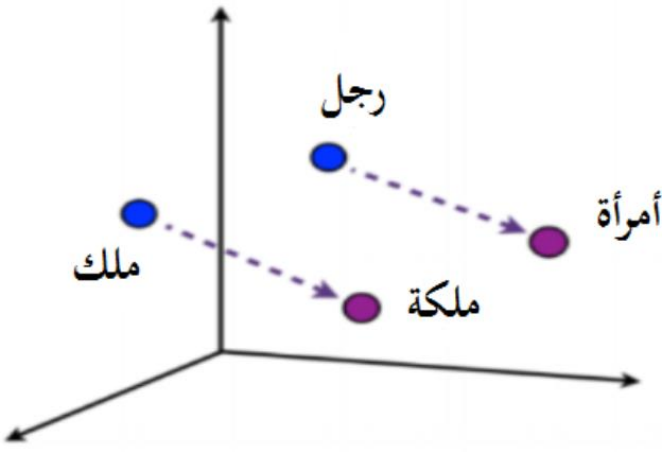
هي مجموعة من النّماذج تعمل على تمثيل الكلمة اعتماداً على سياقها في متّجه أبعاده N، تمّ تطوير هذا الأسلوب من قبل توماس ميكولوف Tomas Mikolov في العام 2013 لجعل التّدريب في الشبكات العصبونيّة مبنيّ على التّضمين، والذي يحقق أكثر كفاءة ومنذ ذلك الحين أصبح المعيار الواقعيّ لتطوير تضمين الكلمة.

نماذج الكلمة إلى متّجه تجيبنا على السؤال التالي:

العلاقة بين الكلمتين "رجل" و "أمرأة" هي بمثابة العلاقة بين الكلمتين "ملك" و؟؟ وماذا؟؟

$$e_{\text{رجل}} - e_{\text{أمرأة}} = e_{\text{ملك}} - e_{?}$$

إنّ طرح شعاع الكلمة "رجل" من شعاع الكلمة "ملك" ثمّ إضافة شعاع الكلمة "أمرة" ينتج عنه شعاع الكلمة "ملكة"، وهو أقرب ما يكون إلى شعاع الكلمة "ملك" وبالتالي هي إجابة صحيحة منطقيّاً. كما هو موضّح في الشكل 2.



الشكل 9: العلاقة بين الكلمات

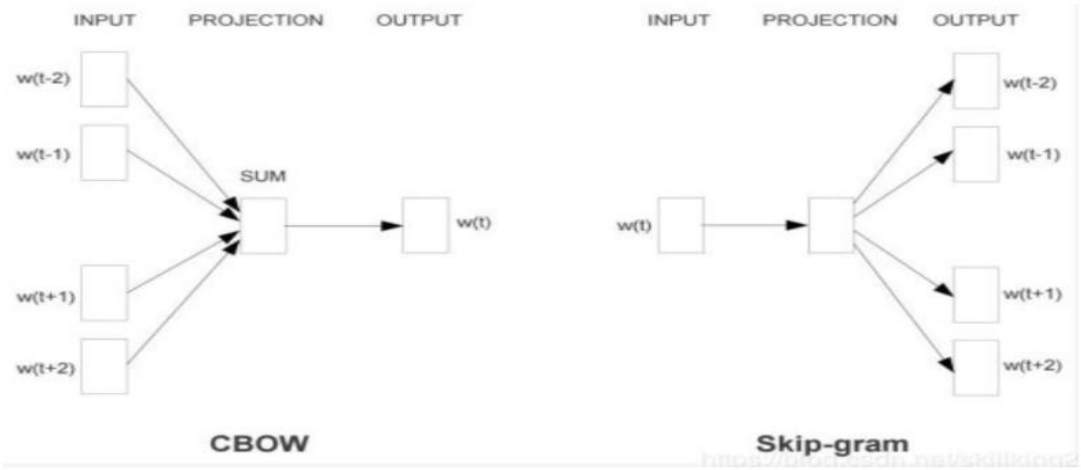
نجد أنّ هذه التمثيلات جيدة بشكل مدهش في التقاط التراكيب النحويّة والدلاليّة في اللّغة. سوف نستعرض في هذا المقال نموذجين مختلفين للتدريب يندرجان تحت فكرة نماذج الكلمة إلى متّجه لتعلّم تضمين الكلمات وهما:

_نموذج تخطّي الكلمة(Skip-Gram Model)

_نموذج حقيبة الكلمات المستمرّة (CBOW (Continuous Bag of Words

6-5- نموذجا تخطّي الكلمات Skip-Gram وحقيبة الكلمات المستمرّةCBOW

نماذج الكلمة إلى متّجه تتضمّن بشكل أساسيّ النموذجين تخطّي الكلمة وحقيبة الكلمات المستمرّة، حيث يتنبّأ نموذج تخطّي الكلمة بالسّياق استنادًا إلى كلمة الدّخل، بينما يعمل نموذج حقيبة الكلمات المستمرّة بشكل مغاير؛ حيث يتنبّأ بالكلمة استناداً على السّياق. الشكل 3 يوضّح الاختلاف بين نموذجي تخطّي الكلمة وحقيبة الكلمات المستمرّة.



الشكل 10: بنية نموذجي تخطّي الكلمة و حقيبة الكلمات المستمرّة

تتشابه عمليّة النّمذجة في نماذج "الكلمة إلى متّجه" مع فكرة المشفّر التلقائيّ autoencoder، فالهدف هو الحصول على مصفوفة أوزان الطبقة المخفية التي تحقق متجهات الكلمة word vectors وللحصول على ذلك نقوم بإنشاء مهمة مزيفة fake task لتدريب الشبكة العصبونية حيث لا نهتم بالدخل والخرج للشبكة وإنما يهمنا مصفوفة أوزان الطبقة الخفية فقط .[1]

لنلقِ نظرة على كيفيّة تدريب الشّبكة العصبونيّة. بفرض لدينا الجملة التالية:

"الثعلب البني السريع يقفز فوق الكلب الكسول"

أوّلاً نختار كلمة الدّخل من منتصف الجملة، على سبيل المثال نختار الكلمة "البني."

مع كلمة الدّخل نحدّد معامل يسمى نافذة التخطي skip_window يمثّل حجم النّافذة التي تعرّف المجال الذي يمكن اختيار كلمات الخرج منه (البعد عن كلمة الدخل)، بفرض وضعنا القيمة skip_window = 2 يكون حجم النافذة كلمتين على اليسار وأيضاً كلمتين على اليمين.

هناك معامل آخر يدعىnum_skips ، والذي يمثّل عدد الكلمات المختلفة التي نختارها بشكل عشوائيّ من النّافذة بأكملها (تكون هذه الكلمات هي المخرجات).

عندما يكون2 = skip_window ، num_skips = 2، فإننا سنحصل على مجموعتين من بيانات التّدريب بالصيغة التاليةcontext) ، target).

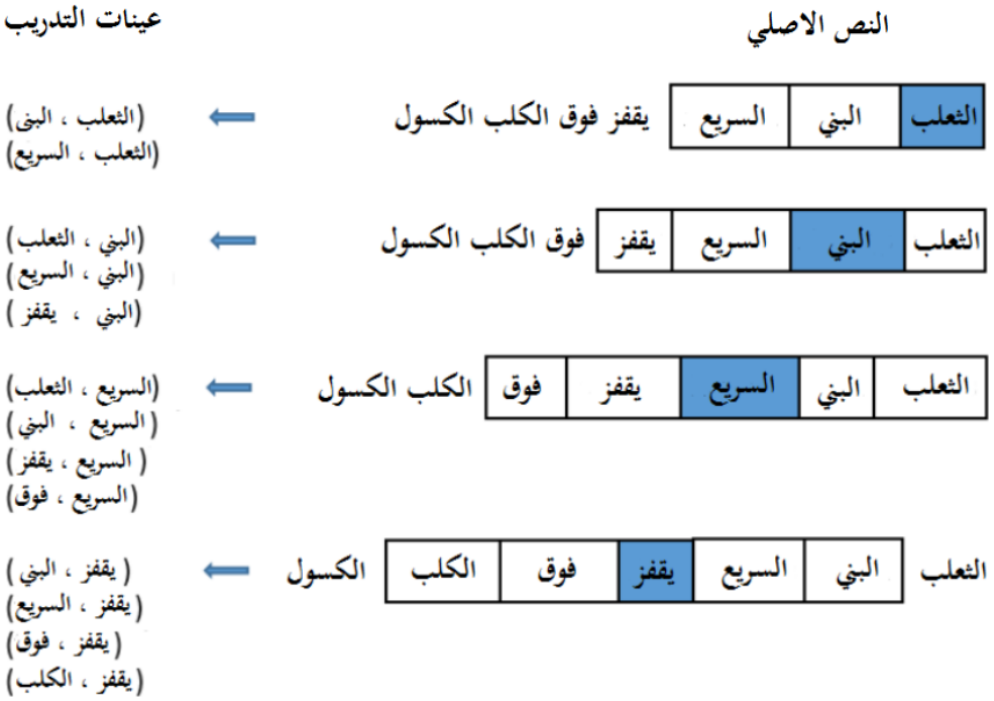
على سبيل المثال تكون المجموعتان (“بني” ، “ثعلب”) (“بني” ، “سريع”)

ستقوم الشبكة العصبونيّة بإخراج الاحتمالات استنادًا إلى بيانات التّدريب.

يُظهرالشكل 4 بعض الأمثلة من عينات التّدريب، نختار الجملة

“الثعلب البني السريع يقفز فوق الكلب الكسول”

يمثّل اللون الأزرق كلمة الدّخل، ويمثّل المربع الكلمات في نّافذة عينات التّدريب (المدخلات والمخرجات)



الشكل 11: بعض الأمثلة عن عيّنات تدريب نموذج skip-Gram

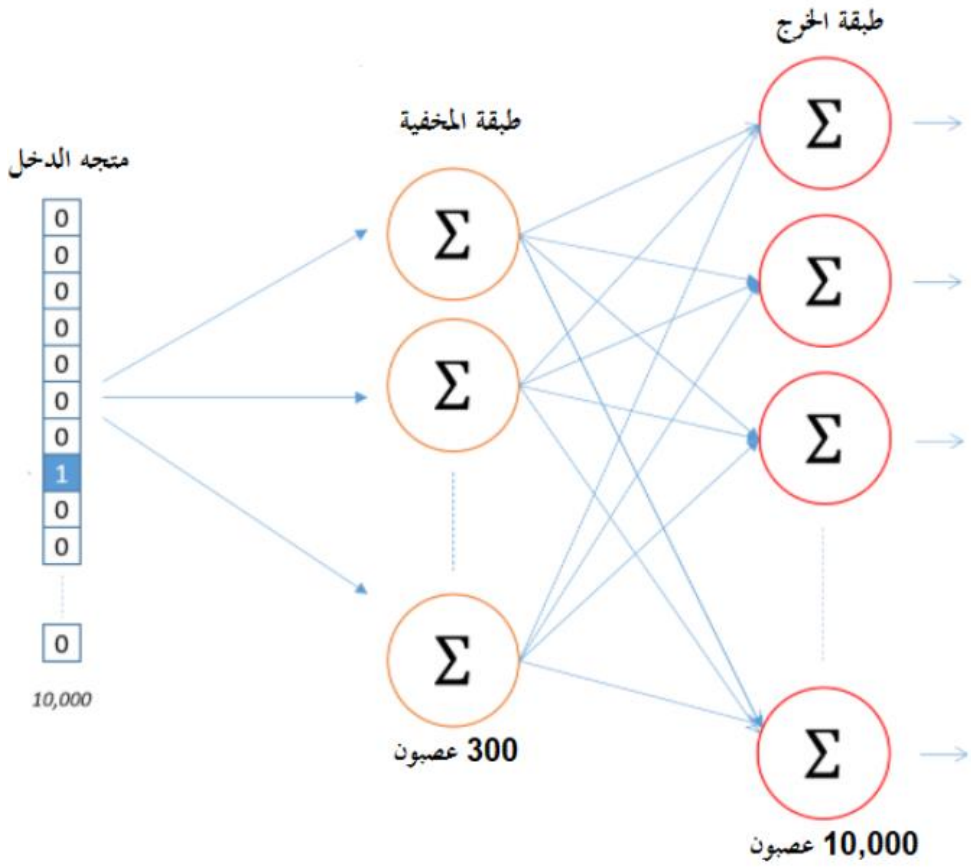
5-7- تفاصيل نموذج تخطّي الكلماتSkip-Gram

نموذج تخطّي الكلمات ما هو إلا شبكة عصبونيّة نرغب في تدريبها من أجل الحصول على مصفوفة أوزان الطبقة المخفيّة، والتي تمثل أشعة التّضمين (كما سنوضح لاحقاً)، لكن قبل تدريب هذه الشّبكة نحتاج إلى تمثيل الكلمات النصيّة بشكل عدديّ فكما ذكرنا أنّ الشّبكة العصبونيّة لا تقبل سوى المدخلات العدديّة، وأبسط طريقة لترميز الكلمات هي طريقة ون هوت لذلك سوف نختارها .

ملاحظة :في نموذج تخطّي الكلمة استخدمنا ترميز ون هوت كطريقة فقط لنصل إلى ترميز التّضمين.

بفرض تمّ استخراج 10000 كلمة فريدة من نوعها من بيانات التّدريب لتشكيل قاموس المفردات، نقوم بإجراء ترميز ون هوت على هذه الكلمات البالغ عددها 10000، وكلّ كلمة تمّ الحصول عليها هي متّجه بعشرة آلاف، بالتالي سيكون بُعد طبقة الدّخل في نموذج تخطّي الكلمة هو 10 آلاف وأيضاً طبقة الخرج بنفس البُعد.

الشكل التالي يوضح معماريّة الشبكة العصبونيّة في نموذج تخطّي الكلمة:



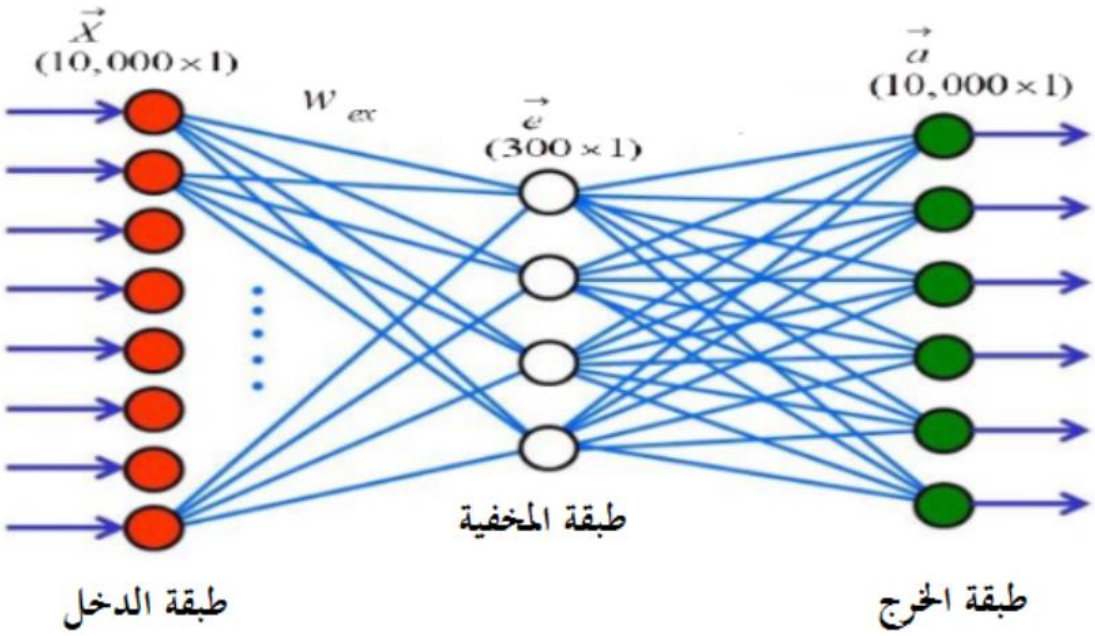
الشكل 12: الشبكة العصبونيّة لنموذج تخطي الكلمات

تستخدم الطبقة الخفيّة التابع الخطّي Linear كـ تابع تنشيط، لكنّ طبقة الإخراج تستخدم تابع سوفت ماكس softmax .

نقوم بتدريب الشبكة العصبونيّة على أساس أزواج من الكلمات، وعيّنات التدريب هي أزواج من الكلمات مثل (كلمة الإدخال، كلمة الإخراج) وكلّ من مدخلات الشبكة ومخرجاتها هي متّجهات مرمّزة باستخدام طريقة ون هوت.

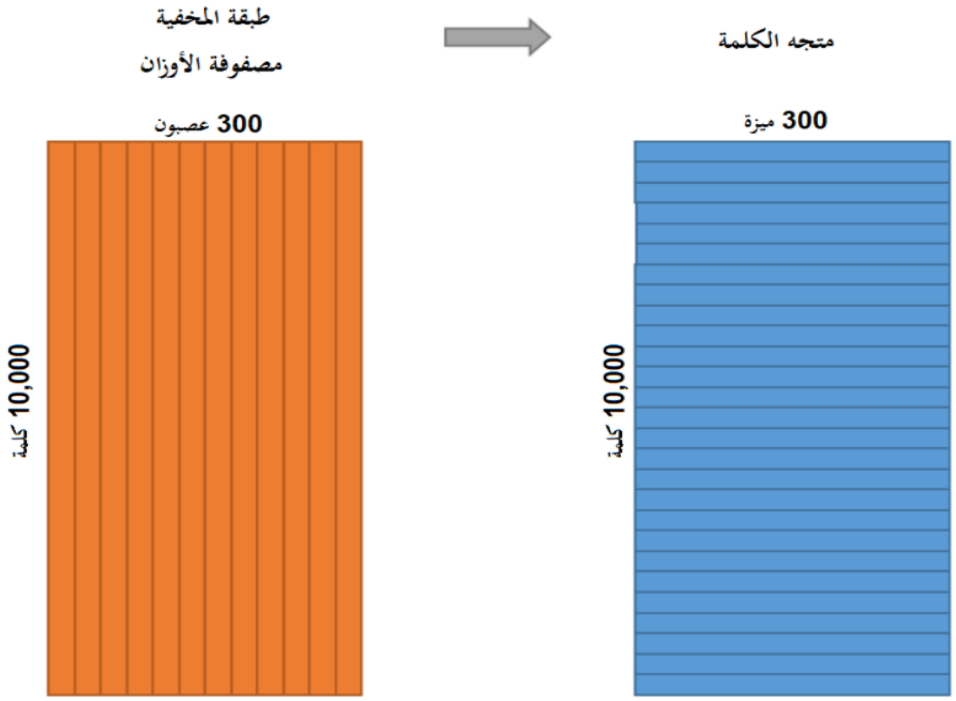
5-8- الطبقة الخفيّة لنموذج تخطّي الكلمات Skip-Gram :

بعد الحديث عن ترميز الكلمات واختيار عينات التّدريب، نلقي نظرة على الطبقة الخفيّة. إذا أردنا الآن تمثيل كلمة بها 300 ميزة، بالتالي يجب أن تحتوي الطبقة الخفيّة على 300 عصبون (أي البُعد هو 300)، تصبح مصفوفة أوزان الطبقة الخفيّة هي 10000×300 (أي 10000 صف و300 عمود) كما يوضّح الشكل 6.



الشكل 13: معماريّة الشبكة العصبونيّة لنموذج تخطّي الكلمات

بالنّظر إلى الشكل 7 تمثل الصورتان اليمنى واليسرى على التّوالي مصفوفة أوزان الطّبقة الخفيّة المصفوفة (Wax)، يمثل كلّ عمود في الصورة اليسرى متّجهًا يتكوّن من 10000 قيمة متصلة بخليّة عصبونيّة واحدة في الطبقة الخفيّة . في الشكل الموجود على اليمين، كلّ سطر في المصفوفة يمثل متّجه كلمة التّضمين.



الشكل 14: أوزان الطبقة الخفيّة لشبكة النموذج تخطّي الكلمات

لذلك هدفنا النهائيّ هو استخراج مصفوفة الأوزان للطبقة الخفيّة، وبعد ذلك يتمّ الحصول على أشعة التّضمين بإجراء عمليّة الضرب للمتّجه 1 × 10000 مع المصفوفة 10000 × 300، هذا الأمر سوف يسبّب استهلاكاً كبيراً للموارد الحاسوبيّة، ولكي يتمّ استخراج شعاع التّضمين بكفاءة دون الحاجة لعمليّة الضرب، فإنه سيتمّ تحديد صفّ الفهرس فقط بقيمة البعد 1 في المتّجه المقابل في المصفوفة كما يوضّح الشكل التالي:

$$[0 \quad 0 \quad 0 \quad 1 \quad 0] \times \begin{bmatrix} 17 & 24 & 1 \\ 23 & 5 & 7 \\ 4 & 6 & 13 \\ 10 & 12 & 19 \\ 11 & 18 & 25 \end{bmatrix} = [10 \quad 12 \quad 19]$$

الشكل 15: تحديد صفّ الفهرس فقط بقيمة البعد 1 في المتّجه المقابل في المصفوفة

كما ذكرنا أعلاه أنّ كلمة الدّخل ستكون مرمّزة بأسلوب ون هوت ومعظم قيم الشعاع تأخذ القيمة 0، موضع واحد فقط يكون 1.

يمكننا أن نرى أنّ نتيجة الضرب للمصفوفة هي في الحقيقة مؤشّر القيمة 1 في المتّجه المقابل للمصفوفة.

في المثال أعلاه البعد المقابل للقيمة 1 في المتّجه الأيسر هو 3 ثمّ تكون نتيجة الحساب هي الصفّ الثالث من المصفوفة، لذلك عند استخراج شعاع التّضمين يتمّ البحث مباشرة عن قيم الوزن المقابلة للقيمة 1 في متّجه الدّخل.

العلاقة الرّياضيّة التالية توضّح خرج الطبقة الخفيّة لكلّ كلمة دخل.

ex = Wax X

حيث أنّ:

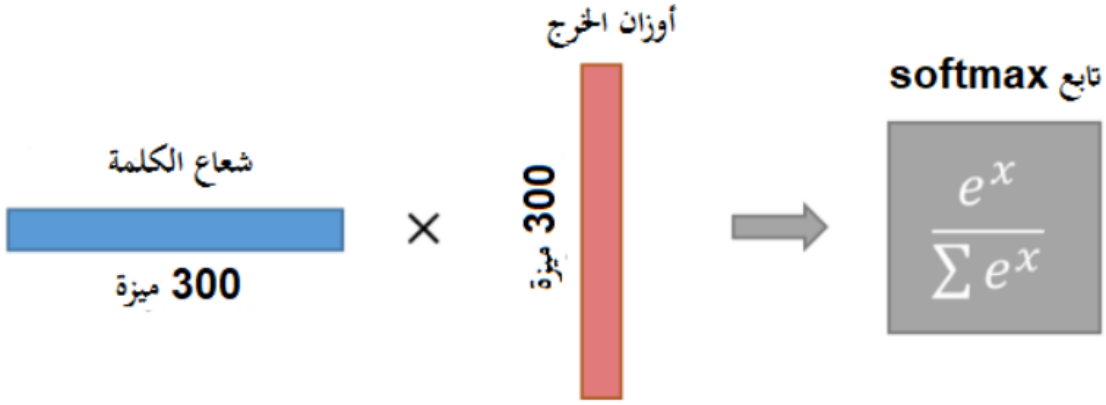
ex: شعاع الكلمة المضمّن

Wax: مصفوفة أوزان الطبقة الخفيّة

X: شعاع كلمة الدّخل المرمز بترميز ون هوت.

5-9- طبقة الخرج لنموذج تخطّي الكلمات Skip-Gram :

بعد حساب الطبقة الخفيّة للشبكة العصبونيّة ستتغيّر الكلمة من متّجه 1 × 10000 إلى متّجه 1 × 300 ثمّ يتمّ إدخالها في طبقة المخرجات، طبقة الخرج عبارة عن طبقة تصنيف بتابع تنشيط سوفت ماكس، حيث ستنتج كلّ عقدة منها قيمة احتماليّة ضمن المجال.[0,1]



الشكل 16: خرج الشبكة العصبونيّة في نموذج تخطّي الكلمة

5-10- تدريب نموذج تخطّي الكلماتSkip-Gram :

تقوم الشّبكة في البداية بتهيئة أوزانها بشكل عشوائيّ، ثمّ تقوم بضبط أوزانها بشكل متكرّر أثناء التّدريب لتقليل الخطأ الذي تحدثه عند التنبّؤ بكلمات السياق. ولتوضيح فكرة التّدريب نأخذ المثال التالي:

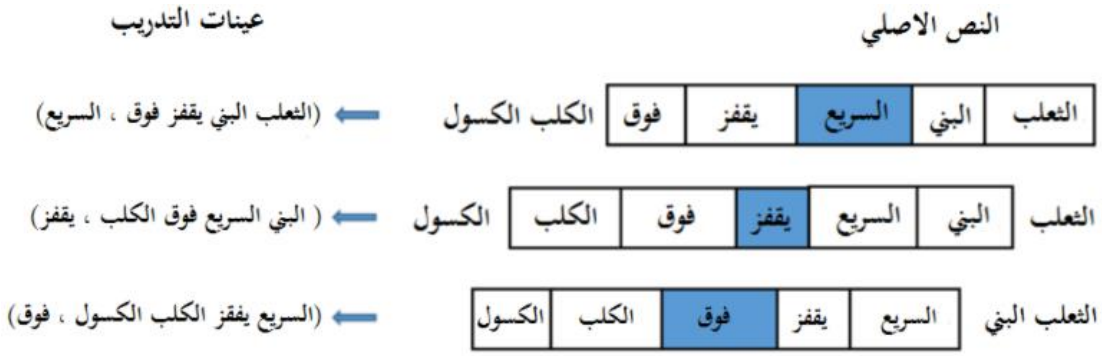
بفرض لدينا الجملة

يوم سعيد يا صديقي!

ففي حال كان لدينا العيّنة على سبيل المثال (“يوم” ، “سعيد”) وقمنا بإدخال الكلمة “سعيد” إلى الشبكة العصبونيّة، وكان خرج شعاع ون هوت يمثل الكلمة “جحيم” وهذا معناه أنّ الكلمة “جحيم” تظهر بكثرة في سياق كلمة “سعيد” وهذا غير صحيح، وبمقارنته مع الخرج الحقيقيّ يظهر خطأ كبير، بالتالي سوف يتمّ استخدام هذا الخطأ في استراتيجية التسلسل الخلفي back propagation لتعديل الأوزان من أجل تقليل احتماليّة الكلمة “جحيم” وتزويد احتماليّة للكلمة “يوم” ستكون الكلمات ذات المعنى المماثل قريبة من بعضها البعض، وهذا يحدث بمفرده لأنّ الكلمات المتشابهة ستكون محاطة أيضًا بسياق مشابه، وهذا يؤدّي إلى تعديل مشابه للأوزان. بعد الانتهاء من عمليّة التّدريب يتمّ وضع مصفوفة الأوزان الخفيّة Waxفي الواجهة الأماميّة للشّبكة العصبونيّة الارتجاعيّة Recurrent Neural Network.

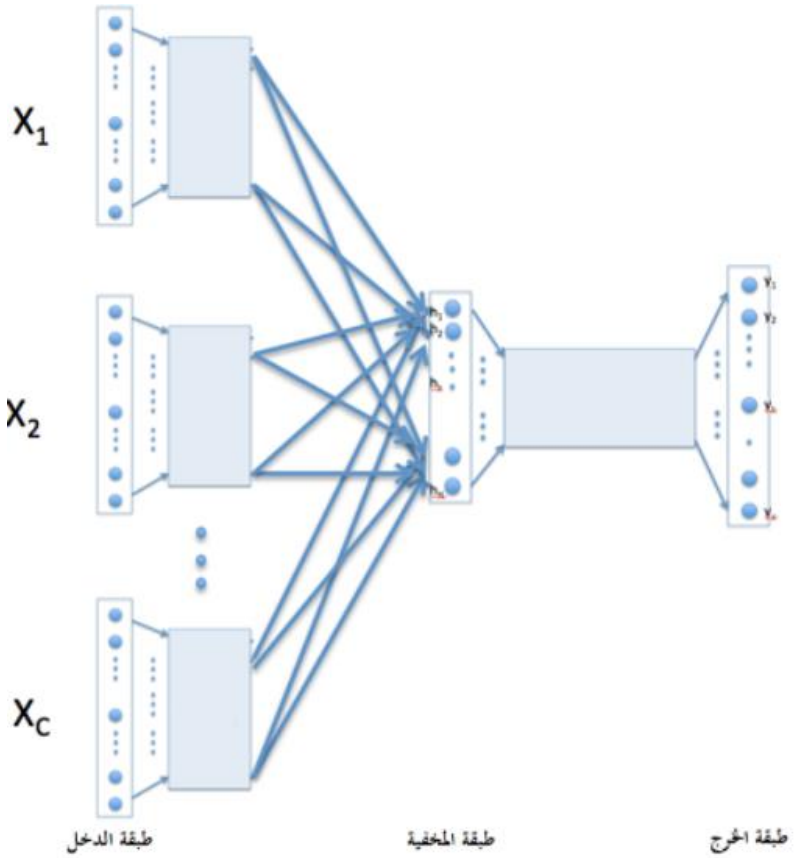
5-11- نموذج حقيبة الكلمات المستمرّة CBOW :

يعمل نموذج حقيبة الكلمات المستمرّة بشكل مغاير لنموذج تخطّي الكلمة، حيث يستخدم للتنبّؤ بالكلمة الهدف الحالية (الكلمة المركزية) بناءً على كلمات السياق (الكلمات المحيطة)؛ أي تشكيل أزواج .(context, target) كما يوضّح الشكل 17.



الشكل 17: بعض الأمثلة عن عيّنات تدريب نموذج حقيبة الكلمات المستمرّة

يوضّح الشكل التالي بنية الشبكة لنموذج حقيبة الكلمات المستمرّة.



الشكل 18:بنية الشبكة العصبونيّة في نموذج حقيبة الكلمات المستمرّة

x1، x2، Xc ....في الشكل 18 تمثّل كلمات السّياق بالنسبة لطبقة الخرج والطبقة الخفيّة وآليّة تدريب نموذج حقيبة الكلمات المستمرّة، فهي مشابهة تماماً لنموذج تخطّي الكلمة.

مقارنة سريعة بين نموذجي تخطّي الكلمات وحقيبة الكلمات المستمرّة:

نموذج تخطّي الكلمة: يعمل بشكل جيّد مع بيانات التّدريب التي حجمها صغير ويمثّل بشكل جيّد حتى الكلمات أو العبارات النّادرة.

نموذج حقيبة الكلمات المستمرّة: أسرع بعدّة مرّات من ناحية التّدريب مقارنة مع تخطّي الكلمة.

مساوئ نموذجي تخطّي الكلمات Skip-Gram وحقيبة الكلمات المستمرّةCBOW

على سبيل المثال في حال كان لدينا مجموعة بيانات مكونة من 10,000 كلمة، وأردنا إجراء تضمين لكلمات التّدريب بحيث تحتوي الطبقة الخفيّة على 300 عصبون، فسنحصل على مصفوفة أوزان بالنسبة للطبقة الخفيّة 10000 × 300 أي 3 ملايين معامل (بمعنى 3 ملايين علاقة رياضيّة)، التّصنيف في مثل هذه الشبكة العصبونيّة الضّخمة بطيء للغاية، ونحتاج إلى الكثير من بيانات التّدريب لضبط أوزان هذه الشبكة بشكل مثالي. الملايين من القيم في مصفوفات الأوزان والآلاف من عيّنات التّدريب تعني أنّ تدريب هذا النموذج سيكون صعباً جداً.

6- الفصل السادس:

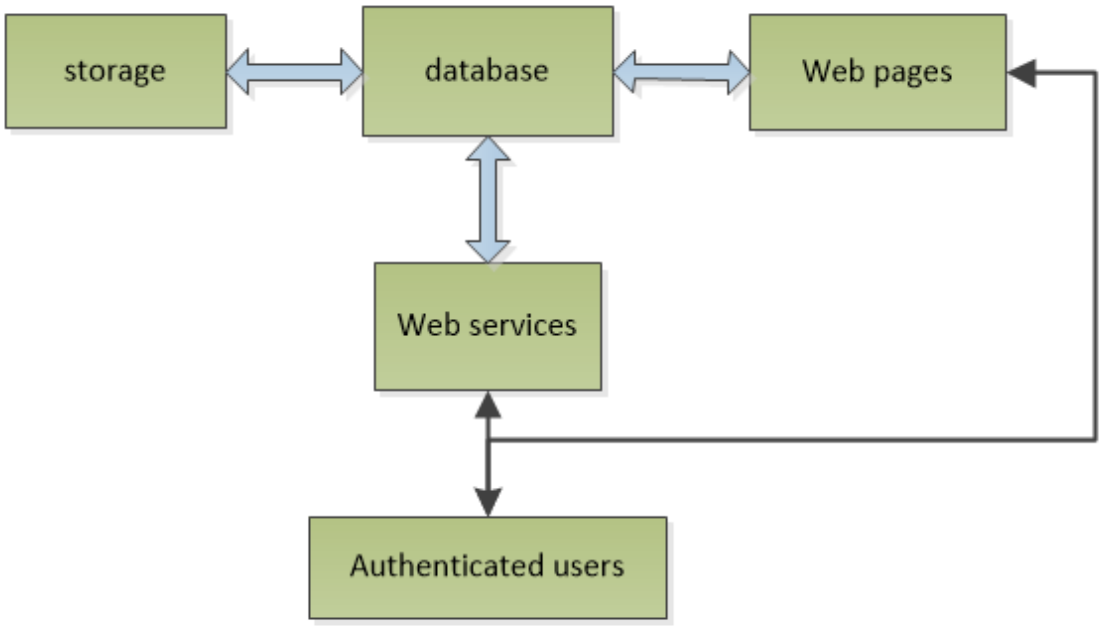
بنية المنظومة المقترحة:

6-1- مقدمة:

يتطلب تصميم الشبكة الاجتماعية المقترحة بنية تحقق كل متطلبات التصميم وتضع في الحسبان كل الأفكار التطويرية اللاحقة التي يمكن إضافتها إلى الشبكة المقترحة، حيث أن التعديلات التي يمكن أن تنفذ لاحقاً على المنظومة قد تصل إلى طرق مسدودة إذا لم يكن التصميم الأولي يحقق المطلوب، كما أن البنية المقترحة يجب أن تعتمد على التصاميم الحديثة التي تتعامل مع التقنيات الحالية مثل الحوسبة السحابية والبيانات الضخمة، لذلك نحن بحاجة إلى تحديد عدة نقاط يجب مراعاتها أثناء التصميم منها:

البنية العامة للمنظومة المقترحة، آلية المواثقة المستخدمة، تحديد مجال العمل وأنواع المستخدمين، نوع قاعدة البيانات المستخدمة، نوع المفاهيم التي سيتم اعتمادها في تصميم الشبكة.

6-2- البنية العامة للمنظومة



الشكل 19: البنية العامة للمنظومة المقترحة

يوضح الشكل 19 البنية العامة للمنظومة المقترحة، حيث يمكن تحديد الكتل الأساسية المكونة للمنظومة كما يلي:

-Database: يتم فيها تخزين بيانات المنظومة المقترحة، هنا يجب الأخذ بعين الاعتبار أن قاعدة البيانات سيتم الاتصال بها من قبل عدد كبير من المستخدمين، تم اقتراح قاعدة بيانات Oracle لتحقيق البنية المطلوبة، مع مراعاة عدم تخزين البيانات الكبيرة الحجم ضمن قاعدة البيانات، وإنما اتباع تخزين الملفات ضمن مجلدات منفصلة لمنع التحميل الزائد على قاعدة البيانات عند التعامل مع الملفات الضخمة مثل المقالات أو مقاطع الفيديو ، يمكن ملاحظة

إمكانية الوصول إلى قاعدة البيانات من خلال صفحات المنظومة web pages، كما يمكن الوصول إلى قاعدة البيانات من خلال الخدمات المقدمة من المنظومة web services، وذلك من خلال مستخدمين معتمدين authenticated.

-Storage: يتم فيه تخزين الملفات المتبادلة بين المستخدمين، يمكن أن يحوي مختلف أنواع الملفات المتبادلة بين المستخدمين أو ضمن الشبكة، مثل الصور والصوت والفيديو، هنا يجب اتباع آلية نسخ احتياطي وحماية لمنع ضياع البيانات أو احتمال حصول مشكلة في التجهيزات، لذلك يفضل اتباع تخزين هذه الملفات على Storage SAN لضمان عدم الفشل في التجهيزات.

-Web Pages: يمكن تمثيل البرامج المصممة ضمن المنظومة من خلال صفحات ويب تتعامل مع المستخدمين المعتمدين، وتعتبر هذه الصفحات هي الوسيط بين المستخدم وقاعدة البيانات، يمكن تحقيق هذه الميزة من خلال اتباع تقنية MVC (Model – View -Controller) التي تحقق بيئة مناسبة للتعامل مع قواعد البيانات بسهولة وعزل المستخدم النهائي عن قاعدة البيانات.

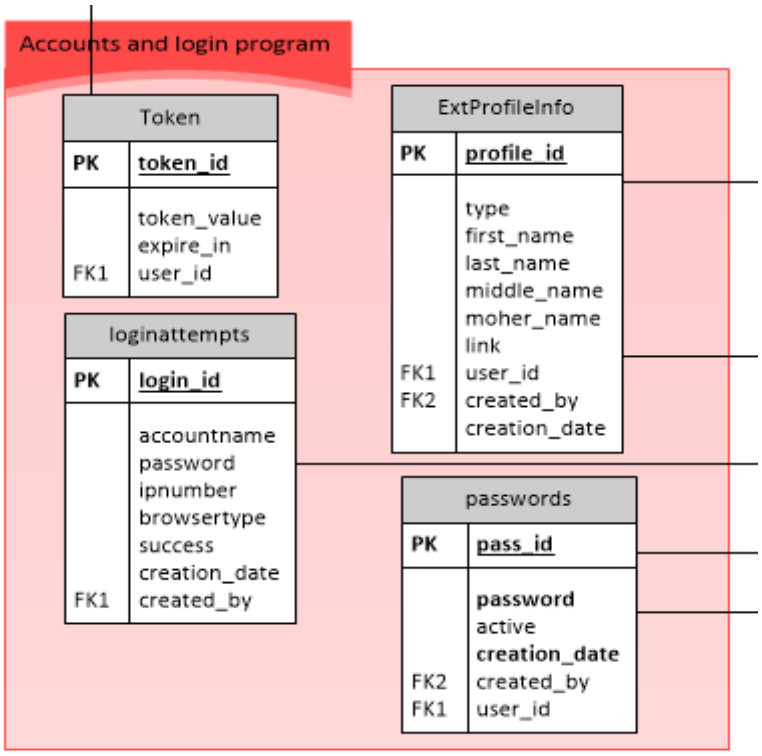
-الخدمات Web Services: قد يتم إضافة خدمات تحققها هذه المنظومة لكي تتلائم مع الشبكات الأكاديمية العالمية مثل : خدمة طلب الملف الشخصي profile التي تتضمن معلومات عن مستخدم محدد مثل الصورة الشخصية وبعض المعلومات مثل الاسم والميلاد والشهادات، كذلك خدمة الاقتباسات المرتبطة بمرجع معين لمعرفة المراجع المرتبطة بمقالة معينة، وغيرها من الخدمات التي تساعد في تسريع العمل وتبادل البيانات مع البرامج والتطبيقات الخارجية، يجب التزام مبدأ الموثوقية عند التعامل مع هذه الخدمات، بحيث لايمكن استثمار هذه الخدمات إلا من خلال مستخدمين معتمدين.

- **المستخدمين المعتمدين Authenticated Users:** يمكن التعامل مع المنظومة المقترحة من خلال مستخدمين معتمدين حسب الصلاحيات الممنوحة لهم، هنا يجب التفريق بين الصلاحيات على ثلاث مستويات، المستوى الأول هو الصلاحيات التي يمكن أن تمنح لأي مستثمر للمنظومة بصلاحيات ضيف guest ، مثل الاطلاع على المناقشات العامة أو عرض ملخصات عن المقالات أو استعراض احصائيات خاصة بمقالة معينة، أو استعراض برامج وإعلانات على مستوى عام، المستوى الثاني هو الصلاحيات أو العمليات التي يمكن أن تمنح لطالب مثل المشاركة في المناقشات وتحميل الملفات، المستوى الثالث من الصلاحيات هي الصلاحيات التي يمكن أن تمنح لمدراء التطبيقات أو مدراء الشبكة مثل إمكانية الحذف والتعديل على السجلات، مع الانتباه إلى ضرورة استخدام آلية مواثقة مضمونة تمنع الاختراق وتقلل ما أمكن من الهجمات الخارجية.

6-3- المواثقة:

يمكن استخدام OAuth كآلية مواثقة في المشروع المقترح، يعتبر OAuth اختصار لكلمة Open Authorization وهو يتيح للمستخدم امكانية مشاركة المحتوى الخاص بالشبكة المقترحة إلى مواقع أو برامج خارجيه بطريقه سريه وآمنه بدون الحاجة الى ادخال اي بيانات سرية كتسجيل الدخول او عمل حساب جديد او

غيره، حيث يمكن استخدام هذا البروتوكول من طلب الخدمات من الشبكة المصممة عن طريق التطبيقات المباشرة الخاصة بالشبكة، أو عن طريق تطبيقات وسيطة تتعامل مع الخدمات المصممة لهذا الغرض،



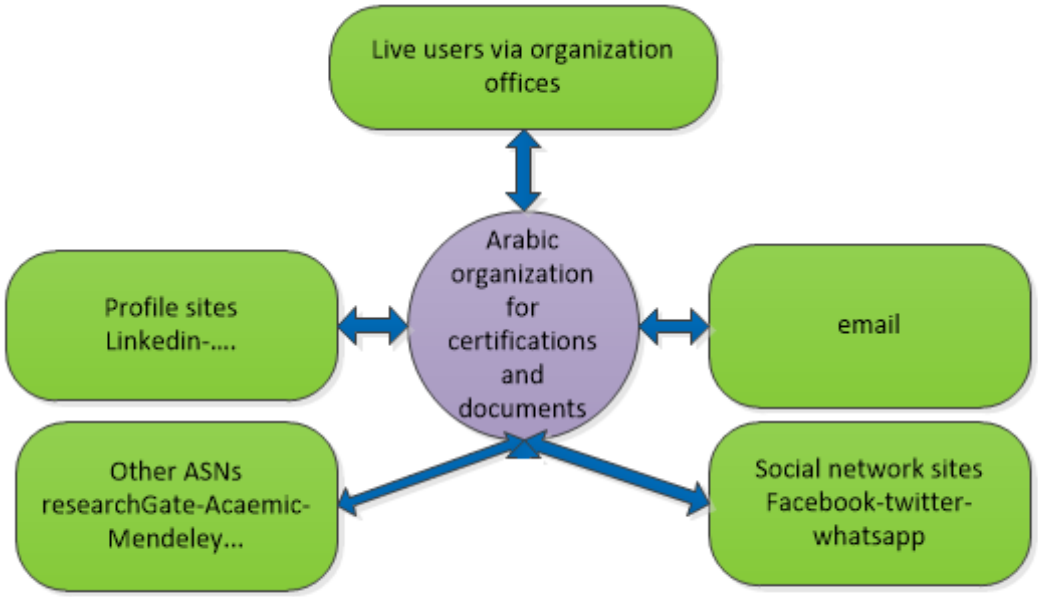
الشكل 20: بنية الجداول المخصصة للمواثقة

6-3-1- آلية عمل بروتوكول الـ OAuth؟

الكثير من المستخدمين يمتلك حسابات على المواقع الإجتماعية كموقع تويتر، لاستخدام موقع تويتر أو وضع رسالة جديدة tweet على الحساب الخاص بالمستخدم يمكن الدخول مباشرة الى موقع تويتر وكتابة رساله جديده وستظهر لمستخدمي حسابات تويتر الآخرين. [22]

الطريقه الأخرى هي استخدام واحد من البرامج الوسيطة بين المستخدم وبين تويتر كمقدم خدمه مثل Tweetdeck او Gwibber أو غيره من البرامج المعروفة.

يمكن لهذا البرنامج القيام بهذه العمليه في حالة منح المستخدم التصريح لهذا البرنامج صلاحيات معينه للقيام بمهام بالنيابه عن المستخدم نفسه، تكمن الحاجة إلى هذا النوع من المواثقة من خلال أن تصميم هذه المنظومة سوف يتبعه تصميم تطبيقات لاحقة تستخدم بعض الخدمات التي يمكن أن توفرها المنظومة.



الشكل 21: طرق التسجيل في المنظومة:

6-3-2- طرق التسجيل في المنظومة:

يوضح الشكل 21 طرق التسجيل في المنظومة، حيث يهدف استخدام نوع المواثقة oAuth إلى استخدام عدة أنواع من الحسابات في إنشاء مستخدم وتسجيل الدخول إلى الشبكة المقترحة، يظهر الشكل آلية تسجيل الدخول وإنشاء المستخدم من خلال:

1-المستخدم المباشر: يمكن أن يكون من خلال مراكز تسجيل معتمدة، حيث يتم طلب وثائق معينة مثل الشهادة العليا المكتسبة (شهادة جامعية – ماجستير – دكتوراه)، وذلك للوصول إلى مستوى معين في مستخدمي المنظومة، كما يمكن أن يتاح المجال لمستخدمين عامين (بدون متطلبات شرطية على التسجيل)، وذلك للاطلاع على ميزات الشبكة واكتساب الشهرة والموثوقية.

2-المواثقة مع مستخدمين من مواقع أخرى مثل الشبكات الأكاديمية الاجتماعية الأخرى مثل ResearchGate، Academic، Mendeley، Linkedin، مواقع التواصل الاجتماعي مثل facebook، twitter، whatsapp.

3-المواثقة من خلال حسابات الايميل مثل Gmail.

6-4- أنواع الحسابات من حيث الصلاحيات:

يوضح الشكل 22 أنواع الحسابات المتوفرة في المنظومة من حيث الصلاحيات، حيث أن الشبكة المقترحة يمكن أن يتم استخدامها من قبل عدد كبير من المستخدمين لذلك يمكن تقسيم المستخدمين حسب نوع الصلاحيات إلى:

1-visitor: يعتبر هذا النوع من المستخدمين هو الأقل صلاحية في الشبكة المقترحة، لايحتاج إلى تسجيل معلومات ضمن الشبكة، حيث يمكن أن تكون الصلاحيات مقتصرة على استعراض مراكز التسجيل، ومعلومات التواصل، وشرح مبسط عن الشبكة.

2-user: يمكن التسجيل في هذه الشبكة من خلال موقع الشبكة، ويمكن لهذا المستخدم أن يتمتع ببعض الصلاحيات مثل الاطلاع على معلومات عامة عن المقالات والمناقشات.

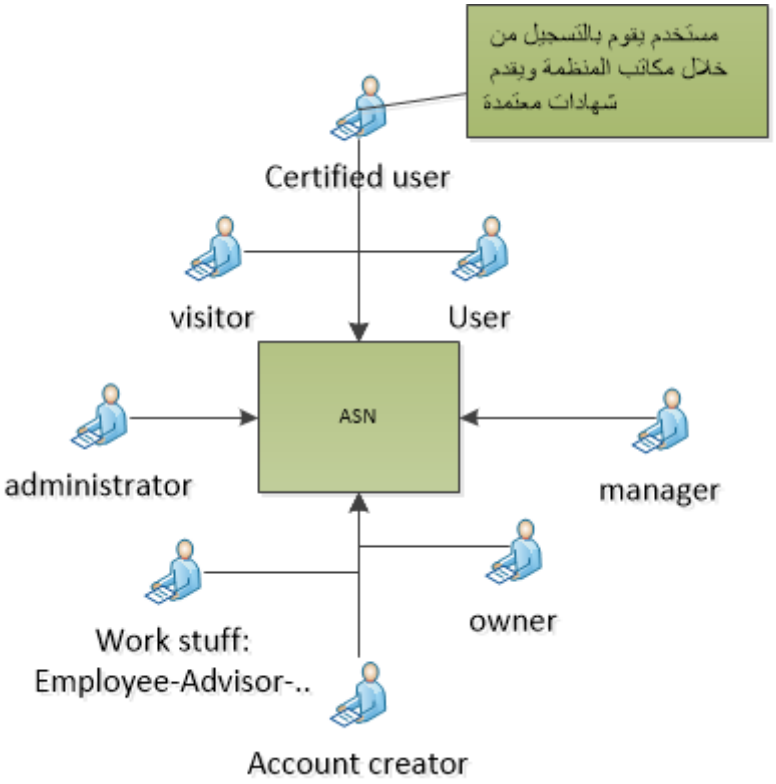
3-certified user: يمنح هذا المستوى من الصلاحيات من خلال مراكز التسجيل، حيث يتم تقديم وثائق تتعلق بمستوى الدراسة، ويمكن أن تكون الصلاحيات عبارة عن إمكانية المشاركة بالمناقشات والمؤتمرات وتنزيل المقالات والتعليق عليها.

4-work stuff: يمكن أن يكون هناك صلاحيات خاصة تتعلق بالعاملين ضمن الشبكة مثل موظفي مراكز التسجيل - مدققي المقالات وغيرها، يمكن أن تكون هذه الصلاحيات هي نفس صلاحيات certified users بالإضافة إلى صلاحيات مثل استعراض معلومات مراكز التسجيل وتدقيق البيانات الخاصة بطلبات التسجيل وغيرها.

5-account creator: هي صلاحيات خاصة بالموظفين الذين يقومون بإنشاء مستخدمين جدد، مثل إضافة مستخدمين وتعديل معلوماتهم.

6-manager: هي صلاحيات خاصة بمدراء المراكز.

7-administrator: صلاحيات مدير المنظومة.



الشكل 22: أنواع الحسابات المتوفرة في المنظومة من حيث الصلاحيات

7- الفصل السابع:

التطبيقات المقترحة ضمن المنظومة المراد تنفيذها

7-1- مقدمة:

تمت السيطرة على المجتمع العربي من خلال العبودية العلمية والفكرية والثقافية للغرب، هذه العبودية تظهر في العديد من الأشكال، حيث أن أي نوع من أنواع التطوير العلمي تمر من خلال المواقع الغربية، سواء من خلال المحادثات النصية أو المؤتمرات الصوتية والفيديوية أو تبادل الملفات، كما أن مواقع تبادل الأفكار والتساؤلات هي مواقع غربية، عدا عن ذلك، يجب أن لا نغفل أن العديد من جامعات الوطن العربي تعتبر النشر في مجلة غربية شرط أساسي للقبول في الماجستير والدكتوراه، هذه العبودية العلمية والفكرية، نشأت بوجود منظومة استخباراتية عالمية غربية تهدف بشكل أساسي إلى السيطرة على الكفاءات ومعرفة الخبرات وجذب العلماء وبالتالي السيطرة على القطاع العلمي والفكري بكافة مجالاته.

من خلال ماسبق ظهرت الحاجة إلى إنشاء شبكة أكاديمية متكاملة مع وجود مهام وبرامج إضافية لهذه الشبكة بحيث تؤمن البديل من استخدام كافة الأدوات الموجودة حالياً، بحيث يمنح استقلالية في المجال العلمي يدفع نحو التطور العلمي للمجتمع العربي، بعيداً عن كافة المواقع التي تسيطر حالياً على المسيرة العلمية، ومن خلال هذا البحث تم اقتراح عدة تطبيقات تعتبر أساسية في الشبكة الاجتماعية الاكاديمية المقترحة وهي:

1. برنامج المحادثة.
2. برنامج المؤتمرات.
3. برنامج الاستفسارات.
4. تطبيق مراكز التسجيل.
5. تطبيق الصلاحيات.
6. تطبيق المقالات.
7. تطبيق التحكيم.

7-2- برنامج المحادثة Chatting Application:

تطبيق يهدف إلى تسجيل المحادثات المتبادلة بين الأشخاص المسجلين ضمن المنظومة، والابتعاد قدر الإمكان عن برامج المحادثة الشائعة مثل whatsapp، لكن يجب مراعاة عدة نقاط في التطبيق المقترح:

-يحقق كافة المزايا التي تحققها برامج المحادثة التقليدية (ارسال الملفات والصور- ارسال ملفات صوتية)

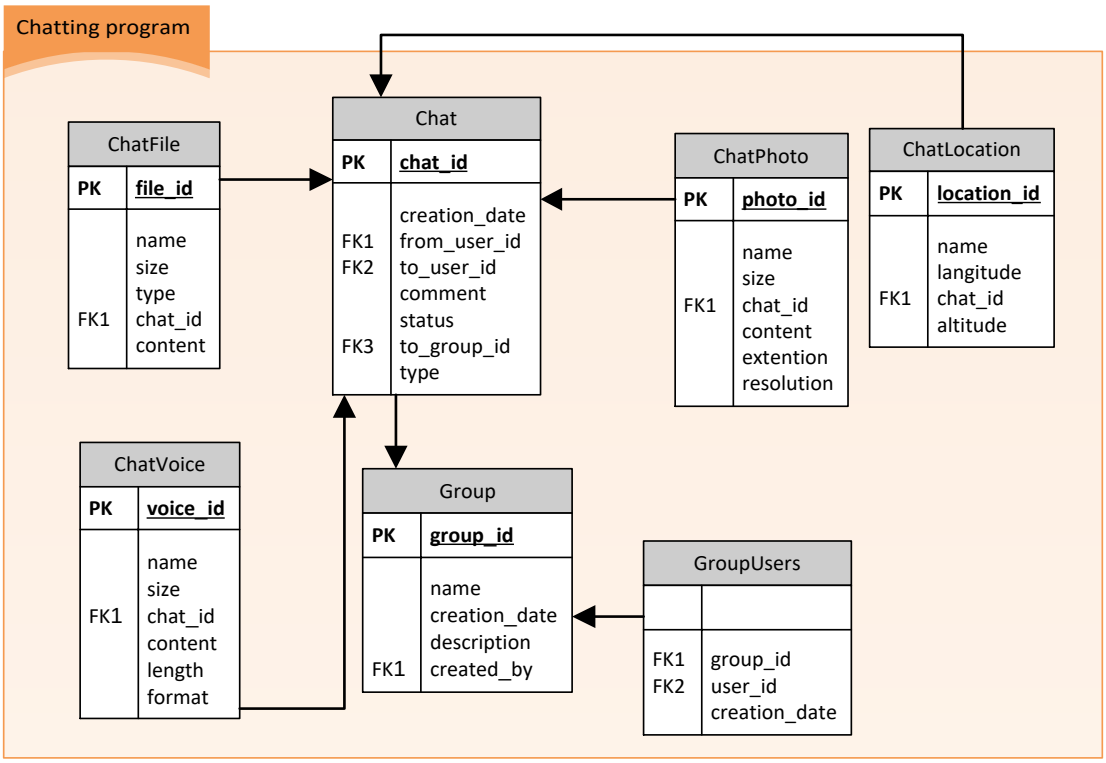
-توفير الحماية الكافية للمحادثات من خلال تشفير المحادثات بين الطرفين

-الاحتفاظ بسجلات المحادثة حتى فترة زمنية محددة حيث يمكن تحديد هذه الفترة الزمنية اعتمادا على حجوم وسائط التخزين.

-امكانية احداث مجموعات محادثة نصية بوجود مدير للمحادثة أو عدة مدراء، ويمكن اعتبار هذه الخطوة هي بديل للصفوف والمحادثات والملفات المتبادلة بين المدرس والطالب لقاعات التدريب بحيث يكون مدير المجموعة هو مدرس المحاضرة،

7-2-1- بنية قاعدة البيانات:

يوضح الشكل 23 بنية قاعدة البيانات لبرنامج المحادثة، حيث تحوي قاعدة البيانات Database الخاصة بالبرنامج على الجداول التالية:



الشكل 23: بنية قاعدة البيانات لبرنامج المحادثة

-**الجدول Chat:**جدول يحتوي معلومات المحادثة، هنا كل سطر في هذا الجدول يعبر عن سطر في المحادثة، يخزن هذا الجدول المعلومات التالية:

1-chat_id: معرف أساسي.

2-creation_date: توقيت وتاريخ المحادثة.

3-from_user_id: معرف المستخدم منشيء المحادثة.

4-to_user_id: معرف المستخدم مستقبل المحادثة.

5-comment: نص الرسالة.

6-status: حالة الرسالة(تم الارسال ولم يتم الاستلام-تم الاستلام ولم يتم القراءة-تمت القراءة-محذوفة)

7-to_group_id:معرف المجموعة التي المراد توجيه المحادثة اليها، يمكن لهذا الحقل أن يحوي قيمة تعبر عن معرف المجموعة، بحيث تظهر المحادثة لكل المجموعة وليس فقط لمستخدم محدد.

8-type:نوع سطر المحادثة، حيث يمكن أن يكون (نصي- ملف -صوت - صورة- موقع).

-**الجدول Group**: جدول يحتوي معلومات المجموعات التي يتم انشاؤها، يخزن هذا الجدول المعلومات التالية:

1-Group_id: معرف المجموعة.

2-Name: اسم المجموعة.

3-creation_date: تاريخ انشاء المجموعة.

4-created_by: مدير المجموعة.

5-description: توصيف لعمل المجموعة.

-**الجدول GroupUsers**: هو جدول يربط المجموعة مع أعضاءها، لايمكن تعديله إلا من فبل مدير المجموعة.

-**الجدول ChatFile**: جدول يحوي معلومات الملفات المتبادلة في المحادثة، يخزن هذا الجدول المعلومات التالية: (file_id معرف الملف-Name اسم الملف-Size حجم الملف-content محتويات الملف كبايتات-chat_id معرف المحادثة التي يتبع لها هذا الملف، type نوع الملف).

-**الجدول ChatPhoto**: جدول يحوي معلومات الصور المتبادلة في المحادثة، يخزن هذا الجدول المعلومات التالية: (Photo_id معرف الصورة-Name اسم ملف الصورة-Size حجم الصورة-content محتويات الصورة كبايتات-chat_id معرف المحادثة التي تتبع لها هذه الصورة-extention امتداد أو صيغة الصورة – resolution دقة الصورة).

-**الجدول ChatVoice**: جدول يحوي معلومات الملفات الصوتية المتبادلة في المحادثة، يخزن هذا الجدول المعلومات التالية: (المعرف voice_id، اسم الملف الصوتي Name، الحجم Size، معرف المحادثة التي تتبع لها الملف الصوتي chat_id، المحتويات كبايتات content،طول الملف بالثواني length،صيغة الملف format)

-**الجدول ChatLocation**: جدول يحوي معلومات الموقع المتبادلة في المحادثة، يخزن هذا الجدول المعلومات التالية: (المعرف location_id، اسم الموقع Name، خط الطول والعرض Longitude and Altitude، معرف المحادثة التي تتبع لها الملف الصوتي chat_id).

-يمكن إضافة جداول أخرى حسب الحاجة في حال وجود أنواع أخرى من الكائنات التي يمكن ارسالها في المحادثة مثل (ملف فيديو Video file -ملف سيرة ذاتية CV file- ملف معلومات اتصال contact file).

7-3- برنامج المؤتمرات Conference Application:

عادة يتم استخدام برامج المحادثات الشائعة في حال الحاجة إلى اجراء مقابلة صوتية أو فيديوية مثل google meet،zoom وغيرها، هذه البرامج تعتبر ذات أداء عالي ولكن تفتقد إلى الخصوصية، فإن الحاجة ماسة لإنشاء اداة أو تطبيق يعمل عمل هذه التطبيقات، حيث يمكننا هذا التطبيق من إلقاء محاضرات فيديوية عبر الشبكة يكون فيها المدرس هو مصدر الصورة والصوت، والحضور هم الطلاب، ومن الميزات التي يمكن ادراجها ضمن برنامج المؤتمرات:

1-إمكانية تشارك مؤتمر فيديوي يوجد فيه مصدر واحد للصورة والصوت وعدة مشاهدين، وهذه الميزة هامة جداً لإلقاء المحاضرات الجامعية على الطلاب.

2-إمكانية تشارك مؤتمر فيديوي فيه كل مشارك هو مصدر لصورة وصوت، بحيث يمكن استخدام هذه الميزة في المؤتمرات التي يكون فيها كافة الحضور فاعلين في المؤتمر، هذه الميزة تتأثر بشكل كبير بجودة خدمة الشبكة عند الحضور.

3-إمكانية مشاركة عدة مصادر للصورة (مشاركة شاشة حاسب – مشاركة شاشة موبايل – مشاركة واجهة تطبيق مفتوح – مشاركة كاميرا) وهذا يساعد على استخدام كافة وسائل الإيضاح التي يمكن أن يستخدمها المشارك في هذا المؤتمر.

4-إمكانية مشاركة عدة مصادر للصوت (مشاركة ملف صوتي – مشاركة أداة صوت(Mic) للحاسب- مشاركة أداة صوت(Mic) للموبايل).

5- مشاركة لوحة ملاحظات يمكن اعتبارها بديل عن اللوح الذي يتم استخدامه في المؤتمرات الحقيقية بهدف التوضيح.

6- وجود شاشة محادثة نصية مشتركة

7-إمكانية تسجيل المؤتمر بالكامل، أو تسجيل مصدر الصوت والصورة.

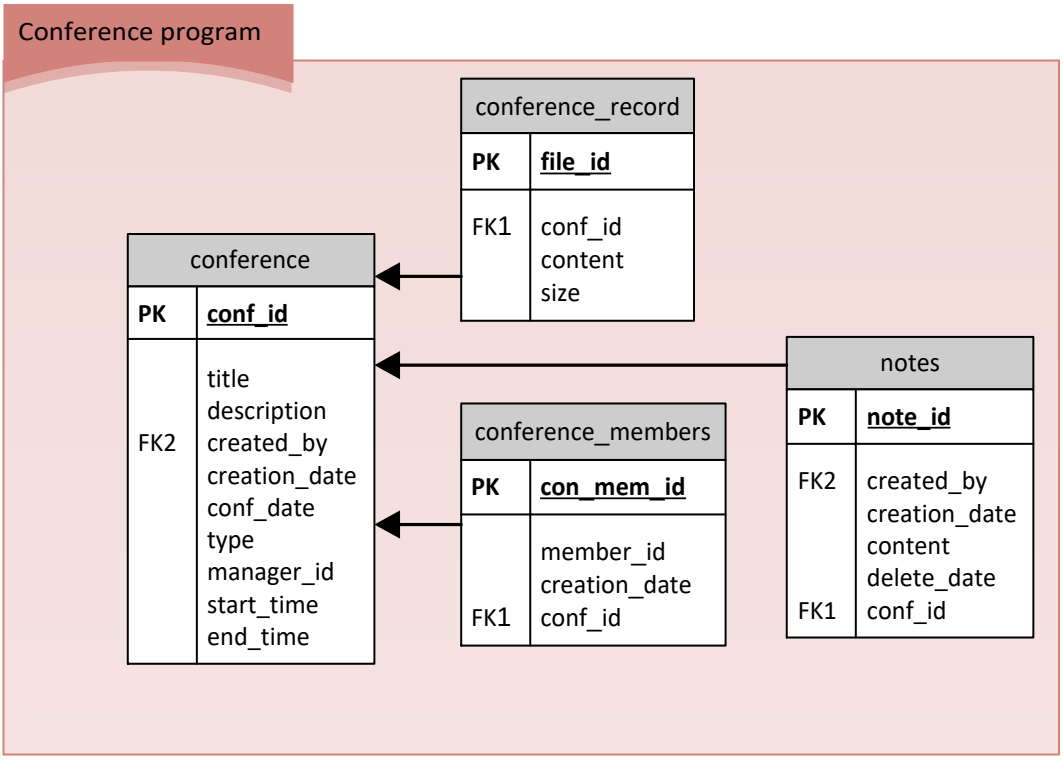
8-إمكانية تنزيل المؤتمر كفيديو أو صوت.

7-3-1- بنية قاعدة البيانات:

يوضح الشكل 24 بنية قاعدة البيانات ويتألف من الجداول التالية:

-الجدول Conferance: جدول يحتوي معلومات المؤتمر حيث أن كل مؤتمر يتميز بمدير يقوم بإدارة فعاليات المؤتمر من حيث إضافة أو منع أعضاء من المشاركة، يتألف من الحقول التالية:

(معرف المؤتمر conf_id – عنوان المؤتمر title – وصف عن المؤتمر description – تاريخ المؤتمر conf_date-تصنيف المؤتمر type- معرف المدير manager_id- توقيت البداية start_time-توقيت النهاية end_time-تاريخ الانشاء creation_date-منشىء المؤتمر created_by).



الشكل 24 :بنية قاعدة البيانات لبرنامج المؤتمرات

-الجدول Conference_record: جدول يتم فيه تخزين ملفات تسجيل المؤتمر، يخزن هذا الجدول المعلومات التالية: (file_id معرف الملف-Name اسم الملف-Size حجم الملف-content محتويات الملف كبايتات-conf_id معرف المؤتمر التي يتبع لها هذا الملف).

-الجدول Notes: جدول يحتوي المحادثات المتبادلة اثناء حصول المؤتمر، يخزن هذا الجدول المعلومات التالية: (معرف الملاحظة note_id، معرف المنشئ created_by، تاريخ الانشاء creation_date، محتوى الملاحظة content، تاريخ الحذف (في حالة حذف الرسالة) delete_date، معرف المؤتمر conf_id).

-الجدول Conference_member:هو جدول يربط المؤتمر مع أعضاءه، لايمكن تعديله إلا من قبل مدير المؤتمر.

7-4- برنامج الاستفسارات Questions Application:

هو برنامج يتم من خلاله نشر الاستفسارات والاجوبة المتعلقة بهذه الاستفسارات، الهدف من هذا الموقع هو انشاء نسخة محلية من هذا البرنامج على غرار مواقع عالمية مثل stackoverflow.com وذلك بهدف نشر الاستفسارات الاكاديمية التي تساهم في تقدم المستوى العلمي والاحتفاظ بها وانشاء مجتمع أكاديمي عربي، يجب أن يحقق التطبيق مايلي:

1-انشاء تساؤلات بمواضيع أكاديمية علمية.

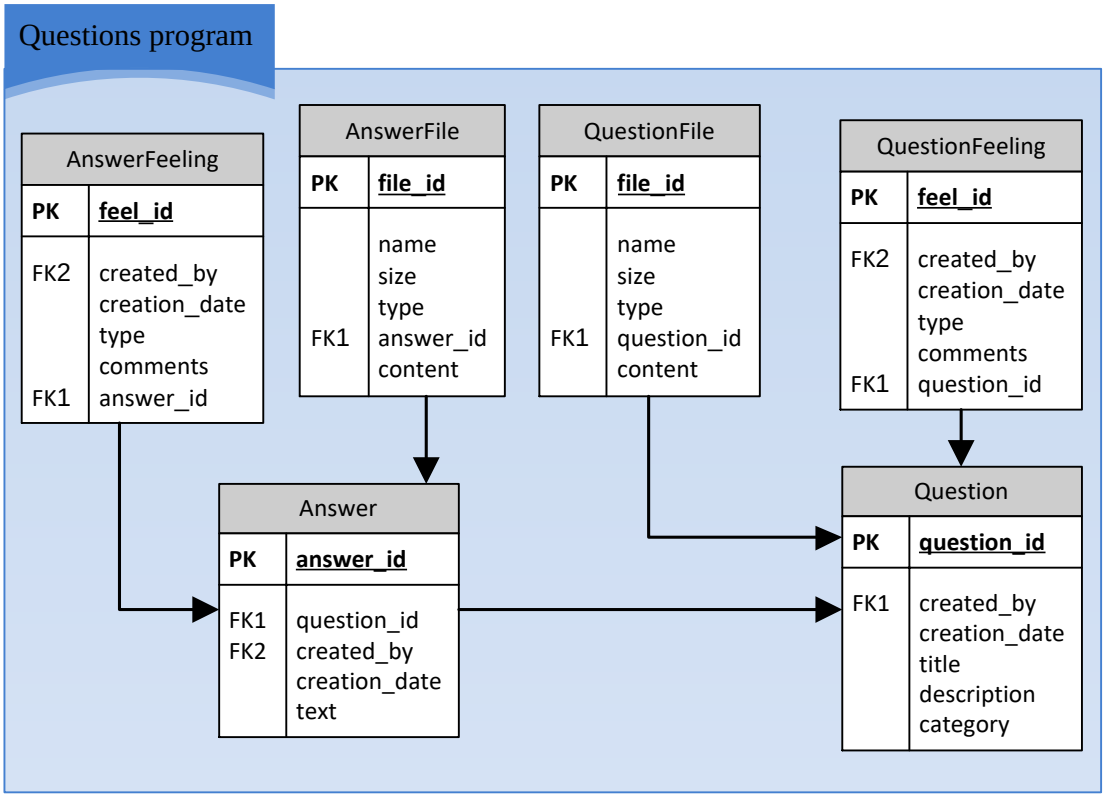
2-إضافة مناقشات خاصة بهذه التساؤلات.

3-إمكانية منح تقييمات للمناقشات والتساؤلات وذلك من خلال امكانية اضافة تقييم من المستخدم للسؤال أو أحد الأجوبة.

4-إمكانية تحميل ملفات مساعدة مع التساؤلات أو الإجابات.

7-4-1- بنية قاعدة البيانات:

يظهر الشكل 25 بنية قاعدة البيانات اللازمة لهذا التطبيق حيث يحوي الجداول التالية:



الشكل 25: بنية قاعدة البيانات لبرنامج الاستفسارات

-الجدول Question: جدول يحوي معلومات التساؤلات التي يتم طرحها من قبل المستخدمين يحتوي المعلومات التالية:

1-question_id: معرف السؤال.

2-title: عنوان السؤال.

3-creation_date: تاريخ انشاء السؤال.

4-created_by: منشىء السؤال.

5-description: شرح موسع للسؤال.

6-category: تصنيف السؤال، حيث يمكن انشاء تصنيفات عامة تدرج ضمنها التساؤلات (مثلاً bigdate-security-cloud-manufacturing).

-الجدول Answer: جدول يحوي الإجابات والتعليقات على التساؤلات التي يتم طرحها من قبل المستخدمين يحتوي المعلومات التالية:

1-answer_id: معرف الجواب.

2-question_id: معرف السؤال الذي يتضمن هذه الإجابة.

3-creation_date: تاريخ انشاء الجواب.

4-created_by: منشىء الجواب.

5-text: نص الجواب أو التعليق.

-الجدول QuestionFile: جدول يحوي معلومات الملفات الموجودة ضمن الأسئلة، يخزن هذا الجدول المعلومات التالية: (file_id معرف الملف-Name اسم الملف-Size حجم الملف-content محتويات الملف كبايتات-question_id معرف السؤال التي يتبع لها هذا الملف، type نوع الملف).

-الجدول AnswerFile: جدول يحوي معلومات الملفات الموجودة ضمن الأجوبة، يخزن هذا الجدول المعلومات التالية: (file_id معرف الملف-Name اسم الملف-Size حجم الملف-content محتويات الملف كبايتات-answer_id معرف الجواب الذي يتبع له هذا الملف، type نوع الملف).

-الجدول QuestionFeeling: جدول يعبر عن التفاعلات بالمشاعر أو العبارات التي يمكن أن تضاف من قبل المستخدمين على التساؤلات المطروحة، حيث يمكن أن تأخذ هذه المشاعر أنواع مختلفة (مثلا like-dislike-agree-refuse)، وقد تكون هذه المشاعر عبارات نصية(مثلا أشكرك على هذا التساؤل)، يخزن هذا الجدول المعلومات التالية: (feel_id معرف التفاعل-question_id معرف السؤال التي يتبع لها هذا التفاعل- type نوع التفاعل حيث يأخذ القيم التالية(..agree-disagree-like-dislike-etc)-creation_date تاريخ انشاء التفاعل-created_by منشىء التفاعل-comment التعليق النصي للتفاعل في حال وجوده).

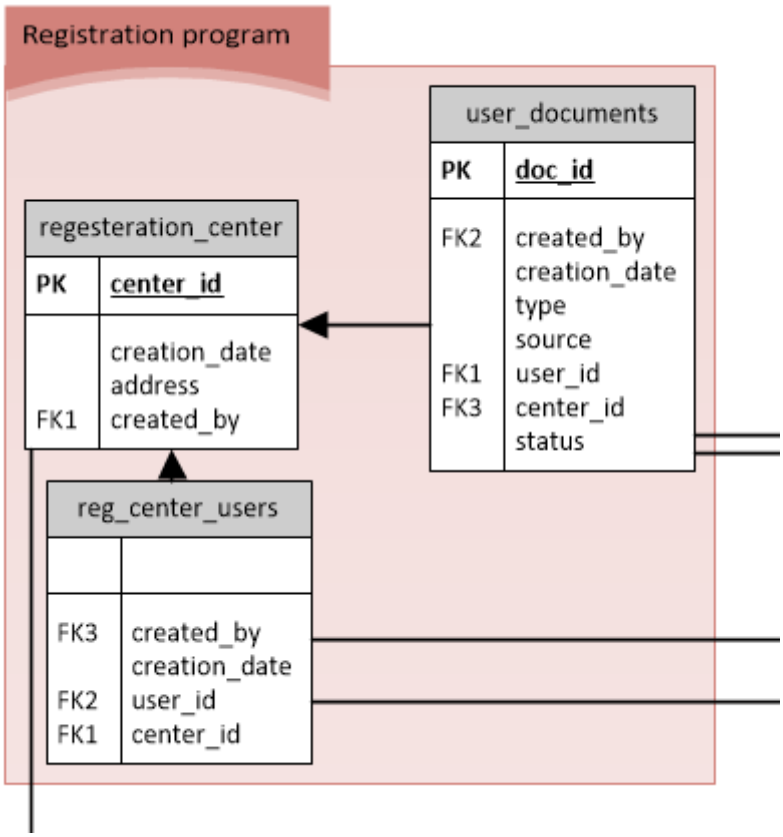
-الجدول AnswerFeeling: جدول يعبر عن التفاعلات بالمشاعر أو العبارات التي يمكن أن تضاف من قبل المستخدمين على الأجوبة المطروحة، حيث يمكن أن تأخذ هذه المشاعر أنواع مختلفة (مثلا like-dislike-agree-refuse)، وقد تكون هذه المشاعر عبارات نصية(مثلا أشكرك على هذا التساؤل)، يخزن هذا الجدول المعلومات التالية: (feel_id معرف التفاعل-answer_id معرف الجواب الذي يتبع له هذا التفاعل- type نوع التفاعل حيث

يأخذ القيم التالية(..agree-disagree-like-dislike-etc)-creation_date تاريخ انشاء التفاعل-created_by منشىء التفاعل-comment التعليق النصي للتفاعل في حال وجوده).

7-5- تطبيق مراكز التسجيل Registeration Places Manager Application:

إن تنفيذ هذه المنظومة على مستوى اقليمي كالوطن العربي يحتاج إلى وجود مراكز تسجيل ضمن البلاد تسمح بتسجيل الطلبات المقدمة من الأكاديميين في مختلف البلاد للتسجيل ضمن الشبكة المقترحة، تقوم هذه المراكز بتدقيق الوثائق المقدمة والتأكد منها، وذلك للوصول إلى شبكة موجهة إلى الشريحة الأكاديمية من المستخدمين، من أهم الوثائق التي يجب تقديمها في هذه المراكز والتأكد منها هو مستوى التعليم والمهارات والشهادات المكتسبة، هنا تظهر الحاجة إلى تطبيق يتم من خلاله إدارة مراكز التسجيل ضمن المنظومة المقترحة والموافقة عليهم، يجب أن يحقق التطبيق مايلي:

1- انشاء المستخدمين وتحميل الورقيات الخاصة بهم والتحقق منها.
2- تعديل الوثائق المقدمة.
3- إضافة تراخيص لمراكز تسجيل جديدة.
4- الموافقة على الطلبات الواردة من مراكز التسجيل.
5- تدقيق الوثائق المسجلة ضمن مراكز التسجيل



الشكل 26: بنية قاعدة البيانات لتطبيق مراكز التسجيل

7-5-1- بنية قاعدة البيانات:

يظهر الشكل 26 بنية قاعدة البيانات اللازمة لهذا التطبيق حيث يحوي الجداول التالية:

-الجدول Regesteration_center: جدول يحوي معلومات مراكز التسجيل المنتشرة ممكن أن تكون هذه المعلومات هي: عنوان المركز، وتاريخ الافتتاح، واسم المركز، وتوصيف المركز.

-الجدول Reg_center_users: هو جدول يربط الجدول الأساسي users مع جدول المراكز regesteration_center يهدف هذا الجدول إلى تجميع المستخدمين ضمن مراكز من خلال روابط مشتركة، قد تكون هذه الروابط: مكان الإقامة، الإختصاصات، الاهتمامات، وغيرها.

-الجدول User_documents: جدول يضم معلومات عن الوثائق التي تم تقديمها من قبل المستخدم، من هذه المعلومات نوع الوثيقة type، مصدر الوثيقة source، وحالة الوثيقة status اعتمادا على مبدأ إمكانية إلغاء اعتماد الوثيقة لسبب معين.

7-6- تطبيق **الصلاحيات:Permissions Program :**

إن تنفيذ شبكة اجتماعية أكاديمية على مستوى اقليمي كالوطن العربي يحتاج إلى وجود تطبيق لإدارة كافة موارد وصلاحيات الشبكة الإجتماعية الأكاديمية، يربط بين المستخدمين والتطبيقات من خلال صلاحيات rights وأدوار roles. يتم من خلال هذا التطبيق مايلي:

1-تحديد التطبيقات الأساسية العاملة ضمن الشبكة المقترحة.

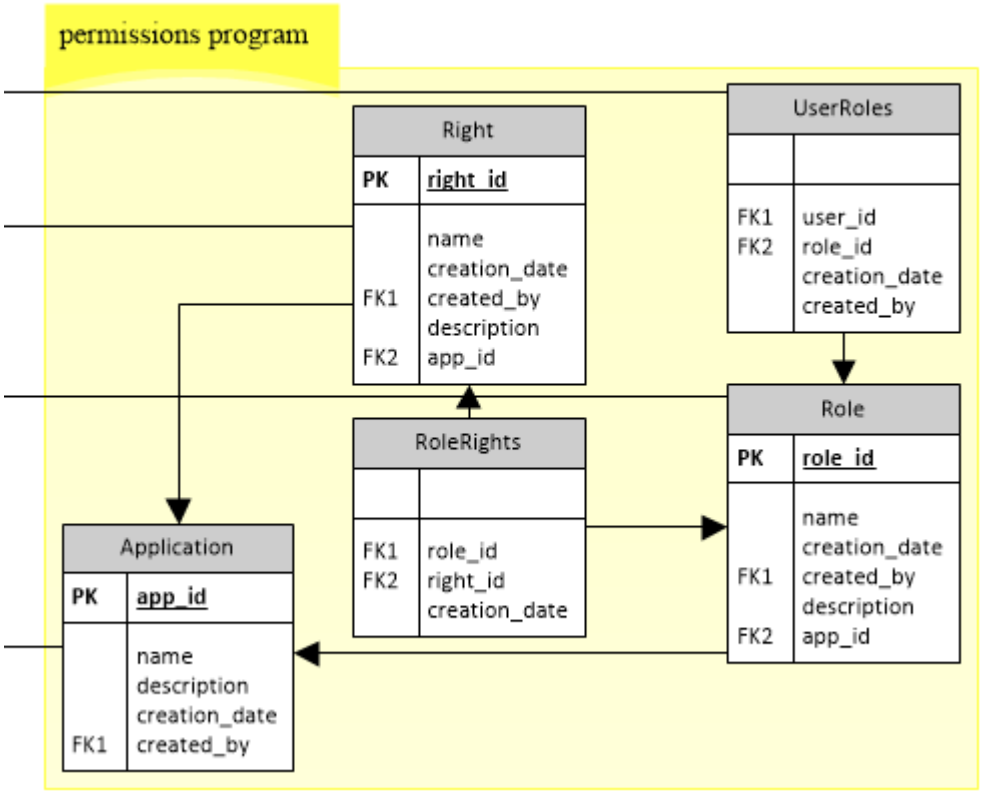
2- تحديد الأدوار الأساسية roles الخاصة بكل تطبيق.

3-تحديد الصلاحيات الأساسية rights التابعة لكل دور.

4-منح الأدوار للمستخدمين حسب الوظائف المحددة لكل مستخدم.

7-6-1- بنية قاعدة البيانات:

يظهر الشكل 27 بنية قاعدة البيانات اللازمة لهذا التطبيق حيث يحوي الجداول التالية:



الشكل 27: بنية قاعدة البيانات لتطبيق الصلاحيات

- **الجدول Application:** يتم فيه تحديد التطبيقات العامة التي سيتم إضافتها إلى المنظومة المقترحة، يمكن اعتبار البرامج التالية هي برامج يجب إضافتها إلى المنظومة المقترحة (المحادثة chat – الاسئلة questions –المؤتمرات)
- **الجدول Right:** يتم فيه تحديد الصلاحيات التي يمكن أن تكون ضمن المنظومة، في الحالة العامة يمكن جعل هذه الصلاحيات على مستوى الجداول أو الوظائف، بحيث يمكن اضافة الصلاحيات (اضافة – تعديل – حذف) لكل جدول من الجداول، كما يمكن اضافة صلاحية إلى وظائف محددة قد لاتحتاج العمل على أي جداول (مثلا تحميل كتاب).
- **الجدول Roles:** يتم فيه تجميع الصلاحيات ضمن كل تطبيق ومنحها بشكل مجمع إلى المستخدم.
- **الجدول UserRoles:** يتم من خلاله تحديد الأدوار لكل مستخدم.

7-7- تطبيق المقالاتDocuments Program :

يمكن إضافة خدمات إضافية إلى الشبكة الاجتماعية المقترحة، ومن هذه الخدمات (تطبيق المقالات) وهو تطبيق يهدف إلى نشر المقالات من قبل الطلاب بهدف إنشاء مجتمع علمي على أساس المناقشات التي يمكن أن تتم حول هذه المقالات، مما ينمي روح الإبداع والتعاون والتشاركية، ويجب أت يحوي التطبيق الميزات التالية:

1-إضافة مقالات خاصة بالطلاب.

2-عرض إحصائيات تبين مشاعر الآخرين حول المقالات (قبول – عدم قبول) وعمليات التنزيل وعدد المشاهدات.

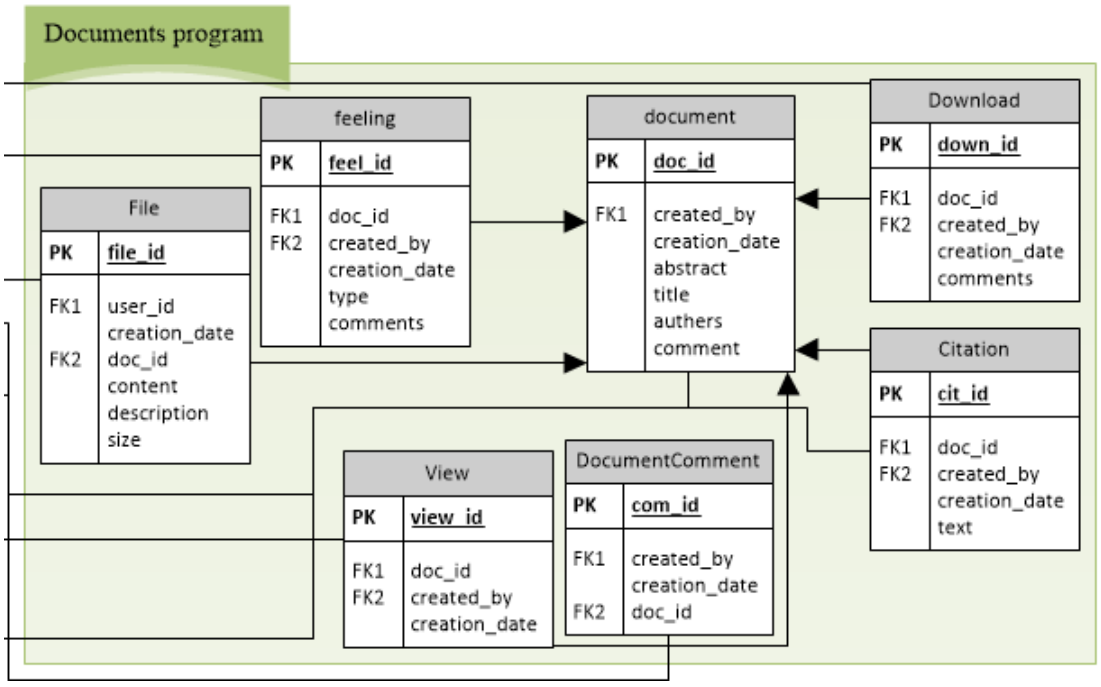
3-إمكانية تنزيل المقالات من قبل مستخدمين آخرين.

4-إمكانية إضافة إقتباسات من المقالة إلى مقالات أخرى، وإضافة معلومات هذه الاقتباسات.

5-إمكانية حصول مناقشات حول هذه المقالات والاحتفاظ بهذه المناقشات.

7-7-1- بنية قاعدة البيانات:

يظهر الشكل 28 بنية قاعدة البيانات اللازمة لهذا التطبيق حيث يحوي الجداول التالية:



الشكل 28: بنية قاعدة البيانات لتطبيق المقالات

-جدول Document: جدول يهدف إلى تخزين المعلومات الخاصة بالمقالات مثل العنوان title، الخلاصة abstract، المؤلفون authers، وملاحظات notes حول هذه المقالة.

-جدول File: جدول يتم فيه تخزين معلومات الملفات الخاصة بالمقالات وهنا يمكن أن يكون لدينا خيارين، الخيار الأول هو تخزين الملفات مباشرة إلى قاعدة البيانات، والخيار الثاني هو تخزين ارتباطات إلى الملفات وتخزين الملفات على وسائط تخزين مستقلة.

-جدول Download: يتم فيه تخزين معلومات التنزيل مثل: المستخدم الذي قام بعملية التنزيل-وتاريخ التنزيل والملاحظات على عملية التنزيل.

-جدول Citation: جدول يتم فيه تخزين معلومات عن عمليات الاقتباس من المقالة المنفذة من قبل المستخدمين يخزن فيه: المستخدم الذي قام بعملية الاقتباس-تاريخ الاقتباس – النص الذي تم اقتباسه.

-**جدول View:** جدول يسجل معلومات الاستعراض الخاصة بالمقالة من قبل المستخدمين، ويحوي معلومات مثل: المستخدم الذي قام بعملية الاستعراض، تاريخ الاستعراض، ومعرف الوثيقة التي تم استعراضها.

-**جدول Feeling:** جدول يتم فيه تخزين المشاعر التي يمكن أن يعبر عنها المستخدمون حول مقالة معينة، ويحوي معلومات مثل: المستخدم صاحب الشعور، نوع الشعور (رضا – عدم رضا – غضب – فرح..) وتاريخ الشعور.

-**جدول DocumentComment:** يتم فيه تخزين المناقشات والتعليقات التي يتم تبادلها حول مقالة معينة، وتحوي معلومات مثل: صاحب التعليق – المقالة التي تم التعليق عليها – نص التعليق – تاريخ التعليق)

7-8- تطبيق التحكيم Judging Program:

يمكن تطوير مهام الشبكة المقترحة من خلال إضافة لجان تحكيمية لتحكيم المقالات التي يتم اضافتها من قبل الطلاب، بحيث يتم اعتماد هذه المقالات كمقالات قابلة للنشر أو كمقالات محكمة، بحيث يمكن تطوير هذه العملية إلى مجلة علمية عربية تقوم بنشر كافة المقالات المحكمة والمعتمدة من قبل الشبكة المقترحة، يمكن تنفيذ هذه المهام من قبل تطبيق التحكيم الذي يوفر البيئة المناسبة بتنفيذ مهام التحكيم، يمكن أن يوفر هذا التطبيق المهام التالية:

1-إمكانية تحميل مقالات بهدف التحكيم.

2-إمكانية تنفيذ مراحل تحكيم على المقالة المطلوبة.

3-تخزين المناقشات المتبادلة من قبل الحكام حول المقالة.

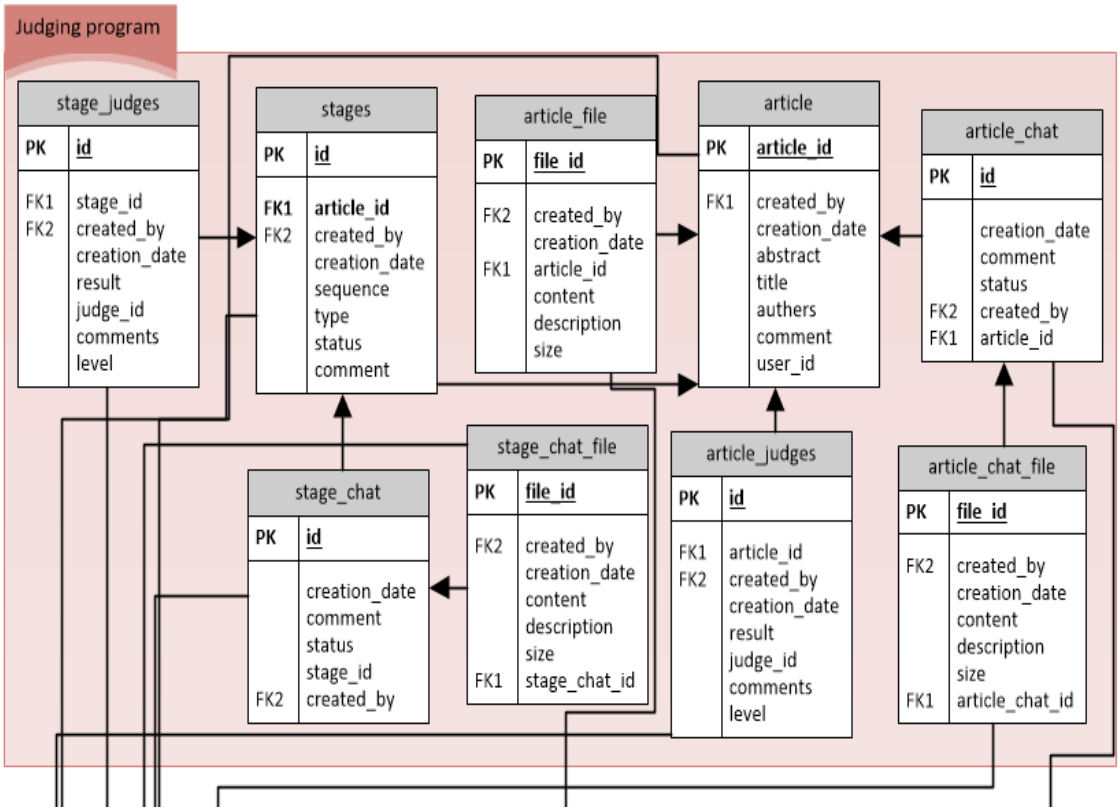
4-تبادل المناقشات مع صاحب المقالة وتخزينها.

5-إمكانية تبادل الملفات بين اللجنة وصاحب المقالة بهدف إيضاح التعليقات والتحكيمات.

6-إمكانية تبادل الملفات بين المحكمين بهدف إيضاح التعليقات والتحكيمات.

7-8-1- بنية قاعدة البيانات:

يظهر الشكل 29 بنية قاعدة البيانات اللازمة لهذا التطبيق حيث يحوي الجداول التالية:



الشكل 29: بنية قاعدة البيانات لتطبيق التحكيم

-الجدول Article: جدول يتم من خلاله انشاء المقالات من قبل الطلاب ويعتبر الجدول الأساسي لتحديد معلومات المقالة مثل: العنوان title، الملخص abstract، المؤلفون authers.

-الجدول Article_file: لتحديد الملفات المرتبطة بالمقالة ويتم تحديد معلومات الملفات مثل: الحجم والمحتوى والعنوان ومسار التخزين والمقالة التابع لها.

-الجدول Article_chat_file: لتحديد الملفات المرتبطة بالمناقشات الخاصة بالمقالة ويتم تحديد معلومات الملفات مثل: الحجم والمحتوى والعنوان ومسار التخزين والمناقشة التابع لها article_chat_id.

-الجدول Stage_chat_file: لتحديد الملفات المرتبطة بالمناقشات الخاصة بالمرحلة التحكيمية ويتم تحديد معلومات الملفات مثل: الحجم والمحتوى والعنوان ومسار التخزين والمناقشة التابع لها stage_chat_id.

-الجدول Stages: يحدد المراحل التي يمكن أن تمر بها المقالة ويحوي معلومات مثل نوع المرحلة type مثل (تسجيل- تحكيم- تنقيح- نشر)، ترتيب المرحلة ضمن المقالة sequence، الحالة status (فعالة – غير فعالة)، المقالة التابعة لها المرحلة article_id.

-الجدول Article_judges: جدول يحدد الأحكام والقرارات التي يمكن أن تشمل المقالة بالكامل ويشمل المعلومات التالية: article_id ويحدد معرف المقالة، نتيجة التحكيم resault، مستوى التحكيم level، ملاحظات.

-الجدول Stage_judges: جدول يحدد الأحكام والقرارات التي يمكن أن تشمل المقالة في مرحلة محددة ويشمل المعلومات التالية: judge_id ويحدد معرف المحكم، نتيجة التحكيم resault، مستوى التحكيم level، ملاحظات.

-الجدول Article_chat: يحوي المناقشات المرتبطة بالمقالة ويحوي المعلومات التالية: منشئ التعليق created_by، معرف المقالة، محتويات التعليق، حالة التعليق (فعال – غير فعال).

-الجدول Stage_chat: يحوي المناقشات المرتبطة بالمرحلة التحكيمية ويحوي المعلومات التالية: منشئ التعليق created_by، معرف المرحلة التحكيمية، محتويات التعليق، حالة التعليق (فعال – غير فعال).

8- الفصل الثامن:

تصميم نموذج تعلم آلي لتصنيف المقالات العلمية

8-1- المقدمة:

تزخر المكتبات الرقمية بمقالات علمية هائلة تغطي مجموعة واسعة من المجالات العلمية والبحثية. ومع تزايد حجم هذه المعلومات، يصبح من الضروري تطوير أدوات فعالة تسهل عملية تصنيف هذه المقالات وتوجيه الباحثين والمهتمين إلى الموارد المناسبة. هنا يأتي دور تصميم نموذج تعلم آلي قادر على تصنيف المقالات العلمية إلى مجال علمي محدد. يعتمد هذا النموذج على القوة الحسابية لتحليل البيانات الكبيرة والتعلم الآلي، حيث يقوم بتحليل مضمون المقالات واستخراج الأنماط والمعلومات الرئيسية التي تميز كل مجال علمي عن غيره. بفضل هذا التقنيات، يمكن للنموذج تقديم تصنيف دقيق وسريع للمقالات، مما يوفر وقتًا وجهدًا للباحثين ويساهم في تعزيز التفاعل العلمي والتطور في مختلف المجالات.

8-2- أهمية النموذج المقترح وأسباب القيام به:

تنبع أهمية النموذج المقترح من النقاط التالية:

إدارة السحابة وتوفير الموارد: يمكن أن يقلل التصنيف الآلي للمقالات من الحاجة إلى البحث اليدوي والإدارة الضخمة للمحتوى، مما يخفف العبء على الباحثين ويزيد من كفاءة استخدام موارد الحوسبة السحابية من خلال تصنيف المقالات وتوجيهها إلى الخوادم ذات الحاجة الفعلية (توفير موارد الحوسبة حسب الطلب).

تسهيل البحث العلمي: تطورت تقنيات التعلم الآلي بشكل كبير في السنوات الأخيرة، وأصبحت تلعب دورًا محوريًا في تسهيل العديد من العمليات والمهام، بما في ذلك البحث العلمي. يعتبر تصميم نموذج تعلم آلي قادر على تصنيف المقالات العلمية إلى مجال علمي محدد من بين التطورات الرئيسية التي أسهمت في تسهيل عملية البحث العلمي. فبدلاً من الاعتماد على البحث اليدوي المضني في المكتبات الرقمية الضخمة، يمكن لهذه التقنية أن تقدم طريقة آلية وفعالة لتصنيف المقالات وتوجيه الباحثين نحو المصادر المناسبة بسرعة ودقة أكبر. يقوم النموذج بتحليل المحتوى العلمي للمقالات واستخراج السمات الرئيسية لكل مجال علمي، مما يسهل على الباحثين تحديد المصادر ذات الصلة بمواضيع بحوثهم بشكل أكثر فعالية. وبهذه الطريقة، يسهم تصميم هذا النموذج في توفير الوقت والجهد للباحثين، ويعزز التفاعل العلمي والتطور في مختلف المجالات العلمية.

تحسين تجربة المستخدم: تمثل المقالات العلمية مصدرًا هامًا للمعرفة والمعلومات في مختلف المجالات العلمية، ومع زيادة حجم هذه المقالات على الإنترنت، يصبح التصفح والبحث فيها أمرًا متعبًا ومعقدًا للغاية. هنا يأتي دور تصميم نموذج تعلم آلي قادر على تصنيف المقالات العلمية إلى مجال علمي محدد في تحسين تجربة المستخدم. بفضل هذا النموذج، يمكن للمستخدمين الآن الوصول بسهولة إلى المقالات المتعلقة بمجالاتهم المهنية أو الأكاديمية بشكل أسرع وأكثر فعالية، دون الحاجة إلى الغوص في الكثير من المواد غير المرتبطة. بواسطة تحليل البيانات واستخراج الأنماط من المقالات، يمكن للنموذج توفير قائمة مرتبة ومنظمة للمقالات المتنوعة، مما يساعد المستخدمين على اكتشاف المصادر البحثية المناسبة بسهولة. وبهذه الطريقة، يلعب تصميم النموذج الذكي هذا دورًا حيويًا في تحسين تجربة المستخدم وجعل عملية البحث العلمي أكثر فعالية وإنتاجية.

التحليل الاستراتيجي للبيانات: في ساحة البحث العلمي، تمثل المقالات العلمية ثروة من المعرفة، ولكن التحليل الاستراتيجي لهذه المعلومات يمكن أن يكون تحديًا كبيرًا. هنا يأتي دور تصميم نموذج تعلم آلي قادر على تصنيف المقالات العلمية إلى مجال علمي محدد في التحليل الاستراتيجي للبيانات. بفضل هذا النموذج، يمكن توفير وسيلة فعالة لتحليل البيانات العلمية بطريقة تستند إلى الذكاء الاصطناعي وتعلم الآلة. يستخدم النموذج تقنيات متقدمة لتحليل مضمون المقالات وتحديد مجالاتها العلمية بدقة، مما يمكن الباحثين والمحللين من استخراج الأنماط والاتجاهات الرئيسية في البيانات العلمية بشكل فعال. بالتالي، يساعد هذا التصميم الذكي في تحسين التحليل الاستراتيجي للبيانات واتخاذ القرارات المستنيرة في مجالات البحث والتطوير، مما يسهم في التقدم العلمي والابتكار في مختلف المجالات.

8-3- النموذج المقترح:

ضمن دراسة النموذج المقترح سيتم دراسة النقاط التالية:

- بنية النموذج المقترح.
- تجميع العينات.
- استخراج السمات عبر تقنية Word2Vec.
- تطبيق خوارزمية MLP في التصنيف.

8-3-1- بنية النموذج المقترح:

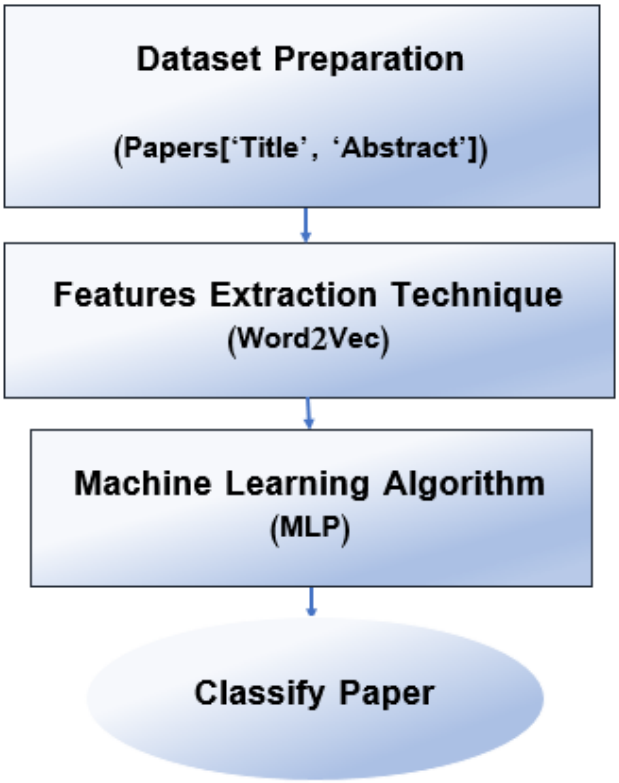
يتطلب تصميم نموذج تعلم آلي بناء نموذج متكامل تنفيذ الخطوات التالية:

- تحضير مجموعة البيانات بحيث ينتج لدينا لائحة من عناوين وملخصات المقالات: يبدأ النموذج بمرحلة تحضير مجموعة البيانات، حيث يتم جمع مجموعة كبيرة ومتنوعة من المقالات العلمية المصنفة حسب المجالات العلمية المختلفة.

-استخدام تقنيات معالجة اللغة الطبيعية لاستخراج المعلومات من عناوين وملخصات هذه المقالات: مما يتيح للنموذج فهم محتوى الأوراق العلمية بشكل أفضل وهنا تم استخدام تقنية Word2Vec.

- تدريب النموذج باستخدام خوارزميات تعلم الآلة المناسبة: حيث يهدف التدريب إلى تمييز الأنماط والمعالم المميزة لكل مجال علمي هنا تم استخدام خوارزمية MLP.

-تصنيف المقالات الجديدة إلى المجالات العلمية المحددة: بناءً على تحليل عناوين وملخصاتها بطريقة فعالة ودقيقة. من خلال هذه العملية، يساهم تصميم النموذج في تسهيل التحليل الاستراتيجي للبيانات العلمية وتقديم نتائج ذات جودة عالية تعزز فهمنا للمجالات العلمية المختلفة. يوضح الشكل 1 النموذج المقترح



الشكل 30: : مخطط يوضح بنية النموذج المقترح

8-3-2- تجميع العينات:

تم جلب البيانات من المرجع [24]، يتم تجميع العينات حسب الخطوات التالية:

- تجميع محتوى المقالات العلمية في ملف CSV.
- كل ورقة بحثية لديها عنوان Title وملخص Abstract.
- كل ورقة علمية تنتمي إلى صف محدد Label، مثل الفيزياء، الرياضيات، علوم الحاسب. الخ.

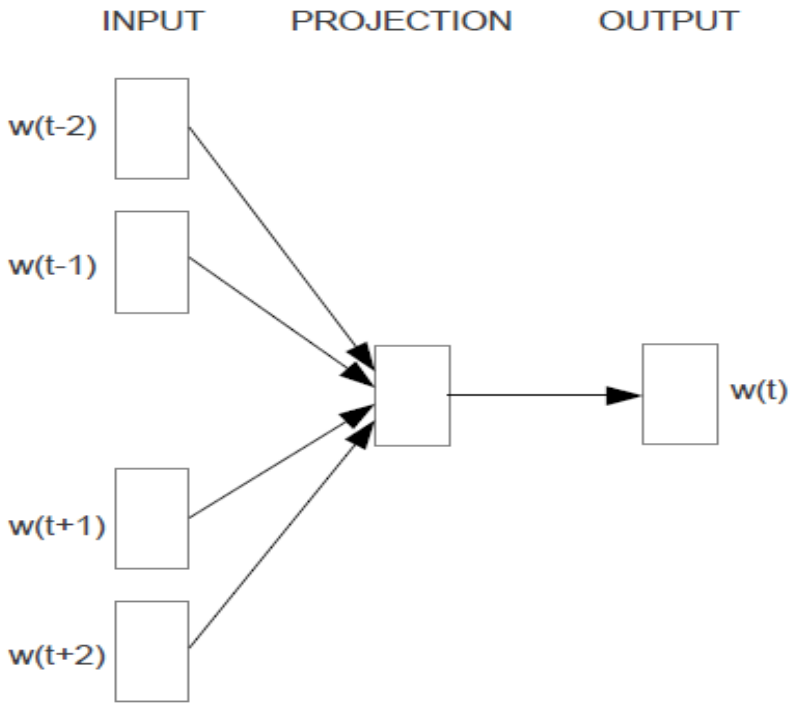
8-3-3- استخراج السمات عبر تقنية Word2Vec [25]:

إن تصميم نموذج تعلم آلي قادر على تصنيف المقالات العلمية إلى مجال علمي محدد يعتمد على مجموعة متنوعة من التقنيات المتقدمة، ومن بينها تقنية Word2Vec التي تُستخدم لاستخراج السمات الرئيسية من النصوص. يعتبر Word2Vec إحدى تقنيات معالجة اللغة الطبيعية التي تسمح بتحويل الكلمات إلى متجهات رياضية ذات بعد معين، مما يمكننا من فهم العلاقات الدلالية بين الكلمات في النصوص. يقوم النموذج بتحليل النصوص المتضمنة في عناوين وملخصات المقالات العلمية باستخدام تقنية Word2Vec لاستخراج سمات مميزة تمثل مضمون المقالات. على سبيل المثال، يمكن للنموذج أن يتعرف على الكلمات المترابطة بشكل وثيق بمجال علمي معين، مثل الكلمات المتعلقة بالطب أو الفيزياء أو الهندسة. من خلال هذا التحليل الدقيق، يمكن للنموذج بناء تمثيلات رياضية فعالة للمقالات العلمية، مما يمكنه من تصنيفها بدقة إلى المجالات العلمية

المناسبة. استخدام تقنية Word2Vec في تصميم النموذج يسهم بشكل كبير في تحسين عملية التصنيف وزيادة دقتها، مما يجعل التحليل الاستراتيجي للبيانات العلمية أكثر فعالية، تتميز هذه التقنية بمايلي: [26]

- إحدى طرق Word Embedding.
- عبارة عن شبكة عصبونية من طبقة خفية واحدة تقوم بتمثيل الكلمات بقيم عددية لتحديد معناها ومعرفة مدى التقارب أو الابتعاد فيما بينها.
- لها نوعان: 1- Continuous Bag Of Words (CBOW)، 2- Skip Gram
- إن الهدف من Word2Vecهو التدرب على تعلم أوجه التشابه بين الكلمات من خلال مجموعة كبيرة من النصوص، حيث تقوم بتمثيل كل كلمة في نهاية المطاف بشعاع رقمي Dense Vector.
- ومن خلال حساب المسافة بين شعاعين يمكن معرفة درجة التقارب والتشابه بين الكلمات.

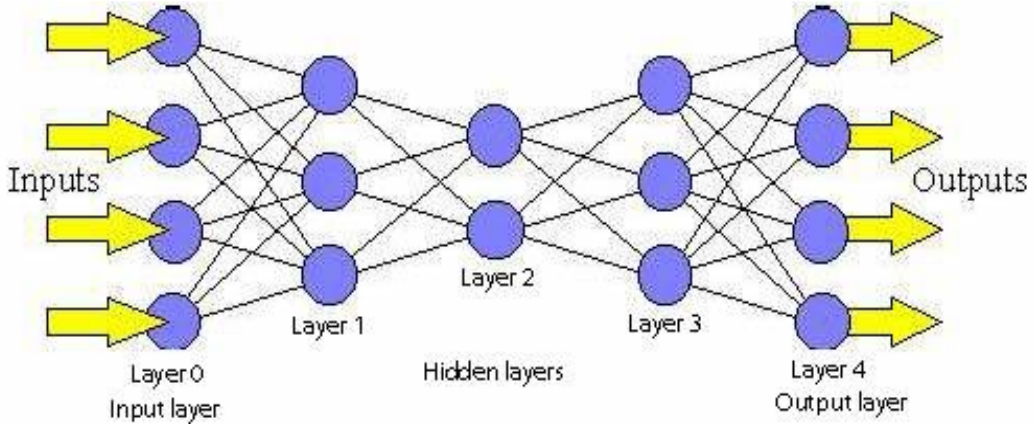
يوضح الشكل 31 آلية عمل تقنية Word2Vec.



الشكل 31: :آلية عمل تقنية Word2Vec

8-3-4- تطبيق خوارزمية MLP في التصنيف[27]

تعتمد خوارزمية MLP على شبكة من العقد الصناعية (النيورونات) التي تعمل في طبقات متعددة، مما يمكنها من تعلم وتمثيل العلاقات المعقدة بين المتغيرات في البيانات. يستخدم النموذج المبني على خوارزمية MLP البيانات المعدة سلفًا والتي تتضمن معلومات عن عناوين وملخصات المقالات العلمية، ويقوم بتدريبها لتحديد الأنماط والمعالم المميزة لكل مجال علمي. يتم ذلك من خلال تكرار تقديم مجموعة متنوعة من الأمثلة للنموذج، وتعديل الوزن والارتباطات بين العقد الصناعية لتحسين دقة التصنيف. بفضل قدرتها على التعامل مع البيانات غير الخطية والعمل في بيئات غير متوازنة، تُعتبر خوارزمية MLP خيارًا مثاليًا لتصميم نموذج تصنيف المقالات العلمية. تطبيق هذه الخوارزمية يسهم بشكل كبير في تحسين جودة التحليل الاستراتيجي للبيانات العلمية، ويساعد في تعزيز فهمنا للمجالات العلمية المختلفة بدقة أكبر. يوضح الشكل 32 آلية عمل خوارزمية MLP في النموذج المقترح:



الشكل 32: آلية عمل خوارزمية MLP في النموذج المقترح

8-4- العمل المنجز:

- ❑ دراسة الأطر والدراسات المرجعية ذات الصلة.
- ❑ تجميع وتحضير العينة Data Preparation.
- ❑ تجهيز وبرمجة تقنية Word2Vec.

8-4-1- دراسة الأطر والدراسات المرجعية ذات الصلة:

- دراسة التعلم الآلي الخاضع للإشراف.
- دراسة الشبكات العصبية Neural Network.
- دراسة تقنيات NLP مثل Word2Vec.

8-4-2- تجميع وتحضير العينة:

- تم تجميع العينة الأولية والتي تتألف إجمالا من (10000) ورقة بحثية تُصنف إلى عشرة أصناف [1] ، وهي:

1. Computer Science
2. High Energy Physics - Phenomenology
3. High Energy Physics - Theory
4. Astrophysics
5. Condensed Matter
6. Mathematics
7. Electrical Engineering and Systems Science
8. Quantum Physics
9. Physics
10. Statistics

- يوضح الشكل 33 العينة ضمن جدول Pandas بعد التجميع:

```
data.head()
```

		title	abstract	label
0	0	Critical state model for explosive emission pl...	A model has been proposed for estimation of ...	physics
1	1	Operational quantification of continuous-varia...	The diverse range of resources which underli...	quant-ph
2	2	Soliton fractional charges in graphene nanorib...	An introductory overview of current research...	cond-mat
3	3	Hit-time and hit-position reconstruction in st...	In this article, a new method for the recons...	physics
4	4	Smoothing splines on Riemannian manifolds, wit...	There has been increasing interest in statis...	stat

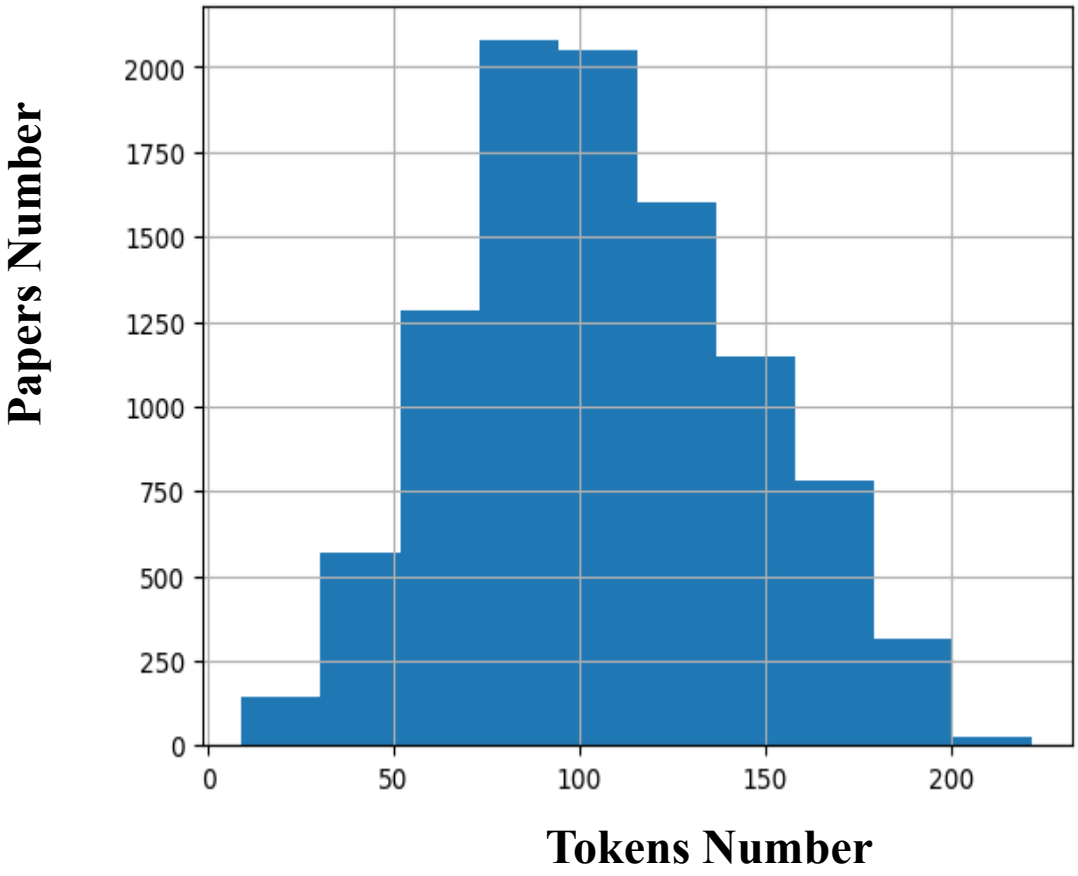
الشكل 33: العينة ضمن جدول Pandas بعد التجميع

- يتم بعد ذلك عدة خطوات منها دمج الحقلين Titleو Abstract للورقة البحثية بحقل واحد، ثم إزالة كلمات التوقف Stop Words، أخذ حالة الأحرف الصغيرة لكل الكلمات، إزالة الأقواس وعلامات الترقيم والأرقام، إزالة الفراغات الزائدة بين الكلمات.

- التقطيع Tokenization: تحويل ملخص وعنوان كل مقالة علمية إلى سلسلة من المفردات Tokens ضمن قائمة list. يوضح الشكل 5 مخطط يوضح عدد التقطيع مع عدد المقالات

- تحويل الأصناف (المجال العلمي للورقة البحثية) من نص إلى ترميز One-Hot Vector.

- عدد الأوراق البحثية بالنسبة لطول العنوان والملخص بعد تنظيف العينة:



الشكل 34::مخطط يوضح عدد التقطيع مع عدد المقالات

8-4-3- تجهيز وبرمجة تقنية Word2Vec:

تم تنفيذ الخطوات التالية فيما يتعلق بتجهيز تقنية Word2Vec:

- تم برمجة Word2Vec وتحديد وضبط المعلمات مثل معدل التعلم وعدد مرات التدريب وطول الشعاع المطلوب للكلمة.
- الحصول على شعاع Dense لكل كلمة في العينة بعد نهاية التدريب.
- تشكيل شعاع فريد لكل ورقة وذلك عبر حساب المتوسط الحسابي لكافة أشعة الكلمات التي يتألف منها حقلي العنوان والملخص (الحقل المدمج).

8-4-4- تجهيز وبرمجة خوارزمية MLP:

يتم تجهيز الخوارزمية للتصنيف من خلال تحضير البيانات وتنسيقها بشكل مناسب، مثل تمثيل الكلمات بصورة رياضية باستخدام تقنيات مثلWord2Vec ، وتحديد المعلمات الرئيسية وتهيئة الشبكة الصناعية بطريقة مناسبة لمجموعة البيانات المعينة. يتم بعد ذلك تدريب الخوارزمية على البيانات التدريبية، حيث تقوم بضبط وزن

الارتباطات بين النيورونات لتحسين أداء التصنيف. بفضل قدرتها على التعامل مع البيانات غير الخطية والتعلم من البيانات.

8-4-5- تحضير وتجهيز مجموعة التدريب والاختبار:

في هذه المرحلة تتم الخطوات التالية:

- التأكد من أن العينة جاهزة بشكل نهائي.
- بعثرة العينة.
- تقسيم العينة بشكل عشوائي إلى مجموعتين:

1- مجموعة التدريب Train Dataset بنسبة (67%).

2- مجموعة الاختبار Test Dataset بنسبة (33%).

8-4-6- ضبط بارامترات النموذج:

- تم تحديد نطاق معين لعدد من البارامترات المهمة مثل معدل التعلم وحجم الدفعة وتدريب النموذج على جزء عشوائي صغير من العينة بالاستناد لتلك النطاقات.
- ومن ثم تم اختيار القيم التي حققت الدقة الأفضل من أجل استخدامها لتدريب النموذج على العينة ككل.

8-4-7- المقياس:

- المقياس الذي تم اعتماده لتقييم أداء النموذج هو Categorical Accuracy.
- مقياس مناسب على اعتبار أن العينة تحتوي على فئات متوازنة.
- يقوم بحساب الدقة لكل فئة على حده ومن ثم يقوم بأخذ المتوسط الحسابي لدقة جميع الفئات.

5-8- النتائج والمناقشة:

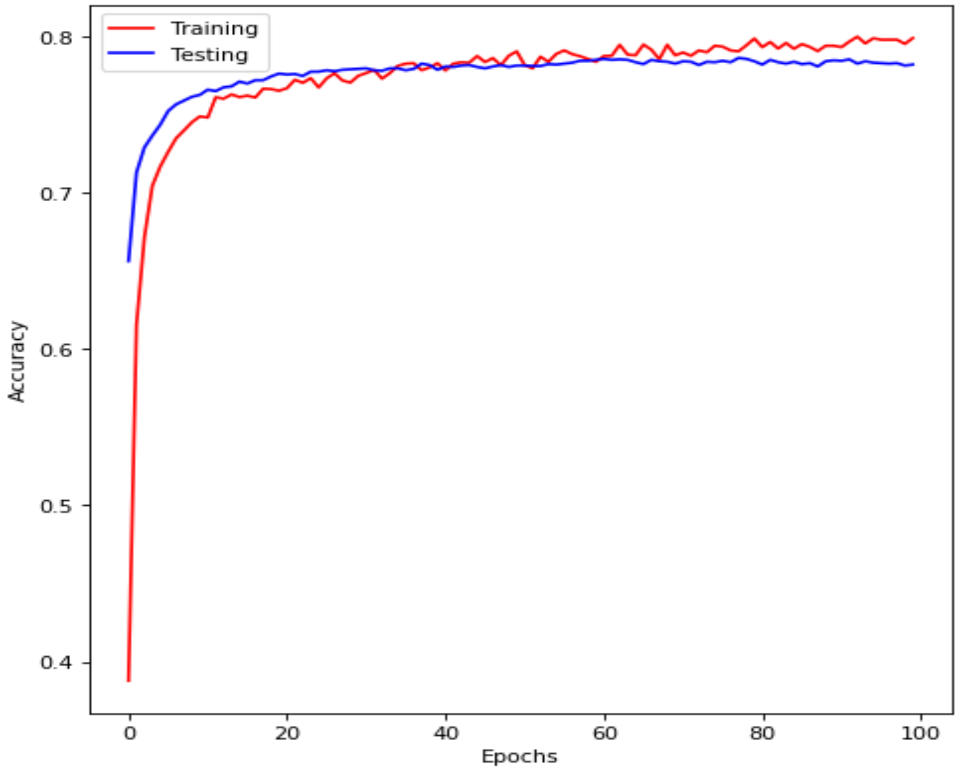
يوضح الجدول (35) النتائج النهائية للنموذج المقترح بعد تدريب النموذج على مجموعة التدريب واختباره على مجموعة الاختبار.

المجموعة	الخسارة	الدقة
مجموعة التدريب	0.5474	%79.91
مجموعة الاختبار	0.6055	%78.21

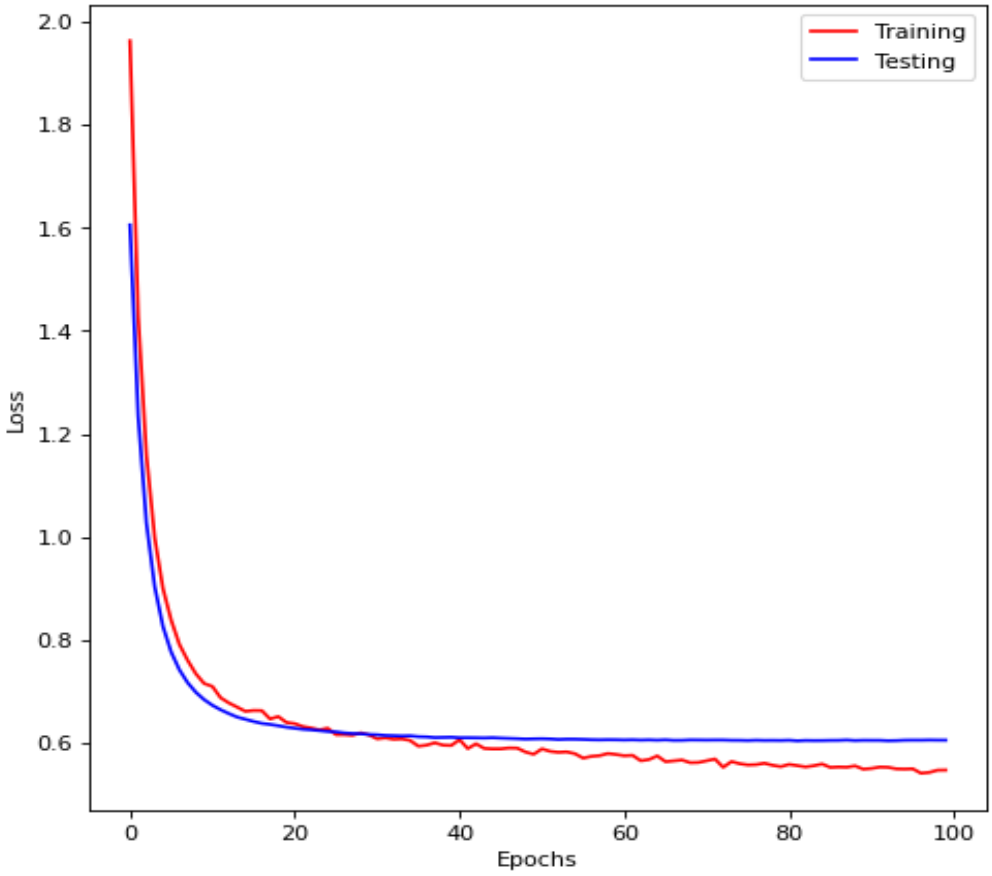
الجدول 35: نتائج النموذج المقترح

كما يوضح الجدول أعلاه أعطى النموذج أداء جيداً بدقة تصل لـ 78.21% وخسارة بلغت 0.6055 عند تطبيقه على مجموعة الاختبار، وهي دقة معقولة وجيدة قياساً بحجم التداخل الكبير بين الفئات المدروسة ضمن مجموعة البيانات.

تبين الرسوم البيانية أدناه في الشكل (36) والشكل (37) العلاقة بين عدد أدوار التدريب وتحسن الدقة والخسارة على التوالي، حيث تظهر تلك الأشكال بوضوح أن النموذج المقترح مناسب جدا لحل مشكلة البحث بدون الوقوع في مشكلة فرط الملاءمة.



الشكل 36: دقة النتائج لبيانات التدريب و الاختبار



الشكل 37: الخسارة في النتائج لبيانات التدريب و الاختبار

مع ذلك وعلى الرغم من تجريب بنى وإعدادات مختلفة للنموذج ككل بالنسبة لكل من تقنية Word2Vec وخوارزمية MLP معاً، لم تتحسن الدقة بشكل أفضل، من المحتمل أن يعود السبب في ذلك إلى التداخل الكبير بين بعض فئات التصنيف الذي لم يسمح للنموذج بتمييز تلك الفئات فيما بينها بشكل كبير، حيث أن بعض هذه الفئات تشترك بالمفردات والمصطلحات العلمية ذاتها.

على سبيل المثال، تبين عند فحص العينة أن فئة الفيزياء Physics متداخلة بشدة مع فئات أخرى، مثل:

Astrophysics

Quantum Physics

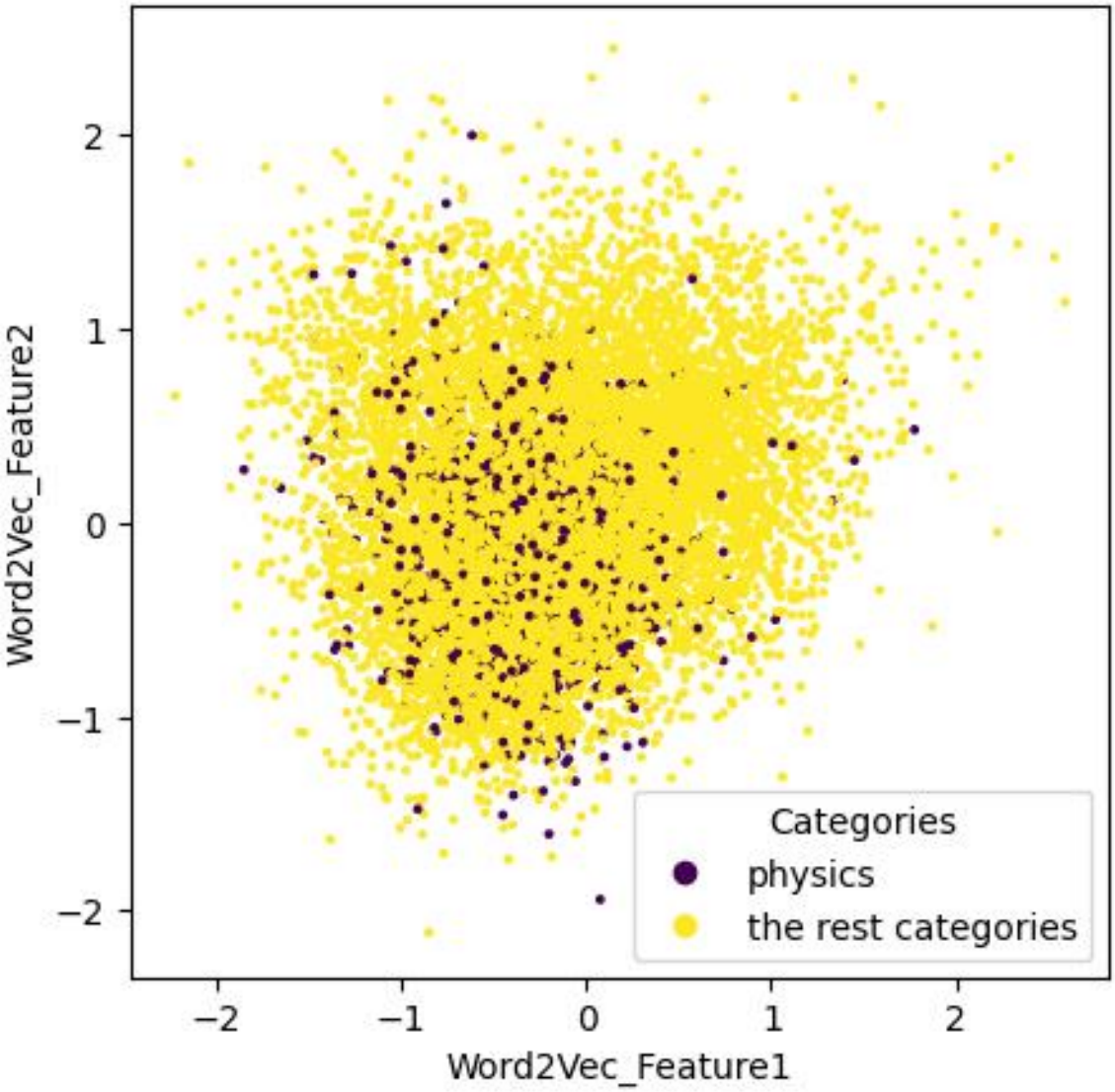
Condensed Matter

High Energy Physics – Phenomenology

High Energy Physics – Theory

حيث أن جميع تلك الفئات تنتمي جميعا لمجال علمي واحد.

يبين الشكل (38) التداخل الحاصل بين فئة الفيزياء وبقية الفئات بالنسبة لأول سمتين ضمن شعاع Word2Vec النهائي الممثل للورقة البحثية.



الشكل 38: التداخل الحاصل بين فئة الفيزياء وبقية الفئات بالنسبة لأول سمتين

8-5-1- الآفاق المستقبلية لتطوير البحث:

من الممكن أن يركز الجهد المستقبلي لتطوير البحث على نقطتين أساسيتين:

1- زيادة حجم العينة المدروسة وتضمين متن الورقة العلمية ضمن حقول العينة: نظرا للتداخل الكبير بين بعض فئات التصنيف كما يوضحه الشكل (38)، من المتوقع في حال زيادة عدد العينات وتضمين محتوى الورقة البحثية بشكل جزئي أو كلي (بالإضافة لحقلي العنوان والملخص) أن يعطي النموذج أداء أفضل.

2- تجريب واختبار خوارزميات تعلم آلي حديثة مخصصة للتعامل مع النصوص والجمل مثلTransformers، ومقارنة أدائها مع أداء النموذج المقترح من أجل الحصول على الدقة الأفضل.

9- المراجع References :

[1] M. Mohammad, Y. M. Lazim and S. Rosle, "Academic Social Network Sites: Opportunities and Challenges," *International Journal of Engineering & Technology,* vol. 7, pp. 133-136, 2018.

[2] K. Jordan, "From social networks to publishing platforms: a review of the history and scholarship of academic social network sites," *Frontiers in Digital Humanities*, 2019.

[3] "What is meant by cloud computing?," [Online]. Available: https://www.oracle.com/eg-ar/cloud/what-is-cloud-computing/.

[4] "What is cloud computing?," [Online]. Available: https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing.

[5] W. chai and S. Bigelow, "cloud computing," [Online]. Available: https://www.techtarget.com/searchcloudcomputing/definition/cloud-computing.

[6] H. Tabrizchi and M. K. Rafsanjani, "A survey on security challenges in cloud computing: issues,threats, and solutions," *The Journal of Supercomputing*, 2020.

[7] Verma and S. Kaushal, "Cloud Computing Security Issues and Challenges: A Survey," pp. 445-454, 2011.

[8] I. Odun-Ayo, A. M., F. Agono and R. Goddy-Worlu, "Cloud Computing Architecture: A Critical Analysis," *CLOUD COMPUTING ARCHITECTURE,* 2018.

[9] Z. N. Rashid, S. R. M. Zebari, K. H. Sharif and K. Jacksi, "Distributed Cloud Computing and Distributed Parallel Computing: A Review," *International Conference on Advanced Science and Engineering*, pp. 167-172, 2018.

[10] D. K. A.-R. Al-Malah, I. A. Aljazaery, H. T. S. Alrikabi and H. A. Mutar, "Cloud Computing and its Impact on Online Education," in *IOP Conference Series: Materials Science and Engineering*, 2021.

[11] Y. Gandhi, A. Singh and R. Jani, "Big Data and Its Application in Healthcare and Medical Field," *SpringerLink,* vol. 52, 2020.

[12] S. Tiwari and K. K. Rana, "Feature Selection in Big Data: Trends and Challenges," *SpringerLink,* vol. 52, 2020.

[13] F. Abukhodair, W. Alsaggaf, A. T. Jamal, S. Abdel-Khalek and R. F. Mansour, "An Intelligent Metaheuristic Binary Pigeon Optimization-Based Feature Selection and Big Data Classification in a MapReduce Environment," *MDPI,* vol. 9, no. 20, 2021.

[14] S. Meera and C. Sundar, "RETRACTED ARTICLE: A hybrid metaheuristic approach for efficient feature selection methods in big data," *J Ambient Intell Human Comput,* vol. 12, 2020.

[15] M. Prajapati and S. Patel, "A Review on Big Data with Data Mining," *SpringerLink,* vol. 52, 2020.

[16] O. AlFarraj, A. AlZubi and A. Tolba, "Optimized feature selection algorithm based on fireflies with gravitational ant colony algorithm for big data predictive analytics," *eural Comput & Applic,* vol. 31, 2019.

[17] د. علي بن ذيب الأكلبي، "البيانات الضخمة واتخاذ القرار في جامعة الملك سعود: دراسة تقييمية لنظام اتقان," *Journal of Information Studies & Technology (JIS&T),* vol. 2018, no. 2, 2019.

[18] G. Shobha and S. Rangaswamy, "Chapter 8 - Machine Learning," *Handbook of Statistics,* vol. 38, pp. 197-228, 2018.

[19] T. O. Ayodele, "Types of Machine Learning Algorithms," in *New Advances in Machine Learning*, InTech, 2010, pp. 19-48.

[20] https://vijay-choubey.medium.com/, "how-to-evaluate-the-performance-of-a-machine-learning-model," [Online]. Available: https://vijay-choubey.medium.com/how-to-evaluate-the-performance-of-a-machine-learning-model-d12ce920c365.

[21] www.analyticsvidhya.com, "Improve machine learning results," [Online]. Available: https://www.analyticsvidhya.com/blog/2015/12/improve-machine-learning-results/.

[22] "learn about oAuth," 2011. [Online]. Available: https://arabhardware.net/articles/oauth-introduction.

[23] S. J. Bigelow, "An introduction to big data in the cloud," 2021. [Online]. Available: https://www.techtarget.com/searchcloudcomputing/tip/An-introduction-to-big-data-in-the-cloud.

[24] F. e. al, "arXiv-10 dataset," 2022. [Online]. Available: https://paperswithcode.com/dataset/arxiv-10.

[25] X. Rong, "word2vec Parameter Learning Explained," *Computer Science*, vol. 1, no. Computation and Language, 2016.

[26] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *Computer Science*, no. Computation and Language, 2023.

[27] "neural networks architecture," 2020. [Online]. Available: https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Architecture/feedforward.html.

[28] H. Al-Samarraie and N. Saeed, "A scoping review of cloud computing tools for collaborative learning: Opportunities and challenges to the blended-learning environment," *Computers & Education*, 2018.

[29] K. B. S. K. A. M. Yamin MM, "Implementation of insider threat detection system," in *Future of Information and Communication*, Berlin, 2019.

[30] D. S. M. G. R. L. Coppolino L, "Cloud security: emerging threats and current solutions," *Comput Electr Eng*, pp. 126-140, 2017.

[31] Z. Y. M. Q. X. H. Y. Z. Y. M. W. X. L. L. D. H. Zhang X, "An empirical study of web resource manipulation in real-world mobile applications," in *Proceedings of the 27th Security Symposium (Security 18)*, 2018.

[32] E.-K. M. K. B. S. H. Butun I, "Cloud-centric multi-level authentication as a service for secure public safety device networks," *IEEE Communication Magazine*, vol. 4, pp. 47-53, 2016.

[33] K. Artem, N. Kunanets, R. Holoshchuk, V. Pasichnik and A. Rzheuskyi, "Information Support Of The Virtual Research Community Activities Based On Cloud Computing," in *IEEE CSIT*, Lviv, Ukraine, 2018.

[34] A. Majeed, J. Lv and T. Peng, "A framework for big data driven process analysis and optimization for additive manufacturing," *Rapid Prototyping Journal*, 2018.

[35] J. Fan, F. Han and H. Liu, "Challenges of Big Data analysis," *National Science Review*, vol. 1, pp. 293-314, 2014.

Abstract:

With the development of science, new concepts emerged that changed traditional society and changed the nature of social relationships. Among these concepts is the possibility of social communication without the need for direct communication. Social networking sites can be defined as a group of network services that allow people to create a public personal file and determine who can participate in communication and allow or prohibit them. These parties have access to what is shared, but with the emergence of these concepts, the need emerged to develop the level of reliability of online relationships in social networks that are based on real-world relationships to form a dynamic social cloud and to enable users to share resources in the context of a social network, so the concept of the network appeared. Academic Social Network (ASNs), which includes a variety of effective platforms, aims to benefit from social networks for special academic purposes.

In this research, a private local academic social network was designed to ensure the participation of actual academics. A general structure for applications, authentication methods, and types of users within the network was also developed. A machine learning model was also designed to classify scientific articles.

Syrian Arab Republic

Damascus University

Faculty of Mechanical and Electrical Engineering

Department of computer and Automation Engineering

Developing a Social Academic Network

Based on Cloud Computing

A dessertation submitted in partial fulfilment of the requirements for the degree of Master in

Computers Engineering and its Networks

By:

Eng. Amjad Makhlouf

Supervisor:

Prof. Samir Karaman

2023_2024