نظرة عامة لتعلم الآلة وبيانات تدريبها

أ.د. محمد محمد الهادي
أكاديمية السادات للعلوم الإدارية

المستخلص

تبني تطبيقات تعلم الآلة صار منتشرا علي نطاق واسع بدءا من عام ٢٠١٧. وطبقا لدراسة مسحية عن حالة الذكاء الاصطناعي والتكنولوجيات المعرفية لمنشآت الأعمال الكبيرة قامت بها شركة Deloitte عام ٢٠١٧ وهي دراسة مسحة بينت فيها أن عدد الشركات المستخدمة التعلم الآلي ٥٨٪ من مجموع الشركات التي اشتمل عليها المسح، التي نمت في إطار حالات الاستخدام لحوالي ٦٣٪ وذلك بزيادة ٥٪ في عام ٢٠١٨، مما يوضح مدي سرعة الأعمال التي تتبني تطبيق تعلم الآلة. كما أن مركز تطوير المعلومات الشهير IDC تنبأ أن الذكاء الاصطناعي الذي يحتضن تعلم الآلة اشتمل علي أيرادات من تعلم الآلة التي يتنبأ أن تنمو لحوالي ٢، ٥٢ بليون دولار أمريكي في عام ٢٠٢١، بينما كان معدل نموها في عام ٢٠١٦ يربو علي حوالي ٤٦ بليون دولار فقط. وتعني هذه التنبؤات أن هناك نموا مضطردا بما يعادل نصف مرة تقريبا سنويا في الأعوام الخمسة من ٢٠١٦ إلي ٢٠١٨. كما وضح أحد بحوث شركة O'Rleilly في عام ٢٠١٨ (Lorica & Nathan, 2018) أن ٨١٪ من الأعمال توظف علماء البيانات، ٣٩٪ في المنشآت الممسوحة توظف مهندسي تعلم الآلة بحوالي ٤١٪ من الشركات، وأن ٢٠٪ منها توظف مهندسي التعلم العميق. بالإضافة لكل ذلك استنتجت دراسة حديثة لشركة Deloitte لعام ٢٠١٩ أن الوصل بين تطبيق تعلم الآلة الناجح والميزة التنافسية ينمو بحوالي ١١٪ كما وضحة المجيبين علي استبيان المسح الذين وضحوا فيه تبنيهم تطبيقات الذكاء الاصطناعي وصولا للميزة التنافسية في الوقت الحالي، كما أن ٥٦٪ من المستجوبين اعتقدوا أن الذكاء الاصطناعي سوف يكون موضوعا حرجا للتطبيق في السنوات الثلاث القادمة. إلي جانب ذلك، توجد إمكانية توسيع الإستراتيجيات والمهارات الخاصة بالذكاء الاصطناعي لتبني تطبيقات تعلم الآلة التي صارت منتشرة علي نطاق واسع عالميا وخاصة في السنوات القليلة الماضية.

في مجال علم الحاسبات،إن ما بكل من التعرف علي النمط، استرجاع المعلومات، تعلم الآلة، تنقيب البيانات، وذكاء الويب يحتاج لإعداد جودة بياناته من خلال معالجة البيانات الخام المسبقة. وفي الممارسة الفعلية وجد أن تنقية/تنظيف البيانات وإعدادها بصفة عامة يستغرق تقريبا ٨٠٪ من الجهد المكرس لهندسة البيانات. وبذلك صار إعداد البيانات يمثل موضوعا بحثيا جوهريا يجب التطرق له في مبادرات ومشروعات تعلم

الآلـة والذكاء الاصطناعي المبنية علي توافر جودة البيانات وخاصة فيما يتعلق ببيانات تدريب خوارزميات تعلم الآلة.

يوضح هذا العمل نشوء ومفهوم تعلم الآلة والتعلم العميق؛ أبعاد تعلم الآلة التي ترتبط بكل أنوعها، تطبيقاتها الأكثر شيوعا؛ تحدياتها؛ خطوات تنفيذها؛ و أدوات وصيانة تعلم الآلـة وبائعيها ومصادر برمجياتها المفتوحة؛ تبني تعلم الآلة نمو الذكاء الذكاء الاصطناعي؛ ثم يتطرق العمل المقدم لكل من خصائص بيانات التدريب المهمة المحتاج لها تعلم الآلة وجودتها وقضياها، أغراض إعداد بيانات التدريب لنماذجها، عملية إعـداد بيانات التدريب وتسميتها ، توافر البشر لأعداد البيانات، تحديات إعداد البيانات، نشوء سوق إعداد البيانات.

الكلمات الآساسية: تعلم الآلة، التعلم العميق، أنواع تعلم الآلة تطبيقات تعلم الآلة، تحديات تعلم الآلة، خصائص بيانات التدريب، جودة بيانات التدريب، عملية إعداد بيانات التدريب.

١. المقدمة:

التعلم الآلي صار مستخدما في كثير من تطبيقات الذكاء الاصطناعي بزيادة واضحة، من خلال تحسين خبرة المستخدمين إلي تطوير منتجات وخدمات جديدة لم تكن متوافرة من قبل. ولا يوجد أي مجال أعمال في المنشآت المختلفة في الدول المتقدمة لم يتطرق لتعلم الألة حديثا.

تعلم الآلة هو مسار مهم لأنشاء الذكاء الاصطناعي ذاته، الذي بدوره يعتبر أحد دوافع ومحركات تعلم الآلة في أي منشأة تستخدمه. وفي هذا السياق، يوجد بعض الخلاف حـول طبيعة العلاقة بين الذكاء الاصطناعي وتعلم الآلة. البعض يري تعلم الآلة كمجال فرعي من الذكاء الاصطناعي، بينما البعض الآخر يري الذكاء الاصطناعي كمجال فرعي من تعلم الآلة بشكل أساسي. وبشكل عام، الذكاء الاصطناعي يهدف لمحاكاة وتقليد بعض جوانب الإدراك البشري أو اتخاذ القرار، بينما تعلم الآلة يمكن أن تستخدم لتعزيز أو لآلية أي مهمة افتراضيا، وليست مجرد علاقة واحدة بالمعرفة البشرية فقط. هذان الرأيان يعتبرا متصلان معا ويغذيان بعهما البعض في مدي إنتشار كل منهما علي أي حال. كما أن ممارسة تعلم الآلة تتضمن أخذ البيانات، فحصها للأنماط والاتجاهات، تطوير نوع من التنبؤ بالنتائج المستقلة من خلال تغذية خوارزمية بيانات أكثر. وفي هذا الإطار يمكن لعلماء البيانات شحذ تنبؤات نموذج تعلم الآلة المحتاج له.

وفي مجال علم الحاسبات كل من التعرف علي النمط، استرجاع المعلومات، تعلم الآلة، تنقيب البيانات، وذكاء الويب يحتاج لإعداد جودة بيانته من خـلال معالجة البيانات الخام مسبقا. وفي الممارسة الفعلية، وجد أن تنقية/تنظيف البيانات، وإعدادها يستغرق تقريبا ٨٠٪ من الجهد البشري المكرس لهندسة البيانا بصفة عامة. وبذلك صار إعداد البيانات يمثل موضوعا بحثيا جوهريا يجب التطرق له في مبادرات ومشروعات تعلم الآلة والذكاء الاصطناعي المبنية علي توافر جودة البيانات، وتمثل بيانات تدريب خوارزميات تعلم الآلة.

مما سبق، يتضح أن خوارزمية تعلم الآلة تعتمد علي إعـداد بيانات التدريب الدقيقة المسماة أو المعلمة جيدا التي تمثل مهام جمع البيانات، هيكلتها وتنظيمها لإمكانية تحليلها كجزء أساسي من تصور تعلم الآلة وتحليلاتها وتطبيقاتها. وتشتمل عملية إعداد بيانات تدريب تعلم الآلة علي المعالجة المسبقة للبيانات، تنميطها، تطهيرها/تنظيفها، التحقق منها، وتحويلها. وهذا يعتمد علي إمكانية

استخلاص البيانات المحتاج لها من نظم داخلية عديدة، بالإضافة لمصادر خارجية متنوعة. ويتم عمل إعداد بيانات تدريب تعلم الآلة من خلال تكنولوجيا المعلومات وفرق ذكاء الأعمال عند دمجهم مجموعات بيانات التدريب للتحميل في مستودع البيانات، بحيرة البيانات Data Lake، أو قاعدة بيانات تطبق ما يطلق عليه لا لغة التساؤل المهيكلة NoSQL، بالإضافة لمحللي البيانات الذين يستخدمون أدوات إعداد البيانات للخدمة الذاتية Self-service ولجمع وإعداد البيانات المجمعة والمهيئة للتحليل ولاستخدام عروض البيانات المرئية كما في كل من لوحات العرض Dashboard و Tableau.

والعرض التالي يوضح نشوء ومفهوم تعلم ألالة والتعلم العميق؛ أبعاد تعلم الآلة التي ترتبط بكل أنوعها، تطبيقاتها الأكثر شيوعا؛ تحدياتها؛ خطوات تنفيذها؛ و أدوات وصيانة تعلم الآلة وبائعيها ومصادر برمجياتها المفتوحة؛ تبني تعلم الآلة نمو الذكاء الذكاء الاصطناعي؛ ثم يتطرق لكل من خصائص بيانات التدريب المهمة المحتاج لها تعلم الآلة وجودتها وقضياها، أغراض إعداد بيانات التدريب، عملية إعداد بيانات التدريب وتسميتها ، توافر البشر لأعداد البيانات، تحديات إعداد البيانات، نشوء سوق إعداد البيانات.

٢- تطور ومفهوم تعلم الآلة:

في السنوات القليلة الماضية كثر الحديث والنقاش الكثيف عن تعلم الآلـة، التعلم العميق والذكاء الاصطناعي الذي صار موثقا ومتاحا في عدد لا يحصي من المقالات العلمية وتقارير البحوث المنشورة المهنية ذات التفكير التكنولوجي. وكل ذلك تنبأ بمستقبل واعد من برامج الدردشة الذكية Chatbots، المركبات ذاتية القيادة Driverless Vehicles، المساعدين الافتراضيين Virtual Assistants، النظم الذكية، الخ مما يشكل مستقبل قد يرسم في ضوء قاتم بعض الأحيان، وكمثالي في أحيان أخري. حيث ستكون الوظائف البشرية نادرة، ومعظم النشاط التجاري سيتم التعامل معه من خلال الروبوتات أو وكلاء الذكاء الاصطناعي. كما أن الممارس المستقبلي أو الحالي للتعلم الآلي من المهم أن يقدر علي التعرف علي الاشـارة في الضوضاء المحيطة حتي يتمكن من معرفة التطورات العالمية المتغيرة من الأخبار المنشورة التي قد يكون مبالغا فيها لحد ما.

مما سوف يجعل المجتمع علي بيان دائم ومستمر لأن المستقبل سيكون للفرد فيه دورا نشطا ليؤديه.

وفيما يلي الإجابة على السؤال المطرح ما الذكاء الاصطناعي، تعلم الآلـة والتعلم العميق؟ الذي يوضح الشكل التالي مدي الترابط والتلاحم بين هذه المصطلحات.

١/٢ الذكاء الاصطناعي:

نشأ مصطلح الـذكـاء الاصطناعي الحالي في خمسينيات القرن العشرين الماضي، عندما بدأت حفنة قليلة من الـرواد في مجال علم الحاسب الآلي المنشأ حديثا التساؤل عن التالي: ما الذي يمكن أجهزة الحاسب الآلي للتفكير؟ سؤال ما زلنا حتي اليوم نكتشف عواقبه. كما أن مجال الذكاء الاصطناعي صار يفسر في أنه الجهد لآلية المهام الفكرية التي ينفذها البشر بشكل طبيعي، وعلي ذلك يعتبر الذكاء الاصطناعي مجالا عام يحتضن كل من تعلم الآلة، والتعلم العميق إلي جانب مداخل كثيرة جدا لا تتضمن أي تعلم.

٢/٢ تعلم الآلة:

المحرك التحليلي Analytical Engine أول محرك ميكانيكي معروف للأغراض العامة الذي

صمم في ثلاثينات وأربعينيات القرن الثامن عشر، لم يكن يقصد منه جهاز الحاسب الآلي ذو الأغراض العامة، لأن مفهوم الحوسبة ذات الأغراض العامة لم يتم قد تم اختراعها بعد. وكان يعني مفهوم المحرك التحليلي طريقة استخدام العمليات الميكانيكية فقط لآلية حسابات معينة في مجاله، التي انبثق منها ا اسم المحرك التحليلي ولديها أي اختراقات مهما كانت تنتج أي شيء. كما كان في مقدرتها فعل أي شيء يتطلب أداؤه، أي أنها تساعد في جعل ما تتعرف عليه بالفعل متوافرا. هذه الملاحظة القديمة اقتبست أخيرا من قبل ألين تورينج Allen Turing أحد رواد علم الحاسب الآلي والذكاء الاصطناعي الذي قدم اختبار تورينج Turing Test بالإضافة إلي المفاهيم الأساسية للذكاء الاصطناعي. حيث اقتبس مفهوم المحرك التحليلي القديم أثناء التفكير ما إن كانت أجهزة الحاسبات ذات الأغراض العامة يمكنها القدرة علي التعلم والأصالة ؟ واستنتج انه بقدرتها أداء ذلك.

ومن هذا المنطلق طرحت عدة أسئلة عن هل يمكن أجهزة الحاسبات التحاور فيما نأمرها بما تقوم به؟ هل يمكن أن تفاجئنا الحاسبات الآلية أنه بدلا من صياغة المبرمجين قواعد معالجة البيانات يدويا، يمكن الحاسب الآلي تعلم هذه القواعد تلقائيا بالنظر للبيانات المتاحة له؟ هذه التساؤلات صارت تنير الطريق لبحوث برمجة جديدة غير تقليدية. حيث أنه في البرمجة التقليدية الكلاسيكية تكون القواعد والبيانات من إدخال البشر (المبرمجين) للمعالجة طبقا لها في إطار الإجابات الناتجة، كما في الشكل التالي:

شكل (١/٥) البرمجة التقليدية الكلاسيكية

أما مع تعلم الآلة وخوارزميتها تكون المدخلات

شكل (٢/٥) التعلم الآلي نموذج برمجة جديد

أي أن تعلم الآلة يؤدي إدخال بيانات البشر بالإضافة إلي الإجابات المتوقعة من البيانات للمخرج النهائي منها المتمثل في القواعد. هذه القواعد يمكن عندئذ أن تطبيق لبيانات جديدة لإنتاج إجابات أصلية.

ونظام تعلم الآلة متدرب بدلا من برمجة بشكل صريح، حيث يعرض أمثلة كثيرة ذات صلة بمهمة، ويوجد هيكل إحصائي في هذه الأمثلة الذي يسمح للنظم في النهاية التوصل إلي القواعد لآلية المهمة. علي سبيل المثال، عند الرغبة لآلية مهمة وضع علامات علي صور لحفل التخرج أو لعطلة وما شابه ذلك، يمكن عرض نظام تعلم الآلة مع أمثلة كثيرة وضعت علاماتها بواسطة البشر بالفعل، وسوف يتعلم النظام القواعد الإحصائية لربط صور معينة للعلامات المعينة الموضوعة.

وعلي الرغم من أن تعلم الآلة بدأ ينتعش في تسعينيات القرن العشرين الماضي، إلا أنه قد أصبح بسرعة مجال الذكاء الاصطناعي الفرعي الأكثر انتشارا ونجاحا، ألي جانب أنه صار أيضا اتجاها مدفوعا بواسطة توافر أجهزة حاسبات أسرع، ومجموعات بيانات أكبر. علما أن الذكاء الاصطناعي مرتبط بإحكام للإحصائيات الرياضية، لكنه يختلف من الإحصائيات في طرق عديدة مهمة لا تشبه الإحصائيات. كما يتجه تعلم الآلة إلي التعامل مع مجموعات بيانات ضخمة معقدة (مثل مجموعة بيانات ملايين الأشكال التي كل منها تتشكل من عشرات الآلاف من النقاط Pixels) التي سوف يكون التحليل الإحصائي مثل تحليل Bayesian غير عملي لها. نتيجة لذلك، تعلم الآلة والتعلم

العميق يعرضان القليل نسبيا بصفة خاصة، علي الرغم أنه الأكثر حداثة (من النظرية الرياضية الموجهة هندسيا)، كما أنه مجالا يدويا، حيث أن الأفكار تبرهن تجريبيا أكثر من أن تكون نظرية في الغالب.

٣/٢ التعمق في التعلم العميق:

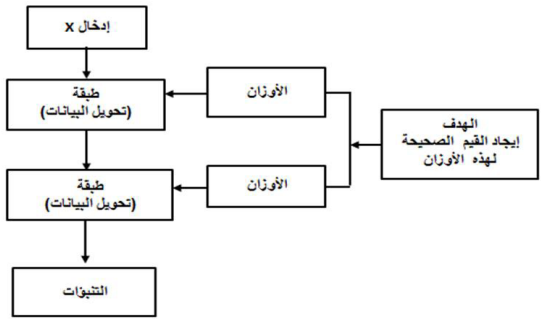
التعلم العميق Deep Learning مجالا فرعيا لتعلم الآلـة. حيث أنـه يمثل نظرة جديدة على تمثيلات التعلم من البيانات التي ترتكز عللي الطبقات المتعاقبة لتمثيلات ذات مغزي متزايد. العمق Depth في التعلم العميق ليس الإشارة إلي أي نوع لفهم أعمق محقق من خلال المدخل، بدلا من أنها تقف علي فكرة الطبقات المتعاقبة هذه. كم عدد الطبقات المتطلبة لنمذجة البيانات التي يطلق عليها «تعمق النموذج»، كما أن من أسماء الملائمة الأخرى «تعلم تمثيلات الطبقات Layered representations Learning»، و «تعلم التمثيلات الهرمية Hierarchical Representations Learning».

ويتضمن التعلم العميق المعاصر عشرات أو حتى مئات الطبقات المتعاقبة غالبا التي تعلمت كلها تلقائيا من التعرض لبيانات التدريب. وفي الوقت نفسه، توجد مداخل أخري لتعلم الآلة تتجه للتركيز علي التعلم لطبقة أو طبقتي تمثيلات بيانات فقط. ومن ثم يطلق عليهم «التعلم الضحل Shadow Learning» في بعض الأحيان.

في التعلم العميق تمثيلات الطبقات هذه تعلمت دائما تقريبا عبر النماذج التي يطلق عليها «الشبكات العصبية " Neural Networks» المهيكلة في طبقات حرفية مكدسة فوق بعضها البعض.

ومصطلح الشبكة العصبية يشير إلي علم الأعصاب Neurobiology ، لكن علي الرغم من أن بعض المفاهيم المركزية طورت جزئيا من خلال الإلهام من فهم المخ البشري، إلا أن نماذج التعلم العميق ليست نماذج المخ البشري. كما لا توجد أدلة أن المخ البشري ينفذ أي شـيء، مثل آليات التعلم المستخدمة في نماذج التعلم العميق المطورة. وقد تصادف بعض المقالات في إطار ما يطلق عليه العلم الشعبي Pop-Science التي تعلن عن أعمال التعلم العميق كالمخ او النمذجة علي غرار المخ البشري، لكنها لا تمثله فعليا مما سيكون مربكا أو عكسيا للمبدئين الجدد في المجال التفكير في التعلم العميق كما هو متعمق بأي شكل من الأشكال بعلم الأعصاب. لذلك لا يحتاج في أن يكون كفن مثل سحر وغموض عقول البشر. مما يجب نسيان أي شيء ربما قد تمت قراءته عن الروابط الافتراضية بين التعلم العميق وعلم البيولوجيا. أي أن التعلم العميق هو إطار عمل رياضي لتمثيلات التعلم من البيانات

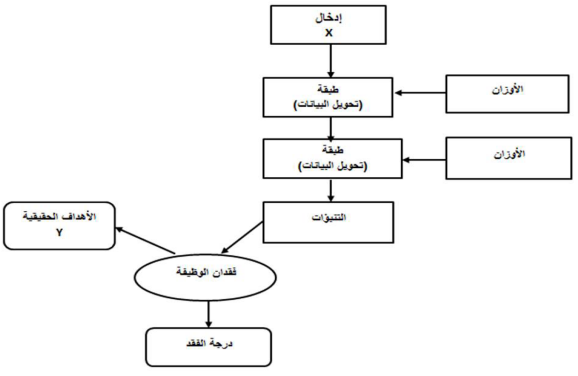
توصيف ما تعمله الطبقة لإدخالها، أن البيانات المخزنة في أوزان Weights الطبقة التي في جوهرها تتضمن حفنة أرقام . من الناحية الفنية، يمكن القول أن التحول المنجز بواسطة الطبقة يوصف او يمثل في ألفاظ ضوابط القياسات التي يطلق عليها معلمات Parameters (ضوابط قياسية كعوامل عددية). من خلال أوزانها، كما في الشكل التالي:



شكل (٣/٥) الشبكات العصبية المعلمة بواسطة أوزانها

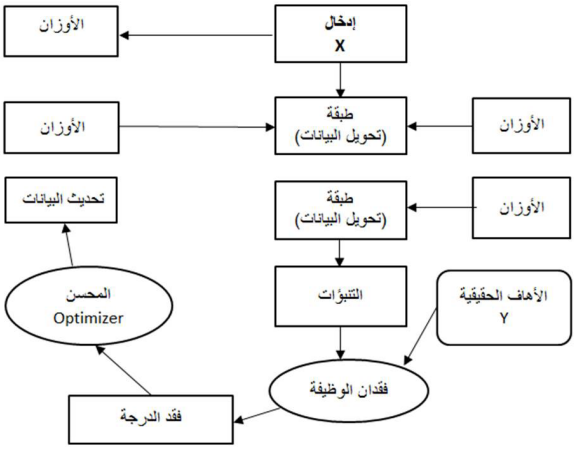
حيث أن الأوزان يطلق عليها معلمات Parameters الطبقة في بعض الأحيان. في هذا السياق، يعني التعلم إيجاد مجموعة من القيم لأوزان كل الطبقات في الشبكة، كتلك التي ستحدد خريطة مدي صحة أمثلة المدخلات لما يرتبط بها من أهداف. إلا أن الشبكة العصبية العميقة DNN التي يمكن أن تتضمن عشرات الملايين من المعلمات العددية لإيجاد القيمة الصحيحة لكل من منها قد تبدو مهمة شاقة خاصة بالنظر إلي قيمة المعلمات المعدلة التي قد تؤثر علي سلوك الآخرين.

كما أنه لرقابة أو عدم رقابة شيء ما، هناك حاجة لتوافر القدرة علي الملاحظة لرقابة ناتج الشبكة العصبية، مما يحتاج إلي توافر القدرة لقياس إلي أي مدي هذا الناتج هو المتوقع فعليا. ويمثل هذا ما يمثل «فقدان الشبكة» والهدف الحقيقي منها (ما يراد إخراجه من الشبكة) الذي يحسب درجة المسافة لالتقاط ما قامت به الشبكة بشكل جيد في هذا المثال كما يبينه الشكل التالي:



شكل (٤/٥) قياسات وظيفة الفقد لجودة الناتج

من الملاحظ أن الحيلة الأساسية في التعلم العميق تتمثل في استخدام هذه الدرجة النابعة من القياسات كإشارة تغذية مرتدة لضبط الأوزان قليلا في الاتجاه الذي سيقلل فقد الدرجة للمثال كما في الشكل التالي:



شكل (٥/٥) فقد الدرجة يستخدم كتغذية مرتدة

في البداية، أوزان الشبكة يتم تعيين قيما عشوائية لها، لذا يجب القيام بتنفيذ سلسلة من التحويلات العشوائية فقط. ومن الطبيعي، أن يكون ذلك ناتجا بعيدا عما يجب أن يكون عليه بشكل مثالي، حتى لا تصبح درجة الخسارة عالية جدا. لكن مع كل مثال لعمليات الشبكة، يتم تعديل الأوزان قليلا في الاتجاه الصحيح وانخفاض درجة الخسارة. ويمثل ذلك، حلقة التدريب التي تكرر عدد مرات كاف (في العادة عشرات التكرارات لألاف الأمثلة) مما يعطي قيم الوزن التي تقلل فقدان الوظيفة. حيث أن الشبكة المدربة مع فقد محدود تكون أهدافها أقرب ما يمكن للأهداف الحقيقية الخاصة بها. مرة أخري، يمثل ذلك آلية بسيطة التي بمجرد انتهائها يطلق عليها كالسحر.

٣ أبعاد تعلم الآلة: الأنواع، حالات التطبيقات، التحديات

١/٣ أنواع تعلم الآلة:

يمكن تحديد أربع أنواع رئيسة طورت لتعلم الآلة، كما يلي:

١) تعلم الآلة تحت الإشراف Supervised Machine Learning: يعتبر هذا النوع شكل

تعلم الآلة الأكثر انتشارا الذي يتضمن خوارزمية كميات بيانات تدريب مسماة أو معلمة ضخمة، وسؤالها عمل التنبؤات علي بيانات لم يسبق لها مثيل عن الارتباطات التي تتعلمها من البيانات المسماة أو المعلمة المتواجدة.

٢) التعلم غير الخاضع للرقابة أو الإشراف Unsupervised Machine Learning: غالبا يكون مستخدما في تطبيقات ذكاء اصطناعي أكبر، ويتضمن بيانات تدريب غير مسماه او غير معلمة. حيث تسألها خوارزمية لالتقاط الصلات بمفردها مهما كانت. هذا النوع من تعلم الآلة منتشر في تطبيقات التجمع Clustering (فعل اكتشاف المجموعات في البيانات المتاحة) والاتصال (تنبؤ القواعد التي تصف البيانات).

٣) التعلم شبه المراقب أو المشرف عليه Semi supervised Machine Learning: في هذا النوع يتم تدريب الخوارزميات علي مجموعة بيانات مسماة أو معلمة صغيرة. وعندئذ كما في التعلم غير الخاضع للمراقبة تطبق البيانات غير المسماة أو غير المعلمة المدخل المستخدم غالبا عند تواجد نقص في جودة البيانات المتاحة.

٤) التعلم المعزز Reinforcement Machine Learning: في هذا النوع تستلم الخوارزميات مجموعة تعليمات وإرشادات لكي تتخذ قراراتها بعدئذ عن كيف يمكن تداول مهمة أو نشاط خلال عملية المحاولة والخطأ. والقرارات سواء بمكافأة أو بعقاب تعتبر كأداة توجيه الذكاء الاصطناعي لعمل أفضل للمشكلة المثارة.

وقد طورت المنشآت مجموعة أساليب وتطبيقات مميزة من خلال هذه الأنواع الأربعة للتعلم الآلي التي تتعلق بكل شيئ من مبيعات بسيطة نسبيا إلي أقصي مدي من أدوات الذكاء الاصطناعي التي تشغل علي نماذج تعلم الآلة الحالية.

٢/٣ حالات تطبيقات تعلم الآلة الأكثر انتشارا وشيوعا:

حالات استخدام تطبيقات تعلم الآلة تنمو بسرعة كبيرة، حيث تستفيد منشآت الأعمال الحالية منها وتضمها في نماذج أعمالها الجديدة التي ساهمت التكنولوجيا الناشئة في أداء مهاما وأنشطة لم تكن محققة سابقا، لا لتوليد الكفاءات فقط لكن أيضا في مراحل إنشاء إعمالا جديدة. ويعتبر نمو استخدام حالات تطبيقات تعلم الآلة مهمة حرجة تنعكس علي مدي استخدامها الكثيف لكي تؤدي دورا مهما محتاج له.

وفيما يلي أمثلة لحالات تطبيقات تعلم الآلة الأكثر إنتشارا في المنشآت المختلفة:

١) محركات التوصية Recommendation Engines: من الشركات الرائدة المتاحة علي الخط التي تستخدم محركات توصية توجه المستهلكين الحصول علي المنتج الصحيح في الوقت الصحيح، كل من شركة أمازون عملاق تجارة التجزئة علي الخط التي تعتبر رائدة هذه التكنولوجيا من بداية الحقبة الحديثة من القرن الحادي والعشرين. ومنذ ذلك الوقت، أصبحت تكنولوجيا محركات التوصية موحدة قياسيا لمواقع التسوق المتاحة علي الخط. هذه الأدوات تشتمل علي تاريخ تصفح العملاء النهائي، وتضاهي الأفضليات المتواجدة بواسطة ذلك التاريخ لمنتجات وخدمات عملاء آخرين أدركوا مدي الاستفادة منها مسبقا.

٢) الكشف عن الغش والخداع Fraud Detection : حيث أن كثيرا من المعاملات المالية التي صارت متاحة علي الخط، إدي لزيادة فرص الغش والخداع تطبيقات اكتشاف الغش هذه تعمل من خلال تعلم خصائص المعاملات الشرعية المفروضة، ثم فحص المعاملات الواردة لتحديد مدي التزامها بهذه الخصائص أو انحرافها عنها،

بعدئذ تقوم بتمييز المعاملات بالقبول أو الرفض.

٣) تحليل العملاء Customer Analysis: معظم الأعمال تجمع مخازن هائلة من البيانات عن عملائهم الحاليين والمتوقعين، التي يطلق عليها حاليا المعلومات الضخمة Big Data التي تتضمن كل شيء عن العميل من تاريخ التصفح علي الخط، إلي نشاط الوسائط الاجتماعية الضخمة والمتنوعة للغاية، وجعل ذلك منطقيا ومفهوما للمستخدمين من تلقاء أنفسهم. وهذا يوضح من أين تعلم الآلة أتي بالفعل. كما أن الخوارزميات يمكن أن تسهل إنشاء ما صار يطلق عليه بحيرات البيانات Data Lakes التي تستخدم تكنولوجيا هادوب Hadoop الخاصة بالبيانات الضخمة لتخزين البيانات الخام المتعلقة بمنشأة الأعمال، وتطوير الرؤي والبصائر الملائمة عن العملاء. وفي نفس الوقت، يمكن تعلم الآلة تطوير استراتيجيات السوق الشخصية التي تستهدف العملاء الأفراد وإبلاغهم بهذه الاستراتيجيات من أجل تحسين خبراتهم.

٤) التجارة المالية Financial Trading : كان سوق المال الأمريكي الذي يطلق عليه Wall Street من أقدم من تنبؤا بتكنولوجيا تعلم الآلة. والسبب في ذلك واضح في عالم متسم بالمخاطر العالية التي تجعل مليارات الدولارات علي المحك. حيث تقدر خوارزميات تعلم الآلة فحص مجموعات بيانات تاريخية هائلة لإيجاد الاتجاهات، الارتباطات والأنماط في أداء المخزون مثلا، وعمل التنبؤات عن كيف بعض الأسهم المرجحة للأداء في المستقبل.

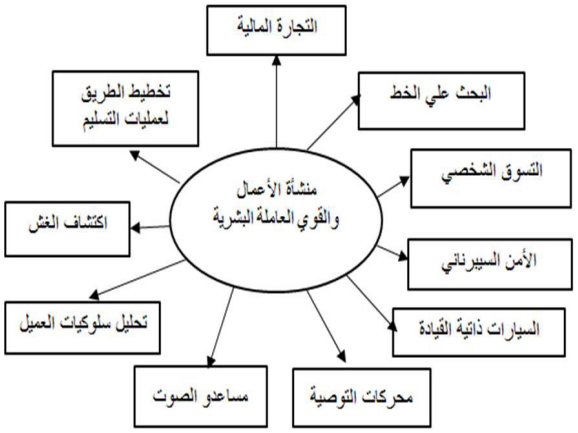
٥) المساعدون الافتراضيون Virtual Assistants: معظم الناس ملمون حاليا بمدي ما تقدمه المساعدون الافتراضيون من خلال شركات التكنولوجيا القائمة بالفعل مثل شركة أبل Apple وشركة جوجل Google من قدرات خلال المساعدين الافتراضيين، التي تمثل امتدادا لأي قوي تعلم الآلة لمستخدميها ويطلق عليها الروبوتات Bots. ويدخل تعلم الآلة في عدد من الطرق المختلفة التي تتضمن التعلم العميق الذي يوضح أسلوبا جديدا من تعلم الآلة المبني علي الشبكات العصبية NN ، ويؤدي دورا مهما في تطوير معالجة اللغة الطبيعية NLP الذي يوضح كيف أن الروبوتات قادرة للتفاعل مع المستخدمين في تعلم أفضلياتهم.

٦) السيارات ذاتية القيادة Self-Driving Cars : حالة التطبيق هذه تبين من أين تعلم الآلة يدخل عالم الذكاء الاصطناعي الذي يهدف لكي يكون علي قدم المساواة مع الذكاء البشري. وتستخدم السيارات ذاتية القيادة الاستقلالية الشبكات العصبة لتعلم تفسير الأشياء المكتشفة بواسطة كل من الكاميرات والمستشعرات الأخرى المتاحة لها من أجل تقرير أي فعل يتخذ لتحريك المركبة علي الطريق الممهد لها. في هذه الطريقة، يمكن لخوارزميات تعلم الآلة أن تستخدم البيانات للاقتراب من المحاكاة والتقليد التي تتماثل مع الإدراك البشري واتخاذ القرار الصائب.

ما سبق عرضه من حالات تطبيقات تعلم الآلة المستخدمة على نطاق واسع، يمثل ما تم عرضه عالية بعض الأمثلة فقط علي الرغم من تواجد الكثير من الحالات الأخرى غيرها. وأي عملية أعمال إما أن تنتج أو تستخدم كميات بيانات ضخمة وعلى وجه الخصوص البيانات المسماة أو المعلمة المهيكلة التي تتسم بالنضوج للآلية المستخدمة تعلم الآلة. وكل المنشآت عبر الصناعات والأعمال وخاصة في المجتمعات المتقدمة تعلمت هذا بالفعل، وصارت تعمل لتنفيذ طرق تعلم الآلة في عملياتها المختلفة.

والشكل التالي يوضح مدي استخدام تلك الطرق

في تطبيقات تعلم الآلة المنتشر استخدامها حاليا في كثير الشركات القائمة في الدول ذات الاقتصاديات المتقدمة:



شكل (٦/٥) حالات استخدام تطبيقات تعلم الآلة

ليس من الصعب رؤيـة لماذا تعلم الآلـة قد تم استخدامه في كثير من المواقف المختلفة التي أدت للمنشآت من تبني تعلم الآلة لحل مشكلات الأعمال بها وجني القيمة المرجوة من تطبيقات الذكاء الاصطناعي المستخدمة لها. وفيما يلي ستة فوائد أعمال أساسية:

- الإنتاجية المرتفعة،
- تكاليف عمل أقل،
- توقعات مالية أفضل،
- فهم أوضح للعملاء،
- مهام مكرر وروتينية أقل للعاملين، و
- مخرج أكثر تقدما وشبيه بشريا.

٣/٣ تحديات تعلم الآلة:

لم يعد السؤال ما إن يستخدم تعلم الآلة، بل كيف يمكن تفعيل تعلم الآلة في طرق تعود بأفضل النتائج الممكنة؟ هذا هو السبب في صعوبة التوجه لتعلم الآلة الذي يمثل تكنولوجيا معقدة تتطلب خبرة كبيرة. علي عكس بعض مجالات التكنولوجيا الأخـري، حيث أن البرمجيات تكون غالبا من أجل التوصيل والتشغيل المباشر، إلا أن تعلم الآلة يدفع المستخدم للتفكير عن لماذ يستخدمها، ومن يبني تلك الأدوات، وما افتراضاتهم، وكيف تطبق تكنولوجياتها. علما بأنه توجد تكنولوجيات أخري لها العديد من نقاط الفشل المرجحة لتطبيقها.

وفيما يلي مجموعة من المشكلات التي تواجه استخدام تعلم الآلة وتمثل تحديات لها، مما يحتم العمل علي حلها:

١) حالة الاستخدام الخطأ: التي تتمثل في تعطل وتوقف تطبيقات تعلم الآلة. وفي بعض الأحيان حيث أن المنشآت تؤدي أعمالها بالتكنولوجيا تبحث عن طرق لتنفيذ تعلم الآلة، بدلا من السماح للمشكل لكي يملي الحلول المتوقعة. وعند استخدام تعلم الآلة يكون مستخدما هامشيا، غالبا ما يفشل في توصيل النتائج المناسبة.

٢) البيانات الخطأ: تؤدي لتوقف نماذج تعلم الالة أسرع من أي شيئ آخر. حيث أن البيانات تمثل شريان حياة تعلم الآلة. حيث أن النماذج تعرف ما تعرضه النماذج فقط، لذلك عندما يتم التدريب علي بيانات غير دقيقة، غير منظمة، غير كاملة، أو متحيزة يكون مخرج النموذج خاطئ لحد كبير.

٣) التحيز: يعيق التحيز تنفيذ تعلم الآلة في كثير من الأحيان. وتوجد مجموعات تحيز عديدة التي ممكن أن تقوض تنفيذ تعلم الآلة الذي قد يحدث في مجموعتين في العادة: مجموعة نوع التحيز الأولي تحدث عندما لا تعكس البيانات المجمعة تدريب الخوارزمية للعالم الحقيقي، البيانات في هذا النوع تكون غير دقيقة، غير كاملة، وليست متنوعة بما فيه الكفاية؛ ومجموعة نوع التحيز الأخرى تنبثق من الطرق المستخدمة لأخذ عينة البيانات، تجميعها، تطبيقها وتعزيزها. في كل من

الحالتين، يمكن أن تنبثق الأخطاء من تحيزات علماء/خبراء البيانات في الإشراف علي التدريب والنتائج المنبثقة من النماذج الغير دقيقة والأسوأ تأثيرا بشكل غير عادل علي مجموعات معينة من المستخدمين.

٤) وظائف الصندوق الأسود Black BoxFunctionality: تمثل أحد أسباب لماذا ينتشر التحيز جدا في تعلم الآلة. وفي هذا الإطار، توجد أنواعا كثيرة من خوارزميات تعلم الآلة التي قد تكون غير خاضعة للرقابة بشكل خاص، وتشغل طرقا مبهمة. وكصندوق أسود للمطور، إن عالم /مهني البيانات يغذي بيانات الخوارزمية، التي تعمل ملاحظات تتعلق بالارتباطات وتنتج نوعا من الإخراج المبني علي هذه الملاحظات. لكن معظم النماذج لا تستطيع الشرح لعالم/خبير البيانات المختص لماذا تنتج المخرجات التي تعمل فعليا. وهذا ما يجعل الأمر صعبا لاكتشاف حالات التحيز أو حالات فشل النماذج الأخرى.

٥) التعقيد الفني: أحد التحديات الأكبر لاستخدام منشأة الأعمال التي تطبق تعلم الآلة ترتبط بالتعقيد الفني الضمني، الذي يمثل مفهوم تغذية بيانات التدريب الأساسي للخوارزمية وجعلها تتعلم خصائص مجموعة البيانات المتاحة. وقد يكون بسيطا بدرجة كافية. لكن يوجد كم هائل من التعقيد الفني المتواجد تحت الغطاء. حيث أن الخوارزميات تبني حول مفاهيم رياضية متقدمة، والكود الذي يشغل عليه الخوارزميات يمكن أن يكون صعبا للتعلم بدرجة كبيرة. كما أنه ليست لكل الأعمال الخبرة الفنية المحتاج لها من أجل تطوير تطبيقات تعلم الآلة الفعالة.

٦) عدم قابلية التعميم Lack of Generalizability: مما يؤدي لمنع تعلم الآلة من زيادة حجم حالات الاستخدام الجديدة. وتطبيقات تعلم الآلة تعرف فقط ما تم تدريبهم عليه بشكل صريح. وهذا يعني أن النموذج لا يمكن أخذ شيء ما تعلم عن مجال معين وتطبيقه لمجال آخر، كالطريقة التي في مقدرة البشر القيام بها. لذلك، تحتاج البرمجيات لكي تكون مدربة من البداية لكل حالة استخدام جديدة علي حدة.

إلي جانب ما سبق من مشكلات توضح مدي التحدي لمواجهتها والتغلب عليها، توجد أيضا ستة مشكلات إضافية تمثل تحديات تواجه منشآت الأعمال مع تعلم الآلة. حيث أن كثيرا من المنشآت بغية تحريك تحميلاتها من علم البيانات إلي السحابة الحاسوبية علي الإنترنت، تواجه هذه التحديات التي تحتاج فرق علم البيانات لمخاطبتها أثناء زيادة عمليات تعلم الآلة. وتتمثل هذه التحديات في التالي:

١) التنسيق بين واجهات التفاعل المتفرقة،

٢) حل التبعيات البيئية،

٣) تأكيد التعاون الوثيق بين كل منتفعي تعلم الآلة،

٤) بناء (أو إيجار) بنية تحتية مناسبة لتعلم الآلة،

٥) التوسع لتلبية متطلبات تعلم الآلة، و

٦) تمكين النشر السلس لتعلم الآلة.

وفيما يتعلق بالتحدي الأول المرتبط بالتنسيق بين واجهات التفاعل التي تواجه مهنيو البيانات يمكن أن ينبثق منها ثلاثة تحديات فرعية يجب مواجهتها والتغلب عليها، التي تتمثل في:

- التعاون بفعالية مع الزملاء،
- الربط مع قواعد البيانات أو مخازن البيانات،
- الربط مع محركات تنفيذ سعة النطاق المتطلبة.

أما التحدي الثاني المرتبط بتبعيات البيئة، إن كثيرا من تحديات علم البيانات تظهر عند تحرك البيانات أو الكود المستخدم بين بيئات مختلفة. وتدل البيئة فيما يخص علم الحاسب الآلي علي اعتمادات معينة

بين إصدارة نظام تشغيل، لغة البرمجة، ومكتبات بناء النماذج المستخدمة. وفي هذا السياق، حددت تحديثين خاصتين بواسطة الخبراء يرتبطان في اختناقات عمليات التطوير DevOps وقضايا قابلية تنقل الأكواد مما يحتم تجنبهما.

وفيما يتعلق بتحدي تأكيد التعاون الوثيق بين كل منتفعي تعلم الآلة، أي التحدي الثالث يمكن تحديده في أن واجهات التفاعل، البيئات، ونماذج تعلم الآلة هي مجرد أجزاء تشكل لغز التعلم الالي لمنشآت الأعمال المختلفة، وخاصة لكل من علماء بيانات المواطنين، محللي البيانات، مهنيي البيانات، ومهندسي تعلم الآلة بغية اتخاذ قرارات أعمال تتسم بالحكمة والرجاحة. ويستوجب هذا التعاون الوثيق بينهم جميعا مع مديري متخذي قرار الأعمال.

أما التحدي الرابع المتعلق ببناء (أو إيجار) بنية تحتية ملائمة لتعلم الآلة، يمكن ملاحظة أن تعلم الآلـة في منشأة الأعمال يتطلب مـوارد بشرية مؤهلة عاليا بشكل كبير عند بناء بنية تحتية قادرة على عمل تطبيقات تعلم الآلة قياسيا. مما يستوجب مراعاة إمـداد المورد بثلاثة عوامل وحـدات المعالجة المركزية CPUs ، وحدات المعالجة الرسومية GPUs، والتخزين. وقد قاد استخدام وحدات المعالة الرسومية بصفة معينة تسريع تدريب النموذج بدرجة كبيرة، لدرجة أن بعض الشركات صارت تبني أجهزتها لتعلم الآلة التي تتجاوز وحدات المعالجة المركزية الرسومية أيضا. وفي هذا الإطـار، صممت شركة جوجل نوع أجهزة أطلقت عليه وحدات معالجة الموتر (Tensor Processing Units (TPUs من أجل تدقيق تدفق الموتر المفترض أنه أسرع جدا من وحدات المعالجة الرسومية وغيرها، إلا أن الأغلبية العظمي من المنشآت التي تتعامل مع تعلم الآلة تقوم بإيجار وحدات المعالجة الرسومية المنتشرة بالفعل. كما أنه علي الرغنم من قيمة إدارة البنية التحتية المتطلبة مع وحدات المعالجة المركزية فإنها تمثل ترتيب حجم أكثر تعقيدا مع وحدات المعالجة الرسومية. كما أن تكلفة إدارة وحدات المعالجة الرسومية مرتفعة لحد كبير مع بطاقات الأسعار التي تصل لعشرات أو حتي مئات آلاف الدولارات في بعض الأحيان. علي الرغم أن هذا التحدي يجب أن يرتكز علي وحدات المعالجة الرسومية نظرا للسرعة العالية المتطلبة من معالم البنية التحتية/الأساسية المتطلبة لحد كبير والتي صارت شائعة لتدريب نماذج تعلم الآلة المتعلقة بكيفية إمدادها، إختيار إطار عمل ملائم، وإدارة تلك البنية التحتية بدون تكلفة باهظة.

أما التحدي الخامس المتصل بالتوسع لتلبية متطلبات تعلم الآلة والمرتبط بقابلية التوسع، فإن أطر عمل تعلم الآلة تؤدي حسابات رياضية معقدة لحد كبير. بالإضافة إلي أنه في كل منظمة لدي الموظفين بها مخططات أطر العمل التي تؤدي لآلية تدفق البيانات. وبذلك عندما تنمو البيانات، تنمو الآلية المستخدمة أيضا وتصبح أكثر تحديا.

وفي الحقيقة، عند النظر لأي مشروع برمجيات تكون إدارة تدفق البيانات من نقطة لأخري مهمة جدا لألية وتطوير تعظيم وتحسين كل مرحلة من مراحلها بطريقة أحسن. ومـع مشروعات علم البيانات هناك حاجة لإيجاد طريقة معينة لاستخدام وقت مهنيي البيانات أحسن لضمان أن أي نموذج نشر يقلل من معدل الأخطاء، لكن في نفس الوقت ، من المهم أيضا توافر عمليات قياسية لإدارة الأخطاء وتخفيف المخاطرة الت قد تحدث في الإنتاج. علي سبيل المثال، حتى يمكن إيجاد طريقة لوضع وظائف الاستخلاص، التحويل والتحميل ETL المتطلبة في معالجة البيانات المسبقة وتتعلق بتنقية/تنظيف البيانات، ووصلها معا، وإعدادها لكي تنمط وتتوحد، ويحسم جودتها لكي تتناسب مع تعلم الآلة.

أي أن قابلية توسع تعلم الآلـة يجب أن تتطلب التخطيط الجيد لها. وهذا يتطلب القدرة علي الحركة بسرعة، معالجة كميات بيانات ضخمة بسرعة، بناء نماذج جديدة لتعلم الآلة لعمل التنبؤات، التوصيات، أو إمداد تعلم الآلة المبني علي البصائر الصائبة، كما يمكن من مرونة قابلية التوسع التي يمكن أن تقدمها المنصات المحلية التي تمثل الحل الأكثر فعالية لتكلفة معالجة تحميلات البيانات الضخمة المتعامل معها. إلي جانب ما تقدم، هناك اعتبار آخر متمثل في الحاجة لاختيار إطار العمل الأكثر ملاءمة لاحتياجات منشأة الأعمال. علي الرغم من تغير مشهد إطار العمل بصفة مستمرة تبعا للتطورات والمتغيرات المجتمعية المتلاحقة. وفي نفس الوقت، تتوافر أطر عمل كثيرة مختلفة تتعامل مع طرق تدريب نماذج تعلم مختلفة أيضا. ومن وجهة نظر علم البيانات، كقضايا قابلية التوسع التي قد تأخذ أشكالا عديدة من بينها الأربع أشكال الأكثر شيوعا التالية:

- القدرة علي دعم علم البيانات علي نطاق واسع،
- تنفيذ المعلمات المفرطة الأقوى،
- مساندة نمو عدد المستخدمين والتطبيقات،
- إداء التدريب الموزع.

ويلاحظ أنه لا أي من تلك التحديات له حلولا سهلة. بل هناك حاجة ملحة لتخطيط قابلية التوسع فيما يتعلق بالأدوات التي تستخدم والعمليات الموضوعة لها. أي أن الحاجة لاعتبار كل ذلك من قبل علماء البيانات الذين يشكلون موارد باهظة الثمن يحتاج لتأكيد استخدام واستغلال وقتهم بشكل جيد.

أما التحدي السادس الـذي يواجه أيضا منشآت الأعمـال، فإنه يختص بإمكانية النشر السلس لمشروعات تعلم الآلة المتوج للتحديات الخمس السابقة بغية نشر منتجات تعلم الآلة علي نطاق واسع. حيث قد يجابه تحقيق حل هذا التحدي قضايا معينة سوف تتضمن هبوط حالات استخدام وأطر يجب أن تساندها. ومما هو محزن، أن عددا قليلا من مشروعات تعلم الالة لا ينجح تطبيقها بسبب بعض الحواجز والعراقيل التي تعوق حلول هذا التحدي ومنها علي سبل المثال لا الحصر ما يلي:

- الدفعة مقابل الوقت الحقيقي ونشر الحافة.
- التجريب مقابل الإنتاج.
- التدقيق المستمر كأمر بالغ الأهمية القصوى.

وبذلك عند مخاطبة وحل كل هذه التحديات من خلال التطرق لحلها ذاتيا من قبل المنشاة التي تسعي لتبني تعلم الالة كطريقة بناء، تعديل، أو إصلاح الأشياء بدون معاونة الخبراء أو المهنيين بصفة مباشرة. ومن المستحسن استخدام منصة تبني خصيصا لعلم بيانات المنشاة، مع العلم أن منصة سحابة البيانات تقدم أيضا إمكانيات النجاح الأحسن من خلال مخاطبة كل هذه التحديات السابق الإشارة لها.

٤. تنفيذ عملية تعلم الآلة:

١/٤ خطوات التنفيذ:

تنفيذ تعلم الآلة عملية متعددة الخطوات التي تتطلب إدخال البيانات من خلال خبراء عديدين. وفيما يلي تلخيص للعملية المتضمنة الستة خطوات الأساسية:

١) أي تنفيذ تعلم الآلـة يجب أن يبدأ **بتعريف المشكلة المثارة.** حيث أن مشروعات تعلم ألاله الأكثر فعالية تعالج تحديات أو فرص الأعمال المعينة المفسرة بوضوح.

٢) بعد مرحلة صياغة المشكلة، يجب علي فرق عمل علم البيانات **اختيار خوارزمياتهم**. حيث توجد خوارزميات تعلم الآلة كثيرة ومختلفة حتي يمكن أن يلائم ما يخار منها التطبيق المراد بطريقة

أحسن. ومن الملاحظ، ان خوارزميات الإنحدار الخطي Linear Regression البسيطة تعمل جيدا في كثير من حالات الاستخدام ، حيث يسعي المستخدم التنبؤ بمتغير غير معروف بناء علي متغير أخر معروف.

٣) كما أن أحدث خوارزميات التعلم العميق تكون أفضل في كثير من الأشياء المعقدة مثل التعرف علي الشكل، أو توليد النص. كما يوجد عشرات من أنواع الخوارزميات الأخري التي تغطي المسافة بين كل من هذه الأمثلة . أي أن الاختيار الصحيح للخوارزميات يعتبر ضروريا لنجاح مشروعات تعلم الآلة.

٤) بمجرد تعريف فريق علم البيانات المختص علي معالجة المشكلة المعينة ، واختيار خوارزمية ملائمة تصبح الخطوة التالية **جمع البيانات** المحتاج لها. وأهمية جمع نوع البيانات الصحيح الكافي غالبا تمثل شريان حياة تعلم الآلة المتطلب. حيث أنه في نفس الوقت، يمكن إمداد الخوارزميات بكل شيء ثمين معروف يمكن أن يفسر ما تسطيع القيام به. وجمع البيانات يتضمن مهاما معقدة مثل تعريف مخازن البيانات، كتابة النصوص لوصل قواعد البيانات مع تطبيقات تعلم الآلة، التحقق من البيانات، تنقية/تنظيف البيانات، تسمية أو تعليم البيانات، وتنظيمها في ملفات لكي يمكن عمل الخوارزمية عليها. وتعمل كل هذه المهام التيقد تكون صعبة ومعقدة، ولا يمكن التغاضي عن أهميتها.

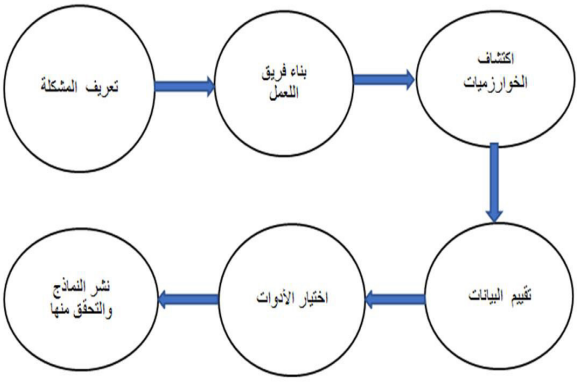
٥) في الخطوة الرابعة، يحين الوقت البدء سحر عملية تعلم الآلة. بحيث أنه بمجرد أن يحصل فريق علم علي كل البيانات التي يحتاجها، يمكن أن يبدأ في **بناء النموذج** المتطلب . هذه الخطوة في عملية تعلم الآلة سوف تختلف جوهريا عن الخطوات السابقة في أنها تعتمد علي ما إن كان الفريق مستخدما خوارزمية تعلم الآلة تحت الإشراف أو الخوارزمية غير الخاضعة للرقابة. أو غير ذلك . حيث أنه عندما يكون التدريب تحت الإشراف، يقوم الفريق بتغذية بيانات الخوارزمية واختبار خواصها للفحص. أما في في إطار مدخل التعلم غير الخاضع للرقابة، يعمل الفريق علي تحويل الخوارزمية الفضاضة علي البيانات إلي أن تنتج هالخوارزمية نموذج شبيه بالبيانات. المدخلة له.

٦) **تتصل الخطوة الخامسة بتطوير التطبيق** ذاته. حيث تطور الخوارزمية نموذج شبيه بالبيانات المغذاة، مما يمكن بناء التعلم المتطلب في التطبيق، الذي يخاطب تحدي أو فرصة الأعمال المعرفة في خطوة العملية الأولي المتعلقة بتعريف المشكلة. وفي بعض الأحيان، يكون هذا بسيطا جدا كما في لوحة عرض البيانات Data Dashboard التي تبين توقعات المبيعات المحدثة المبنية على تغيير الأوضاع الاقتصادية. كما قد تكون محرك توصية التي تعلم تفصيل الاقتراحات بناء علي سلوك العميل السابق، أو تكون مكون أحدث للبرمجيات الطبية التي تستخدم تكنولوجيا التعرف علي الشكل لاكتشاف الخلايا السرطانية مثلا في الأشكال الطبية خلال مرحلة التطوير، كما تمكن المهندسون في اختبار النموذج في مواجهة البيانات الواردة الجديدة للتأكد من أنها تتيح التنبؤات الدقيقة.

٧) علي الرغم من اكتمال عملية تعلم الآلة بشكل أساسي، إلا أن الوقت لا يكون متاحا للابتعاد عن النموذج. هذه الخطوة الأخيرة في عملية تعلم الآلة تعمل علي التحقق من صحة النموذج. لذلك يجب علي علماء/خبراء البيانات التحقق من أن تطبيقهم يمد التنبؤات الدقيقة علي أساس مستمر. وبذلك يصبح من المحتمل تواجد أسباب قليلة لعمل التغييرات المتطلبة، حيث أن أداء النموذج يتحلل بمرور الوقت عادة. ويرجع ذلك في أن الحقيقة الكامنة التي بني النموذج عليها سواء كانت اقتصادية أو متعلقة بميول وأفضليات العملاء

تؤدي للتحول مع مرور الوقت. عند حدوث هذا، يسوء أداء النموذج. عندئذ يبدا بعملية تعلم الآلة مرة أحري من البداية بشكل أساسي.

الشكل التالي يوضح خطوات عملية تعلم الالة هذه تجاه التطبيق الذي يتم في معظم الشركات:



شكل (٧/٥) خطوات عملية تعلم الآلة

٢/٤ إدارة وصيانة تعلم الآلة:

إدارة وصيانة تطبيقات تعلم الآلة في المنشأة تعبر عن منطقة تعطي فيها فترة قصيرة ومحدودة لها أحيانا، لكن يمكن أن تؤثر علي تفعيل أو فشل استخدام تعلم الآلـة. حيث أن وظيفة تغلم الآلة الأساسية تعتمد علي اتجاهات نماذج التعلم التي تمثل سلوك العملاء، أداء المخزون والطلب عليه، وإسقاطها علي المستقبل لإعلام القرارات. لذلك، تتغير الاتجاهات الأساسية باستمرار، إما أحيانا قليلة أو أحيانا كثيرة جوهرية. ويطلق علي هذا «مفهوم الانجراف» الذي إن لم يفسره علماء/ خبراء البيانات في نماذجهم في كثير من الأحيان. بالإضافة إلي أن توقعات النموذج قد تكون خارج القاعدة المعترف بها في النهاية. الطريقة المتبعة لتصحيح هذا الوضع هي عدم رؤية النماذج في الإنتاج بأنها منتهية بل تتطلب حالة تحقق ثابت، إعادة تدريب وإعادة عمل لتأكيد مدي استمراها لإمداد النتائج المحتاج لها.

- **التحقق Verification :** علماء/خبراء البيانات يحتفظون غالبا بجزء من البيانات الواردة الجديدة ثم التحقق من تنبؤات النماذج لتأكيد تأنها تقترب من البيانات الواردة الجديدة.
- **إعادة التدريب Retraining:** نتائج النموذج للانحراف/الانحدار من البيانات الملاحظة فعليا بشكل كبير يحان الوقت للاحتفاظ بها. كما يحتاج علماء/خبراء البيانات إلي مصدر مجموعة بيانات كاملة جديدة تعكس الأوضاع القائمة حاليا.
- **إعـادة البناء Rebuilding:** في بعض الأحيان، من المفترض أن مفهوم نموذج تعلم الآلـة يتنبأ أنه سيتغير كثيرا، بحيث تصبح الافتراضات الأساسية التي ذهبت للنموذج لم تعد صالحة. في هذه الحالات يصبح الوقت متاحا لإعادة بناء النموذج من البداية بصفة أساسية.

٣/٤ عمليات تشغيل تعلم الآلة:

عمليات تشغيل تعلم الآلة التي تتحدد باختصار في اللغة الإنجليزية في الاختصار MLops التي تمثل مفهوما ناشئا يستهدف دورة حياة عملية تعلم الآلة بنشاط. حيق أنه بدلا من المدخل العشوائي التقليدي للتحقق وإعادة البناء عند وضع أدوات عمليات تعلم الآلة موضع تنفيذ كل نموذج ووضع جدولا زمنيا للتطوير، النشر، التحقق وإعادة التدريب، السعي نحو توحيد العمليات بشكل قياسي، مما يمثل ممارسة صارت مهمة كثيرا، إلي جانب جعل تعلم الآلة محورا مكونا لعملياتها في الأساس.

٥. اتجاهات مستقبل تعلم الآلة وبائعيها وبرامجها المفتوحة المصدر:

١/٥ اتجاهات المستقبل:

عند النظر لمستقبل تعلم الآلـة، يسيطر اتجاه

واحد شامل، الذي سوف يستمر في زيادة جلب التكنولوجيا من أحدث الماركات للتيار الرئيسي المفضل لدي المنشآت. وما زال هذا الاتجاه متقدما علي قدم وساق حاليا بالفعل .

ومن خلال مسح صادر عن شركة جاتنر Gartner للاستشارة والبحث عام ٢٠١٩ وجد أن ٣٧٪ من المنشآت الممسوحة تبنت بعض أشكال الذكاء الاصطناعي، وقد مثل ارتفاعا بحوالي ١٠٪ في عام ٢٠١٥. وقد أصبح مسار تعلم الآلة حاليا مناسبا وصالحا تكنولوجيا في كل مكان في السنوات القليلة القادمة في ترتيبه لأعلي ١٠ سياسات وتحليلات من خلال الاتجاهات لعام ٢٠٢٠ الذي وضح الشركات الأسرع والمسؤولة المدعومة بالذكاء الاصطناعي مما يجعلها رائدة ومحتلة القمة. إلي جانب ذلك لاحظ جارتنر مدي أهمية تعلم الآلة وأساليب الذكاء الاصطناعي الأخري ذات الأهمية الحيوية في إمداد البصيرة التي ارتبطت بجائحة تفشي فيروس كورونا المستجد (كوفيد-١٩) الذي انتشر علي نطاق واسع في عالم اليوم ، كما تنبا أنه في عام ٢٠٢٤ سوف يتحول حوالي ٧٥٪ من المنظمات من القيام بالتجريب لمبادرات تطبيقات تعلم الآلة والذكاء الاصطناعي إلي التشغيل مباشرة. وكنتيجة لمعدلات تبني تعلم الآلة العالية في المنشآت، فإن سوق أدوات تعلم الآلة سوف ينمو بسرعة أيضا. كما أن أسواق تنبؤ تعلم الآلة سوف تنموا أيضا لما يقرب من تحقيق إيرادات من ٨، ٨ بليون دولار في عام ٢٠٢٢ أكثر مما كانت عليه في عام ٢١٧ المقدرة بحوالي ٤، ١ بليون دولار فقط.

وقد تكون أسباب هذا النمو في إيرادات سوق تعلم الآلة واضحة، حيث أن المنشآت الأكثر نجاحا في الوقت الحالي مثل شركات كل من جوجل ، أمازون، وأوبر Uber علي سبيل المثال لا الحصر تضع تطبيقات تعلم الآلة في مركز نماذج أعمالها. بدلا من رؤية تعلم الآلة مجرد وظيفة للحصول علي التكنولوجيا. وبذلك تستخدم الشركات الرائدة في الصناعة والأعمال تكنولوجيا تعلم الالة والذكاء الاصطناعي كأمر بالغ الأهمية للحفاظ علي ميزتها التنافسية.

من هذا المنطلق صار التقدم في في التعلم العميق منتشرا علي نطاق واسع الانتشار. والتعلم العميق كما سبق الإشارة له هو نوع من تعلم الالة مبني علي الشبكات العصبية الذي أدي دورا هائلا في جلب الذكاء الاصطناعي للصدارة في المنشآت المختلفة. حيث قد صار استخدام الشبكات العصبية شائعا في تطبيقات المنشآت في الوقت الحالي نسبيا. وتمكن أساليب التعمق العميق المتقدمة النماذج لعمل كل شيئ من التعرف علي الأشياء في الأشكال لإنشاء نص اللغة الطبيعي لأوصاف المنتج والتطبيقات الأخرى. في الوقت الحالي، يوجد عدد من أنواع الشبكات العصبية العديدة المختلفة التي صممت لأداء وظائف معينة. لذلك، يمكن فهم تفرد أنواع الخوارزميات المختلفة التي تمثل المفتاح لتحقيق أقصي استفادة منها.

ومن خلال تطبيقات تعلم الآلة والتعلم العميق سوف يمكن آلية كمية معرفة كبيرة كأمر لا مفر منه، كما أن بعض المجالات الابتكارية سوف تتم من خلال تطبيقات الذكاء الاصطناعي المدفوعة بهما. مما قد يطرح أسئلة عن مستقبل العمل في عالم تكون فيه الآلات ذات قدرة علي إدارة علاقات العملاء، اكتشاف أبعاد مرض السرطان من الأشكال الطبية، إجراء المراجعات القانونية، دفع حاويات الشحن عبر الدول، وإنتاج الأصول الابتكارية الجديدة. مما يطرح سؤال عن مستقبل دور العمالة البشرية في كل ذلك؟ ومن المرحظ أن أنصار الذكاء الاصطناعي يقتنعون أن الآلية سوف تحرر البشر

توجهه نحو متابعة الأنشطة الابتكارية أكثر من خلال القضاء علي المهام الروتينية والمكررة، بينما هناك آخرون يقلقلون من الحراك المتواصل للآلية الذي سوف يترك مساحة ضئيلة جدا لعمل العمالة البشرية، مما قد يؤدي للبطالة وما ينتج من ذلك من مشاكل اجتماعية واقتصادية كبيرة.

٣/٥ البائعون ومنصات تعلم الآلة:

تتجه كثير من المنشآت إلي ما ينشره البائعون والوكلاء عن أبعاد منتجات وخدمات أدوات وبرمجيات تعلم الآلة لكي يمكن التعرف علي كل جديد يؤدي لزيادة كفاءة أعمالهم. وتتميز مساحة تعلم الآلة بمنافسة قوية بين كل من أدوات المصدر المفتوحة المتاحة بدون مقابل، والبرمجيات المبنية والمساندة من البائعين ووكلائهم بمقابل مادي مكلف في كثير من الأحيان. وعلي الرغم من اختيار المنشأة سواء بشراء برمجيات وأدوات تعلم الآلة من البائعين، أو الحصول علي ما يتاح من خلال المصدر المفتوح، إلا أنه من الشائع استضافة التطبيقات المختلفة في بيئات الحوسبة السحابية التي توفر ما هو محتاج له من أجهزة وبرمجيات المقدمة للمنشآت بصفة مشتركة.

وفيما يلي قائمتين مختصرتين تعطيان نظرة عامة لكل من العروض المتاحة من بعض الشركات لبرمجياتها ؛ إلي جانب قائمة لبعض برمجيات المصدر المفتوح.

١) برمجيات وأدوات بعض الشركات التكنولوجية المطورة والمخصصة للتسويق التجاري:

- خدمات سحابة جوجل Google Cloud: مجموعة خدمات سحابية تتراوح من الإدخال والتشغيل حيث تقدم توصيل وتشغيل مكونات تطبيقات ونظم الذكاء الاصطناعي لأدوات علم البيانات.
- تعلم آلة برمجيات واتسون لشركة آيزبي.إم IBM Watson Machine Learning: تم إمدادها لسحابة شركة آي بي إم بحيث تسمح لعلماء/خبراء البيانات بناء تطبيقات تعلم الآلة والتدريب عليها ونشرها.
- إستديو تعلم آلة آزور لشركة مايكروسوفت Microsoft Azure Machine Learning Studio: تعتبر أداة تفاعل رسومية مع المستخدم، تساند بناء نماذج تعلم الآلة ونشرها علي سحابة مايكرو سوفت.
- منقب المنشأة لشركة ساس : SAS Enterprise Miners عرض تعلم الآلة من شركة ساس للتحليلات يرتكز علي بناء تطبيقات تعلم الآلة للمنشأة ويقوم بإنتاجها بسرعة.

٢) أدوات وبرمجيات المصدر المفتوح الحر بدون مقابل:

- أداة Caffe: إطار عمل تم تصميمه لمساندة نماذج التعلم العميق المرتبط بالشبكات العصبية.
- نماذج كود Scikit-Learn: نماذج أكواد لغة Python مفتوحة المصدر تسمح للمستخدمين من القيام بتحميلات تعلم الآلة التقليدية كما في تحليل الانحدار/الانحسار والتجمع بصفة خاصة.
- تدفق تنسور Tensor Flow: مستودع تعلم الآلة مبني علي مصدر مفتوح بواسطة شركة جوجل، من الشائع استخدامه لتطوير الشبكات العصبية.
- مكتبة تيانو Teano : التي صدرت أصلا في عام ٢٠٠٧ ويعتبر أحد مكتبات تعلم الآلة الأقدم

والأكثر موثوقة. وقد تم تطويرها وتحسينها لتشغيل الوظائف علي وحـدات المعالجة الرسومية GPUs التي في مقدرتها أن تنتج في تدريب خوارزمية سريعة.

- مكتبة تورش :Torch مكتبة تعلم آلي محسنة لتدريب الخوارزميات علي وحدات المعالجة الرسومية GPUs، وهي مبنية أصلا لتدريب الشبكات العصبية للتعلم العميق.

بشكل عـام، معظم مستخدمي آلات وبرمجيات تعلم الآلة يعتبرون أدوات المصدر المفتوح أكثر إبداعا وقوة. علي أي حال، لا يزال يوجد حالة قوية لأدوات الملكية الخاصة، حيث يقدم بائعوها التدريب والمساندة الفنية الغائبة من عروض المصدر المفتوح بشكل عام، مما يسمح للمستخدمين الحصول علي الخبرة والأمن والصيانة المستمرة.

٦.بيانات تدريب تعلم الآلة: الخصائص، الأغراض، الجودة والقضايا الكامنة:

البيانات هي محور مشروعات الذكاء الاصطناعي وتعلم الآلة، كما أنها أكثر من كود التطبيق. وتعتبر البيانات هامة وجوهرية في التدريب، الاختبار، التحقيق، ومساندة خوارزميات تعلم الآلة في محور نظم الذكاء الاصطناعي. وجزء من السبب لماذا تعلم الآلة وبالتالي الذكاء الاصطناعي ارتفعت شعبيتهما مرة أخري فإن ذلك يرجع لمزيج من الحوسبة السحابية التي لا حدود لها تقريبا، إلي جانب توافر البيانات الضخمة Big Data لتدريب نماذج تعلم الآلة وتطوير خوارزميات التعلم العميق المعتمدة كليا علي البيانات. حقيقة كلما كانت البيانات التي تغذي خوارزميات الذكاء الاصطناعي أكثر، كلما يتصير نتائج أداء تعلم الالة أكثر أهمية.

كما أنه من المعروف أن حلول تعلم الآلة تحتاج خوارزميات جيدة لتشغيلها، كما تدفع النماذج تعلم الآلة الجديدة نجاح الأعمال في أداء كل شيئ متطلب من التسوق الشخصي إلي الأمن الوطني. وتعتبر البيانات المكون الرئيسي لكل ذلك، بجانب الخوارزميات والنماذج المعدة لتعلم الآلة مما يسهم في جعل الآلة تفهم البيانات التي تتغذي بها.

علي أي حال، ليس كافيا الحصول علي قدر كبير من البيانات بدون جودة بيانات، حيث أن نظم الذكاء الاصطناعي تفشل لحد كبير بدون جودة بيانات. ويرجع جذور الكثير من فشل مشروع تعلم الآلة أن لديها القليل لتعلمه مع خوارزميات أو كود تعلم الآلة، أو حتى اختيار أي بائع تكنولوجيا معينة، حيث تعود المشكلات إلي جودة البيانات وكيفية إعداها بطريقة سليمة. وحتى يمكن لنماذج تعلم الآلة إتمام تدريبها بشكل صحيح بحيث تقدم النتائج الدقيقة والوثوق منها والمتوقعة، تحتاج البيانات أن تكون نقية/نظيفة، مسماة /معلمة، دقيقة وكاملة. إعداد البيانات هذا يعتبر مهما جدا لجودة بيانات تدريب تعلم الآلة. ومن الملاحظ علي نطاق واسع، أن معظم الوقت الذي تستغرقه الشركات علي مشروعات تعلم الآلة والذكاء الاصطناعي يتجه نحو جمع البيانات، تنقيتها/تنظيفها، إعدادها وتسميتها. علما أن المنشآت تحتاج لكي تستثمر أكثر في عملية إعداد البيانات المنبثقة أساسا من علم البيانات، والمشكلة لكل من نموذج التدريب، وأجزاء التشغيل الأخرى.

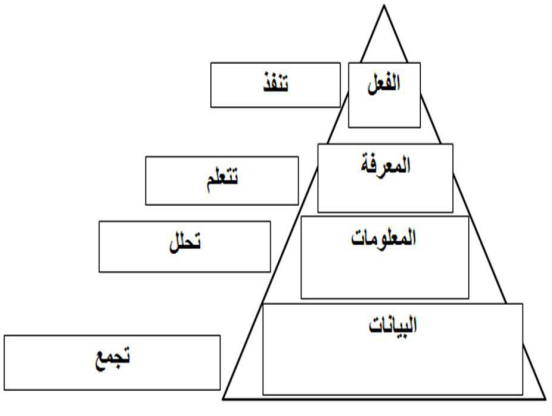
وقد أدي ذلك إلى نمو حقيقي في الطلب المتزايد للأدوات والخدمات التي تساعد في تحقيق جودة البيانات، إعدادها وتسميتها/تعليمها بطريقة سلسة.

والعرض التالي يوضح أبعاد خصائص تدريب كل من الخصائص، الأغراض، الجودة، والقضايا الكامنة:

١/٦ خصائص البيانات المهمة:

علم المعلومات يستخدم مفهوما يطلق عليه «هرم البيانات» لكشف العلاقة بين البيانات، المعلومات، والمعرفة. وبغض النظم قد تجمع البيانات الخام، وبالتبعية تنظم لتحولها لمعلومات لكي يمكن فهمها بشريا. هذه المعلومات يمكن تحويلها بعد ذلك إلي معرفة كتلك المجسدة بشريا بشكل عام. كما أنه في الإمكان التفكير في البيانات المخزنة في وحدات Bits، لكل من المعلومات المتضمنة في الوثائق، والمعرفة الضمنية بشريا أيضا، حيث توجد أسواقا مطورة جيدا وبيئات تنظيم المعلومات (من الكتب، المقالات، صفحات الويب، الموسيقي، الفيديوهات، الخ)، والمعرفة (من خلال أسواق العمل والمستشارين، والأكاديميين مثلا) التي تتاح جميعها لأسواق البيانات في إطار مجموعات من الوحدات Bits الغير منظمة والغير مطورة. ومن المرجح، أن يكون هذا بسبب طبيعة البيانات الخام التي تعتمد بشكل كبير علي السياق في الغالب، و التي قد تكون غير مفيدة في حد ذاتها بدون تواجد سياق محدد لها من أجل التحول لمعلومات مفهومة ومرشدة.

والشكل التالي يوضح معالم هرمية البيانات:



شكل (٨/٥) هرم البيانات

وفيما يتعلق بملكية البيانات والوصول لها، هناك مقولة أن البيانات بمثابة الزيت الجديد لكل الأعمال. بالتأكيد تحتاج البيانات أن تصقل لكي تكون مفيدة، إلا أنه لا يوجد تمييز هام لها. علي سبيل المثال، الزيت يمثل سلعة خاصة مستهلكة عند توافره واستهلاكه يقل وقد ينضب، لكن في حالة البيانات فإن استخدامها لا يقل ولا ينضب.

كما إنه بدلا من التركيز علي ملكية البيانات كمفهوم ملائم للسلع الخاصة، يجب التفكير في إمكانية الوصول للبيانات. وفي هذه الحالة قد تباع البيانات في النادر، كما تباع السلع بالفعل أيضا. إلا أنه بدلا من البيع قد ترخص البيانات لاستخدامات معينة وفقا لما يتطلبه مصدرها. وفي الوقت الحالي، تتوافر مناقشات عديدة عن تحديد سياسة من يملك بيانات المركبات ذاتية القيادة؟ كسؤال مطروح علي سبيل المثال يوضح من له الحق في الوصول لبيانات المركبات ذاتية القيادة والعمل بها؟ والإجابة علي هذا التساؤل توضح مدي التركيز علي كثير من الأبعاد الممكن الوصول لها لتحديد أبعاد ومكونات البيانات في الوقت الحقيقي. ومن وجهة نظر السلامة يصبح السماح لأطراف متعددة الوصول إليها؛ كما يمكن توضيح سهولة بعض نقاط جمع البيانات التي يجب أن تتوافر في المركبات المتعلقة بكل من: المحرك، نظام الإبحار، هواتف المحمول لدي الركاب، الخ. التي تمثل متطلب التفرد الذي يكون بدون سبب جيد لأدائه، والذي يصبح الحد غير الضروري لما يعمل مع البيانات.

بالإضافة لذلك، يصف روس أندرسون Ross Anderson ما سوف يحدث عند اصطدام الطائرات في الجو بدون تواجد مسارات مأمونة ومحددة لخط طيرانها. كما أن هناك أبعادا أخري يجب تضمينها وتكون ملائمة وممهدة لتعلم الآلة مثلا كتشييد الطرق والمسارات والمحددة للسيارات

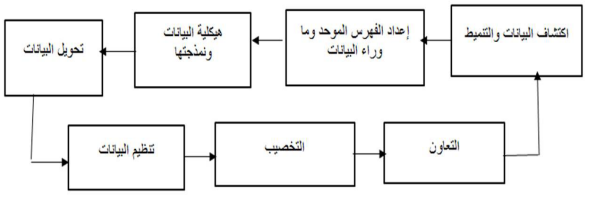
ذاتية القيادة، إلي جانب توافر أطرافا أخري تتضمن في مسؤولية عملها (Anderson, 1993). ويوضح ذلك ما تتطلبه آلية التعلم للسيارات ذاتية القيادة من توافر بعض أنواع المعلومات المحمية من خلال حقوق الملكية الفكرية وحقوق النشر المعمول بها في الوقت الحالي لدي كثير من دول العالم.

إلا أنه علي الرغم من هذا، قد يجمع مقدمو البيانات بعض البيانات و يقومون بعرضها وفقا للترخيص المعين المشروط لأطراف أخري معينة. علي سبيل المثال، توجد شركات بيانات متعددة تدمج بيانات التعداد السكاني للولايات المتحدة مع أنواع بيانات جغرافية/مكانية أخري، وتعرض مدي إمكانية ترخيص إتاحتها لأطراف أخري. هذه المعاملات قد تودي لمنع إعادة البيع أو الترخيص المتطلب للوصول والاستخدام. أي أنه لم تتواجد الملكية الفكرية القابلة للحماية المطلقة حتى الآن التي توضح في شروط التعاقد للحصول علي البيانات .وفقا لأي عقد خاص تحمية القوانين الم نظمة المعمول بها.

٢/٦ أغراض بيانات تدريب تعلم الآلة:

أحد أغراض إعداد بيانات تدريب تعلم الآلة الأصلية هو تاكيد أن المعلومات يجب الاستعداد لتحليلها بدقة وتنسيق ملائم، حتى تسهم في التوصل لنتائج ذكاء الأعمال Business Intelligence و تؤدي لتحليلات صالحة للاستخدام. وقد نشأت البيانات منذ القدم مع قيم قد تكون ناقصة ومفقودة وعديمة الدقة مع أخطاء أخري في غالب الأحيان.، بالإضافة إلي مجموعات البيانات المخزنة في ملفات متفرقة ومنفصلة، أو مجموعات بيانات لديها مميزات مختلفة محتاج لتسويتها وتهيئتها مقدما. أي أن إعداد البيانات يحتاج لعملية تصحيح وتنقية عدم دقة البيانات المتواجدة، والتحقق من صحة مجموعات البيانات الأصلية الصحيحة الموثوق مها لحد كبير.

وفيما يتصل بتطبيقات البيانات الضخمة، يعتبر التحقق من البيانات المتاحة مهمة جوهرية يمكن أدائها آليا بدلا من الاستغراق أعواما عديدة من خلال موظفي تكنولوجيا المعلومات، أو محللي البيانات في تصحيح كل حقل من كل ملف بيانات يقرر استخدامه في التحليل. وفي هذه الحالة، تسرع خوارزميات تعلم الآلة من كل ما يتطلب من الأشياء من خلال فحص حقول البيانات وملء القيم الناقصة تلقائيا، أو إعادة تسمية حقول بيانات معينة لتأكيد التناسق المتطلب عند وصل ملفات البيانات، والشكل التالي يوضح محور أوجه أدوات مهام إعداد البيانات المتطلبة لتوضيح أغراض البيانات:



شكل (٩/٥) مهام بيانات التدريب المتطلبة

أي يجب العمل علي تحديد كل من أغراض إمكانيات كل من وصول بيانات التدريب المناسبة، عيناتها، تنميطها، فهرستها، سماتها وخصائصها، وأنماطها لخوارزميات تعلم الآلة لكي تقوم بما هو مطلوب منها بشكل دقيق وصحيح.

٣/٦ جودة البيانات:

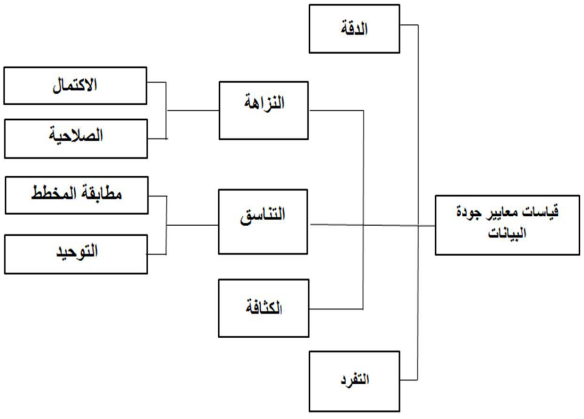
في مجال علم الحاسبات كل من التعرف علي النمط، استرجاع المعلومات، تعلم الآلة، تنقيب البيانات، وذكاء الويب يحتاج لإعداد جودة بيانته من خلال معالجة البيانات الخام المسبقة. وفي الممارسة الفعلية وجد أن تنقية/تنظيف البيانات وإعدادها

بصفة عامة يستغرق وقتا يتراوح تقريبا لحوالي ٨٠٪ من الجهد البشري المكرس لهندسة البيانات. وبذلك صار إعداد البيانات يمثل موضوعا بحثيا جوهريا يجب التطرق له في مبادرات ومشروعات تعلم الآلة والذكاء الاصطناعي المبنية علي توافر جودة البيانات حيث أن بيانات تشكل كل من نماذج تعلم الآلة وبيانات تدريب خوارزمياتها.

وقبل القيام بتحليل البيانات يجب تنظيم تلك البيانات في نسق وشكل ملائم . علما إن إعداد البيانات هو عملية معالجة وتنظيم البيانات الخام تكراريا قبل تحليها. وفي كثير من الأحيان تكون البيانات الخام غير مهيكلة ومشوشة غالبا، لذلك يجب العمل علي إعدادها في نسق وشكل مهيكل لكي يصبح جاهزا للتحليل الإضافي. وتتضمن عملية إعداد البيانات الكلية سلسلة مهام أو أنشطة رئيسية تتضمن إعداد سمات البيانات، تنقيتها، تكاملها، وتحويلها مما يحتم مراعة البيانات المتوافرة بالفعل وتقييم ملاءمتها للمهمة المحددة، مع دراسة جودة جمعها من مصادر البيانات الجديدة التي تتعلق بالمهمة المرغوبة. كما أنه من المهم أيضا تقييم ما إن كان هناك بيانات كافية تسهم في التوصل لناتج ومخرج تعلم الآلة بواقعية.

وبذلك، يمكن التحقق من جودة البيانات مع ملاحظة أن معظم مجموعات البيانات الخام المتاحة للتعلم الآلي تكون ذات جودة منخفضة غالبا وذلك نتيجة طبيعية، لأن معظم جامعي البيانات يدويا قد يكون لديهم قليل من الالتزام بافتراض دقة البيانات، كما يستخدمون اختصارات في إدخال ما جمعوه من بيانات. علي سبيل المثال، عندما يتم توفير القيم الافتراضية بواسطة نظام ما، تقدم بشكل كبير في البيانات المجمعة من عمليات جمع البيانات آليا الذي قد يعيبها ويؤدي لبيانات قد تكون غير صحيحة وغير دقيقة بسبب أن دقة أداة القياس المستخدمة تكون أقل من المتغير الحاصل بالفعل، أو قد تكون البيانات قديمة بحيث لم تعد صالحة للتوظيف الحالي.

وبذلك يعتبر افتراض وتحسين جودة البيانات سببا رئيسيا لمعالجة البيانات مسبقا. و وفي هذا السياق توجد مجموعة معايير مشتركة لقياس وتقييم جود البيانات الممكن جمعها في عنصرين رئيسيين هما: الدقة Accuracy والتفرد Uniqueness ((Muller & Freytag, 2005 كما يوضحه الشكل التالي:



شكل(١٠/٥) قياسات معايير جودة البيانات

وفيما يلي مفهوم كل من معيرر جودة بيانات رئيسي وما يتفرع منه من معايير فرعية:

الدقة Accuracy: وصفت الدقة كمعيار جـودة قيمة مجمعة تتضمن كل من النزاهة، التناسق، والكثافة. وتصف الدقة مدي تمثيل موحد وكامل للبيانات، مع الجوانب التي تصفها البيانات، وفيما يلي وصف كل معيار فرعي من معيار الدقة الرئيسي:

- **النزاهة Integrity:** تمثل النزاهة مجموعة متكاملة تشتمل علي تمثيلات كل المدخلات

التي تتطلب الاكتمال والصلاحية، كما يلي:

o **الاكتمال Completeness:** إكمال البيانات المحتاج لها تـؤدي لتمثيل شامل للموضوع أو الحدث المثار بحيث لا يتضمن قيما مفقودة لا تتواجد.

o **الصلاحية Validity:** تكون البيانات صالحة عندما لا يكون هناك قيودا تنتهكها. وتوجد آليات عديدة لزيادة صلاحية البيانات بما في ذلك حقول البيانات الإلزامية، فرص القيم الفريدة، ومخطط هيكل البيانات.

- **التناسق Consistency:** يختص معيار التناسق بالتخلص من أبعاد الشذوذ النحوي بالإضافة إلي التناقضات. ويتمثل التحدي ارئيسي لتناسق البيانات في اختيار مصادر البيانات الموثوق منها للتوافق بين البيانات عبر مصادر مختلفة.

o **مطابقة المخطط Schema Conformanal:** ويعتبر هذا المعيار حقيقيا لنظم قاعدة البيانات العلاقية ، حيث يعتمد علي الالتزام المتناسق بين كل أبعاد النظام .

o **التوحيد Uniformity:** يرتبط معيار التوحيد مباشرة باستبعاد كل معالم الشذوذ والاختلافات وتوحيد الشكل والمعني لكل كيان أو شيئ مقدم.

- **الكثافة Intensity:** هذا المعيار يخص مدي توافر قيم البيانات المفقودة، حيث قد تتواجد قيما مفقودة أو غير متوافرة الخصائص المعينة التي يجب تمثيلها وتعتبر غير معروفة.

هذه القيم الثلاثة التي تتعلق بكل من النزاهة، التناسق والكثافة تشكل جميعا قياس دقة البيانات.

بينما قياس جودة البيانات الثاني المتمثل في التفرد Uniqueness يجب أيضا الالتزام به ويستبعد أي تكرارات قد تتواجد في البيانات المتاحة.

كما أن معيار التوقيت Timeless يعتبر معيارا آخر لجودة البيانات ويشير لمدي حداثة Currency البيانات .

٧.عملية إعداد بيانات تدريب خوارزميات تعلم الآلة:

منذ بدأ في القيام بإدخال البيانات في الحاسبات الآلية كانت هناك رغبة ملحة لتحليل البيانات المدخلة. وقد أثبت الواقع أن ذلك يمثل مشكلة معقدة لمدخلي ومحللي البيانات علي حد سواء. علي الرغم من قدرة وإمكانية الحاسبات في التعامل مع جداول الأعداد وأداء العمليات الحسابية بسرعة كبيرة جدا. بينما كانت المشكلة المثارة في مدي تقرير مدي كفاءة الأعدداد المقدمة لتطبيق العمليات الحسابية، وكيفية التعامل مع عملية إعداد البيانات.

وفي هذا السياق، صممت قواعد البيانات المختلفة للاحتفاظ بالبيانات وخاصة لغراض التحليلات المتطلبة. إلا أنه وجدت أسباب جيدة لكي تحل ما كان يمثل صعوبة في الماضي لتطبيق البرمجة التحليلية مباشرة لنظم معاملات مصادر البيانات التي منها تنبثق مدخلاتها. وقد أدي ذلك لتحسين معاملات البيانات لتحميلات أعباء العمل المختلطة في قواعد البيانات، وصـولا لاستيعاب كميات بيانات كبيرة مطلوب تحليلها، مما قد يؤدي لبطيء عمليات نظم المعاملات الحرجة. هذا الوضع حتم ضرورة الاهتمام بعملية إعداد البيانات لتدريب خوارزميات تعلم الآلة، وفي نفس الأوقات حل

التحديات التي قد تعوق ذلك.

وتتلخص عملية إعداد البيانات في أنه بعد التحقق من البيانات والتوافق بينها، يمكن استخدام برمجيات وأدوات إعداد البيانات لتشغيل الملفات خلال تدفق العمل Workflow المرتبط بعمليات محددة سب تطبيقها للملفات. علي سبيل المثال، تتضمن هذه الخطوة إنشاء حقل جديد في ملف البيانات الذي يجمع الأغراض من الحقول المسبقة، أو تطبيق صيغة أو معادلة إحصائية، كما في حالة كل من النموذج الخطي أو الانحدار اللوجستي إلي البيانات. بعد الذهاب إلي تدفق عمل البيانات التي تمثل الإخراج في ملف نهائي كامل يحمل في قاعدة البيانات، أو مخزن بيانات آخر حيث يتوافر للتحليل.

وعلي الرغم من ذلك، صارت طرق إعداد البيانات آلية للغاية، حيث صار ممكنا ألا يستغرق أعداد البيانات وقتا طويلا كما كان من قبل، وخاصة عندما يكون حجم البيانات المستخدمة في التحليلات كبيرا ومستمرا في النمو. حيث أنه في الغالب، يشتكي علماء البيانات إستغراق معظم وقتهم الثمين في تحديد مكان تواجد البيانات، وتنقيتها بدلا من تحليلها في الواقع. وقد أدي ذلك جزئيا لازدياد عدد بائعي وموردي برمجيات عملية إعداد البيانات ومحاولتهم معالجة مشكلة إعداد البيانات والتغلب عليها في أدواتهم و برمجياتهم. كما تضع كثير من المنظمات موارد باهظة تجاه التزود بآلية برمجيات إعداد البيانات.

وأحد فوائد تأسيس عملية إعداد البيانات التنسيق الأكبر، حيث يمكن للمستخدمين استغراق وقتا ضئيلا للعثور علي ما يبحثون عنه من بيانات وهيكلتها.

١/٧ تحديات عملية إعداد البيانات:

تواجه كثير من منشآت الأعمال عددا كبيرا من المشكلات عند قيامها بإعداد نظم معاملتها التي قد تتواجد في صوامع أي مخازن بيانات متفرقة مستقلة بعضها عن بعض، ويستخدم كل منها أدوات وبرمجيات مختلفة. ويعني هذا أن ما يحتاج له من بيانات لسياق المعاملات كما في منفذ بيع بالتجزئة ويتعلق بمنتج معين، أو أي عميل تعامل مع هذا المنتج بالشراء، وتم تخزينه في البيانات محليا التي تخص ذلك النظام، قد يؤدي لتكاثر النظم، البيانات عن العملاء، المنتجات/السلع، المواقع، والأصول التي تتكرر وتصيح غير متناسقة معا.

والبيانات غير المتناسقة قد لا تكون مشكلة ضخمة لنظام لكل فرد في حد ذاته، لكنها ذات أهمية كبيرة عند الحاجة لرؤية البيانات عبر المنشأة ككل، كما في حالات حساب المبيعات الإجمالية بواسطة عميل أو منتج معين. وعند التطرق لعملية إعداد البيانات والأرقام المحللة لها، قد تصطدم المنشأة بعدم معرفة العلامات التجارية التي كانت مربحة علي أساس عالمي ومستخدمة لا مركزيا. علاوة علي ذلك، قد تكون جودة البيانات المتاحة مشكوكا فيها، حيث قد يرتكب مدخلوا البيانات يدويا أسماء وعناوين لبعض العملاء، واكواد وأسعار المنتج خطأ علي الرغم من الجهد الشاق المبذول في تصميم نظم إدخال البيانات والتحق من مصادرها العديدة المتعامل معها.

بالإضافة لذلك، قد تكتشف المنظمات الكبيرة بعد قيامها بممارسة تنقية/تنظيف البيانات، أنه بمجرد استبعاد جميع الإدخالات المكررة قد يؤدي ذلك تقليص عدد العملاء المساهمين والحد منهم لحد كبير. وقد يدعو التفكير في الجمع بين بيانات المنشأة الداخلية، البيانات الخارجية من الأطراف

الثالثة مما يسهم في تفاقم المضاعفات التي يمكن مواجهتها.

وقد أدي ذلك لبناء مستودعات البيانات Data Warehouses بواسطة إدارات أو مراكز تكنولوجيا المعلومات لحل القضايا المتعلقة بعملية إعداد البيانات المهيكلة، التي تغذي مستودعات البيانات من خلال نظم المصادر الصاعدة التي تتضمن نظم تخطيط موارد المنشأة ERP ونظم إدارة علاقات العملاء CRM وغيرها من نظم البيانات المهيكلة المتاحة لدي المنشأة التي تستخدم أدوات جودة البيانات لكي تنقي/تنظف تلك البيانات المحملة في مستودع البيانات غالبا. أي أن بناء مجموعة بيانات موثوق منها وموثقة تستخدم لتلبية رضا احتياجات التحليلات وإعداد التقارير الخاصة بالشركة.

مثل هذه المشروعات المتعلقة بإنشاء مستودعات البيانات، قد تعتبر مشروعات ذات نطاق واسع الانتشار وطموح للشركات لحد كبير. وتسهم مستودعات البيانات فى حفظ سلامة وتكامل البيانات عندما تواجه المنظمات تغييرات لعمليات الدمج والاستحواذ.

علي الرغم من أن قواعد البيانات التقليدية شغلت وطبقت جيدا عند إنشائها، إلا أن تغيير هيكل مخططاتها يتغير بمجرد تشغيل تمرين كبير عليها. مع أن تكييف تغذية البيانات في مستودع البيانات لاستيعاب عملية استحواذ جديدة للشركات، قد يستغرق مدة طويلة تصل لشهور عديدة أحيانا . وخلال كل ذلك لا يكون مستودع البيانات غير محدث وغير متاح بالكامل. إلا أنه عندما تحدث تغييرات كافية مع استحواذ الشركات لها، عندئذ يصبح مستودع البيانات ثابتا لحد ما وذلك علي الرغم من عدم قدرة العاملين الحفاظ علي المستودع الحالي بما فيه بدرجة كفاية. وقد يؤدي هذا أن يفقد محللو الأعمال الثقة في مستودع البيانات، ويبدأوا تحميل نسخ بياناتهم لجداول البيانات الإلكترونية Spreadsheets لإصدار تقارير الأعمال المتطلبة. بمجرد حدوث هذه الحلقة المفرغة يصبح مستودع البيانات أقل وثوقا ويستغرق المحللون وقتا طويلا في حفظ نسخ الظل الخاصة بهم من البيانات خارج وصول إدارة أو مركز تكنولوجيا المعلومات. علي الرغم من وجود أدوات إعداد البيانات والتحليلات أكثر ذكاء من أي وقت مضي لإنتاج مخططات رائعة.

ومعظم منشآت الأعمال التي ترغب في الأصل الاحتفاظ بنسخة موثقة واحدة من البيانات أدن لصعود تطوير أدوات إعداد البيانات كبرهان لحل ما تعانيه المنشاة من مشقة وجهد في إعدادها المسبق للبيانات حيث وفرت التكنولوجيا المتطورة وخاصة تعلم الآلة علاجا لأمراض البيانات السابق الإشارة لها، كما سوف يستعرض في البند التالي:

٢/٧ أدوات عملية إعداد البيانات وسوق تسويقها:

معظم بيانات المنشآت تعتبر غير جاهزة للاستخدام المباشر بواسطة تطبيقات تعلم الالة وتتطلب جهدا كبيرا في الإعداد. حيث أن أدوات إعداد البيانات لتعلم الآلة تحتاج للقدرة علي أداء قائمة مهام طويلة، تتضمن توحيد الصيغ والأشكال عبر مصادر البيانات المختلفة، استبعاد أو الإحلال محل بيانات غير صحيحة ومكررة، تأكيد دقة وحداثة البيانات، مساعدة وتعزيز البيانات المحتاج لها، تقليل ضجيج وتشوش البيانات، إصلاح أخطاء هوية البيانات، تطبيع البيانات، السماح لأخذ عينات البيانات المناسبة خاصة عند العمل مع أحجام بيانات كبيرة، والسماح لتطبيق هندسة الميزات والاستخلاص الجيد.

وقدد حدد تقرير صادر من شركة Cognilytica للاستشارات وبحوث الذكاء الاصطناعي أن أدوات إعداد البيانات المتصلة بالذكاء الاصطناعي تقدم طرقا تفاعلية تزامنية تسمح للمستخدمين رؤية تأثير أنشطة إعداد البيانات علي البيانات الضخمة بسرعة. وبعض ميزات هذه الأدوات تسمح باكتشاف الشذوذ في البيانات بسرعة، التعرف علي التكرارات واستبعادها، حال تعارضت البيانات، تطبيع أشكال وصيغ البيانات، إنشاء معالم خطوط توجيهية لاستخلاص البيانات وجمعها من مصادر متعددة، تعزيز البيانات بمزايا إضافية متطلبة للنماذج، إخفاء هوية البيانات عند الضرورة لبعض التطبيقات (Cognilytica, 2019).

كما استنتجت تلك الشركة أن أكثر من ٨٠٪ من وقت الشركات المستغرق في مشروعات الذكاء الاصطناعي يتجه نحو إعداد، تنقية وتسمية البيانات. و كثيرا من الخطوات التي تتضمن في الإعداد والتسمية ترتبط بجمع، وغربلة،و تنقية/ تنظيف، إزالة الأخطاء، بناء النموذج، جوانب نشر البيانات. وفي هذا الإطار يحتل كل من تنقية/ تنظيف البيانات وتسميتها ٥٠ حوالي ٪ من الوقت المخصص لكل هذه المهام.

وقد ظهر في سوق تعلم الآلة فئة جديدة من أدوات إعداد البيانات مبنية لإدارة مجموعات البينات الكبيرة الضخمة وتنميتها لمخاطبة مشكلات مشروعات تعلم الآلة. وقيمت شركة Cognilytica أن إيرادات سوق أدوات أعداد البيانات المرتكزة علي الذكاء الاصطناعي يصل بأكثر من ٥٠٠ مليون دولار أمريكي المتوقع تضاعفها لأكثر من ٢، ١ بليون دولار في نهاية ٢٠٢٣ (Cognilytica, 2019).

وفي الماضي اعتمدت منشآت الأعمال علي مجموعة أدوات معروفة يطلق عليهاETL (Extract, Transform and Load)) أي استخلاص، تحويل وتحميل لتحريك وانتقال البيانات في داخل وخارج مستودعات البيانات Data Warehouses للقيام بكثير من المهام التي تتعلق بكل من: تسهيل التحليلات، إعداد التقارير، ذكاء الأعمال، وكثير من العمليات الأخرى المتطلبة. أما في البيئات المبنية علي الحوسبة السحابية Cloud Computing الموجهة في الأساس نحو نمو البيانات الضخمة الجديدة، صار تحريك وانتقال البيانات داخل وخارج مستودعات البيانات الذي كان يتم باستخدام أدوات ETL غير موائم وغير صالح للاستخدام علي منصات السحابة الحديثة.

وقد أدي ذلك لاتجاه الشركات العمل مع البيانات في كل ما يخصها في الأساس من خلال إنشاء بحيرات البيانات Data Lakes التي بدلا من من استخدام أدوات ETL المتعاملة مع مستودعات البيانات التقليدية الطابع صارت تتطلع لاستخدام أدوات استخلاص المعلومات عند طلبها من مصدر البيانات وتحويلها بمجرد الاستخلاص والتحميل بعدئذ. هذا النمط يمثل أكثر من طريقة أدوات ETL التقليدية مما أدي للتحول لطريقة ELT أي استخلاص، تحميل وتحويل. وقد نفذت كثير من منشآت الأعمال في المجتمعات المتقدمة إنشاء بحيرات البيانات المبنية حول مخازن بيانات تكنولوجيا هادوب Hadoop غالبا، التي فيها تخزن كميات كبيرة من البيانات شبه المهيكلة وغير المهيكلة. وعند حاجة علماء البيانات أو محللي البيانات مجموعة بيانات للتحليل عليهم البحث عن البيانات أولا في بحيرات البيانات المتاحة. ومع عملية إعداد بيانات تنسيق البيانات

في المكان المناسب (بحيرة البيانات) يمكن تغذية البيانات تلقائيا والقيام بالتحليلات المتكررة بدلا من الطلب لإعداد البيانات وتنقيتها كل مرة كما في الطرق التقليدية الأخرى التي ما زال معمولا بها علي طاق واسع في مستودعات البيانات.

كما صارت أدوات إعداد البيانات موجهة نحو الخدمة الذاتية ومبنية علي السحابة ، مما أدي إلي أن يطبق مستخدمي منشأة الأعمال تلك الأدوات ذاتية الخدمة والاعتراف بأهميتها في استراتيجيات ذكاء الأعمال BI ، التي صارت تعتبر ذات أهمية أولوية عالية لدي مستخدميها والمستجيبين لها الذين هم في الأساس مهنيو تكنولوجيا المعلومات، حيث أن ذكاء الأعمال أو وظيفة الإدارة التنفيذية إلي جانب إعداد البيانات المرئية التي تعتبر كمهمة أكثر ألفة كما في حالة الحوسبة السحابية، البيانات الضخمة وإنترنت الأشياء IoT. وقد أدي ذلك لتنوع مستخدمين أكثر، خاصة مستخدمو الأعمال المتسمين بالذكاء والمهارة في التعامل مع التكنولوجيا الحديثة الذين صاروا يستخدمون الخدمات الذاتية وأدوات إعداد البيانات لمدخلات تعلم الآلة والذكاء الاصطناعي.

وعلي الرغم من توافر أدوات إعداد وتسمية البيانات قد صار متاحا آليا، إلا أن حماس المستخدمين لها لا يزال محدودا ومنخفضا لحد ما وخاصة في منشآت الدول النامية التي من بينها مصر، علي الرغم من أن بعض المؤسسات في الدول المتقدمة كما في حالة شركة Ericson لا تفكر في هذا التوجه. وقد يشير ذلك لخطة فاصلة لإعداد البيانات. كما ظهرت أولويات أخري في الصدارة التي تم استيعابها وأصبحت أكثر نضجا. وكل ذلك يعني أن إعداد البيانات يمكن أن يصبح سلعيا وتجاريا علي الرغم من أنه ما زال في النشوء.

كما أن معماريات البيانات وأنواع البيانات الجديدة وردت بالفعل إلي السوق التي استوعبتها وتعاملت معها، مما يتطلب أهمية مراقبة ما سوف تتطلبه الاحتياجات المتطورة علي الدوام.

وصارت سوق أدوات إعداد البيانات الناشئة حديثا تساعد في مشاهدة عروض خدمات أدوات إعداد بيانات مبنية علي السحابة أكثر من الخدمات التي كانت متاحة في السابق ومرتكزة في مواقع هذه المنشآت التي صار الكثير منها يمتلك تلك الأدوات علي الرغم من هجرتهم الظاهرة نحو ما توفره خدمات السحابة. إلا أن الطلبات التي ترد لسوق إعداد البيانات تعتمد في كثير من الأحيان علي تلك الخدمات المبنية علي السحابة أكثر. كما صار كثير من بائعي/مقدمي أدوات إعداد و تحليل البيانات المتاحة المتواجدين في السوق يقدمون عروضا لها مبنية علي السحابة ومن بينهم شركات مثل Meissa Data, Trifacta, Paxata, etc التي تعمل من منظور إفتراض أن البيانات موجودة في أشكال / تنسيقات مختلفة خلال المنظمة. وقد أصدرت شركة Tableau في عام ٢٠١٧ عرض مطور لنتائج تحليلات النتائج المرئيةTableau في إطار الخدمة الذاتية باستخدام طرق تعلم الآلة الذي صار بعد ذلك يمثل سوقا رائجة لإعداد البيانات. كما صارت شركة Tableau تتنافس مع شركات أخري مثل شركات ,Trifacta Tibco Software, MS.Power BI, and Unifi Software التي صارت جميعها تقدم القدرات المبنية علي السحابة والخدمة الذاتية في موقع المنشآت، إلي جانب شركة Datamer التي تقدم برمجيات إعداد البيانات ايضا في موقع المنشأة.

٨. الخلاصة:

تعلم الآلة هو أحد المجلات الأسرع نموا من غلم الحاسب الآلي والذكاء الاصطناعي مع تطبيقات بعيدة المدي. مصطلح تعلم الآلة يشير إلي اكتشاف الأنماط الآلية ذات المعني في البيانات المتاحة. وفي العقدين الآخرين أصبح تعلم الآلة أداة شائعة ومشتركة في أي مهمة تقريبا تتطلب استخلاص معلومات من مجموعات بيانات كبيرة. وفي الوقت الحالي، صارت التكنولوجيا المبنية علي تعلم الآلة تحيط بنا من خلال: محركات البحث التي تتعلم كيف تمدنا بالنتائج الأحسن؛ البرمجيات ضد الرسائل غير المرغوب فيها Span التي تتعلم غربلة رسائل البريد الإلكتروني؛ معاملات بطاقات الائتمان تؤمن بواسطة البرمجيات التي تتعلم كيف تكتشف الغش والخداع؛ الكاميرات الرقمية التي تتعلم اكتشاف الأوجه؛ وتطبيقات المساعدين الأذكياء الشخصية علي الهواتف الذكية التي تتعلم التعرف علي الأوامر الصوتية. كما أن السيارات صارت مجهزة بنظم منع الحوادث المبنية علي خوارزميات تعلم الآلة المستخدمة. وكذلك صار تعلم الآلة مستخدما علي نطاق واسع في التطبيقات العلمية المختلفة ميل المعلوماتية الحيوية، الطب والفلك، الصناعة التعليم وخلافه.

أي أن تعلم الآلة يمكن أن يمكن أشكال التحليلات التنبؤية وتضمين الذكاء المدفوع بالخوارزميات في تطبيقات البرمجيات، إلا انه لا شيء من ذلك يكون ممكنا بدون توافر بيانات ضخمة دقيقة وموثوق منها ملتقطة ومعالجة بطريقة صحيحة. وتستهلك وتعالج خوارزميات تعلم الآلة أحجام بيانات ضخمة لتعلم الأنماط عن الناس، عمليات الأعمال، المعاملات، الأحداث، الأشياء وهكذا. هذا الذكاء يتضمن بعدئذ في نموذج تنبؤي. ويمكن المقارنات للنموذج أن تكتشف ما إن كان الكيان يشغل مع أنماط مقبولة أو يعرض شذوذا.

وحالي، تعلم الآلة وما يصاحبه من إعداد بيانات يستخدم لحل مهام يحددها بشكل جيد مثل التصنيف والتجميع. ويمكن ملاحظة أن خوارزمية تعلم الآلة تتعلم مما يطلق عليه بيانات التدريب خلال عملية التطوير، كما أنها أيضا تتعلم من نشر بيانات العالم الحقيقي بصفة مستمرة مما يمكن الخوارزمية تحسين نموذجها مع الخبرات المكتسبة. وبذلك لتعلم الآلة شهية لاذعة للبيانات خلال كل من التطوير والإنتاج، وتقديم طلبات بنية تحتية لإدارة بيانات المنظمة. ما يجعل متطلبات تعلم الآلة الناجحة متمثلة في كل من مجموعة بيانات متنوعة ضخمة، وبنية تحتية لإدارة البيانات والمتنوعة والضخمة.

المراجع:

1. Appen (2020). The State of AI and Machine Learning: AI Industry Gains Traction with Increased Executive Visibility and Budgets, An Appen White-paper https://appen.com/wp-content/uploads/2020/06/whitepaper-state-of-ai-2020-final.pdf

2. Chollet, Francois (2018). Deep Learning with Python. Shelter Island, NY: Manning Publication Co.

3. Cinquegrana, Piero & Raza,

ploy-a-machine-learning-model.

9. Gartner (Feb. 2020). Magic Quadrant for Data Science and Machine Learning Platforms. https://www.gartner/com/en/

10. Gartner (2019). "Gartner Survey 37 Percent of Organizations Have Implemented AI in Some Form," https://www.gartner/com/en/newsroom/press-release/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have.

11. Kubat, Miroslav (2015). An Introduction to Machine Learning, 2nd ed. Springer International Publishing

12. Lorica, Ben & Nathan, Paco (2018). AI Adoption in the Enterprise: How Companies are Planning and Prioritizing AI Projects in Practice. O'Reilly Media

13. Muller, Heiko & Freytag, Johan (2005). Problems, Methods and Challenges in Comprehensive Data Cleansing. Professren des Inst. Fur Informatik

14. Nevala, Kimberly (2017). The Machine Learning Primer. Cary, NC: SAS Institute Inc. (Executive Series)

Matheen (2019). Machine Learning at Enterprise Scale: How Real Practitioners Handle Six Common Challenges. Sebastopol, CA: O'Reily Media Inc.

4. Cognilytica (2019). DATA PREP and Labeling for AI. https://www.cognilytica.com/wp-content/uploads/2019/DATA-PREP-AND-LABELING-for-AI.pdf

5. Deloitte Insights (2017). State of AI in the Enterprise, 2nd ed. – Early Adopters Combine Bullish Enterprise with Strategic Investments. Deloitte . https://www2.deloitte.com/content/dam/insights/us/articles/4780_state-of-AI-in-the-enterprise/DI_state-of-AI-in-the-enterprise-2nd-ed./pdf

6. Deloitte Insights (2020). Thriving in the Era of Pervasive AI: Deloitte's State of AI in the Enterprise, 3rd ed. (Deloitte Insights) https://www.deloitte.vom/us/Allinstitute.

7. Figure Eight "The State of AI and Machine Learning," https://www.figure-eight.com/the-state-of-AI-and-machine-learning.

8. Figure Eight "How to build, Train, Test and Deploy a Machine Learning Model,". https://www.figure-eight.com/how-to-build-train-test-and-de-