مجلة جامعة الملك عبد العزيز: علوم وتقنية المعلومات، م ١٠ ع١، ص ص: ٩٧ – ١٢١ (١٤٤٢هـ /٢٠٢١م)
Doi: 10.4197/Comp.10-1.5

تحليل الأداء الأكاديمي للطلاب باستخدام تقنيات تعلم الآلة

محمد عاطف موسي[١]

قطاع تقنية المعلومات، معهد الإدارة العامة، الرياض، المملكة العربية السعودية

mosamo@ipa.edu.sa

المستخلص. يعتبر التنبؤ بأداء الطلاب الأكاديمى من الأشياء الحيوية لنجاح أي نظام تعليمي. كما يعتبر جذب الطلاب وطريقة الحفاظ عليهم والارتقاء بهم جزء أساسيا من ذلك النظام. التنبؤ بأداء الطلاب الأكاديمي واحد من أهم المؤثرات على التصنيف الجامعي، والسمعة التعليمية للمؤسسة، والدعم المالى لها. أصبح مراقبة الأداء الطلابى وكيفية المحافظة عليهم والإرتقاء بهم أحد أهم الأولويات لصناع القرار في مؤسسات التعليم العالي. تبدأ منظومة التحسين وكيفية الحفاظ على الطلاب بفهم شامل للأسباب الكامنة وراء تسربهم وجنوحهم عن العملية التعليمية. مثل هذا الفهم هو أساس التنبؤ الدقيق بالطلاب المعرضين للخطر ومن ثم يكون التدخل المناسب لضمان الإبقاء عليهم. تستخدم العديد من تقنيات التنقيب فى البيانات مثل التجميع، التصنيف، الانحدار، والتنبؤ لمساعدة صناع القرار فى انجاز مهامهم. في هذه الدراسة، تم تقديم نموذج التنبؤ بأداء الطالب الجدد بالاعتماد على الميزات التي لها تأثير كبير على التحصيل الدراسي له. وقد اعتمدت هذه الدراسة في آلية عملها على بعض خوارزميات تعلم الآلة باستخدام بيانات تم جمعها من نظام التعليم الإلكتروني Kalboard 360. حقق النظام الخبير بعد عدد من التجارب العملية وتحليلها ومقارنة عدد من خوارزميات تعلم الآلة دقة تقدر بـ ٨٥,٥٪.

كلمات مفتاحية: التنبؤ، التنقيب في البيانات التعليمية للطلاب، تعلم الآلة، الذكاء الاصطناعي، خوارزم الغابة العشوائية.

١) العنوان الدائم: مركز تجميع واختبار الأقمار الصناعية، وكالة الفضاء المصرية، القاهرة، مصر.

١- المقدمة

أصبح تناقص الطلاب أحد أكبر المشكلات التى باتت تؤرق صناع القرار في المؤسسات الأكاديمية، على الرغم من كل تلك البرامج والخدمات المقدمة من أجل الحفاظ على الطلاب. وفقا لوزارة التعليم الأمريكية، مركز الإحصاء التربوي (nces.ed.gov)، ما يقرب من نصف أولئك الذين يلتحقون بالتعليم العالي هم فقط من يحصلون على درجة البكالوريوس. وعادة ما يؤدي ارتفاع معدل التسرب الطلابى إلى خسارة مالية شاملة، وانخفاض معدلات التخرج، وتدنى سمعة الجامعة أو المؤسسة التعليمية (Gansemer-Topf & Schuh, 2006). يبحث المشرعون وواضعو السياسات الذين يشرفون على التعليم العالي والمسؤولون عن تخصيص الموازنات لهذا الأمر، بل والآباء الذين يدفعون مقابل تعليم أبنائهم من أجل إعدادهم لمستقبل أفضل، والطلاب أنفسهم عن المؤسسات التعليمية ذات الجودة العالية والسمعة الحسنة. ولهذا السبب، فقد أصبحت أولوية قصوى لمسؤولي الكليات والجامعات في المملكة العربية السعودية وغيرها من البلدان المتقدمة حول العالم إدارة عمليات التسجيل وآليات الإبقاء على الطلاب. كي تكون ناجحًا كمؤسسة تعليمية، يجب أن تكون قادرا على تحديد الطلاب المعرضين لخطر التسرب بدقة وكيفية مجابهة الأمر بحكمة وأدوات متطورة. حتى الآن، تم تكريس عددا لا بأس به من الأبحاث التي تدرس تلك الظاهرة (Veenstra, 2009, Romero & Ventura, 2007, Romero & Ventura, 2010, Romero & Ventura, 2008, Miller & Herreid, 2010).

في مجال التنقيب عن البيانات وتطبيقاته الكثيرة لاكتشاف المعارف الدفينة، واحدة من أهم حقول المعرفة هى التنقيب فى بيانات الطلاب الأكاديمية لمساعدة أنظمة التسجيل وأنظمة القبول والتي تساعد الطلاب في كل مرحلة من مراحل دراستهم. في هذه الدراسة تم الحصول على مجموعة البيانات من نظام التعلم الإلكتروني Kalboard 360. وهي عبارة عن مجموعة بيانات تعليمية لموقع التعليم الإلكتروني، وتحتوي مجموعة البيانات على ٥٠٠ سجل ولديها ١٧ سمة/خاصية (بيانات خاصة بالطلاب). بعد ذلك، تم عمل تهيئة وتجهيز للبيانات وتطبيق أربعة من أشهر خوارزميات تعلم الآلة والتنقيب في البيانات وهي: ١) شجرة القرارات، ٢) الغابة العشوائية، ٣) نايف بايز، ٤) والشبكة العصبونية وذلك لتصنيف آدا الطلاب الأكاديمي بناء على خصائصهم. أخيرا، تم تقييم النتائج بمقارنة جميع الخوارزميات باستخدام مقاييس التشتت المختلفة.

تم تنظيم بقية الدراسة على النحو التالى: الفقرة القادمة تخص الدراسات السابقة. والفقرة الثالثة، نناقش فيها المنهجية المقترحة للنظام الخبير. والفقرة الرابعة نناقش فيها النتائج المتعلقة بالخوارزميات الأربع ومقارناتهم. الفقرة الخامسة خاتمة والعمل المستقبلى.

١-١ الدراسات السابقة

على الرغم من الارتفاع المستمر في معدلات الالتحاق بالمؤسسات الأكاديمية فى معظم دول العالم، إلا أن ارتفاع معدلات التسرب لا تزال مستمرة بين الطلاب الجامعيين. بالنسبة للمؤسسات الأكاديمية، تؤدي معدلات التسرب الطلابى العالية إلى اضافة أعباء أخرى على الجهود المبذولة لجذب طلاب جدد. بالنسبة للطلاب، فإن التسرب قبل حصولهم على شهادة جامعية بمثابة عقبة لهم وعائدا منخفضا على استثماراتهم المستقبلية (Mannan, 2007). يشير الأداء الأكاديمي الضعيف غالبا إلى وجود صعوبات كبيرة في التكيف مع الكلية أو الجامعة ونظامها مما يجعل التسرب ملازا مرجحا في نهاية الأمر (Lau, 2003). اصطلاحيا، تم تعريف تناقص الطلاب في الجامعة على أنه عدد الطلاب الذين لا يحصلون على شهادة في تلك المؤسسة. أظهرت الدراسات أن الغالبية العظمى من الطلاب ينسحبون خلال عامهم الأول (Deberard *et al*., 2004 & Hermaniwicz, 2003). عمل الباحثون على التطوير والتحقق من صحة النماذج النظرية بما في ذلك نموذج تكامل الطالب الشهير والتي وضعت من قبل (Tinto, 1993). طور آخرون نماذج خاصة بالتسرب الطلابى باستخدام الدراسات البحثية القائمة على المسح فقط (Berger & Braxton, 1998, Berger & Milem, 1999). على الرغم من أنها وضعت الأساس لهذا المجال، إلا أنها تفتقر للتعميم على المؤسسات الأخرى، بالاضافة الي الكلفة العالية التي تكبدها أدوات المسح تلك (Cabrera *et al*., 1993). أصبح تحليل البيانات المتعلقة بالطلاب فى قواعد البيانات الموجودة بالطرق الإحصائية ومنهجيات الذكاء الاصطناعى هو الطريقة المثلى لحل تلك المشكلات بطرق مبتكرة وفعالة. المعلومات الطلابية بما في ذلك الخلفية التعليمية والاجتماعية والاقتصادية، والتقدم الأكاديمي من أهم السمات/الخصائص التى يعتمد عليها النظام الخبير فى آلية عمله. أظهرت مقارنة بين الأبحاث القائمة على نظام المسح الميدانى والأبحاث المستندة إلى خوارزميات تعلم الآلة تفوق الأخيرة وعموميتها بشكل كبير (Caison, 2007). ولكن في واقع الأمر، فإن هذه التقنيات البحثية (والتى يعتمد أحدهما على الدراسات الاستقصائية ويعتمد الآخر على البيانات المؤسسية والأساليب التحليلية) يكمل كل منهما الآخر (Miller & Tyree, 2009)، أي أن البحث النظري قد يساعد في تحديد متغيرات التنبؤ المهمة التي يجب استخدامها في الدراسات التحليلية. في حين أن الدراسات التحليلية قد تكشف عن علاقات جديدة بين المتغيرات التي قد تؤدي إلى تطوير نظريات جديدة وتحسينها. ترتبط عدد من العوامل الأكاديمية والاجتماعية والاقتصادية وغيرها وتتناغم مع بعضها البعض وفقا لدراسة (Wetzel *et al*., 1999)، حيث تواجه الجامعات التي لديها سياسة قبول أكثر انفتاحا، ولا توجد لديها قائمة انتظار كبيرة للمتقدمين والتحويلات مشاكل استنزاف للطلاب أكثر من تلك الجامعات التي لديها فائض فى

المتقدمين. من ناحية أخرى، فإن الجامعات الأكثر انتقائية لا تتمتع بالضرورة بمعدلات تخرج أعلى، النسبة الأعلى من الاستبقاء الطلابى يتم تحقيقها غالبا عندما يعثر الطلاب في جامعتهم على بيئة مرتبطة بشكل كبير مع مصالحهم (Hermaniwicz, 2003). أوضح (Astin, 1993) أن استبقاء الطلاب يتأثر إلى حد كبير بمستوى ونوعية تفاعلهم مع أقرانهم وكذلك أعضاء هيئة التدريس والموظفين. كما يشير (Tinto, 1993) إلى أن من العوامل الرئيسة المسببة الى التسرب الطلابي من ضمنها صعوبة المناهج الأكاديمية، والافتقار إلى الأهداف الأكاديمية والوظيفية الواضحة، وضعف الاندماج مع مجتمع الكلية أو الجامعة، والعزلة التي يواجهها الطلاب أثناء عامهم الدراسي.

لقد حولت العديد من الدول المتقدمة نظامها التعليمي إلى نظام آلي بالكامل وذلك لمساعدة الطلاب والمعلمين أيضا على المشاركة في عملية تعليمية تفاعلية وناجحة. لتحسين العملية التعليمية للطلاب ودوراتهم التدريبية أعتمد (Quadri & Kalyankar 2010) على خوارزمية شجرة القرار C4.5 لترتيب مجموعة من الصفات التى تؤثر بشكل ملحوظ على الأداء التعليمي للطلاب. كما وتعد الشبكة العصبية واحدة من أكثر الممارسات والمنهجيات المستخدمة في تحليل البيانات التعليمية بشكل عميق. حيث استخدم (Arsad and Buniyamin, 2013) شبكة عصبية للتنبؤ بالتقدم الأكاديمي لطلاب درجة البكالوريوس. كما استخدم (Hien & Haddawy, 2007) خوارزمية نايف بايز Naïve Byes للتنبؤ بمتوسط النقاط التراكمية (CGPA) في وقت القبول والذي كان يعتمد على خلفيتهم الأكاديمية.

تم إجراء العديد من الأبحاث حتى الآن للتنبؤ بالأداء الطلاب الأكاديمي باستخدام تقنيات التنقيب فى البيانات. لكن القليل منهم هم من سلطوا الضوء على السمات/الخصائص المهمة التي تؤثر على الأداء التعليمي للطلاب. في هذا البحث، سنستخدم بعض أهم خوارزميات تعلم الآلة والذكاء الاصطناعي وذلك للتنبؤ بالآداء الأكاديمي للطلاب وأدائه العام.

١-٢ *منهجية العمل*

تتكون منهجية العمل حيال بناء هذا النظام الخبير المقترح من ست خطوات رئيسة: (١) تحديد أهداف الدراسة المراد تحقيقها، (٢) فهم ودراسة البيانات فهما عميقا ووافيا (٣) المعالجة المسبقة للبيانات، وتشمل عمليات التنظيف، والتحويل والإعداد والعرض واستخراج أبرز السمات/الخصائص للبيانات، (٤) تطوير النظام الخبير باستخدام تقنيات الذكاء الاصطناعي وتعلم الآلة، (٥) تقييم النظام الخبير ومقارنة الخوارزميات المقترحة حيال أهداف الدراسة، (٦) نشر النظام الخبير للاستخدام في عمليات صنع القرار. توفر تلك المنهجية طريقة مهيكلة ومنظمة لإجراء عمليات تنقيب بيانات صحيحة وفعالة، وبالتالي زيادة دقة وموثوقية النتائج والنظام ككل. عادة ما يتم استنفاذ ٨٠٪ من إجمالي وقت

المشروع على هذه الخطوات الثلاث الأولى. ولذلك فإن فهم مجال الدراسة والغرض والهدف الرئيس منها بالاضافة لفهم البيانات وإعدادها بالشكل الأمثل يمهد الطريق لتحليل البيانات والتنقيب فيها بطرق مثالية وناجحة. في هذه الدراسة، لقياس دقة نظام التنبؤ الخبير، استخدامت منهجية (10--fold cross validation) (وفقا للمعادلة (١) حيث k تعنى قيمة التكرار والتي هى ١٠). بحيث يتم استخدام مجموعة واحدة فقط من تلك المجموعات العشر مرة واحدة لتقييم النظام الخبير، وذلك بالاعتماد على بيانات الطلاب في التسع مجاميع الأخري واللاتى يستخدمن بغرض تدريب النظام.

$$CV = \frac{1}{K} \sum_{i=1}^{k} PM_i \qquad (١)$$

١-٢-١ وصف البيانات

قاعدة البيانات التي قمنا بتدريب النظام عليها متاحة على الموقع https://www.kaggle.com/ aljarah/xAPI–EDu–Data حيث إنها تحوى بيانات ٥٠٠ طالب و ١٧ سمة لكل طالب من بيانات ديمو جرافيه وأكاديمية وشخصية وفقا لتفاعل الطلاب مع الموقع التعليمي Kalboard 360 e–learning. ووصف البيانات موضح في الجدول ١.

بعد تجميع البيانات، تكون المهمة الأكثر أهمية هي المعالجة المسبقة للبيانات. المعالجة المسبقة للبيانات تشتمل على عدة مراحل أساسية لإعداد البيانات لعملية التنبؤ وهى: عملية استكمال البيانات المفقودة. عملية تنظيف البيانات من القيم الشاذة والقراءات الغير دقيقة. عملية التحويل من نوع لنوع لكي تتلاءم كل أنوع البيانات الخوارزمية المستخدمة التي ستطبق على تلك البيانات. طرق عرض البيانات وتحليلها بالطرق الإحصائية. استخراج مجموعة السمات المؤثرة بالإيجاب على دقة النظام. ثم أخيرا ادخال تلك البيانات المجهزة الى النظام الخبير للتنبؤ بإنجاز الطلاب. الهدف الرئيس من تلك الدراسة هو التنبؤ بحالة طالب من نجاح أو تقصير عن طريق بياناته التاريخية.

١-٢-٢ تحليل وعرض البيانات

عرض البيانات بطرق تحليلية ورسومية هام جدا لاستنباط بعض أنواع المعرفة الدفينة منها والتي لا تظهر إلا بها. عند عرض أعداد الطلاب الحاصلين على تصنيف عال ومتوسط ومتدنى كما بالشكل ١، نجد أن عدد الطلاب الحاصلين على تقدير متدني ١٢٧ طالبًا، بينما الحاصلين على تقدير متوسط هم ٢١١ طالبًا، اما الحاصلين على تقدير عال هم البقية وعددهم ١٤٢ طالبًا.

من المهم في مرحلة تهيئة البيانات هو تحديد ومعرفة مدي اعتمادية الخصائص علي بعضها البعض ومدي تأثير تلك الخصائص في القرار والتصنيف النهائي. سننتقل الى عرض علاقة بعض السمات/الخصائص بالتصنيف وتأثيرها عليه. يعرض الشكل ٢ علاقة تصنيف الطلاب ببعض السمات/الخصائص المهمة التي قد يتأثر بها، فنجد

أن الطلاب المتفاعلين أكثر في الفصل الدراسي بزيادة عدد رفع أيديهم ومشاركتهم في النقاشات وزيارتهم للمحتوى العلمي ومتابعة النشاطات الدراسية هم في معظم الأحيان الطلاب الذين يحصلون على تصنيف ودرجة أعلى من أقرانهم. من الأشكال التي تكسب البيانات مصداقية أكثر، هو شكل توزيع البيانات وقربها من التوزيع الطبيعي، فنجد من خلال الشكل ٣ أن عملية التنبؤ المعتمدة على سمات/خصائص المناقشة وزيارة المحتوى العلمي ستكون أكثر مصداقية من غيرهم وذلك لقلة الحيود في التوزيع الطبيعي البيانات. من ناحية أخرى، يظهر الشكل ٣ تفوق الإناث عن الذكور وارتفاع معدلات التحصيل باستعراض دلالة بعض السمات/الخصائص على بعضها البعض.

من مرحلة عرض البيانات تبين لنا أنه سيكون هناك بعض السمات/الخصائص التي لها التأثير الأكبر على عمل النظام الخبير. من ناحية أخري، نجد أن معظم السمات عبارة عن قيم اسمية وذلك سيكون عائقا أمام الخوارزميات الخاصة بتعلم الآلة حيث أنها تتعامل مع القيم الرقمية فقط، فكانت مرحلة تحويل البيانات من المراحل المهمة لضمان عمل الخوارزمية بشكل صحيح. بالإضافة الى ان كل السمات/الخصائص المتعلقة بالطلاب ليست على وتيرة واحدة من الأهمية، لذا فعملية اختيار السمات التي ستؤثر بالإيجاب على عمل الخوارزمية أمر حيوي. لأن بعض السمات قد تؤثر بالسلب على عمل النظام الخبير ودقته. كما وأظهرت عملية تحليل البيانات أنه لاوجود للقيم الشاذة أو المفقودة للتعامل معها.

١-٢-٣ *تحويل البيانات*

دعونا نرى كيفية التعامل مع البيانات الفئوية، وهي البيانات التي تندرج ضمن عدة فئات ولا تأخذ شكلا عدديا (كالتصنيف، الجنس، تقييم الوالدين، الخ) كما بالشكل رقم (٥). نظرا لأن نماذج التعلم الآلي (machine learning) تعتمد في آلية عملها على معادلات رياضية، لذا فقد تتسبب بعض المشكلات إذا قمنا بالاحتفاظ بالبيانات الاسمية/الفئوية في المعادلات لأننا نريد فقط القيم الرقمية في المعادلات. وحيث أن القيم الاسمية/الفئوية لا يتم تمثيلها بأرقام، لذلك وجب تحويل تلك البيانات الفئوية الي بيانات رقميه لسهولة التعامل معها بواسطة خوارزميات التعلم الالي ومعادلاته لضمان عملها بشكل صحيح.

يتضح لنا من الشكل ٥ وجود سمات يعبر عنها بقيم فئوية كرضا الوالدين، الفئة التي يصنف لها الطلاب في النهاية، عدد أيام الغياب أكير من أم أقل من ٧ أيام. كل هذه القيم الفئوية يتم تحويلها الى قيم رقمية لسهولة التعامل معها في النظام لاحقا. فعلى سبيل المثال ستكون القيمة للسمة/الخاصية الخاصة برضا الوالدين هي (٠ للقيمة bad) و (١ للقيمة good). أيضا ستكون القيمة للسمة/الخاصية الخاصة بمعدل الغياب هي (٠ للطلاب الذين لم تتجاوز عدد أيام غيابهم السبعة أيم) و (١ لمن تجاوز سبعة أيم).

وفى المحصلة النهائية للتنبؤ ستكون قيمة التصنيف الأقل من ٦٩٪ لمعدل الطالب هي التقدير "L"، والقيمة بين ٧٠٪ و٨٩٪ هي التقدير "M"، أما القيمة أبر من ٩٠٪ فهي للتقدير "H".

١-٢-٤ تسوية القيم

خوارزمية التنبؤ الخاصة بتعلم الآلة تعمل وفق منهجيات وآليات معروفة. معظم خوارزميات تعلم الآلة تقوم بعملية وزن لكل سمة/خاصية (أي إعطائها قيمة من ٠ الي ١) وذلك حسب أهميتها. فلو كانت المدي الخاص بسمة/خاصية ما كبيرا وذو قيم كبيرة فإن ذلك سيؤثر على قيم بقية السمات بشكل غير صحيح. في هذه المرحلة يتم تسوية جميع قيم السمات بحيث تكون جميع القيم للسمة الواحدة محصورة بين (٠،١) أو (-١و ١) وذلك وفق عدة منهجيات. في هذه الورقة استخدمنا طريقة (Standard Scaler) بحيث تكون جميع القيم للسمة/للخاصية الواحدة محصورة بين (٠،١). يتم حساب المتوسط $\boldsymbol{mean}\ (\boldsymbol{x})$ والانحراف المعياري $\boldsymbol{stdev}(\boldsymbol{x})$ لكل سمة $\boldsymbol{x_i}$ ويتم تحديث جميع القيم لكل سمة وفقا للمعادلة رقم (٢).

$$\boldsymbol{x_i} = \frac{\boldsymbol{x_i - mean}\ (\boldsymbol{x})}{\boldsymbol{stdev}(\boldsymbol{x})} \qquad (٢)$$

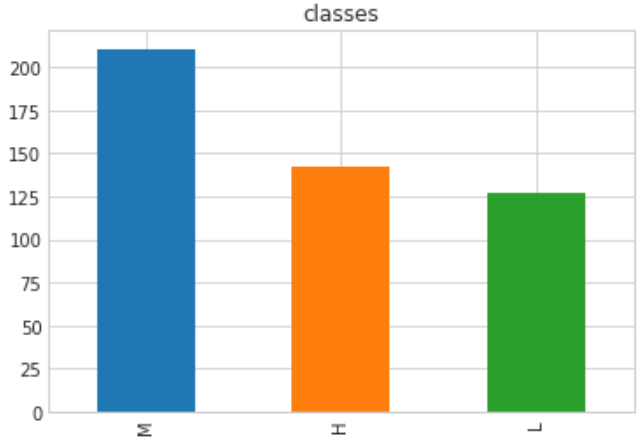
وبذلك نضمن عدم التضحية ببعض السمات على حساب نظائرها.

١-٢-٥ اختيار السمات

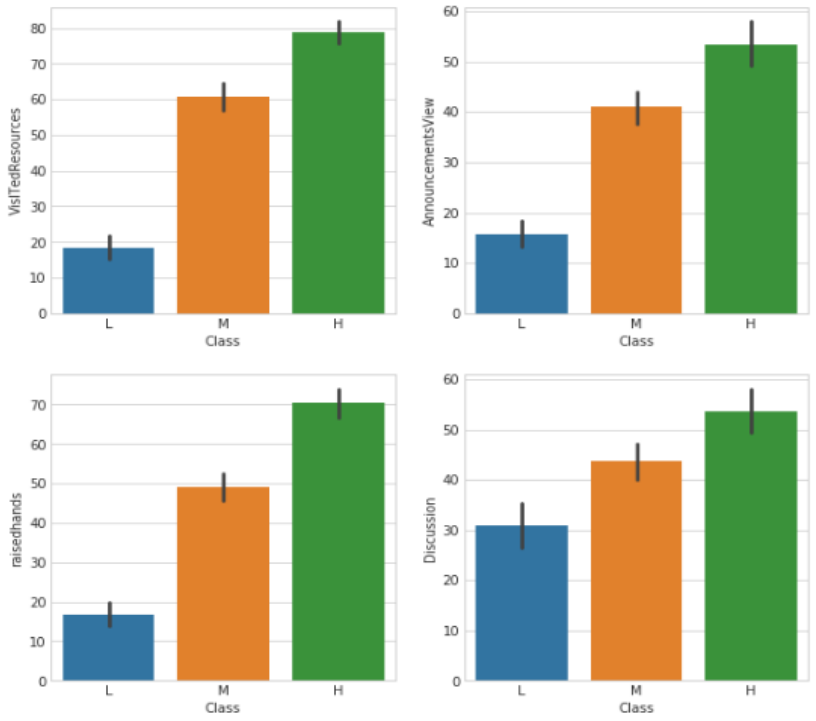
اختيار السمات بشكل صحيح هي العملية الأكثر أهمية في مرحلة المعالجة المسبقة للبيانات. والهدف الرئيس من هذه الخطوة هو اختيار المجموعة الأنسب من السمات بحيث يكون الهدف هو تقليل عدد السمات المختارة الى اقل عدد ممكن من الخصائص والسمات المؤثرة فقط. عند اختيار سمات غير مؤثرة على مخرجات النظام فإنها قدد تتسبب في عملية تضليل (misleading) للنتائج والتشويش على المخرجات. لذا فإن عملية اختيار السمات التي سيتم تدريب الخوارزمية عليها يجب أن تؤثر في المخرجات بشكل ملحوظ. تنقسم طرق اختيار السمات إلى فئتين رئيستين: (١) الأساليب القائمة على التغليف (٢) الأساليب القائمة على التصفية. يتم تطبيق طريقة التصفية لتحديد مجموعة السمات المختارة ذات الصلة مع تجنب الباقي. تقوم هذه الطرق بترتيب السمات باستخدام تقنيات الترتيب بحيث يمكن اختيار السمات عالية الترتيب والتي ستكون أخيرا مدخلات لخوارزميات التعلم.

جدول ١. وصف البيانات.

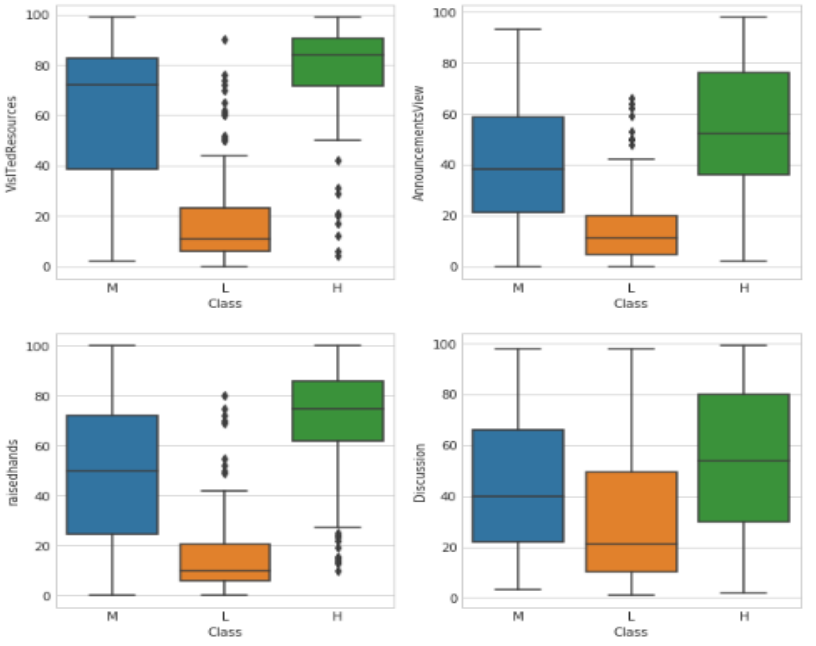
	السمة	الوصف	النوع	التصنيف
١	الجنس - gender	نوع الطالب (ذكر - أنثى)	ثنائية الاسم	سمات ديمو جرافيه
٢	البلد - NationalITy	بلد المنشأ للطالب	متعددة الأسماء	
٣	مكان الميلاد - PlaceofBirth	بلد الميلاد للطالب	متعددة الأسماء	
٤	علاقة المسؤول - Relation	الأب أم الأم	ثنائية الاسم	
٥	المستوى التعليمي - StageID	المراحل التعليمية للطلاب	متعددة الأسماء	سمات أكاديمية
٦	تقدير الطالب - GradeID	تقدير الطالب من (١ الى ١٢)	متعددة الأسماء	
٧	رقم القاعة - SectionID	أ – ب - ج	متعددة الأسماء	
٨	الفصل الدراسي - Semester	الأول – الثانى	ثنائية الاسم	
٩	المادة العملية - Topic	رياضة – تقنية المعلومات – اللغة الإنجليزية – قرآن – علوم ...	متعددة الأسماء	
١٠	الغياب - StudentAbsenceDays	أقل من ٧ أيام أم أكثر	ثنائية الاسم	
١١	مشاركة الآباء - ParentAnsweringSurvey	مشاركة الآباء بالإجابة على الأسئلة الدورية المتعلقة بالطالب	ثنائية الاسم	مشاركة الآباء في متابعة وتقييم العملية التعليمية
١٢	رضا الآباء - ParentschoolSatisfaction	مدى رضا الآباء عن مستوى الطالب (إيجابي - سلبي)	ثنائية الاسم	
١٣	عدد المناقشات الجماعية - Discussion	سمات تعنى بسلوك الطالب خلال تفاعله مع الموقع التعليمي Kalboard 360 e–learning	رقم	سمات شخصية
١٤	عدد مطالعة المحتوى العلمي - VisITedResources		رقم	
١٥	عدد مرات رفع اليد - raisedhands		رقم	
١٦	عدد مطالعة المهمات - AnnouncementsView		رقم	



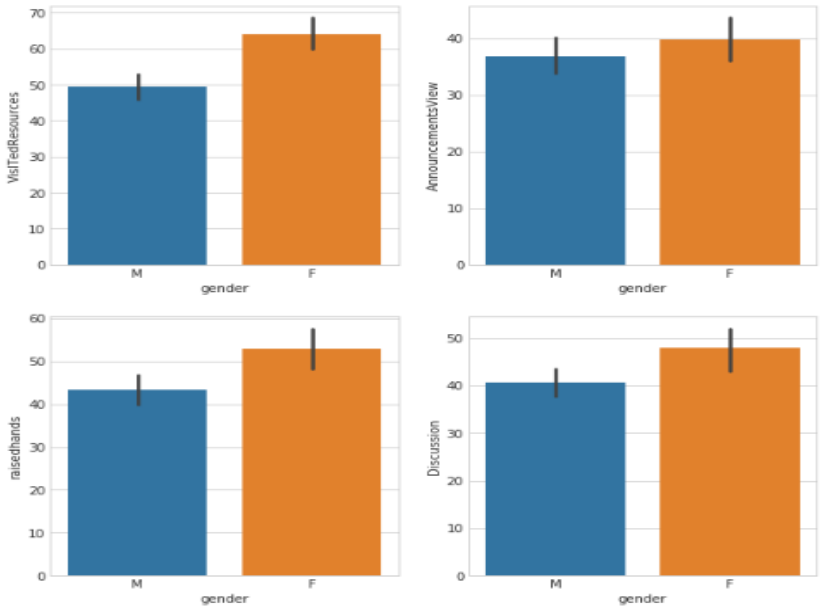
شكل ١. تصنيف الطلاب.



شكل ٢. علاقة بعض السمات بالتصنيف (ا).



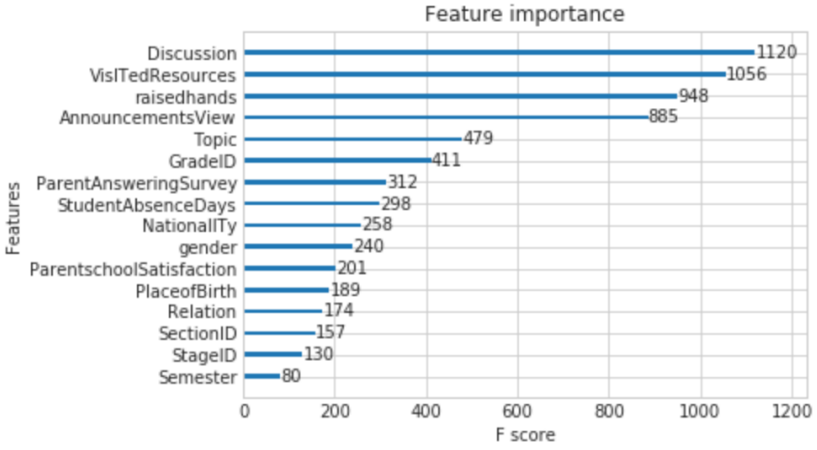
شكل ٣. علاقة بعض السمات بالتصنيف (ب).



شكل ٤. علاقة السمات ببعضها البعض.

VisITedResources	AnnouncementsView	Discussion	ParentAnsweringSurvey	ParentschoolSatisfaction	StudentAbsenceDays	Class
16	2	20	Yes	Good	Under-7	M
20	3	25	Yes	Good	Under-7	M
7	0	30	No	Bad	Above-7	L
25	5	35	No	Bad	Above-7	L
50	12	50	No	Bad	Above-7	M

شكل ٥. أنواع البيانات.



شكل ٦. ترتيب السمات من حيث الأهمية.

١-٣ الأعمال ذات الصلة

سنستعرض في هذا القسم مجموعة من الأبحاث التي استخدمت التعلم العميق (خوارزميات الشبكات العصبية) للكشف عن مرض الكوفيد-١٩، وهي مختزلة في الجدول ٢.

١-٤ المساهمات العلمية للبحث

كل الأبحاث التي وردت في الفقرة السابقة اعتمدت على استخراج السمات بواسطة الشبكات العصبية العميقة فقط، ولكن في عملنا هذا:

١- قمنا بالدمج بين الطرق الأتوماتيكية لاستخراج السمات باستخدام الشبكات العصبية العميقة والطرق التقليدية، وحصلنا على نتائج واعدة.

٢- تم بناء نظام ذكي يتعامل مع نوعي الصور الطبية الشعاعية (صور الأشعة السينية وصور الأشعة المقطعية) للكشف عن مرض الكوفيد-١٩.

٣- استخدمنا مفهوم استرجاع الصور المعتمد على المحتوى ليتم استرجاع الحالات الأكثر تشابهًا للحالة التي يتم اختبارها لتساعد الطبيب وترشده أن هذه الحالة مشابهة جدًا مع بعض الحالات التي قد راجعت المشفى في فترات سابقة، وهذا قد يساعده في اتباع نفس البرتوكول العلاجي الذي تم اتباعه على المرضى السابقين.

١-٥ منهجية البحث

(أ) الفكرة العامة للنظام المقترح

يتكون النظام المقترح من مهمتين أساسيتين، كما هو موضح في الشكل ١:

المهمة الأولى: هي تصنيف الصور (تشخيص المرض)، أما **المهمة الثانية:** فهي استرجاع الصور المعتمد على المحتوى وهو ما يعرف بـ (Content Based Image Retrieval). وبشكل أكثر تفصيلًا، لدينا صورة الدخل (صورة أشعة السينية أو أشعة المقطعية لمنطقة الصدر)، والتي يتم استخراج شعاع السمات لها باستخدام الخوارزمية المقترحة، ليتم استخدامه في تصنيف الصورة (تشخيص المرض) بمساعدة إحدى المصنفات المدربة، وبالتالي فإن خرج هذه المهمة هو الصنف الذي تنتمي إليه صورة الدخل. كما أن شعاع السمات هو أساس مهمة استرجاع الصور المتشابهة والتي تعتمد على قياس نسبة التشابه بينه وبين قاعد السمات وخرج هذه المهمة هو الصور الأكثر تشابهًا لصورة الدخل.

- **مهمة التصنيف**

في مهمة التصنيف، قمنا بتقسيم قاعدة الصور إلى قاعدتين: إحداهما للتدريب والأخرى للاختبار، نسبة التقسيم هي ٨٠٪ من قاعدة الصور للتدريب و٢٠٪ للاختبار. في مرحلة التدريب، يتم استخراج أشعة السمات لجميع صور قاعدة التدريب باستخدام

الخوارزمية المقترحة لنستفيد منها مع تسميات الصور لتدريب المصنف، وفي نهاية هذه المرحلة، سوف نحصل على مصنف مدرب وجاهز لتصنيف أي صورة من قاعدة الاختبار. أما مرحلة الاختبار فتبدأ بإدخال الصورة المراد معرفه الصنف الذي تنتمي إليه (تشخيص الصورة)، ليتم استخراج شعاع السمات لها باستخدام نفس الطريقة المقترحة والذي بدوره سيكون دخل المصنف (المدرب في مرحلة التدريب) الذي سيصنفها إلى الصنف المناسب. استخدمنا في عملنا المقترح كل من خوارزمية (SVM–Support Vector Machine) وخوارزمية الجوار الأقرب (KNN–K Nearest Neighbors).

- **مهمة استرجاع الصور**

تتألف مهمة استرجاع الصور المعتمد على المحتوى (CBIR) من [١٦]: مرحلة التدريب لاستخراج السمات ويطلق عليها أحيانًا مرحلة الـ Off–Line وذلك لكونها تحتاج إلى زمن كبير نسبيًا (حيث أن هذا الزمن يعتمد على عدد الصور الموجودة في قاعدة الصور ويقاس بالدقائق)، ومرحلة الاختبار لاسترجاع الصور ويطلق عليها أحيانًا مرحلة الـ On–Line وذلك لكونها لا تحتاج إلى زمن كبير (ويقاس بالثواني). في مرحلة التدريب، يقوم النظام باستخراج أشعة السمات لكل الصور الموجودة في قاعدة الصور وتخزينها في قاعدة السمات. أما مرحلة الاختبار، فيقوم المستخدم بإدخال صورة إلى هذا النظام كصورة استعلام (دخل)، والذي يقوم باستخراج شعاع السمات من هذه الصورة وقياس نسبة التشابه (باستخدام إحدى المسافات المعيارية) بين شعاع سمات صورة الدخل وجميع أشعة السمات المخزنة في قاعدة السمات والتي تم الحصول عليها في مرحلة التدريب. بعد ذلك يتم ترتيب هذه المسافات بشكل تصاعدي (حيث أن المسافة الأصغر تعني أن التشابه أكبر) واسترجاع أول k صورة كخرج لمرحلة الاختبار.

(ب) المسافات المعيارية المستخدمة في قياس نسب التشابه

سوف نعتمد في قياس التشابه على المسافة الاقليدية (Euclidean Distance) ومسافة الستي بلوك (Cityblock) [١٧]

- المسافة الاقليدية يتم حسابها وفقًا للقانون (١):

$$D_{Euclid}(Q,DB) = \sqrt{\sum_{i=1}^{L}(f_{DB_{ji}} - f_{Q_i})^2} \quad \text{(١) ... ...}$$

- مسافة الستي بلوك يتم حسابها وفقاً للقانون رقم (٢):

$$D_{cityblock}(Q,DB) = \sum_{i=1}^{L}\left|f_{DB_{ji}} - f_{Q_i}\right| \quad \text{(٢) ... ...}$$

f_{Q_i} ترمز إلى شعاع السمات لصورة الدخل في الموقع i؛ حيث أن الموقع iهو متحول يتحرك على مواقع الشعاع بدءً من الموقع الأول وحتى الموقع الأخير من الشعاع f_Q.

$f_{DB_{ji}}$ ترمز إلى شعاع السمات للصورة j (المتحول j يشير ترتيب الصورة في قاعدة الصور) من قاعدة السمات في الموقع i من شعاع السمات.

L ترمز إلى طول شعاع السمات.

(ج) الخطوات الأساسية للخوارزمية المقترحة

تتألف الخوارزمية المقترحة من ست مراحل أساسية، كما هو موضح في الشكل ٢:

١- استخراج شعاع السمات الأول باستخدام شبكة الـ AlexNet (بعد تدريبها على قاعدة التدريب) وذلك من خرج الطبقة "drop7" (وحجمه ٤٠٩٦).

٢- استخراج شعاع السمات الثاني باستخدام حقيبة الكلمات المرئية Bag of Visual Word (وحجمه ٥٠٠).

٣- استخراج شعاع السمات الثالث باستخدام خوارزمية النمط الثنائي المحلي (وحجمه ٢٥٦).

٤ - تنظيم البيانات وتوحيد قيمها بين الصفر والواحد من خلال عملية تعرف بالـ Normalization وإعطاء كل شعاع من السمات وزن محدد وفقًا لأهميته.

٥- دمج هذه الأشعة لنحصل على شعاع سمات طويل نوعًا ما (وحجمه ٤٨٥٢).

٦- تقليص حجم شعاع السمات بتطبيق خوارزمية الـ PCA للحصول على شعاع قصير للسمات.

(د) شبكة الـ AlexNet

تحتوي شبكة الـ AlexNet على ٦٠ مليون بارامتر و٦٥٠ ألف خلية عصبية واستغرقت من خمسة إلى ستة أيام للتدريب على وحدتي معالجة GPX580 3GB، وتتكون بشكل عام من ٨ طبقات؛ خمس طبقات للالتفاف (Convolution Layers) وثلاث طبقات متصلة بالكامل (Full connected layers). وإذا أردنا التفصيل أكثر فإن شبكة الـ AlexNet تتكون من ٢٥ طبقة [١٨]. يتم توزيع طبقات شبكة الـ AlexNet الخمس والعشرين بالطريقة التالية:

- طبقة الدخل حيث تستقبل هذه الطبقة الصور الملونة والتي حجمها هو (٢٢٧ × ٢٢٧).
- سبع طبقات من نوع ReLU.
- طبقتان من نوع "Norm".
- ثلاث طبقات من نوع "Pool".
- طبقتان من نوع "dropout".
- تسمى الطبقة التي تسبق الطبقة الأخيرة طبقة "Prob" أو "Softmax".
- طبقة الخرج التي تعطي الصنف الناتج عن عملية التصنيف.
- خمس طبقات "Conv" وثلاث طبقات من نوع "FC".

حيث تم تدريب شبكة الـ AlexNet على قاعدة الصور الطبية، لتصبح جاهزة لاستخدامها في استخراج

السمات، حيث ندخل صورة بحجم ٢٢٧ × ٢٢٧ إلى الشبكة لنحصل على شعاع السمات من خرج الشبكة وبالتحديد من طبقة "drop7".

(هـ) حقيبة الكلمات المرئية (BoVW)

الفكرة العامة لحقيبة الكلمات المرئية [١٩، ٢٠]: هي تمثيل الصورة كمجموعة من الميزات (السمات) والتي تتكون من النقاط الرئيسية (KeyPoints) والوصف (Descriptors). النقاط الأساسية هي النقاط "البارزة والأكثر أهمية" في الصورة، لذلك بغض النظر عن تدوير الصورة أو تقليصها أو توسيعها، فإن نقاطها الرئيسية ستكون دائمًا هي نفسها. أما الوصف فهو استخراج السمات من هذه النقاط، حيث يتم استخدام النقاط الرئيسية والواصفات لبناء المفردات (Words) ونمثل كل صورة كرسم بياني (Histogram) للميزات الموجودة في الصورة.

بالتالي فإن الـ BoVW تتكون بشكل أساسي من مرحلتين هما:

مرحلة التدريب، والتي تتلخص بالخطوات التالية:

- تنظيم قاعدة الصور على شكل فئات (أصناف).
- من أجل كل صنف، نختار ٦٠٪ من الصور بشكل عشوائي.
- نحدد النقاط الرئيسية للصور KeyPoints، ونستخرج الواصفات من هذه النقاط الرئيسية من أجل كل صورة في كل صنف باستخدام خوارزمية الميزات القوية السريعة (SURF-Speeded up Robust Features).
- الحفاظ على ٨٠٪ من أقوى الميزات لكل صنف.
- بناء قاموس مرئي باستخدام خوارزمية التجميع (K-Means) بحجم ٥٠٠ كلمة.
- استخدام مركز كل عنقود كمفردات القاموس المرئي (Visual Dictionary's Vocabularies).

مرحلة الاختبار، وتتلخص بالتالي:

يتم إدخال صورة الدخل ليتم تحديد النقاط المفتاحية في الصورة، واستخراج الواصفات من هذه النقاط، ليتم تجميع الواصفات بواسطة خوارزمية العنقدة K-means، ثم بناء الهستغرام بطول (٥٠٠ قيمة) وذلك من خلال مقارنة نتيجة العنقدة لواصفات الصورة مع حقيبة (قاموس) الكلمات المرئية الذي قمنا ببنائه في مرحلة التدريب والذي سيكون شعاع السمات الناتج عن تطبيق فكرة حقيبة الكلمات المرئية.

(و) خوارزمية النمط الثنائي المحلي (Local Binary Pattern)

خوارزمية الـ LBP هي الخوارزمية المقترحة للصور البنيوية (Texture Images) ذات الملمس المتشابه [٢١-٢٤]. وهي ميزة قوية لتصنيف واسترجاع هكذا نوع من الصور. يتم بناء شعاع السمات الخاص بخوارزمية الـ LBP بالطريقة التالية:

من أجل كل بكسل في الصورة:

- نقارن البكسل مع كل من جيرانه الثمانية (بدءً من بكسل اليسار العلوي ثم البكسل الذي يوافق اتجاه عقارب الساعة).

- عندما تكون قيمة البكسل المركزي أكبر من قيمة الجار، نستبدل قيمة الجار بـ "٠". خلاف ذلك، نستبدل قيمته بـ "١".

- هذا يعطي رقم ثنائي مكون من ٨ أرقام (والذي يتم تحويله إلى عدد عشري) وقيم هذه الأرقام بين ٠ و٢٥٥.

- بناء الهستغرام (الرسم البياني) للصورة والذي سيكون بحجم ٢٥٦ قيمة.

١-٦ منهجية العمل وبناء النظام الخبير

فى هذه المرحلة وبعد تهيئة البيانات بحيث أصبحت جاهزة أن تكون مدخلا للخوارزميات الخاصة بتعلم الآلة، سنقوم باستخدام بعض الجوريذمات تعلم الآلة وتطبيقها على البيانات ومقارنتها من حيث الكفاءة والدقة. سيتم الاعتماد لحل هذه المشكلة على خوارزم نايف بايز (Naïve Bayes Classifier)، خواريزم شجرة القرار (Decision Tree)، خواريزم الغابة لاعشوائية (Random Forest Classifier)، خواريزم الشبكات العصبية (Neural Network Classifier). تم تقسيم البيانات الي جزئين رئيسين: الجزء الأول خاص بتدريب الخوارزميات الأربع، والقسم الثاني خاص باختبار الأنظمة الخبيرة المقترحة.

١-٦-١ *خوارزم نايف بايز (Naïve Bayes Classifier)*

يعتبر خوارزم نايف بايز من أشهر خوازمات تعلم الآلة (Machine Learning) وتحليل البيانات (Data Analytics) والذي يتم استخدامه في مشاكل التصنيف (Classification) على وجه التحديد، حيث يتميز بالسرعة في المعالجة والكفاءة في عمليات التنبؤ. يعتمد هذا الأسلوب على المفهوم الإحصائي Bayes' theorem والذي يحسب احتمالية حدوث نتيجة معينة بتحقق ماهو متاح ومعروف ويسمى (ساذج) Naive لأنه يعتمد مبدأ الفروض المستقلة (Independence Assumptions)، بحيث يعتمد استقلالية العلاقة بين الخصائص (Attributes Features) وبعضها البعض. بمعنى أن النموذج لا يعير اهتماما للعلاقة بين الخصائص إن وجدت فجميعهم يساهمون في حساب الإحتمال والنتيجة النهائية ستكون رقما لا يحمل معنى من حيث توضيح اعتماد خاصية على أخرى أو قيمة للترتيب.

نموذج بايز للتصنيف يتميز بسهولة البناء والتطوير والقدرة على معالجة البيانات الكبيرة ويتفوق في ذلك على عدد من الخوارزميات المتعقدة والمتقدمة. يعمل بنوع من الكفاءة عندما يكون حجم البيانات كبير الي حد ما (Nikam, 2015). بحيث يتم تدريب النموذج بالبيانات وخصائصها المتاحة في قواعد البيانات ومن ثم يقوم النموذج بتحديد نوع السجلات

الجديدة وتصنيفها بالإعتماد على البيانات والإحصاءات المتوفرة سابقا لديه. هناك عدة أساليب متفرعة من هذه الطريقة وتستخدم في كثير من الأنظمة على سبيل المثال في التعرف على الرسائل المؤذية Spam ، وفي تصنيف الوثائق مثل في مواقع الأخبار لتوقع نوع الوثيقة (سياسة، رياضة، تقنية) Text Classification، التعرف على وجهات النظر والمشاعر في محتوى النص (سلبي، إيجابي، متفائل) Sentiment Analysis وفي استخدامات أخرى كالتعرف على الوجه في الصور. في هذه الخوارزمية، لا يرتبط وجود ميزة معينة في الفصل الدراسي بوجود أي ميزة أخرى كما ذكرنا آنفا. فكل الخصائص تساهم بشكل مستقل في احتمال الحكم على آداء الطالب آكاديما. يتم حساب خوارزم نايف بايز من المعادلة رقم (٣).

$$\boldsymbol{p(c \mid x)} = \frac{\prod \boldsymbol{p(x \mid c)\, p(c)}}{\boldsymbol{p(x)}} \qquad (٣)$$

حيث يمثل الحد $\boldsymbol{p(c \mid x)}$ قيمة الاحتمالية أو مدي اعتمادية الصنف/المخرج c علي الخاصية x، حيث أنها احدي السمات التي تشكل الخرج. وهذا مانبحث عنه، نريد معرفة مدي تأثير هذه الخاصية أو السمة على الخرج. كما يمثل الحد $\boldsymbol{p(c)}$ قيمة الاحتمالية للمخرج. فلو كانت قيمة المخرج تتمثل في أن الطالب سينجح أو سيفشل، فتكون القيمة نسبة النجاح الي العدد الكلي ونسبة الفشل الي العدد الكلي. $\boldsymbol{p(x \mid c)}$ تمثل مدي تأثر الخرج c بالخاصية x. $\boldsymbol{p(x)}$ احتمالية قيمة الخاصية x بالنسبة لمجموع قيمها. يتم حساب تلك المعادلة بضرب جميع قيم احتماليات تأثير الخصائص المستقلة على الصنف/المُخرج في نسبة الصنف ويقسم على احتمالية قيم الخصائص وتكون القيمة الأعلى للاحتمالية هي من تنتمي لتلك الصنف. أي أن هذه المعادلة سيتم تطبيقها عددا من المرات مساوٍ لعدد الأصناف.

١-٦-٢ *خوارزم شجرة القرارات (Decision Tree Classifier)*

تعد شجرة القرارات من أهم التقنيات التي يتم من خلالها استنتاج المعرفة الكامنة في كميات هائلة من البيانات، والوصول الى حالات معرفية تدعم اتخاذ القرار. أشجار القرار هي طريقة تعلم غير خاضع للإشراف، تستخدم للتصنيف والانحدار (Regression) على حد سواء. الهدف منها خلق نموذج للتنبؤ بقيمة معينة عن طريق تعلم قواعد بسيطة مستنتجة من سمات/خصائص البيانات. تطبق عملية التصنيف عن طريق مجموعة من القواعد أو الشروط التي تحدد المسار الذي سيتبع ابتداء من عقدة الجذر وانتهاء بإحدى العقد النهائية التي تمثل القرار النهائي. وينبغي عند كل العقد غير النهائية اتخاذ قرار حول مسار العقد التالية. والقرار في حد ذاته هو اختيار حل من بين عدة حلول لمشكلة معينة. وعليه فإن اتخاذ القرار هو اختيار أحد البدائل المتاحة، لذلك فعملية اتخاذ القرار هي مجموعة متتالية من الخطوات والإجراءات التي تؤدي في نهايتها إلى اختيار أفضل الحلول البديلة. يتميز هذا الأسلوب من تمكين متخذ القرار من

رؤية البدائل المتاحة والأخطار والنتائج المتوقعة لكل منها بوضوح. ويستعمل أسلوب شجرة القرار في حل المشكلات ذات البدائل المتعددة، وكذلك الحالات المتعددة المحتمل مواجهتها، خاصة عندما تكون المشكلة متعلقة بعنصر المخاطرة وعدم التأكد. تعرف شجرة القرار أيضا بأنها: مجموعة محددة من العقد بحيث أن: هنالك عقدة مميزة تقع في أعلى الشجرة تدعى بجذر الشجرة، وكل شجرة من المجاميع أسفلها عبارة عن شجرة تدعى الأشجار الفرعية (Subtree) للجذر. القاعدة الأساسية في بناء شجرة القرار هي إيجاد أفضل سؤال عند كل فرع من فروع الشجرة بحيث يقسم هذا السؤال البيانات إلى قسمين، القسم الأول منها ينطبق عليهم السؤال والقسم الثاني لا ينطبق، وهكذا يتم من خلال سلسلة من الأسئلة بناء شجرة القرار بفروعها المتسلسلة. في هذه الورقة تم استخدم خوارزمية DecisionTreeClassifier لتصنيف الآداء الأكاديمي للطلاب.

١-٦-٣ خوارزم الغابة لاعشوائية (Random Forest Classifier)

تعتبر الغابات العشوائية واحدة من أكثر تقنيات التعلم الآلي قوة واعتمادية . تعتبر الغابات العشوائية كذلك أداة تجسد قوة أشجار القرار بالإضافة إلى التعلم الجمعي (Ensemble Learning) لإنتاج نماذج تنبؤية مدهشة بدقتها. لبنة البناء الأساسية للغابة العشوائية مستوحاة من شجرة القرار (Classification and Regression Trees)، وهي إحدى طرق تعلم الآلة لبناء نماذج تنبؤ من البيانات، إذ يتم الحصول على النماذج من خلال تقسيم البيانات وبناء نموذج بسيط للتنبؤ داخل كل قسم. وضعت الغابات العشوائية في الأصل من قبل (Breiman, 1999)، بناء على عمر من المساهمات المؤثرة، والتي من بينها شجرة القرار CART. أشجار القرار هي طريقة تعلم غير خاضع للإشراف كما ذكرنا أنفا، تستخدم للتصنيف والتنبؤ معا. الهدف منها خلق نموذج لتنبؤ قيمة متغيرة أو التصنيف الي أحد الأصناف المذكورة، عن طريق تعلم قواعد بسيطة مستقاة من خصائص البيانات. تطبق عملية التصنيف عن طريق مجموعة من القواعد أو الشروط التي تحدد المسار الذي سيتبع ابتداء من عقدة الجذر وانتهاء بإحدى العقد النهائية التي تمثل الصنف المراد الوصول اليه، وينبغي عند كل العقد غير النهائية اتخاذ قرار ترجيحي حول مسار العقد التالية.

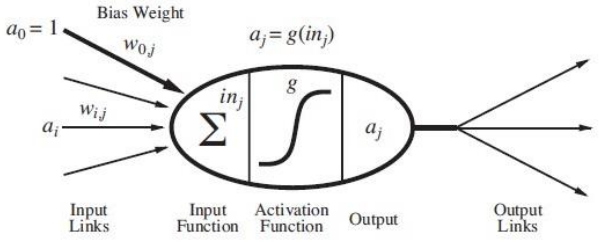
تعتبر الغابة العشوائية هي مجموعة من أشجار القرار كل منها تقوم بعملها مستقلة وقد تكون البيانات مختلفة الي حد ما. لنتخيل سحب عينة عشوائية من قاعدة البيانات الرئيسية وبناء شجرة قرار على أساس هذه العينة العشوائية. ولنفترض أن هذه العينة تستخدم نصف البيانات المتاحة، حتى لو كان من الممكن أن يكون هذا النصف مختلفا عن قاعدة البيانات الرئيسية ككل. الآن نكرر العملية السابقة باختيار عينة عشوائية أخري مختلفة عن سابقتها لنبني شجرة القرار الثانية، بكل تأكيد ستكون التوقعات التي تقدمها شجرة القرار الثانية مختلفة (على الأقل قليلا) عن الشجرة الأولى.

وتستمر تلك العملية وغالبا مايكون اختيار عدد الشجر بالمئات أو الآلاف لإعطاء نتائج جيدة مختلفة نسبيا. وفي النهاية يتم الجمع بين كل هذه التنبؤات المنفصلة باستخدام إما المتوسط أو التصويت الأكثر عددا (أي اختيار القرار الذي اختارته معظم الشجرات). الهدف من تلك العملية هو اتحاد خوارزمية قد تبدو ضعيفة وتوظيفها بصورة ما لتبدو قوية وذات موثوقية عالية. عند اجتماع الشجر بأكمله أو معظمه على تصنيف أو تنبأ ما، فمن الصعب أن يكون هذا التصنيف أو التنبؤ خاطئ بنسبة كبيرة. في هذه الورقة تم استخدام النموذج المقترح من (Chen & Guestrin, 2016) جامعة واشنطن تحت مسمي (XGBoost)، وقد تم اقتباس هذا البحث في أكثر من 4000 مقالة علمية حيث أثبت ناجحا عاليا في فترة قصيرة.

١-٦-٤ خوارزم الشبكات العصبونية (*Artificial Neural Network Classifier)*

في دراسة بحثية (Amrieh, el at., 2015) استخدمت نهج الشبكة العصبية ANN التي تستخدم للتنقيب في البيانات لتحقيق دقة عالية في العديد من مشاكل التصنيف والتنبؤ المعقدة. يستخدم إطار ANN لإنشاء أنماط كثيرة وذلك لحل مشاكل التنبؤ والتصنيف المختلفة. إن ما يحدث في دماغنا كنشاط دائم هو نشاط كهربائي كيميائي بين شبكة من خلايا الدماغ والتي تدعى بالخلايا العصبية أو العصبونات، وبناء على ذلك اتجه علم الذكاء الاصطناعي إلى بناء شبكات عصبونية صنعية تحاكي الدماغ البشري.

تتألف الشبكات العصبونية الصنعية من وحدات مرتبطة مع بعضها البعض عبر وصلات؛ إذ تستخدم هذه الوصلات لنقل النشاط بين هذه الوحدات، وتملك كل وصلة ثقلا معينا Weight يزداد بازدياد قوة الاتصال بين الوحدتين المرتبطتين عبر هذه الوصلة. أي أنه كلما كان الاتصال قويا بين وحدتين ما فإن الثقل بينهم يزداد. توضع المعلومات التي نريد معالجتها عند الطبقة الأولى من الوحدات، وقد يكون خرج كل عصبون دخلا لعصبون آخر، كما تملك كل وحدة دخلا وهميا تساوي قيمته الواحد، ينتقل عبر وصلة مثقلة بثقل بدئي كذلك كما هو موضح بالشكل رقم (٧).



شكل ٧. بنية خلية عصبية.

بعدها يتم بناء الشبكة عبر ربط عدد من هذه العصبونات ببعضها، وهنا يمكن تمييز طريقتين مختلفتين للقيام بذلك: ١) شبكة التغذية إلى الأمام: وفيها تستقبل كل وحدة المعلومات من الوحدات السابقة وتوصلها للوحدات التالية فلا يمكن أن تعود المعلومات بالاتجاه المعاكس أبدا. ٢) شبكة متكررة: وفيها بعد أن تتم معالجة المعلومات يتم إعادة إدخالها من جديد إلى الوحدات لتتم معالجتها مرة أخرى بشكل ارتجاعي.

لاحظ أنه وفي هذه الحالة، فإنها تشبه الذاكرة قصيرة الأمد في دماغنا. عادة يتم ترتيب الوحدات في طبقات، إذ تستلم الوحدات المعلومات من الوحدات في الطبقة السابقة. قد تتكون الشبكة من طبقة من الوحدات تصل الدخل بالخرج وتسمى بالطبقات الخفية، وفي الغالب كلما زاد عدد الطبقات الخفية كلما زاد أداء عمل الشبكة وقلت سرعتها. في هذه الورقة استخدم الخوارزمية MLPClassifier لتصنيف الآداء الأكاديمي للطلاب.

٢- التجارب والمقارنات

تم تحليل النتائج وتدريب النظم الخبير المقترحة بالاختبار على أداة Python مفتوحة المصدر والتى تعد الأداة الأولي من حيث الانتشار والتأثير والمتخصصة في علوم البيانات والذكاء الاصطناعي وتعلم الآلة خصوصا. لتقييم النماذج والمقارنات والنتائج المقترحة تم تقسيم البيانات الى مجموعتين: الأولى خاصة بالتدريب والثانية خاصة بالإختبار وذلك باستخدام (10 folds cross) لأن هذا الخيار يستخدم على نطاق واسع، خاصة إذا كان لدينا كمية محدودة من مجموعة البيانات. تنقسم مجموعة البيانات بشكل عشوائي إلى عشرة مجموعات فرعية. تستخدم أداة Python المجموعة ١ لغرض الاختبار والمجموعات التسع المتبقية لغرض التدريب. تكرر تلك العملية عشر مرات بحيث يتم وضع كل مجموعة من المجموعات العشرة في خانة الاختبار والمجموعات المتبقية في خانة التدريب. في النهاية، يتم احتساب متوسط ناتج المجموعات العشر. تتم تلك العملية للأربع خوارزميات المقترحة والمقارنة من حيث الكفاءة بينهم.

٢-١ *طريقة التقييم*

لتقييم جودة تقنيات التصنيف المختلفة المطبقة على نموذج الأداء الأكاديمي للطلاب، نستخدم ثلاثة مقاييس مختلفة هي (recall, precision, and f–measure) وهو ما يطلق عليها مقاييس مصفوفة التشتت. يوضح الجدول ٢ والمعادلات ٤،٥،٦ كيفية حساب تلك المقاييس.

جدول ٢. مقاييس مصفوفة التشتت.

		Predicted	
		Yes	No
Actual	Yes	TP	FN
	No	FP	TN

$$Recall = \frac{TP}{TP+FN} \quad (٤)$$

$$Percission = \frac{TP}{TP+FP} \quad (٥)$$

$$F-measure = \frac{2\,Percission*Recall}{Percission+Recall} \quad (٦)$$

حيث يدل TP على التصنيف الصحيح الايجابي (أي أن النتيجة الحقيقية هي ايجابي والقيمة المتوقعة كانت ايجابي أيضا)، TN فإنه يدل على التصنيف الصحيح السلبي (أي أن النتيجة الحقيقية هي سلبي والقيمة المتوقعة كانت سلبي أيضا)، FN فإنه يدل على التصنيف الخاطئ السلبي (أي أن النتيجة الحقيقية هي سلبي والقيمة المتوقعة كانت ايجابى). FP فإنه يدل علي التصنيف الخاطيء الايجابي (أي أن النتيجة الحقيقية هي ايجابي والقيمة المتوقعة كانت سلبي). يتم احتساب معاملات التقييم الأربع بناء علي هذه

البيانات. في مشكلتنا نمتلك ثلاث قيم للتصنيف: هي A, B, and C. فتصبح مصفوفة التشتت كما هو مبين بالجدول رقم ٣.

جدول ٣. مصفوفة التشتت لعدد ٣ أصناف.

		Predicted		
		A	B	C
Actual	A	TP_A	Q_{AB}	Q_{AC}
	B	Q_{BA}	TP_B	Q_{BC}
	C	Q_{CA}	Q_{CB}	TP_C

بالنسبة لإجمالي قيم التصنيف السالبة الخاطئة (FN) تعتبر إضافة جميع القيم في الصف المعني باستثناء القيم الإيجابية الموجبة (TP). أما بالنسبة لإجمالي قيم التصنيف الإيجابية الخاطئة (FP) تعتبر إضافة جميع القيم في العمود المعني باستثناء القيم الإيجابية الموجبة (TP). وأخيرا إجمالي القيم السلبية الحقيقية (TN) لأي فئة هو إضافة جميع الأعمدة والصفوف باستثناء الصف والعمود من تلك الفئة. يظهر في المعادلات ٨، ٩، ١٠، ١١، ١٢، ١٣، ١٤ حساب مصفوفة التشتت بشكلها الجديد.

$$\boldsymbol{Recall_A} = \frac{\boldsymbol{TP_A}}{\boldsymbol{TP_A + Q_{AB} + Q_{AC}}} \quad (٧)$$

$$\boldsymbol{Recall_B} = \frac{\boldsymbol{TP_B}}{\boldsymbol{TP_B + Q_{BA} + Q_{BC}}} \quad (٨)$$

$$\boldsymbol{Recall_C} = \frac{\boldsymbol{TP_C}}{\boldsymbol{TP_C + Q_{CA} + Q_{CB}}} \quad (٩)$$

$$\boldsymbol{Percission_A} = \frac{\boldsymbol{TP_A}}{\boldsymbol{TP_A + Q_{BA} + Q_{CA}}} \quad (١٠)$$

$$\boldsymbol{Percission_B} = \frac{\boldsymbol{TP_B}}{\boldsymbol{TP_B + Q_{AB} + Q_{CB}}} \quad (١١)$$

$$\boldsymbol{Percission_C} = \frac{\boldsymbol{TP_C}}{\boldsymbol{TP_C + Q_{AC} + Q_{BC}}} \quad (١٢)$$

$$\boldsymbol{F - measure} = \frac{\boldsymbol{2\ Percission * Recall}}{\boldsymbol{Percission + Recall}} \quad (١٣)$$

٢-١-١ التقييم

تم فحص النتائج المختلفة بناءً على الأربعة خوارزميات الخاصة بعملية التصنيف والتي تم تطبيقها على مجموعة بيانات الطلاب بغرض التنبؤ بالأداء الأكاديمي لهم. يوضح الجدول ٤، ٥، ٦، ٧ مقاييس مصفوفة التشويش المختلفة لتلك الأربعة نظم الخبيرة.

من النتائج الموضحة أعلاه يتضح لنا أن آداء خوارزمية الشبكة العصبونية وشجرة القرار كان الي حد ما جيد وكانت نسب التقييم على التوالى هي: ٧٨٪، ٧٩٪. ربما يرجع عدم جودة دقة الشبكة العصبونية الي تقارب قيم السمات/الخصائص الي حد كبير في القيم الخاصة بهم، أو قلة كمية البيانات (عدد الطلاب) التي تدرب عليها الخوارزمية. من ناحية أخرى، يتضح لنا أن آداء خواريزم نايف بايز كان أعلي منها حيث وصلت دقته الى ٨١٪ وهي دقة جيدة الي حد ما. ومن المعلوم أن خواريزم نايف بايز يضع في اعتباره مبدا الاستقلالية (استقلالية السمات/الخصائص) عند تعامله مع خصائص البيانات. وفي النهاية نجد الآداء المتميز للغابة العشوائية حيث وصلت دقة النظام الخبير المعتمد في آلية عملة على XGBoost، وذلك للتنبؤ بالسلوك الأكاديمي للطلاب الي ٨٥,٥٪، وهي الي حد ما نسبة مرضية جدا. من المعلوم أن الغابة العشوائية تعتمد في آلية عملها على مئات أو آلاف من أشجار القرار المستقلة في آلية عملها. تعمل كل شجرة من تلك الأشجار على كمية محددة من البيانات. في النهاية نقوم بأخذ عدد الأصوات وفي النهاية يتم اعتماد رأى

الأغلبية. اعتماد خوارزيم الغابة العشوائية علي خوارزميات ضعيفة في الأصل لكن كثيرة جدا، فمن الصعب أن يتفقوا فيما بينهم على اختيار خاطئ.

يوضح الشكل رقم (٨) مقارنة بين الخوارزميات الأربع المقترحة بعرض قيم مقاييس التشتت الثلاث المختلفة.

جدول ٤. مصفوفة التشتت لخوارزيم نايف بايز.

		Predicted		
		precision	recall	f1-score
Actual	A	0.82 %	0.89 %	0.85 %
	B	0.79 %	0.86 %	0.83 %
	C	0.82 %	0.72 %	0.77 %
	Avg. total	0.81 %	0.81 %	0.81 %

جدول ٥. مصفوفة التشتت لخوارزيم شجرة القرار.

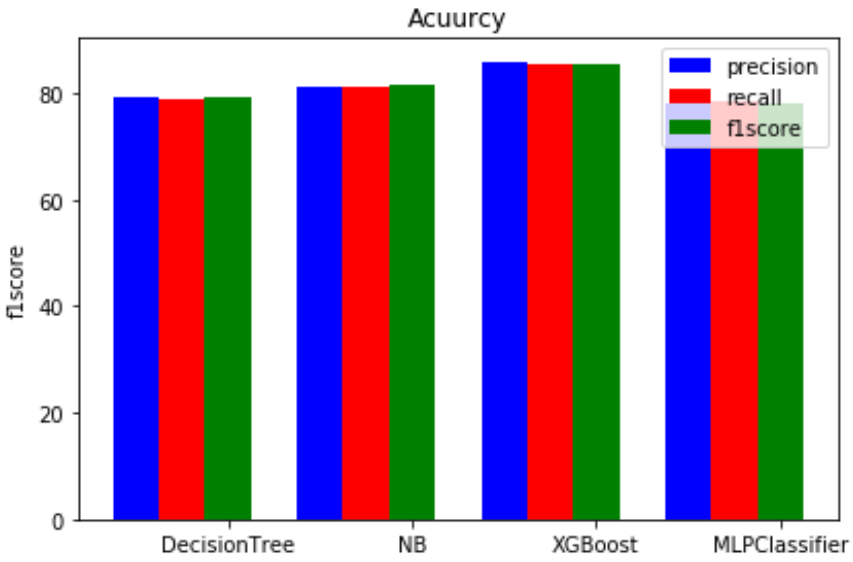
		Predicted		
		precision	recall	f1-score
Actual	A	0.76 %	0.89 %	0.82 %
	B	0.83 %	0.91 %	0.87 %
	C	0.81 %	0.64 %	0.71 %
	Avg. total	0.79 %	0.79 %	0.79 %

جدول ٦. مصفوفة التشتت لخوارزيم الغابة العشوائية.

		Predicted		
		precision	recall	f1-score
Actual	A	0.90 %	0.80 %	0.85 %
	B	0.85 %	1.00 %	0.92 %
	C	0.82 %	0.82 %	0.82 %
	Avg. total	0.86 %	0.85 %	0.85.5 %

جدول ٧. مصفوفة التشتت لخوارزيم الشبكة العصبونية.

		Predicted		
		precision	recall	f1-score
Actual	A	0.76 %	0.80 %	0.78 %
	B	0.84 %	0.95 %	0.89 %
	C	0.76 %	0.67 %	0.71 %
	Avg. total	0.78 %	0.78 %	0.78 %



شكل ٨. قيم الدقة للخوارزميات الأربع المقترحة.

٣- الخلاصة

يمثل الأداء الأكاديمي للطلاب مجال اهتمام كبير لجميع المؤسسات الأكاديمية في جميع أنحاء العالم، حيث يعتبر دعامة لمستقبل الطلاب والمؤسسات الأكاديمية على حد سواء. معظم الدول المتقدمة حولت نظامها التعليمي إلى أنظمة مميكنة بالكامل أو بشكل جزئي لأن هذا النظام يولد ويحتفظ بكمية هائلة من البيانات التي تحتوي على معارف وأنماط مخفية يمكن استخدامها والتنقيب فيها لاحقا لانتاج نظم خبيرة قادرة على تطوير العملية التعليمية برمتها. حيث أن هذه النظم تعتبر بمثابة منبه للطالب، فهي قد تساعدهم على تحسين الدرجات والإنجازات الأكاديمية في المستقبل. في هذا البحث، قدم عددًا من النظم الخبيرة المبنية على خوارزميات تعلم الآلة للتنبؤ بالأداء الأكاديمي للطلاب بالاعتماد على البيانات الأكاديمية والاجتماعية وبعض البيانات الإحصائية الخاصة بهم. تم بناء النظم الخبيرة المقترحة للتنبؤ بالأداء الأكاديمي للطلاب بالاعتماد على أربعة خوارزميات خاصة بتعلم الآلة: وهي خوارزمية نايف بايز، خوارزمية شجرة القرار، خوارزمية الغابة العشوائية، خوارزمية الشبكات العصبية. أظهرت النتائج أن هذه الميزات لها تأثير قوي على النجاح الأكاديمي للطالاب. وكان أفضل أداء في التصنيف من نصيب خوارزم الغابة العشوائية XGBoost حيث وصل الي دقة ٨٥,٥%.

في الأعمال المستقبلية، سيتم الإعتماد في حل هذه المشكلة على خوارزميات التعلييم العميق لنصل الي دقة اعلي ولكن قد يتطلب الأمر الي توفير أو الاعتماد على قاعدة بيانات كبيرة، حيث أن كفاءة

خوارزميات التعليم العميق لا تظهر إلا مع البيانات ذات الأحجام الكبيرة والسمات/الخصائص العديدة.

المراجع

[1] **Astin (**1993**)**. *What Matters in College*? Four Critical Years Revisited, Jossey-Bass, San Francisco.

[2] **Cabrera, A.F., Nora, A.** and **Castaneda**, **M.A.** (1993). College persistence: structural equations modeling test of an integrated model of student retention, *Journal of Higher Education*, **64** (2): 123– 139.

[3] **Caison, A.L.** (2007)**.** Analysis of institutionally specifi c retention research: a comparison between survey and institution adatabase methods, *Research in Higher Education*, **48** (4): 435–449.

[4] **Gansemer-Topf, A.M.** and **Schuh**, **J.H.** (2006)**.** Institutional selectivity and institutional expenditures: examining organizational factors that contribute to retention and graduation, *Research in Higher Education* **47** (6): 613–642.

[5] **Amrieh, E. A., Hamtini, T.** and **Aljarah, I.** (2015). Preprocessing and analyzing educational data set using X–API for improving student's performance. In *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)* (pp. 1–5). IEEE.

[6] **Arsad, P. M.** and **Buniyamin, N.** (2013). A neural network students' performance prediction model (NNSPPM). *In 2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)* (pp: 1–5).

[7] **Breiman, L.** (1999). *Random forests.* UC Berkeley TR567.

[8] **Veenstra, C.P.** (2009). A strategy for improving freshman college retention, *Journal for Quality and Participation*, **31** (4): 19–23.

[9] **Chen, T.** and **Guestrin, C.** (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp: 785-794). ACM.

[10] **Hien, N. T. N.** and **Haddawy, P.** (2007). A decision support system for evaluating international student applications. In *2007 37th annual frontiers in education conference–global engineering: knowledge without borders, opportunities without passports* (pp. F2A–1)

[11] **Berger, J.B.** and **Milem**, **J.F.** (1999). The role of student involveme nt and perceptions of integration in a causal model of stu dent persistence, *Research in Higher Education*, **40** (6): 641– 664.

[12] **Berger, J.B.** and **Brax Ton, J.M.** (1998). Revising Tinto's interactionalist theory of student departure through theory elaboration: examining the role of organizational attributes in the persistence process, *Research in Higher Education*, **39** (2): 103– 119 .

[13] **Hermaniwicz, J.C.** (2003). *College Attrition at American Research Universities: Comparitive Case Studies,* Agathon Press, New York.

[14] **Wetzel, J.N.** and **S., D. O'Toole,** (1999). Peterson, Factors affecting student retention probabilities: a case study, *Journal of Economics and Finance*, **23** (1): 45–55.

[15] **Lau, L.K.** (2003), Institutional factors affecting student retention, *Education*, **124** (1): 126–137.

[16] **Mannan, M.A.** (2007), Student attrition and academic and social integration: application of Tinto's model at the university of Papua New Guinea, *Higher Education*, **53** (2): 147–165.

[17] **Mosa, M. A., Hamouda, A.** and **Marei, M.** (2017). Graph coloring and ACO based summarization for social networks. *Expert Systems with Applications*, **74**: 115-126.

[18] **Quadri, M. M.** and **Kalyankar, N. V.** (2010). Drop out feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology*.

[19] **Romero, C.** and **Ventura, S.** (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, **33**(1): 135–146. doi: http:// dx.doi.org/10.1016/j.eswa.2006.04.005

[20] **Romero, C.** and **Ventura, S.** (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **40**(6): 601–618. doi: http://dx.doi.org/10.1109/TSMCC.2010.2053532

[21] **Romero, C., Ventura, S.** and **García, E.** (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, **51**(1): 368–384. doi: http://dx.doi.org/10.1016/j.compedu.2007. 05.016

[22] **Nikam, S. S.** (2015). "Acomparative Study of Classification Techniques in Data Mining Algorithms," *Orient. J. Comput. Sci. Technol*., **8**(1): 13–19.

[23] **Deberard, S.M.** and **Julka, G.I. L.** (2004). Deana, Predictors of academic achievement and retention among college freshmen: a longitudinal study, *College Student Journal*, **38** (1): 66–81.

[24] **Miller, T.E.** and **Herreid**, **C.H.** (2010). Analysis of variables: predicting sophomore persistence using logistic regression analysis at the University of South Florida, *College and University*, **85** (1): 2–11.

[25] **Miller, T.E.** and **Tyree**, **T.M. ()**, Using a model that predicts individual student attrition to intervene with

those who are most at risk, *College and University*, **84** (3): (2009) 12–21.

[26] **Tinto, V.** (1993), *Leaving college: Rethinking the Causes and Cures of Student Attrition*, Second ed. The University of Chicago Press, Chicago.

[27] **Tinto, V.** (1987). Leaving College: Rethinking the Causes and Cures of Student Attrition, University of Chicago Press, Chicago.

Analyzing Students' Academic Performance Using Machine Learning Techniques

Mohamed Atef Mosa

Institute of Public Administration, Department of Information Technology, Riyadh, Saudi Arabia

mosamo@ipa.edu.sa

Abstract. Predicting students 'academic performance is vital to the success of any education system. Attracting students and the way to preserve and elevate them is an essential part of that system. As it affects the university rank, the educational reputation of the institution, and financial support for it. Monitoring student performance and how to maintain and improve them has become one of the most important priorities for decision-makers in higher education institutions. The system of improvement and how to maintain students begins with a comprehensive understanding of the reasons behind their dropping out and delinquency from the educational process. Such an understanding is the basis for accurate prediction of students at risk and thus appropriate intervention to ensure that they are retained. Many data mining techniques are used such as aggregation, classification, regression, and prediction to help decision-makers accomplish their tasks. In this study, a new student performance prediction model was introduced based on features that have a major impact on academic achievement. This study has relied on its mechanism of work on some machine learning algorithms using data collected from the Kalboard 360 e-learning system. The expert system achieved after several practical experiments, analysis, and comparison of several machine learning algorithms with an accuracy of 85.2%.

Keyword: Prediction, Students Educational Data Mapping, Machine Learning, Artificial Intelligence, Random Forest Algorithms.