Dirasat: *Educational Sciences*, Volume **52**, No. 3, 2025, 10870

Deanship of Scientific Research

Predictive Accuracy of Models for Evaluating Student Performance in PISA Mathematics Test in Jordan Using Explanatory Item Response Models and Machine Learning Models: A Comparative Study

Hanan Housny Ahmad Abu Rashed *, ***Nedal Kamal Mohammad Al-Shraifin***

Counseling and Education Psychology, Yarmouk University, Irbid, Jordan

Received: 26/2/2025
Revised: 14/4/2025
Accepted: 29/5/2025
Published: 29/6/2025

* *Corresponding author:*
hanan_housny@yahoo.com

Citation: Abu Rashed, H. H. A., & Al-Shraifin, N. K. M. (2025). Predictive Accuracy of Models for Evaluating Student Performance in PISA Mathematics Test in Jordan Using Explanatory Item Response Models and Machine Learning Models: A Comparative Study. *Dirasat: Educational Sciences*, 10870. https://doi.org/10.35516/Edu.2025.10870

Abstract

Objectives: The study aimed to compare the predictive accuracy of student performance assessment models on the PISA 2022 mathematics test in Jordan using explanatory item response models (EIRMs) and machine learning models: Random Forest, Artificial Neural Networks, Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor.

Methods: A descriptive analysis method was used, based on data from 7,799 Jordanian students randomly selected from 260 schools that participated in the test. Ten-fold cross-validation was employed to compare model predictive accuracy. Predictive variables included item difficulty, and student-related factors: gender, supervisory authority, socioeconomic status, bullying, use of digital applications outside school, availability of internet-connected devices in schools, teachers' digital skills, and use of digital resources in math classes.

Results: The Naïve Bayes model achieved the highest predictive accuracy (0.718), while the EIRM showed strong discriminatory power with an AUC of 0.693, outperforming machine learning models in distinguishing between student responses. Item difficulty emerged as the most influential predictor.

Conclusions: The study recommends further research incorporating new variables and broader application of the studied predictive models to other assessments or countries to validate and generalize findings, and to explore additional machine learning techniques.

Keywords: Explanatory Item Response Models, Machine Learning Models, Predictive Accuracy, PISA test.

الدقة التنبؤية لنماذج تقييم أداء الطلبة في اختبار بيزا في الرياضيات في الأردن باستخدام نماذج الفقرة التفسيرية ونماذج التعلّم الآلي: دراسة مقارنة

حنان حسني أحمد أبو راشد* ، نضال كمال محمد الشريفين

قسم الإرشاد النفسي والتربوي، كلية العلوم التربوية، جامعة اليرموك، إربد، الأردن

ملخّص

الأهداف: هدفت الدراسة إلى بناء نماذج تنبؤية لتقييم أداء الطلبة الأردنيين في اختبار بيزا في الرياضيات في دورة 2022 ، ومقارنة الدقة التنبؤية لها باستخدام نماذج استجابة الفقرة التفسيرية، ونماذج التعلّم الآلي: الغابة العشوائية، الشبكات العصبية الاصطناعية، نايف بييز، آلة دعم المتجه، الجوار الأقرب-كي.

المنهجية: تم استخدام المنهج الوصفي التحليلي؛ لتحقيق أهداف الدراسة، واستخدام بيانات بيزا 2022 لاختبار الرياضيات في الأردن، وتكونت عينة الدراسة من نتائج (7799) طالبا وطالبة، تم اختيارهم من (260) مدرسة بشكل عشوائي شاركوا في الاختبار، وتم استخدام طريقة الصدق التقاطعي بتكرار كي = 10 لمقارنة الدقة التنبؤية للنماذج ، واستُخدمت المتغيرات التنبؤية من جانب الفقرة (صعوبة الفقرة)، ومن جانب الطلبة: (الجنس، السلطة المشرفة، المستوى الاقتصادي والاجتماعي للأسرة، التعرض للتنمر، استخدام التطبيقات الرقمية خارج المدرسة، توافر أجهزة رقمية موصولة بالإنترنت في المدرسة، مهارات المعلمين الرقمية، استخدام المصادر الرقمية في حصص الرياضيات.

النتائج: أظهرت النتائج أن نموذج نايف بييز سجل أعلى مؤشر دقة بين النماذج المستخدمة بدقة بلغت (0.718) في التنبؤ باستجابات الطلبة، وكان نموذج استجابة الفقرة التفسيرية (EIRM) من أفضل النماذج في قدرته على التمييز بين استجابات الطلبة حيث بلغ مؤشر (AUC) له (0.693)، وتفوّق على نماذج تعلّم الآلة من هذه الناحية، وكان لمتغير صعوبة الفقرة أكبر تأثير في التنبؤ باستجابات الطلبة.

الخلاصة: توصي الدراسة بإجراء مزيد من الدراسات البحثية بمتغيرات جديدة، وتوسيع نطاق تطبيق النماذج التنبؤية التي توصلت إليها الدراسة ليشمل اختبارات أخرى أو دول أخرى لتعميم النتائج واختبار نماذج جديدة من نماذج التعلم الآلي.

الكلمات الدالة: نماذج استجابة الفقرة التفسيرية، نماذج التعلّم الآلي، الدقة التنبؤية، اختبار بيزا.

مقدمة

يمتاز الوقت الحالي بضخامة البيانات وسهولة الحصول عليها، بالإضافة إلى غزو التكنولوجيا لحياة الأفراد والمجتمعات بشكل فاق التوقعات وخصوصًا تطبيقات الذكاء الاصطناعي، وربما تكون طريقة التعامل مع هذه البيانات وطرائق تحليلها أهمّ بكثير من كمية البيانات المطروحة، وإمكانية الحصول عليها، فالهدف يكمن في كيفية الاستفادة من هذه البيانات في صنع القرارات والتخطيط والتنبؤ بالمستقبل، ويعدُّ استخدام هذه البيانات في التنبؤ أمرًا منطقيًا يمكن اعتباره جزءًا من التخطيط للحياة في جميع المجالات وعلى كافة المستويات، وخاصةً مجال التعليم الذي يُعدُّ من أكثر المجالات تأثيرًا في تقدّم الحياة وازدهارها.

تعتبر التقييمات الدولية بما فيها البرنامج الدولي لتقييم الطلبة (PISA) مولّدة لبياناتٍ حول تحصيل الطلبة وعلاقتها مع بعض المتغيرات مثل خلفية الطالب الاجتماعية والاقتصادية وخصائص المدرسة وخصائص المنزل. (Anderson et al., 2007) كما أن المصدر الغني للبيانات التي يقدمها البرنامج من استجابات للطلبة ومعلومات حول المعلمين والمنزل والطالب والمدرسة يساعد في فهم نتائج أداء الطلبة والتنبؤ بها في التقييم التعليمي وتحسين استراتيجيات التدريس. (Park et al., 2023)

ويُعدُّ البرنامج الدولي لتقييم الطلبة (PISA) أداةً تُستخدم لتقييم أداء الطلبة في مناطق كثيرة حول العالم أطلقته منظمة التعاون الاقتصادي والتنمية (OECD) كردّ على الحاجة إلى أدلة قابلة للمقارنة دوليًا حول أداء الطلبة، وإتاحة الفرصة للدول المشاركة أن تراقب وتقيّم مخرجات أنظمتها التعليمية من خلال قياس مهارات وأداء الطلبة ضمن إطارٍ يتم الاتفاق عليه بين الدول المشاركة من أجل إصلاح التعليم، وهو تقييم يتم كلَّ ثلاث سنوات، تم إطلاقه سنة (1997) لأول مرة للطلاب البالغين من العمر 15 سنة، والذين يقتربون من نهاية تعليمهم الإلزامي من جميع أنحاء العالم لتقييم مدى اكتسابهم للمعرفة والمهارات الأساسية الضرورية للمشاركة الكاملة في الحياة الاجتماعية والاقتصادية، كما أنه يدرس مدى قدرة الطلبة على استقراء ما تعلموه، وإمكانية تطبيق تلك المعرفة في بيئاتٍ غير مألوفة داخل المدرسة وخارجها، ويعكس هذا النهج حقيقةً مفادها أن الاقتصادات الحديثة لا تكافئ الأفراد على ما يعرفونه؛ بل على ما يمكنهم فعله بما يعرفونه. (OECD, 2023a).

وفي كل دورة يتم التركيز على إحدى المجالات الأساسية وهي (الرياضيات، القراءة، والعلوم) حيث يستحوذ المجال الذي يتم التركيز عليه على ما يَقرُب من نصف إجمالي وقت الاختبار، وكانت الرياضيات هي المجال الرئيسي للتقييم في عام 2022 كما كان الحال في عاميّ 2012 و2003 حيث كانت (70%) من فقرات الاختبار تخصّ الرياضيات، وكانت القراءة هي المجال الرئيسي في الأعوام 2000 و2009 و2018، وكانت العلوم هي المجال الرئيسي في عامي 2006 و2015، ومع هذا الجدول الزمني المتناوب للمجالات الرئيسة يمكن إجراء تحليل شامل يتم تقديم الإنجاز في كل مجال من المجالات الأساسية الثلاثة كلّ تسع سنوات.(OECD, 2023b).

وأشار تقرير المركز الوطني لتنمية الموارد البشرية (2023)، حول مستوى أداء طلبة الأردن في البرنامج أن ترتيب الأردن في الرياضيات كان (73) من بين (81) دولة أو نظام تعليميّ، والترتيب (71) في العلوم والترتيب (78) في القراءة، ويبين الجدول (1) التغيّر في الترتيب الدوليّ والرتب المئينية لطلاب الأردن بين دورتيّ 2018 و2022 حيث يلاحظ تدني الترتيب مقارنة بعام 2018.

الجدول (1): التغيّر في الترتيب الدوليّ والرتب المئينية لطلاب الأردن بين دورتيّ 2018 و2022

	الرياضيات الرتبة المئينية	العلوم الرتبة المئينية	القراءة الرتبة المئينية
2018 (78) دولة	78/65 %17	78/51 %35	78/55 %29
2022 (81) دولة	81/73 %10	81/71 %13	81/78 %4

ويساعد استخدام هذه البيانات بالإضافة إلى استجابات الطلبة في الاختبارات على فهم نتائج أداء الطلبة والتنبؤ بها، وبالتالي تحسين استراتيجيات التدريس والتعليم، وللحصول على تغذية راجعة أكثر دقة ومتانة من نتائج تقييم الطلبة يجب على الباحثين والممارسين التربويين التفكير بشكل نقدي في ما إذا كانت الأساليب الحالية لتحليل البيانات قادرة على استرجاع المعلومات المقدمة في قاعدة البيانات.(Park et al,. 2023) وبما أن البيانات الضخمة (Big Data) أصبحت سمة العصر الحالي، فإن البحث باستخدام البيانات الضخمة ينبع من منظورين هما: البحث المبنيُّ على البيانات (data-driven research) والبحث المبنيُّ على النظرية (theory-driven research)، فالبحث المبنيُّ على البيانات هو نهج استكشافي يقوم بتحليل البيانات لاستخلاص رؤى مثيرة للاهتمام علميًا مثل الأنماط، من خلال تطبيق التقنيات التحليلية المختلفة، وأما البحث القائم على النظرية فهو نهج أكثر تقليدية لإجراء البحث العلميّ الذي يبدأ بتطوير الفرضيات، يليه جمع البيانات وتحليلها لاختبار هذه الفرضيات، واستخلاص

استنتاجات نظرية بناءً على النتائج، وأدرك العلماء أن وجهات النظر البحثية القائمة على البيانات والقائمة على النظرية يجب أن تعزز بعضها البعض في عصر البيانات الضخمة. (Maass, et al., 2018) وتُعدُّ نماذج نظرية استجابة الفقرة (Item Response Theory Models IRT) نماذج مبنية على النظرية (theory-based approach) ونماذج التعلّم الآلي (machine learning) نماذج مبنية على البيانات (data-driven approach) (Park et al., 2023)، وفي ما يلي استعراضٌ عامٌ لنماذج كلا النهجين:

النهج الأول: نماذج مبينة على النظرية: نماذج نظرية استجابة الفقرة (IRT)

في ثمانينات القرن العشرين (1980s) كانت نظرية استجابة الفقرة (IRT) واحدةً من أكثر المواضيع سيطرّة بين أخصائيي القياس؛ لأنها قدمت إطاراً لحل العديد من مشاكل القياس التي عانت منها نظرية القياس التقليدية. (Hambleton et al., 1991) وتتعامل نماذج استجابة الفقرة (IRT) مع البيانات الواردة من أداة مثل اختبار، أو مقياس اتجاهات، أو استبيان، والتي تُسمى بـ "بيانات الاختبار" والغرض الرئيسيّ من استخدام نماذج استجابة الفقرة هو قياس السمات وعلى وجه التحديد: القدرات أو المهارات أو مستويات التحصيل أو الاتجاهات وما إلى ذلك، وفي نمذجة استجابة الفقرة القياسية، يتم نمذجة كل فقرة بمجموعة المعالم الخاصة بها ويتم بها أيضًا تقدير معالم الفرد ويطلق على ذلك نهج "القياس" (Measurement approach) وفي نهج أوسع، لا يتعارض مع نهج القياس ولكنه يكمّله، قد يرغب المرء في وضع نموذج يبين كيف أن خصائص الفقرات وخصائص الأفراد تؤدي إلى استجابات الأفراد على الفقرات ويطلق على هذا النهج "بالنهج التفسيري" (Explanatory approach) حيث أن فهم كيفية إنشاء استجابات الفقرات يعني أنه يمكن تفسير الاستجابات من حيث خصائص الفقرات وخصائص الأفراد. (Wilson et al., 2006)

وقد صاغ دي بويك وويلسون (De Boeck & Wilson, 2004) مصطلح نماذج استجابة الفقرة التفسيرية (Explanatory Item Response Models EIRM) لوصف استخدام نماذج استجابة الفقرة IRT كأداة لكل من القياس والتفسير. وقد تم تقديم نماذج استجابة الفقرة التفسيرية (ERIM) كبديل لنماذج نظرية استجابة الفقرة القياسية (IRT) حيث تسمح هذه النماذج بتحويل نماذج نظرية استجابة الفقرة القياسية (IRT) إلى نماذج تفسيرية بمتنبئات إضافية، وتهدف إلى تفسير كيفية تأثير خصائص الفقرة (مثل: المحتوى، درجة التعقيد المعرفي، وجود عناصر بصرية....الخ) وخصائص الفرد (مثل: الجنس، العرق...الخ) على استجابات الفرد على الفقرة (Bulut, 2020).

وقُدّمت نماذج استجابة الفقرة التفسيرية (EIRM) في إطار إحصائي يسمى بالنماذج الخطية المعمّمة المختلطة (Generalized Linear Mixed Models GLMMs).(McCulloch & Searle, 2001) حيث إن النماذج الخطية المعمّمة المختلطة لها العديد من الفوائد التي تفوق النهج القياسي، وهي واضحةٌ وبسيطة جداً في تصورها الأساسي الذي يستند إلى نموذج الانحدار الخطي المألوف، ووضوحها يعود؛ لأن أوجه الشبة والاختلاف بين النماذج يمكن وصفها من حيث أنواع المتنبئات (على سبيل المثال: متنبئات الأفراد ومتنبئات الفقرة، وتفاعلات متنبئات الفقرة الفرد معاً) بالإضافة إلى نوع الأوزان في المتنبئات للمشاهدات (مثلا: ثابتة أو عشوائية) كالموجودة في نموذج الانحدار المألوف. (De Boeck & Wilson, 2004) وربما كانت الميزة الأكثر أهمية لهذا النهج الأوسع هي أنها تسهل تنفيذ وجهة النظر التفسيرية التي تم وصفها، ومن المزايا المهمة الإضافية أن هذا النهج هو نهجٌ عام، وبالتالي فهو مرن أيضًا، ويربط القياس النفسي بقوةٍ بمجال الإحصاء حيث يوفر قاعدة معرفية وأدب نظري أوسع. (Wilson et al., 2006).

وقد أشار دي بويك وويلسون (De Boeck & Wilson, 2004) إلى أربعة أنواع من نماذج نظرية استجابة الفقرة التفسيرية والجدول (2) يبين هذه النماذج حيث تختلف فيما بينها فيما إذا كانت وصفية أم تفسيرية من جانب الفرد ومن جانب الفقرة، وتختلف فيما يتعلق بأنواع المتنبِئات التي يتم تضمينها، فهناك نوعان من متنبئات الفقرة: مؤشرات الفقرة وخصائص الفقرة، وهناك أيضًا نوعان من متنبئات الفرد: مؤشرات الفرد وخصائص الفرد.

الجدول (2): نماذج استجابة الفقرة التفسيرية EIRM كدوال للتنبؤ

متنبئات الفقرة (Item Predictors)	متنبئات الفرد (Person predictors)	
	غياب الخصائص (مؤشرات الفرد) (person indicators)	تضمين الخصائص (خصائص الفرد) (person properties)
غياب الخصائص (مؤشرات الفقرة) (Item indicators)	النموذج الوصفي المضاعف (Doubly descriptive)	النموذج التفسيري للفرد (Person Explanatory)
تضمين الخصائص (خصائص الفقرة) (Item properties)	النموذج التفسيري للفقرة (Item Explanatory)	النموذج التفسيري المضاعف (Doubly Explanatory)

النهج الثاني: نماذج مبنية على البيانات: نماذج التعلّم الآلي (Machine Learning Models)

أشار البيدن (Alpaydin, 2005) أن الفكرة الرئيسة للتعلّم الآلي أن الحاسوب يستطيع أن يتعلّم أتوماتيكياً من الخبرة. لذلك تعتبر نماذج التعلّم الآلي مبنية على البيانات (data-driven approach) لتطوير خوارزمية تنبؤية دقيقة وفعّالة من الناحية الحسابية. (Shmueli, 2010) وتضم عملية التعلّم مجموعة بيانات (DataSet) مكوّنة من مخرجات ملاحظة ومتنبئات وخوارزمية تحدد العلاقة بين المتنبئات والمخرجات.(Gonzalez, 2021) وأشار بيشوب (Bishop, 2006) إلى أن التعلّم الآلي ينقسم بشكل عام إلى نوعين هما:

الأول: التعلّم الخاضع للإشراف (Supervised learning) و يُطلق عليه أيضاً التصنيفيّ أو التنبؤيّ والذي يركز على أنماط التعلّم عبر ربط العلاقة بين المتغيرات والمخرجات المعروفة، ويعمل مع مجموعات بيانات مصنفة، كما يعمل على تغذية الآلة بعينة بيانات بسمات مختلفة، وقيمة المخرجات الصحيحة للبيانات وتكون قيم المخرجات وقيم السمات معروفة، وتؤهل قاعدة البيانات لتكون "مصنفة"، بعد ذلك تقوم الخوارزمية بفك تشفير الأنماط الموجودة في البيانات، وإنشاء نموذج يمكنه إعادة إنتاج نفس القواعد الأساسية باستخدام بيانات جديدة، وتأتي البيانات المدخلة مع بنية فئة معروفة (Mohri et al., 2012). وتُعرف بيانات الإدخال هذه ببيانات التدريب، وتكون مهمة الخوارزمية إنشاء نموذج يمكنه التنبؤ بإحدى الخصائص باستخدام خصائص أخرى، وبعد إنشاء النموذج يتم استخدامه لمعالجة البيانات التي لها نفس بنية الفئة الخاصة ببيانات الإدخال ومن أمثلتها: الغابة العشوائية (Random Forest RF)، الجوار الأقرب-كي (K-Nearest Neighbors K-NN)، الشبكات العصبية الاصطناعية (Artificial Neural Networks ANN)، آلات دعم المتجه (Support Vector Machines SVM)، ناييف بيز (Naïve Bayes NB).

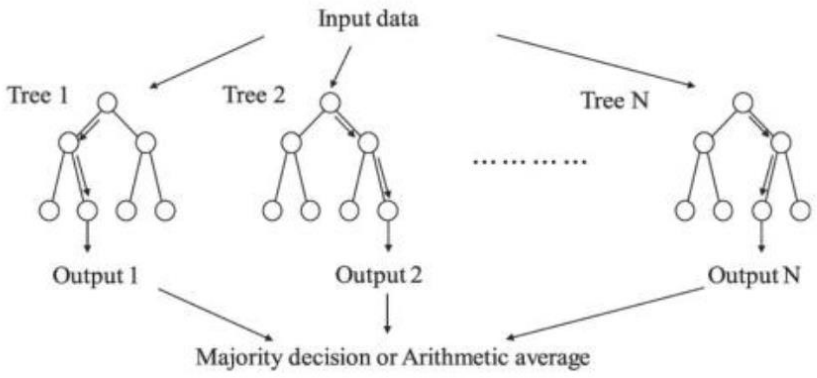
الثاني: التعلّم غير الخاضع للإشراف (Unsupervised learning) و يُطلق عليه أيضاً التجميعيّ أو الوصفيّ، في هذا النوع لا يتم تصنيف كل المتغيرات وأنماط البيانات، وبدلًا من ذلك، يجب على الآلة أن تكشف عن الأنماط المخفية، وإنشاء تسميات من خلال استخدام خوارزميات التعلّم غير الخاضع للاشراف، فلا تحتوي البيانات المدخلة على بنية فئة معروفة، وتكون مهمة الخوارزمية الكشف عن بنية في البيانات. (Sugiyama, 2015) مثل خوارزمية تجميع متوسطات-كي (K-Means Clustring).

ويتم تقسيم البيانات في التعلّم الآلي إلى قسمين: بيانات التدريب (training data) وبيانات الاختبار(testing data)، حيث تُستخدم بيانات التدريب لتطوير النموذج، ويتم إضافة قواعد أو تعديلات معينة حسب ما تقتضيه الحاجة، وبعد النجاح في تطوير النموذج بناء على بيانات التدريب والرضا عن مدى دقته، يتم اختبار النموذج على البيانات المتبقية والمعروفة باسم بيانات الاختبار، وبمجرد الرضا عن بيانات التدريب وبيانات الاختبار يصبح نموذج التعلّم الآلي جاهزاً لتصفية البيانات الجديدة واتخاذ القرارات بشأن تصنيفها ويكون جاهزاً للتطبيق على بيانات العالم الحقيقي. (Bishop, 2006)

خوارزميات التعلم الخاضع للإشراف (Supervised Learning)

- **الغابة العشوائية (RF):**

وهي عبارة عن تجمّع من تنبؤات أشجار القرار، حيث أن كل شجرة تعتمد على قيم متجه عشوائي يتم أخذ عينات منه بشكل مستقل وبنفس التوزيع لكلّ الأشجار في الغابة، وبعد إنشاء عدد كبير من الأشجار يتم التصويت للفئة الأكثر شعبية وتسمى هذه الإجراءات بالغابة العشوائية. (Breiman, 2001) وللغابة العشوائية القدرة على التعامل مع مشاكل الانحدار والتصنيف، وتستخدم عدة نماذج للتعلّم الآلي لعمل التنبؤات، وتستخدم الغابة العشوائية أشجار قرار متعددة. (Tomar & Verma, 2021) ويوضح الشكل (1) تمثيلًا مرئيًا للغابة العشوائية.



الشكل (1): تمثيل مرئي للغابة العشوائية (RF)

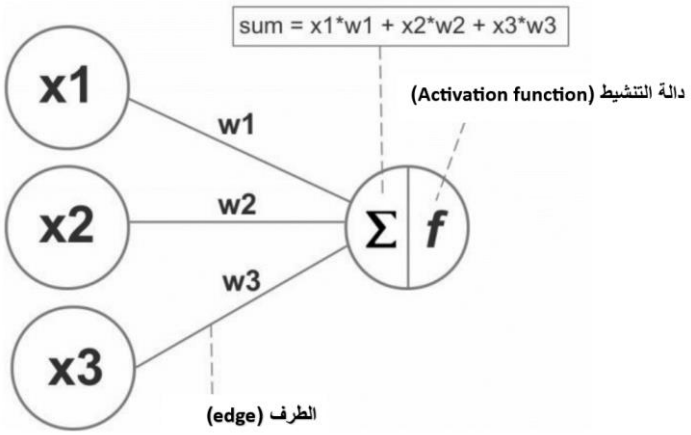
- **الجوار الأقرب – كي (K-NN):**

على النقيض من أساليب التعلّم القائمة على النموذج، والتي تستنتج دالّة التنبؤ من بيانات التدريب، يتم تصنيف خوارزمية (KNN) على أنها خوارزمية تعلّم كسولة، فهي تشكل التنبؤات من خلال تحليل بنية البيانات في الوقت الفعلي عند تقديم حالات جديدة دون الحاجة إلى مرحلة تدريب صريحة سابقة، وتعمل على مبدأ احتمالية التشابه، فهي تفترض أن نقاط البيانات المتشابهة تميل إلى التجمّع بالقرب من بعضها في الفضاء، وبالتالي فإن التنبؤ يمثل بيانات جديدة تعتمد على قربها من الحالات الموجودة في مجموعة التدريب، وخطوات خوارزمية (KNN) تتضمن:

1. تحديد عدد الجيران (K) وهو مَعلمة فرعية بالغة الأهمية يمكن تعديلها بناء على الخصائص المحددة لمجموعة البيانات، وتحديد القيمة المثلى لعدد الجيران (K) ضروري جدًا لدقة التنبؤ.
2. حساب المسافات: يجب حساب المسافة بين النقطة الجديدة وجميع النقاط في مجموعة البيانات التدريبية، وتشمل المسافات الشائعة: المسافة الإقليدية، ومانهاتن ومينوفسكي.
3. تحديد أقرب الجيران: حيث يتم ترتيب النقاط في مجموعة التدريب من الأقرب إلى الأبعد عن النقطة الجديدة. (Halder et al., 2024)

- **الشبكات العصبية الاصطناعية (ANN):**

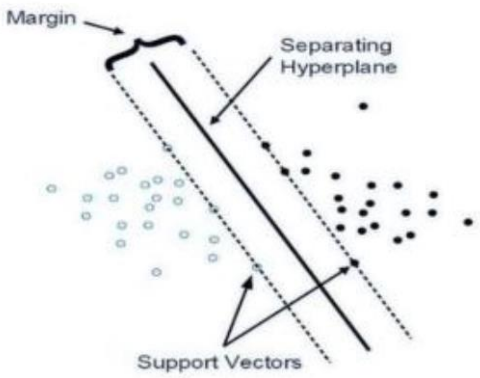
تتكون الشبكات العصبية الاصطناعية على غرار الخلايا العصبية في الدماغ البشري من خلايا عصبية مترابطة تُسمى بالعقد (nodes)، وتتفاعل عبر محاور عصبية تسمى الأطراف (edges). (Haykin, 2009) ويتم تجميع العقد في طبقات، وتبدأ بشكل عام بقاعدة عريضة، وتتكون الطبقة الأولى من البيانات الأولية ويتم تقسيمها إلى عقد حيث تقوم كل عقدة بإرسال المعلومات إلى الطبقة التالية من العقد عبر أطراف الشبكة، ويكون لكل طرف وزن رقمي يمكن أن يتغير أو يصاغ اعتمادًا على الخبرة، وإذا كان مجموع الأطراف المتصلة يفي بالحد الأدنى المحدّد، ويعرف باسم دالة التنشيط (activation function)، فسيتم تنشيط خلية عصبية في الطبقة التالية، وإذا لم يصل مجموع الأطراف المتصلة إلى الحد المحدّد فلن يتم تنشيط الخلية العصبية، والأوزان على طول الحواف فريدة لضمان أن العقد تطلق الإشارات بشكل مختلف كما هو مبين في شكل (2)، ولا ترجع جميعها نفس النتائج. (Theobald, 2017)



الشكل(2): تمثيل مرئي للشبكة العصبية الاصطناعية

- **آلة دعم المتجه (SVM):**

تم تقديم مبدأ آلة دعم المتجه (SVM) من قبل فابنك (Vapnik, 1995) في أوائل السبعينيات 1970's، ومنذ ذلك الوقت تم تطبيق آلة دعم المتجه في العديد من الحقول المتنوعة، والسبب في انتشارها هي أنها قامت على أساس رياضي صلب مُحاولةً حل مشاكل تصنيف عالمية، فإذا تم إعطاء مجموعة من المتجهات مفصولات بمستوى فائق (hyperline) فإن الخوارزمية تعمل على إيجاد مستوى فاصل بأعلى هامش وبدقة كبيرة، وتكون المسافة بين المتجهه الداعم الأقرب إلى المستوى الفاصل هي العظمى، ويمكن استخدام المستوى الفاصل لتحديد المجموعة التي يعود إليها المتجه غير المصنف، ويُعدُّ مستوى الفصل الأفضل (Optimal Separating hyperplane) هو ذلك المستوى الذي يفصل البيانات بأعلى هامش(Maximal Margin)، ويوضح شكل (3) مثالًا لنقاط في فضاء ثنائي الأبعاد، حيث يتم فصل المجموعات المختلفة عن طريق منطقة تسمى الهامش (margin). (Moukhafi et al., 2020)



الشكل(3): تمثيل مرئي يوضح مثال كلاسيكي ل SVM linear classifier

- **نايف بيز (NB)**

تستند خوارزمية نايف بيز إلى قاعدة بيز التي تعتمد على الاحتمالية، وتفترض هذه الخوارزمية افتراضين: الأول هو الاستقلال، وافتراض أن الأحداث مستقلة في الحياة الواقعية هو بالتأكيد افتراض تبسيطي، ولكن بالرغم من اسم الخوارزمية إلا أنها تعمل بشكل جيد للغاية عند اختبارها على مجموعة بيانات فعلية. (Witten & Frank, 2000)، وتعتبر خوارزمية نايف بيز أن كل سمة (متغير) تساهم بشكل مستقل في الاحتمالية.(Ahmed, 2024) وأما الافتراض الثاني فهو أنه لا توجد سمات مخفية يمكن أن تؤثر على عملية التنبؤ، وتمثل هذه الخوارزمية نهجًا واعدًا لاكتشاف المعرفة الاحتمالية، وتوفر خوارزمية فعالة لتصنيف البيانات. (Osmanbegovic & Suljic, 2012)

ويمكن استخدام نظرية بيز (Bayes' Theorem) للتنبؤ بناءً على المعرفة السابقة والأدلة الحالية. (Efron, 2013) وأشار زهانغ (Zhang, 2016) أنه مع تراكم الأدلة فإن التنبؤ يتغير، والتنبؤ هو الاحتمال اللاحق الذي يكون موضع الاهتمام، والمعرفة السابقة تُسمى بالاحتمالية السابقة التي تعكس التخمين الأكثر احتمالية للنتيجة دون أدلة إضافية، ويتم التعبير عن نظرية بيز بالمعادلة الآتية:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

حيث: P(A) و P(B) هما احتمالات الحدثين (A) و (B) دون اعتبار كل منهما للآخر، وأما $P(A|B)$ فهي احتمال (A) المشروط بـ (B) و $P(B|A)$ هو احتمالية (B) المشروط بـ (A)، وتعتبر احداث (A) فئوية وأما (B) فهي سلسلة من المتنبئات المستقلة عن بعضها البعض والمشروطة بنفس قيمة النتيجة، لذلك يمكن كتابة $P(b1, b2, b3|A)$ على النحو الآتي: $P(b1|A) \times P(b2|A) \times P(b3|A)$.

مشكلة الدراسة وأسئلتها

من خلال استعراض الدراسات السابقة والأدب النظري لُوحظ عدم وجود دراسات على مستوى الأردن تناولت مقارنة الدقة التنبؤية لنماذج استجابة الفقرة التفسيرية ونماذج التعلّم الآلي لاستجابات الأفراد على البرنامج الدولي لتقييم الطلبة PISA 2022 – في حدود علم الباحثة- مع وجود العديد من الدراسات التي تناولت استخدام نماذج التعلّم الآلي للتنبؤ بتحصيل الطلبة أو التي تناولت مقارنة دقة مجموعة من نماذج التعلّم الآلي للتنبؤ بالتحصيل الأكاديمي للطلاب، أو تلك التي تناولت استخدام نماذج استجابة الفقرة التفسيرية في مجال التنبؤ بتحصيل الطلبة، كما أن هناك ندرة في الدراسات التي قامت باستخدام نتائج بيزا 2022 للطلاب الأردنيين لبناء نماذج تنبؤية تساعد في فهم أسباب تدني ترتيب الأردن في هذا التقييم بناء على متغيرات تخص الأفراد (كالجنس، الوضع الاقتصادي للطالب، المستوى التعليمي للوالدين،...الخ) أو متغيرات تخص الفقرة (كصعوبة الفقرة، والمحتوى....الخ) أو متغيرات تخص كلا الجانبين معًا.

لذا تأتي هذه الدراسة لبناء نماذج تنبؤية باستخدام نماذج استجابة الفقرة التفسيرية ونماذج التعلّم الآلي، ومقارنة الدقة التنبؤية لهذه النماذج لاستجابات الطلبة على فقرات البرنامج الدولي لتقييم الطلبة في الأردن PISA 2022 في الرياضيات بهدف الكشف عن أسباب تدني ترتيب الأردن في اختبار بيزا 2022 والحصول على النماذج التنبؤية الأفضل للتنبؤ بأداء الطلبة.

وحاولت الدراسة الإجابة عن الأسئلة الآتية:

- ما النموذج التنبؤي لتقييم أداء الطلبة في اختبار بيزا في الرياضيات للعام 2022 باستخدام نماذج استجابة الفقرة التفسيرية (النموذج التفسيري المضاعف)؟
- ما النموذج التنبؤي لتقييم أداء الطلبة في اختبار بيزا في الرياضيات للعام 2022 باستخدام نماذج التعلّم الآلي؟
- أيّ النماذج هو الأكثر دقة (فاعلية) في التنبؤ بأداء الطلبة في اختبار بيزا في الرياضيات للعام 2022؟

أهداف الدراسة

هدفت الدراسة الحالية إلى بناء نماذج تنبؤية باستخدام نماذج استجابة الفقرة التفسيرية (النموذج التفسيري المضاعف) ونماذج التعلّم الآلي للتنبؤ بأداء الطلبة، وهي: الغابة العشوائية، الشبكات العصبية الاصطناعية، نايف بيز، آلة دعم المتجه، والجوار الأقرب-كي، ومقارنة الدقة التنبؤية لهذه النماذج لاستجابات الطلبة على فقرات البرنامج الدولي لتقييم الطلبة (PISA) للعام 2022 في الرياضيات لاستخلاص النموذج الأفضل من حيث الدقة التنبؤية، كما هدفت إلى التعرّف إلى المتغيرات الإضافية (المتنبِئات) التي تتنبأ بأداء الطلبة من جانب الفقرة ومن جانب الطالب.

أهمية الدراسة:

تبرز أهمية الدراسة من الناحية النظرية في تناولها لنتائج البرنامج الدوليّ لتقييم الطلبة (PISA) لعام 2022 في الرياضيات، والذي يشكّل أهمية كبيرة في إلقاء الضوء على أسباب تدني أداء الطلبة وتراجعه حيث كان متوسط أداء الطلبة الأردنيين للعام 2022 أقل من أي تقييم لجميع الدورات السابقة التي شارك فيها الأردن في التقييم منذ العام 2006، مما يضع الجميع موضع مسؤولية عن هذا التدني الحاد.

كما تُكمن أهمية هذه الدراسة أنها تناولت بناء نماذج تنبؤية لاستجابات الأفراد باستخدام نماذج استجابة الفقرة التفسيرية (EIRM)، حيث إن الاهتمام في هذه الدراسة منصبٌّ على النهج التفسيريّ، وليس نهج القياس، وتعد نماذج نظرية استجابة الفقرة التفسيرية (EIRM) النماذج الأنسب التي تهدف إلى تفسير كيفية تأثير خصائص الفقرة وخصائص الفرد على استجابات الفرد على فقرات الاختبار كما تهدف إلى التنبؤ بالمعالم وتفسيرها إما من جانب الفرد أو من جانب الفقرة أو من كلا الجانبين.

وتُكمن أيضًا أهمية هذه الدراسة في أنها غطت المنظورين اللذين ينبع منهما البحث في البيانات وهما: البحث المبني على البيانات (data-driven research) والذي تم تطبيقه من خلال نماذج التعلّم الآلي، والبحث المبني على النظرية (theory-driven research) والذي تم تطبيقه من خلال نماذج استجابة الفقرة التفسيرية (EIRM)، والمقارنة بين الدقة التنبؤية لهذه النماذج، حيث أن استخدام المنظورين معًا يعزز بعضه بعضًا في عصر البيانات نظرًا للبيانات الضخمة الناتجة من نتائج البرنامج الدولي لتقييم الطلبة PISA 2022 .

وأما من الناحية العملية، فإن أهمية هذه الدراسة تبرز في توفير نماذج تنبؤية باستخدام نماذج التعلّم الآلي ونموذج استجابة الفقرة التفسيرية؛ ليستفيد منها صنّاع السياسات، ومطوريّ المناهج والمختصين لاتخاذ القرارات المناسبة والإجراءات التي تُعنى برفع سُوية التعليم، ومعالجة الظروف المحيطة التي أدت إلى تدني تحصيل الطلبة في المهارات الأساسية، وبالأخص الرياضيات، ورفع ترتيب أداء الطلبة في التقييمات الدولية والذي سينعكس على أداء الطلبة في التقييمات الوطنية، وبالتالي اللحاق بركب التطورات في المجال التعليمي.

التعريفات الإجرائية

- **نماذج استجابة الفقرة التفسيرية (EIRM):** نماذج إحصائية تفسيرية تهدف للتنبؤ باستجابات الطلبة على فقرات البرنامج الدولي لتقييم الطلبة PISA 2022 في الرياضيات في الأردن باستخدام النموذج التفسيري المضاعف.
- **نماذج التعلّم الآلي**: فرع من فروع الذكاء الاصطناعي تهدف إلى بناء نماذج تنبؤية للتنبؤ باستجابات الطلبة على فقرات البرنامج الدولي لتقييم الطلبة PISA 2022 في الرياضيات في الأردن باستخدام الخوارزميات الآتية: الغابة العشوائية، الشبكات العصبية الاصطناعية ، نايف بيز، آلة دعم المتجه، الجوار الأقرب – كي.
- **اختبار PISA 2022** : هو اختبار لتقييم أداء الطلبة البالغين (15) سنة من العمر في الأردن للعام 2022 لتقييم مدى اكتسابهم للمعرفة والمهارات والكفايات الأساسية الضرورية في مجال الرياضيات والعلوم والقراءة، ويعقد كل ثلاث سنوات ويكون التركيز كل سنة على مجال محدد.
- **الدقة التنبؤية**: مقياس رئيس في تقييم أداء النماذج التنبؤية يقيس مدى قدرة هذه النماذج على التنبؤ بشكل صحيح بالمخرجات استنادًا إلى بيانات الإدخال ومؤشراته هي: الدقة (Accuracy)، الإحكام (Precision)، الاستدعاء (Recall)، مؤشر F1، والمساحة تحت منحنى ROC (AUC).

محددات الدراسة

اقتصرت الدراسة الحالية على:

- طلاب مدارس الأردن بعمر(15) سنة.
- نتائج البرنامج الدولي لتقييم الطلبة PISA 2022 المرتبطة بالسياق الزمني الذي أجري فيه الاختبار وهو العام 2022.
- نتائج دراسة البرنامج الدولي لتقييم الطلبة (PISA) للعام 2022 في مادة الرياضيات، وهو المجال الذي تم التركيز عليه في اختبار PISA 2022.
- المتغيرات التي تخص الطلبة وهي: الجنس (gender)، السلطة المشرفة (strm)، المستوى الاقتصادي والاجتماعي للأسرة (ses)، التعرض للتنمر (bullied)، استخدام التطبيقات الرقمية خارج المدرسة (usingApp)، توافر أجهزة رقمية في المدرسة (DigSchool)، مهارات المعلمين الرقمية (teacherskill)، استخدام المصادر الرقمية في حصص الرياضيات (digitalmath).
- المتغيرات التي تخص الفقرة وهي: صعوبة الفقرة (diff).

منهج الدراسة

تم استخدام المنهج الوصفي التحليلي (التنقيب في البيانات) لهذه الدراسة؛ لأنه الأكثر ملاءمةً لمثل هذا النوع من الدراسات حيث تناولت نتائج التنبؤ التي تم الحصول عليها من خلال نماذج التعلّم الآلي، ونماذج استجابة الفقرة التفسيرية والمفاضلة بينها من حيث الدقة التنبؤية، بالإضافة إلى تحديد أفضل المتنبئات لاستجابات الطلاب في البرنامج الدولي لتقييم الطلاب الأردنيين في الرياضيات PISA 2022.

مجتمع الدراسة

تكوّن مجتمع الدراسة لاختبار PISA 2022 من الطلبة من ذوي الفئة العمرية (15) سنة وثلاثة شهور و(16) سنة وشهرين عند بداية إجراء التقييم (بإضافة أو إزالة شهر) والذين يلتحقون بالصفوف من السابع وأعلى بغض النظر ما إذا كان هؤلاء الطلاب مسجلين بدوام كامل في النظام التعليمي أو دوام جزئي، وقد شارك ما يقارب (700,000) طالب وطالبة من (81) دولة في الاختبار تم اختيارهم بشكل عشوائي من (29) مليون طالب وطالبة في العالم. (OECD, 2023e)

عينة الدراسة

تكونت عينة الدراسة حسب منظمة التعاون الاقتصادي والتنمية (OECD, 2023e) من (7799) طالب وطالبة تم اختيارهم بشكل عشوائي في (260) مدرسة من الأردن شاركوا في البرنامج الدولي لتقييم الطلبة (بيزا) للعام 2022، وأكملوا التقييم في الرياضيات والقراءة والعلوم.

أدوات الدراسة

تكونت أدوات الدراسة من قاعدة بيانات البرنامج الدولي لتقييم الطلاب PISA 2022 والمكونة من مجموعة من الاستبانات كما وردت في قاعدة بيانات منظمة التعاون الاقتصادي والتنمية في تقريرها التقني (OECD, 2023f) وهي: استبيان الطالب واستبيان المعلم واستبيان مدير المدرسة واستبيان ولي الأمر واستبيان المدرسة، واستبيان استخدام تكنولوجيا المعلومات الاتصالات واستبيان المهارات المالية للطالب، واستبيان الصحة النفسية للطالب بالإضافة الى الاختبار.

الخصائص السيكومترية لأدوات الدراسة

تم استخدام معامل كرونباخ ألفا للتحقق من الاتساق الداخلي لكل مقياس داخل البلدان والأنظمة التعليمية المشاركة ومقارنتها بالبلدان المشاركة الأخرى، وتشير القيم المتشابهة والعالية عبر البلدان المشاركة والأنظمة التعليمية إلى اتساق داخليّ عال، حيث تُعدُّ القيمة 0.9 مؤشرًا ممتازًا ، و0.8 مؤشرًا جيدًا، و0.7 مؤشرًا مقبولًا ، وقد بلغ معامل الاتساق الداخلي للفقرات عبر الدول المشاركة 0.95 لمجال الرياضيات و0.95 لمجال القراءة و0.94 لمجال العلوم و0.93 لمجال المهارات المالية و0.86 لمجال التفكير الإبداعي، وتُعدُّ هذه القيم قيمًا تدل على اتساق ممتاز.

الدراسات السابقة

تَبيّن بعد الرجوع إلى الدراسات السابقة أن هناك العديد من الدراسات العربية والأجنبية مرتبطة بموضوع الدراسة الحالية منها:

دراسة باشولي وبورمانس (Pachouly & Bormance, 2025) التي هدفت إلى التنبؤ بأداء الطلاب باستخدام خوارزميات التعلّم الآلي الأتية: آلة دعم المتجه (SVM)، الغابة العشوائية (Random Forest)، الجوار الأقرب كي (KNN)، الشبكات العصبية الاصطناعية (ANN) وخوارزمية أدا بوست

(AdaBoost) كما هدفت إلى استكشاف المتغيرات التي تؤثر على أداء الطلاب، تم استخدام مجموعة بيانات من (xAPI-Edu-Data) من موقع (Kaggle) مفتوح المصدر و تحتوي على (480) طالبًا وطالبة وتم استخدام المتغيرات الآتية: الجنسية، الصف، الجنس، المرحلة التعليمية، وعدد ومرات تكرار استخدام المنصة التعليمية، ومعلومات حول أدائهم في مواضيع مختلفة، بالإضافة إلى علامتهم النهائية، وبينت النتائج أن خوارزمية الغابة العشوائية والشبكات العصبية كانت من أفضل الخوارزميات من حيث الدقة التنبؤية حيث بلغت الدقة التنبؤية لهما (76.04%) و (75.63%) على التوالي ، يليهما في الدقة خوارزمية آلة دعم المتجه والتي بلغت الدقة التنبؤية لها (71.46%) وأظهرت خوارزميتيّ الجوار الأقرب كي وأدابوست (AdaBoost) الدقة الأقل حيث بلغ مؤشر الدقة لهما (69.79%) و (70.63%) على التوالي.

ودراسة أحمد (Ahmed, 2024) التي هدفت إلى تطبيق نماذج التعلّم الآلي الآتية للتنبؤ بأداء الطلاب: آلة دعم المتجه (SVM)، شجرة القرار (Decision Tree)، نايف بيز (Naïve Bayes)، والجوار الأقرب-كي (KNN)، وتم استخدام منهجية التنقيب في البيانات، وتكونت عينة الدراسة من (32005) طالب وطالبة من طلاب جامعة وولو ومعهد كومبولتشا للتكنولوجيا في أثيوبيا للأعوام 2017-2022، وتم استخدام الخصائص الآتية: الجنس، موطن الطالب الأصلي، نتيجة امتحان القبول الذي خاضه الطالب، عدد المحاولات السابقة للطالب في الامتحان، عدد الساعات المعتمدة التي أكملها الطالب ويدرسها حاليًا، وجود إعاقة، النتيجة الأكاديمية السابقة التي حققها الطالب، وتم استخدام طريقة الصدق التقاطعي بتكرار كي=10 (10 folds Cross Validation) للحكم على فاعلية النماذج، وأظهرت النتائج أن خوارزمية آلة دعم المتجه (SVM) كانت من أعلى الخوارزمية دقة وأفضلها في التنبؤ بأداء الطلاب حيث بلغت الدقة التنبؤية لها (96.03%)، قيّمة، وكانت خوارزمية نايف بيز أقلها دقة فقد بلغ مؤشر الدقة لها (83.32%).

وأما دراسة كيم وآخرون (Kim et al., 2024) فقد هدفت إلى تطبيق نماذج استجابة الفقرة التفسيرية (EIRM) للكشف عن الفروق في الأداء اللغوي بين الطلبة، وتم استخدام خصائص الطالب، وخصائص الفقرة لتحقيق هدف الدراسة، طُبقت الدراسة على عينة مكونة من (208) منهم (57%) إناث، و (62.5%) ذكور من طلبة الصف الأول ثنائيي اللغة (اسباني-انجليزي) من خمس مدارس في الولايات المتحدة الأمريكية، وأظهرت النتائج أن التباين الكبير بين اللغتين يُعزى إلى التأثيرات العشوائية للفقرة بنسبة (74%) والطالب بنسبة (26%).

ودراسة بارك وآخرون (Park et al., 2023) والتي سعت إلى مقارنة الدقة التنبؤية لنموذج استجابة الفقرة التفسيرية (EIRM) و ست من خوارزميات التعلّم الآلي الخاضعة للإشراف للتنبؤ باستجابات الأفراد على الفقرة في التقييم التربوي، آخذًا بعين الاعتبار المعلومات المتعلقة بالطالب (كالقدرة، والمعلومات الخلفية للطلاب) والفقرة (كالصعوبة)، ونماذج التعلّم الآلي الستة المستخدمة في الدراسة بجانب نموذج نظرية استجابة الفقرة التفسيرية (EIRM) هي: شجرة القرار (Decision Tree)، الغابة العشوائية (Random Forest)، التعزيز التدريجي (Gradient Boosting)، الجوار الأقرب كي (KNN)، التحليل التمييزي التربيعي (Quadratic Discriminant analysis)، الشبكات العصبية الاصطناعية (ANN)، وتكونت عينة الدراسة من: عينة المحاكاة ومثاليين لبيانات تقييم واقعية فأما عينة المحاكاة تم توليد بيانات بأحجام (N = 1000) وعدد الفقرات = (110)، وحجم (N = 100) وعدد فقرات = (10)، وأما قاعدة البيانات الواقعية فالأولى تكونت من العلامات النهائية لطلاب المرحلة الثانوية في بلجيكا من (2044) طالباً و (20) فقرة، وقاعدة البيانات الثانية تكونت من (1950) طالبًا من بيانات التقييم الوطني بعدد فقرات (22) في فرنسا للعام (2018)، وتم الحكم على فاعلية الطرائق المستخدمة بإجراء طريقة الصدق التقاطعي بكرار كي = 10 (10 fold Cross Validation)، وأظهرت النتائج أن نهج استجابة الفقرة التفسيرية (EIRM) ينافس كأفضل نهج مقارنة مع طرائق التعلّم الآلي بالتنبؤ بأداء الطلاب، كما بينت النتائج أن المعلومات المتعلقة بخلفية الطالب وصعوبة الفقرة كان لهما التأثير الأكبر على دقة التنبؤ في نموذج هجين.

ودراسة خليلية وآخرون (2020) التي هدفت إلى تطبيق بعض من نماذج التعلّم الآلي (شجرة القرار مع مقياس اكتساب المعلومات وخوارزمية شجرة القرار بمقياس مؤشر "جيني"، وخوارزمية نايف بيز)، وكانت مجموعة البيانات التي طُبّقت عليها الدراسة هي بيانات طلاب قسم هندسة الحاسوب والحوسبة التطبيقية في جامعة فلسطين التقنية -خضوري طولكرم، وبلغ عددهم (422) سجل طلابي من الطلبة الذين سجلوا في تخصص هندسة الحاسوب منذ 2017 إلى 2019، تم استخدام متغيرات مثل حقل تخصص الطالب (علمي، أدبي)، وتخصص الطالب في الجامعة، وما إذا كان الطالب حاصلًا على منحة دراسية أم لا للتنبؤ بأداء الطالب وتوقع أداء الطلاب من خلال قياس معدله التراكمي، وأشارت نتائج الدارسة أن خوارزمية شجرة القرار مع مقياس اكتساب المعلومات تفوقت على الخوارزميات الأخرى في الدقة التنبؤية حيث بلغ مؤشر الدقة (Accuracy) لها (66%)، وبلغ لخوارزمية شجرة القرار بمقياس مؤشر جيني (61%)، أما مؤشر الدقة لخوارزمية نايف بيز فكان الأقل وبلغ (50%)، وكان متغيرا حقل تخصص الطالب (علمي، أدبي) وتخصص الطالب في الجامعة ذات دلالة إحصائية، ولهما تأثير على التنبؤ بأداء الطالب.

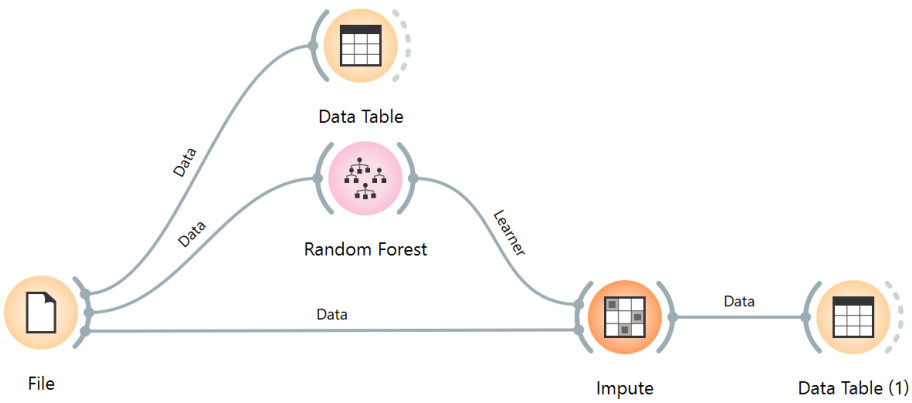
وأما دراسة بلياكوس وآخرون (Pliakos et al., 2019) فقد سعت إلى اقتراح نظام يجمع بين نظرية استجابة الفقرة (IRT) والتعلّم الآلي (ML)، وتم استخدام خوارزميات التعلّم الآلي الآتية: أشجار الإنحدار (Regression Tree)، الغابة العشوائية (Random Forest)، آلة دعم المتجه (SVM)، الانحدار الخطي (Linear Regression)، الجوار الأقرب كي (KNN)، وتكونت عينة الدراسة من مجموعتين من البيانات التعليمية أحدها تم

استخدامها كبيانات تدريب والثانية كبيانات اختبار، وتكونت مجموعة البيانات الأولى من (2044) طالب و(20) فقرة، والثانية كانت مكونة من (229) طالب و(145) فقرة ، وتم استخدام (23) متغير من متغيرات الخلفية للطالب وخلُصت نتائج الدراسة إلى أن أفضل نظام هجين من بين النماذج كان النظام الذي جمع بين نظرية استجابة الفقرة والغابة العشوائية.

إجراءات الدراسة

قبل بناء النماذج التنبؤية والمقارنة بينها لتحديد أفضلها دقة هناك مجموعة من الخطوات المهمة التي تم اتباعها لبناء النماذج التنبؤية وهي كما أشار (Gupta at al., 2024):

- جمع البيانات (Data collection): تم اعتماد بيانات بيزا 2022 لطلاب الأردن في مادة الرياضيات.
- معالجة البيانات (Data Processing):
  - تم اختيار الفقرات الخاصة بالرياضيات، واستبعاد الفقرات الخاصة بالقراءة والعلوم.
  - حُذف (505) طالب وطالبة لم يتعرضوا لفقرات الاختبار بسبب الغياب، ولم يتوفر لهم رقم للمدرسة أو رقم السلطة المشرفة أو رقم لنموذج الاختبار في البيانات أو إجابات ليصل عدد الطلبة من (7799) إلى (7294).
  - تم استبعاد الفقرات ذات الإجابة القصيرة والاقتصار على فقرات الاختيار من متعدد فقط؛ ليكون عدد الفقرات إلى (157).
  - معالجة البيانات المتناثرة (Sparse Data) في بيانات بيزا باستخدام أداة Random Forest من برمجية Orange 3.38.1 ويبين الشكل (4) آلية معالجة البيانات المتناثرة باستخدام أداة (RandomForest)



الشكل(4): آلية معالجة البيانات المتناثرة في بيانات بيزا 2022 لاختبار الرياضيات باستخدام برمجية Orange 3.38.1

يُلاحظ من شكل (6) أنه تم استخدام أداة Random Forest ووصلها بملف البيانات (File)، ولعرض البيانات قبل المعالجة تم وصل أداة (Data Table) بأداة (File) والذي يحتوي على بيانات بيزا 2022 لاختبار الرياضيات، ثم استخدام أداة (impute) ووصلها مع أداة Random Forest وأداة (File)، وخُزنت البيانات بعد المعالجة في ((Data Table(1).

- تم اختيار نموذج (1) (BOOKID 1).
- اختيار الخصائص أو المتغيرات (Feature Selection)، وهي من ناحية الفقرة: صعوبة الفقرة (diff) ، ومن ناحية الطالب: الجنس (gender)، السلطة المشرفة (strm)، المستوى الاقتصادي والاجتماعي للأسرة (ses)، التعرض للتنمر (bullied)، استخدام التطبيقات الرقمية خارج المدرسة (usingApp)، توافر أجهزة رقمية موصولة بالإنترنت في المدرسة (DigSchool)، مهارات المعلمين الرقمية (teacherskill)، استخدام المصادر الرقمية في حصص الرياضيات (digitalmath)، ليكون مجموع الخصائص (9) خصائص.
- تقسيم البيانات إلى بيانات اختبار، وبيانات تدريب (Data Splitting): حيث تم بناء نموذج باستخدام جزء من البيانات (يشار إليها ببيانات التدريب training data) ثم تقييم النموذج باستخدام جزء مختلف من البيانات (يشار إليه ببيانات الاختبار testing data). (Gonzalez, 2020) وتم تقسيم البيانات إلى 70% بيانات تدريب و 30% بيانات اختبار وتُعدُّ النسبة 30:70 من أكثر النسب فاعلية وتُعتبر هذه النسبة معيارية في العديد من

التطبيقات ومن أفضل التقسيمات للبيانات. (Nguyen et al., 2021)

- اختيار النموذج (Model Selection) : والنماذج المعتمدة التي تم اختيارها للدراسة هي: نموذج استجابة الفقرة التفسيرية (EIRM) (النموذج التفسيري المضاعف)، ونماذج التعلّم الآلي الخاضعة للإشراف الآتية: نموذج الغابة العشوائية نموذج الشبكة العصبية الاصطناعية، نموذج نايف بيز، نموذج آلة دعم المتجه، نموذج الجوار الأقرب-كي.
- تدريب النموذج (Model Training): تم تدريب النماذج الستة باستخدام برمجية (R).
- ضبط المعالم الفائقة لنماذج تعلّم الآلة (Hyperparameter Tuning): تم ضبط المعالم الفائقة الخاصة بكل نموذج من نماذج التعلّم الآلي للحصول على أفضل أداء، حيث يوجد في نماذج التعلّم الآلي العديد من المَعلمات والمَعلمات الفائقة (HyperParameters)، وعادة يتم تعلّم المَعلمات أثناء تدريب النموذج، أما المَعلمات الفائقة، فيجب تعيينها قبل تدريب النموذج حيث تُحدِد المَعلمات الفائقة كيف ولماذا يمكن للنموذج أن يتعلّم، كما أنها تؤثر بشكل حاسم على مدى جودة أداء نماذج التعلّم الآلي على البيانات خارج العينة، وتساهم في بناء الثقة في أداء النموذج، كما أنها تلعب دورًا رئيسًا وتحدد قدرة هذه النماذج على التعميم.(Arnold et al., 2024)
- تقويم النموذج (Model Evaluation): لتقويم النماذج الستة والحكم على دقتها تم اعتماد الصدق التقاطعي بتكرار- كي = 10 (10 fold Cross Validation) وهو شبيه بطريقة سحب العينات العشوائية حيث لا تتقاطع أي مجموعتي اختبار، وتقسّم البيانات إلى مجموعات صغيرة (subsets) بعدد K ذات أحجام متساوية تقريبًا، وتُسمى كل مجموعة منها بالطبقة (fold) وهذا التقسيم يتم بشكل عشوائي دون ارجاع، يتم تدريب النموذج باستخدام (k-1) مجموعة فرعية، والتي مع بعضها تمثل مجموعة التدريب (training set)، ثم يطبّق النموذج على باقي المجموعة والتي يشار إليها بمجموعة الاختبار أو مجموعة التحقّق (Validation set) ويتم قياس أداء النموذج، هذه الطريقة يتم تكرارها (10) مرات حتى تخدم كل مجموعة من المجموعات الفرعية كمجموعة تحقق أو اختبار، ويؤخذ معدل الأداء للمجموعات (k).
- حُسبت مؤشرات الصدق التقاطعي للحكم على دقة النماذج وهذه المؤشرات هي: Accuracy (الدقة) ،Precision (الإحكام)، Recall (الاستدعاء)، مؤشر F1 ، ومؤشر المساحة تحت منحنى ROC (AUC)، وتعتمد حساب هذه المؤشرات على ما يُعرف بمصفوفة الالتباس (Confusion Matrix) والتي هي عبارة عن مصفوفة بحجم (N X N) حيث (N) ترمز إلى فئات المخرجات، كلُّ صف في هذه المصفوفة يمثل عدد الحالات للفئة المتنبأ بها، وكل عمود يمثل عدد الحالات للفئة الفعلية، وهذا يوفر معلومات تفصيلية لكل فئة من التنبؤات الدقيقة وغير الدقيقة، والمؤشرات التي يمكن اشتقاقها من مصفوفة الالتباس يمكن أن تساعد على اختيار أداء النموذج الأفضل دقة، وتُستخدم هذه المصفوفات مع النماذج الخاضعة للإشراف (supervised learning methods)، ويمثل جدول (3) البناء العام لمصفوفة الالتباس. (Swaminathan & Tantri, 2024)

الجدول (3): البناء العام لمصفوفة الالتباس (Confusion Matrix)

		Actual Values	
		Positive	**Negative**
Predicted Values	**Positive**	**True Positive (TP)**	**False Positive (Type error) (FP)**
	Negative	**False Negative (Type II error) (FN)**	**True Negative (TN)**

وأما المؤشرات التي يمكن حسابها من هذه المصفوفة فهي كالآتي، (Khor, 2019):

- مؤشر الدقة (Accuracy): وهي نسبة التنبؤات الصحيحة من بين جميع التنبؤات وتُحسب حسب المعادلة الآتية:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN}$$

- مؤشر الإحكام (Precision): يقيس النسبة بين التنبؤات الصحيحة الإيجابية إلى جميع التنبؤات الإيجابية، فهو يركز على دقة التنبؤات الإيجابية فقط ويُحسب حسب المعادلة الآتية:

$$\text{Precision} = \frac{TP}{TP+FP}$$

- مؤشر الاستدعاء (Recall): يقيس هذا المؤشر مدى دقة النموذج في التنبؤ بالحالات الإيجابية، ويُحسب حسب المعادلة الآتية:

$$\text{Recall} = \frac{TP}{TP+FN}$$

- مؤشر F1: والذي يجمع بين الدقة والاستدعاء وهو المتوسط التوافقي للدقة والاستدعاء الذي يأخذ في الاعتبار كل من النتائج السلبية الخاطئة والإيجابية الخاطئة، ويُحسب حسب المعادلة الآتية:

$$F1 = 2 * \frac{Percision*Recall}{Percision+Recall}$$

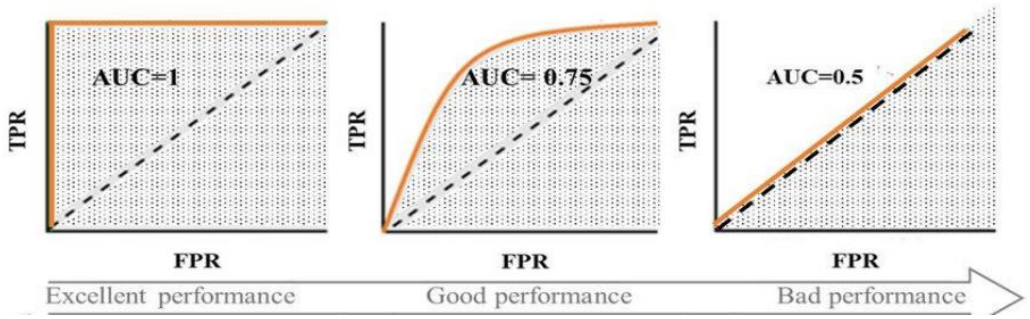
- مؤشر المساحة تحت منحنى ROC (AUC): ومنحنى ROC هو اختصار ل منحنى خاصية المستقبِل (Receiver Operating Characteristic)، وهو تمثيل بياني لمعلمتين هما: True Positive Rate (TPR) نسبة القيم الإيجابية الصحيحة على المحور الصادي و False Positive Rate (FPR) نسبة القيم الإيجابية الخاطئة على المحور السيني حيث:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

ويقيس مؤشر AUC المساحة تحت منحنى ROC من القيمة (0,0) إلى (1,1)، حيث يوفر هذا المنحنى مقياس إجمالي لإداء النموذج عبر كل عتبات التصنيف. (Swaminathan & Tantri, 2024)

ويوضح شكل (5) عتبات الحكم على مؤشر المساحة تحت منحنى ROC (AUC) للحكم على جودة النموذج.



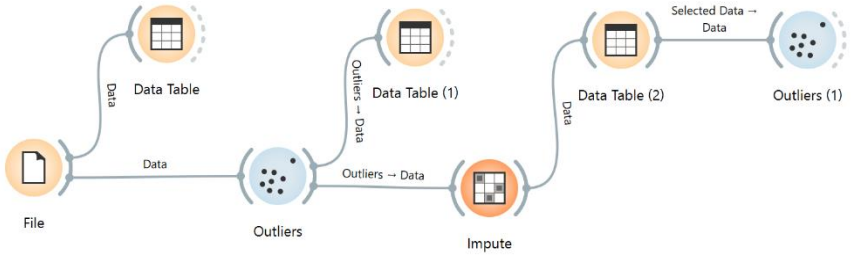
الشكل (5): عتبات الحكم على مؤشر المساحة تحت منحنى ROC (AUC)

يُلاحظ من شكل (5) أن قيمة (AUC) عند القيمة (1) تعتبر ممتازة، وعند (0.75) تعتبر جيدة أما إذا كانت تساوي (0.5) فأقل فهي تدل على أداء سيء للنموذج التنبؤي.

النتائج المتعلقة بالسؤال الأول:

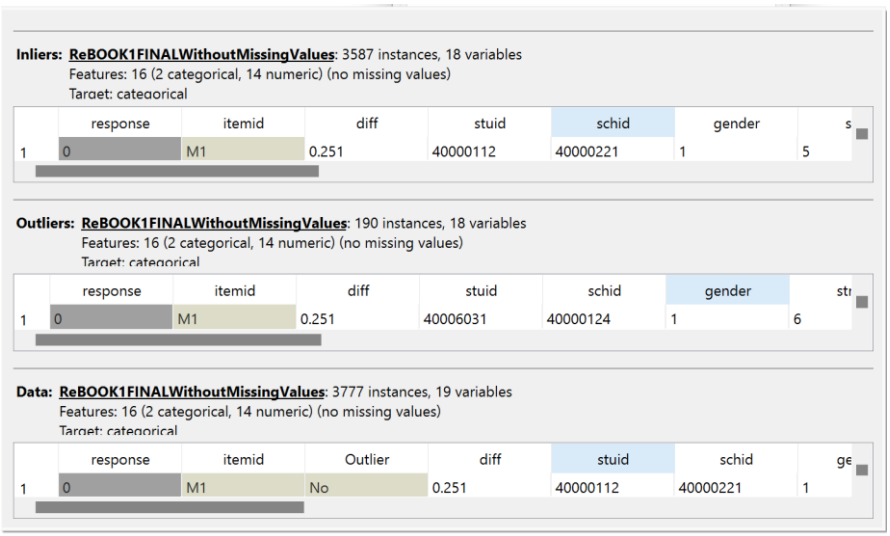
ما النموذج التنبؤي لتقييم أداء الطلبة في اختبار بيزا في الرياضيات للعام 2022 باستخدام نماذج استجابة الفقرة التفسيرية (EIRM) (النموذج التفسيري المضاعف)؟

تم الكشف عن القيم الشاذة (Outliers) للبيانات باستخدام برمجية Orange 3.38.1، عن طريق الأداة (Widget) الخاصة بتحديد القيم الشاذة تلقائيًا وهي أداة (Outliers) وتم اختيار خوارزمية: "Local Outlier Factor" باستخدام المسافة الإقليدية وربطها مع ملف البيانات، ويبين شكل (6) آلية الكشف عن القيم الشاذة ومعالجتها في ملف البيانات.



الشكل(6): آلية الكشف عن القيم الشاذة ومعالجتها في ملف البيانات باستخدام برمجية Orange 3.38.1

وكانت نتيجة التحليل أن هناك (190) قيمة شاذة حسب نتائج التحليل ، ويبين الشكل (7) ناتج التحليل:



الشكل(7): ناتج تحليل الكشف عن القيم الشاذة في ملف البيانات ومعالجتها عبر برمجية Orange 3.38.1

ويُلاحظُ من شكل (7) أن قيم (Outliers) بلغت (190) حالة من بين (3777) حالة أي أن (5%) من البيانات هي بيانات شاذة، ونظرًا لعدم الرغبة في فقدان مزيد من الحالات فقد تمت معالجة هذه القيم عن طريق التعويض (Impute) حيث إن معالجة القيم الشاذة بطرائق مثل الإزالة (Removal) أو التعويض (Imputation) عندما تكون النسب صغيرة هي الأنسب. (Barnett & Lewis, 1994) ويبين شكل (8) قيم مؤشرات الكشف عن القيم الشاذة بعد المعالجة.

الشكل(8): ناتج تحليل معالجة القيم الشاذة في ملف البيانات عبر برمجية Orange 3.38.1

ويُلاحظ من الشكل (8) أن عدد القيم الشاذة (Outliers) أصبح (0) حالة بعد المعالجة باستخدام أيقونة (Impute).

بعد الكشف عن القيم الشاذة ومعالجتها، أجريَّ التحليل باستخدام برمجية (R) لنظرية استجابة الفقرة التفسيرية (EIRM) لبناء النموذج التفسيري المضاعف، واستُخدمت حزمة (lme4) من برمجية R لبناء النموذج التنبؤي لنظرية استجابة الفقرة التفسيرية (EIRM) ومكتبة (glmer) وهي مجموعة من التعليمات البرمجية الجاهزة لبناء النماذج التنبؤية، وتبين المعادلة الآتية التأثيرات العشوائية والتأثيرات الثابتة ونوع التوزيع في النموذج:

model <- glmer (response ~ diff + gender + strm + ses + bullied + usingApp + DigSchool + teacherskill + digitalmath + (1 | schid), data = pisa_machine, family = binomial)

ويوضح جدول (4) القيم المقدرة للتأثيرات الثابتة والخطأ المعياري لها وقيمة z بالإضافة إلى قيمة P للنموذج التنبؤي لنظرية استجابة الفقرة التفسيرية (EIRM)

الجدول(4): القيم المقدرة للتأثيرات الثابتة والخطأ المعياري لها وقيمة z بالإضافة إلى قيمة P لنموذج (EIRM)

	Estimate	Std.Error	z-value	Pr(> |z|)
(Intercept)	-0.290	0.220	-1.314	0.189
صعوبة الفقرة (diff)	-0.983	0.058	-16.960	0.000*
gender(الجنس)	0.180	0.077	2.338	0.019 *
السلطة المشرفة (strm)	0.010	0.019	0.546	0.585
مستوى االاقتصادي والاجتماعي للأسرة (ses)	0.032	0.015	2.069	0.039 *
التعرض للتنمر (bullied)	-0.047	0.048	-0.991	0.322
استخدام التطبيقات الرقمية خارج المدرسة (usingApp)	0.045	0.030	1.507	0.132
توافر أجهزة رقمية موصولة بالإنترنت في المدرسة (DigSchool)	-0.179	0.053	-3.397	0.001 *
مهارات المعلمين الرقمية (teacherskill)	-0.062	0.047	-1.312	0.189
استخدام المصادر الرقمية في حصص الرياضيات (digitalmath)	0.110	0.030	3.672	0.000 *

*** التأثير معنوي عند 0.05**

ويُلاحظُ من جدول (4) أن هناك بعض المتغيرات كانت ذات دلالة معنوية عند 0.05 ، وهناك متغيرات كانت غير معنوية حيث بلغت قيمة P لها أعلى من (0.05) وهي السلطة المشرفة (strm)، التعرض للتنمر (bullied) ، استخدام البرمجيات التعليمية (usingApp)، ومهارات المعلم الرقمية (teacherskill).

لذا تم إعادة التحليل بعد إزالة التأثيرات غير المعنوية؛ ليصبح النموذج كما هو مبين في المعادلة:

model <- glmer(response ~ diff + gender + ses + DigSchool + digitalmath + (1 | schid), data = pisa_machine, family = binomial)

ويوضح جدول (5) القيم المقدرة للتأثيرات الثابتة والخطأ المعياري لها وقيمة z بالإضافة إلى قيمة P للنموذج التنبؤي لنظرية استجابة الفقرة التفسيرية بعد إزالة التأثيرات غير المعنوية للمتغيرات المذكورة.

الجدول(5): القيم المقدرة للتأثيرات الثابتة والخطأ المعياري لها وقيمة z بالإضافة إلى قيمة P لنموذج بعد إزالة التأثيرات غير المعنوية

	Estimate	Std.Error	z-value	Pr(> |z|)
(Intercept)	-0.327	0.198	-1.647	0.099
صعوبة الفقرة (diff)	-0.981	0.058	-16.962	0.000*
الجنس (gender)	0.164	0.075	2.190	0.029 *
المستوى الاقتصادي والاجتماعي للأسرة (ses)	0.032	0.015	2.131	0.033 *

	Estimate	Std.Error	z-value	Pr(> |z|)
توافر اجهزة رقمية موصولة بالإنترنت في المدرسة (DigSchool)	-0.183	0.049	-3.759	0.000 *
استخدام المصادر الرقمية في حصص الرياضيات (digitalmath)	0.120	0.028	3.874	0.000 *

*** التأثير معنوي عند 0.05**

ويُلاحظ من جدول (5) أن متغير صعوبة الفقرة (diff) هو الأكثر تأثيرًا وكانت قيم z له سالبة مما يشير إلى أنه كلما زادت صعوبة الفقرة انخفض احتمال الإجابة الصحيحة عن السؤال، بينما ارتبط متغير الجنس ومستوى الاقتصادي والاجتماعي للأسرة، واستخدام المصادر الرقمية في حصص الرياضيات ارتباطًا إيجابيًا بأداء الطلبة، في حين كان لمتغير وجود أجهزة رقمية في المدرسة أثر سلبي، وربما يعود ذلك إلى أن توفر الأجهزة الرقمية وحدها لا يكفي؛ لتحسين أداء الطلبة؛ بل يجب أن يرافقه استخدامٌ فعّال يعزز مهارات الطلبة.

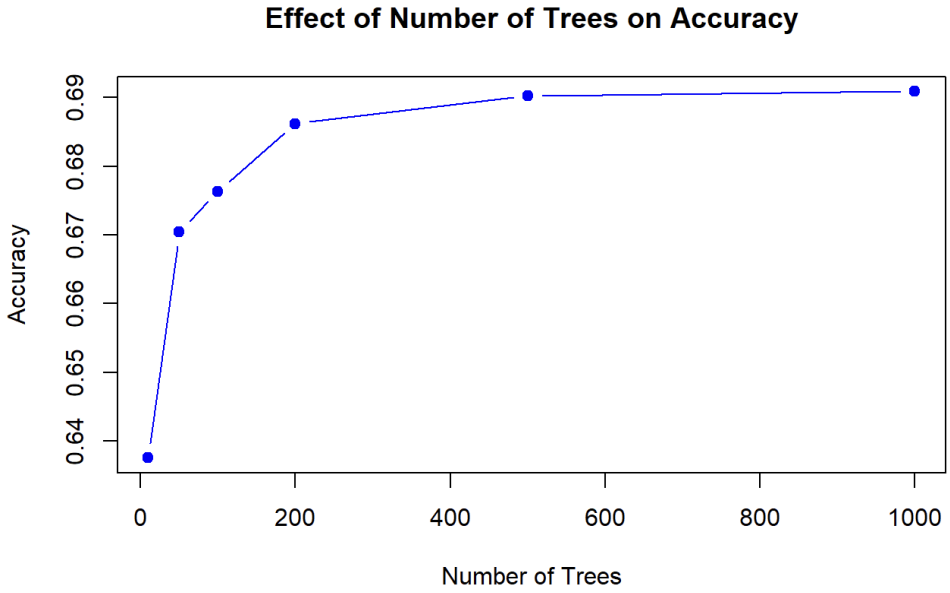
النتائج المتعلقة بالسؤال الثاني:

ما النموذج التنبؤي لتقييم أداء الطلبة في اختبار بيزا في الرياضيات 2022 للعام باستخدام نماذج التعلّم الآلي؟

1. النموذج التنبؤي باستخدام الغابة العشوائية (RF)

قبل البدء ببناء نموذج تنبؤي للغابة العشوائية تم ضبط المعالم الفائقة الخاصة بالنموذج، وهي استخدام عدد كاف من الأشجار في الخوارزمية لضمان استقرار نموذج الغابة العشوائية؛ لأن عدد الأشجار الكبير يقلل من أخطاء التعميم (Generalization Error)، وقد يصل العدد إلى مرحلة يصبح بعدها العدد الكبير غير فعال ويتقارب عنده خطأ التعميم، بالإضافة إلى ضبط مَعلمة عدد المتغيرات في كل تقسيم والذي يُرمز له بالرمز (m_{try}). ويعني (number of variable to try) (Breiman, 2001)

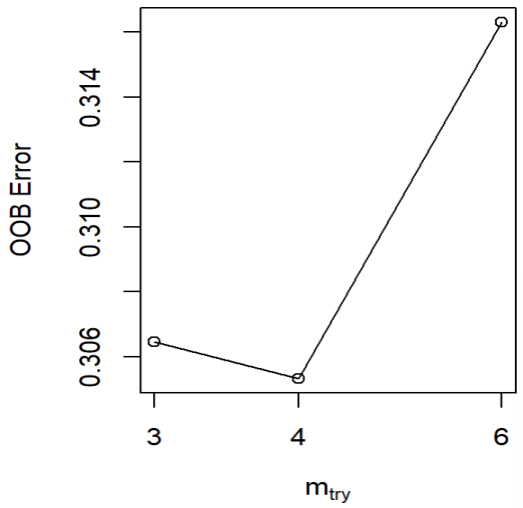
لذلك اختُبر تأثير عدد الاشجار على أداء النموذج باستخدام برمجية R حيث تم استخدام أعدد الأشجار (10, 50, 100, 200, 500, 1000) ثم رسمت العلاقة بين عدد الأشجار ودقة النموذج، وكانت النتائج كما هي مبينة في الشكل (9).



الشكل (9): أثر عدد الأشجار على دقة النموذج في الغابة العشوائية

ويُلاحظ من الشكل (9) أن عدد الأشجار التي بلغت (500) كان من أفضل الخيارات لدقة النموذج، إذ أن عدد الأشجار بعد الرقم (500) أصبح تأثيره لا يُذكر؛ لأن الخط أصبح أفقيًا.

وأما عن ضبط مَعلمة عدد المتغيرات المختارة لكل تقسيم (m_{try}) فقد اختُبرت أعداد مختلفة باستخدام مكتبة (tuneRF) لتحديد أفضل قيمة، وتقليل نسبة الخطا في التنبؤ (OOB) وهو الخطأ الذي يتم حسابه باستخدام البيانات التي لم يتم تضمينها في بناء كل شجرة أثناء التدريب وتسمى (Out-of-Bag Data). (Liaw & Wiener, 2002) ويبين الشكل (10) تمثيلًا بيانيًا لعدد المتغيرات لكل تقسيم (m_{try}) ونسبة الخطأ في التنبؤ (OOB).



الشكل (10): التمثيل البياني لعدد المتغيرات لكل تقسيم (m_{try}) ونسبة الخطأ في التنبؤ (OOB) باستخدام مكتبة (tuneRF) من برمجية (R)

ويُلاحظ من الشكل (10) أن أفضل عدد متغيرات مختار لكل تقسيم (m_{try}) والذي يعطي أقل نسبة خطا للتنبؤ هو (4) متغيرات، لذا تم اعتماده لبناء النموذج التنبؤي للغابة العشوائية.

وحُسبت أهمية المتغيرات التسعة الداخلة في نموذج الغابة العشوائية باستخدام مكتبة varImpPlot من برمجية (R) ومُثّلت بيانيًا باستخدام مؤشريّ متوسط انخفاض معامل جيني (MeanDecreaseGini) ومتوسط انخفاض الدقة (MeanDecreaseAccuracy)، ويبين جدول (6) ترتيب أهمية هذه المتغيرات.

الجدول (6): أهمية المتغيرات في الغابة العشوائية اعتمادًا على مؤشريّ متوسط انخفاض معامل جيني (MeanDecreaseGini) ومتوسط انخفاض الدقة (MeanDecreaseAccuracy)

Variable	MeanDecreaseGini	MeanDecreaseAccuracy
صعوبة الفقرة (diff)	683.200	47.920
مستوى الأسرة الاقتصادي والاجتماعي (ses)	205.600	10.850
استخدام التطبيقات الرقمية خارج المدرسة (usingApp)	148.700	13.590
استخدام التطبيقات الرقمية في حصص الرياضيات (digitalmath)	131.200	10.820
امتلاك المعلمين للمهارات الرقمية (teacherskill)	104.400	10.230
السلطة المشرفة (strm)	97.040	5.495
توافر أجهزة رقمية موصولة بالإنترنت في المدرسة (DigSchool)	95.130	16.380
التعرض للتنمر (bullied)	64.670	5.542
الجنس (Gender)	51.270	4.002

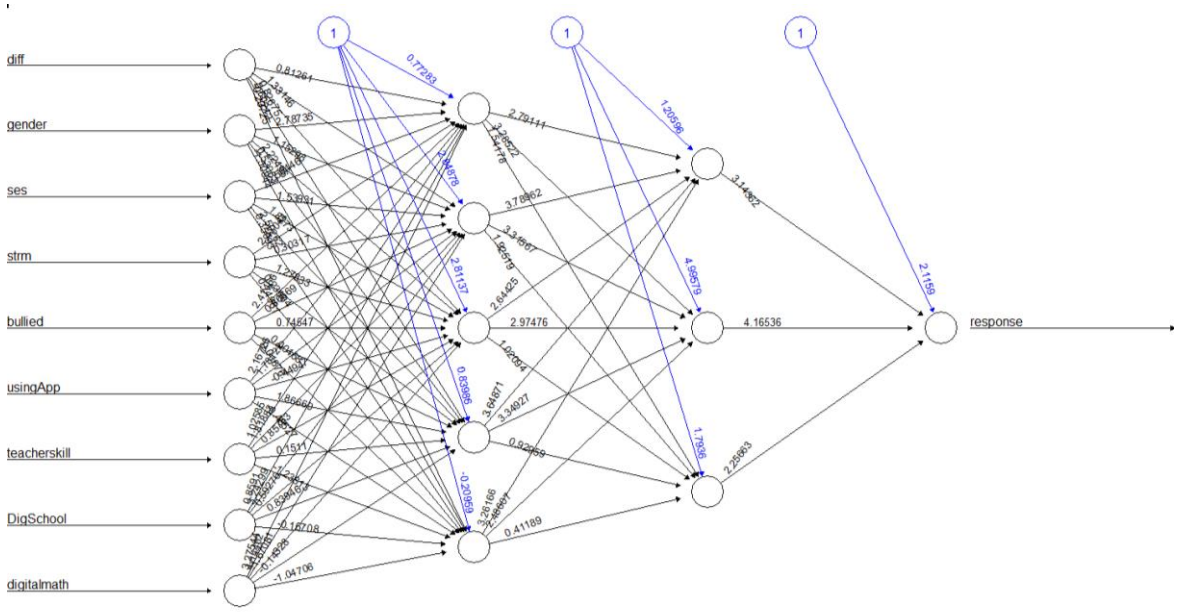
يُلاحظ من جدول (6) أن أعلى قيمة لمتوسط انخفاض معامل جيني (MeanDecreaseGini) و لمتوسط انخفاض الدقة (MeanDecreaseAccuracy) كانت لمتغير صعوبة الفقرة (diff)، والذي بلغ (683.2) و(47.920) على التوالي، وأقل قيمة كانت لمتغير الجنس (gender) وقد بلغ (51.27) و (4.002) على التوالي.

2. **النموذج التنبؤي باستخدام الشبكات العصبية الاصطناعية (ANN)**

تم ضبط عدد الطبقات المخفية التي لها تأثير على القدرة التنبؤية لهذا النموذج. (Haykin, 2009) واستُخدم مؤشر Accuracy لتحديد عدد الطبقات المخفية التي توفر أفضل أداء للنموذج بناءً على دقة التنبؤ باستخدام برمجية R، وكانت النتيجة أن الدقة بلغت عند استخدام طبقة مخفية واحدة مكونة من (5) عقد عصبية (0.6891)، بينما بلغت الدقة عند استخدام طبقتين مخفيتين، الأولى تحتوي على (5) عقد عصبية والثانية تحتوي على (3) عقد عصبية (0.6812)، لذا تم استخدام الشكل القياسي للشبكة، وهو طبقتين مخفيتين والتي لها قدرة أكبر على تمثيل العلاقات غير الخطية المعقدة في البيانات مما يعزز قابليتها للتعميم على بيانات جديدة حتى وإن كانت دقتها في بيانات التدريب أقل. (Bishop, 1995) واستُخدمت المكتبات (neuralnet) و(NeuralNetTools) من برمجية R لبناء النموذج التنبؤي للشبكات العصبية الاصطناعية، وكانت طبقة

المدخلات مكونة من (9) متغيرات أو مدخلات وهي المتغيرات المستقلة الداخلة في الدراسة، وتم استخدام (5) عقد عصبية في الطبقة المخفية الأولى، وبلغ عدد المعالم (الأوزان Wights) بين طبقة المدخلات وبين الطبقة المخفية الأولى 50 حيث 5+(5*9) ، وبلغت عدد العقد العصبية في الطبقة المخفية الثانية (3) وعدد المعالم بين الطبقة المخفية الأولى والطبقة المخفية الثانية بلغت 18 حيث 3+ (3*5) ، وبلغ عدد المعالم بين الطبقة المخفية الثانية وطبقة المخرجات 4 حيث 1+ (1*3) ، وأما طبقة المخرجات؛ فهي مكونة من عقدة واحدة (لأن النموذج تصنيفي من مخرج واحد إما 0 أو 1)فيكون بذلك عدد المعالم الكليّة في النموذج 72 حيث 4+18+50.

ويوضح شكل (11) تمثيل مرئي لنموذج الشبكة العصبية الاصطناعية بعد تدريبه.



الشكل (11): تمثيل مرئي لنموذج الشبكة العصبية الاصطناعية بعد تدريبه (ANN)

واستُخدمت مكتبة knitr من برمجية R لاستخراج أهمية المتغيرات في نموذج الشبكة العصبية الاصطناعية باستخدام طريقة غارسون وكانت النتائج كما هي مبينة في جدول (7).

الجدول (7): أهمية المتغيرات في نموذج الشبكة العصبية الاصطناعية باستخدام طريقة غارسون (Garson's Method)

Variable	Garson's Method
صعوبة الفقرة (diff)	0.223
الجنس (gender)	0.130
التعرض للتنمر (bullied)	0.121
امتلاك المعلمين للمهارات الرقمية (teacherskill)	0.112
مرات استخدام التطبيقات الرقمية في حصص الرياضيات (digitalmath)	0.111
توافر أجهزة رقمية موصولة بالإنترنت في المدرسة(DigSchool)	0.091
استخدام التطبيقات الرقمية خارج المدرسة (usingApp)	0.086
السلطة المشرفة (strm)	0.064
مستوى الأسرة الاقتصادي والاجتماعي (ses)	0.061

ويُلاحظ من جدول (7) أن المتغير الأهم في التنبؤ باستجابات الطلبة في اختبار بيزا 2022 باستخدام نموذج الشبكات العصبية الاصطناعية كان متغير صعوبة الفقرة (diff)، وكانت السلطة المشرفة (strm)، والحالة الاقتصادية والاجتماعية للطالب (ses) أقل المتغيرات أهمية في هذا النموذج.

3. النموذج التنبؤي باستخدام نموذج نايف بيز (NB):

ضُبطت المَعلمة الفائقة لهذا النموذج وهي مَعلمة التسوية (smoothing parameter) حيث تُستخدم هذه المَعلمة لمعالجة مشكلة احتمالات الصفر، ومن أهم أنواعها مَعلمة تسوية لابلاس (Laplace) حيث يتم إضافة قيمة صغيرة (عادةً 1) إلى جميع الترددات لتجنب القيم الصفرية. (Kilimci & Ganiz, 2015)

لتطبيق خوارزمية ناييف بييز تم استخدام المكتبات (e1071) و (caret) وتقسيم البيانات إلى بيانات اختبار وبيانات تدريب، ثم تدريب النموذج وبعدها إجراء التوقعات وحساب مصفوفة الالتباس والمؤشرات عبر الطبقات العشرة، وحُسبت أهمية المتغيرات في خوارزمية ناييف بييز (NB) باستخدام مكتبتيّ: (caret) و (klaR) باستخدام دالّة (varImp) حيث كانت النتيجة كما هي مبينة في جدول (8).

الجدول (8): أهمية المتغيرات في نموذج ناييف بييز (Naïve Bayes)

Variable	Importance
صعوبة الفقرة (diff)	100.000
مرات استخدام التطبيقات الرقمية في حصص الرياضيات (digitalmath)	12.458
توافر أجهزة رقمية موصولة بالإنترنت في المدرسة (DigSchool)	9.237
الجنس (gender)	7.976
استخدام التطبيقات الرقمية خارج المدرسة (usingApp)	6.767
مستوى الأسرة الاقتصادي والاجتماعي (ses)	6.209
امتلاك المعلمين للمهارات الرقمية (teacherskill)	3.442
التعرض للتنمر (bullied)	2.725
السلطة المشرفة (strm)	0.000

ويُلاحظ من جدول (8) أن المتغير الأعلى أهمية هو متغير صعوبة الفقرة (diff) حيث بلغت أهميته (100)، بينما احتل متغير التعرض للتنمر (bullied) والسلطة المشرفة (strm) أقل المتغيرات أهمية.

4. النموذج التنبؤي باستخدام آلة دعم المتجه (SVM)

استُخدمت المكتبات (e1071) و(caret) و(ggplot2) من برمجية R لبناء النموذج التنبؤي لخوارزمية آلة دعم المتجه (SVM)، حيث تم تدريب النموذج، وضُبطت المعالم الفائقة (HyperParameters) في هذا النموذج وهي: نوع النواة المستخدمة (Kernal) وقد كانت الخطية (linear) حيث يتم تطبيق نوعين من الأنوية، وهي: الخطي (Linear) للبيانات التي يمكن فصلها بخط مستقيم ومكونة من فئتين، والشعاعي (Radial) للبيانات التي لا يمكن فصلها بخط مستقيم وتكون متعددة الفئات.(Vijayalakshmi & Venkatachalapathy, 2019)، وحُسبت أهمية المتغيرات لهذا النموذج باستخدام مكتبة (caret) باستخدام دالة (vatImp) وكانت النتائج كما هي مبينة في جدول (9).

الجدول (9): أهمية المتغيرات في نموذج آلة دعم المتجه SVM

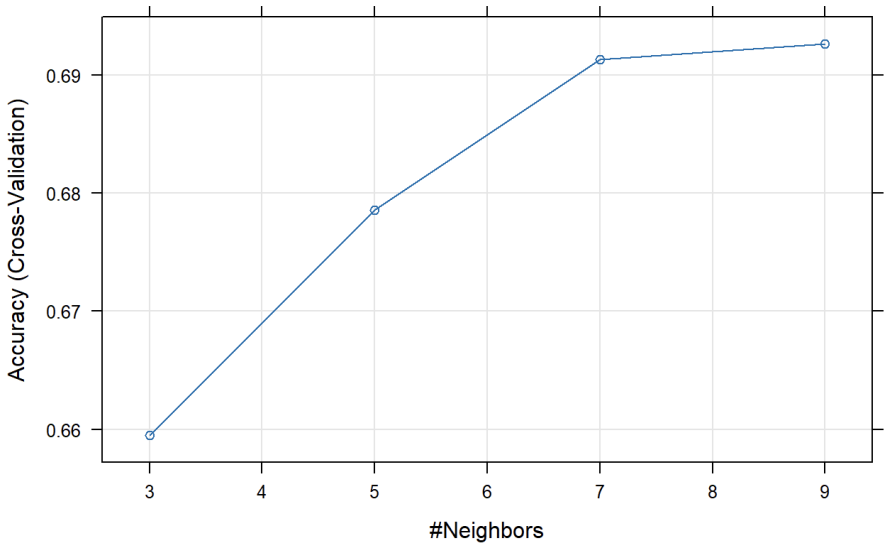
Variable	Importance
صعوبة الفقرة (diff)	0.6784
مرات استخدام التطبيقات الرقمية في حصص الرياضيات (digitalmath)	0.5249
توافر أجهزة رقمية موصولة بالإنترنت في المدرسة (digschool)	0.5193
الجنس (gender)	0.5173
استخدام التطبيقات الرقمية خارج المدرسة (usingapp)	0.5152
مستوى الأسرة الاقتصادي والاجتماعي (ses)	0.5142
امتلاك المعلمين للمهارات الرقمية (teacherskill)	0.5091
التعرض للتنمر (bullied)	0.5079
السلطة المشرفة (strm)	0.5031

ويُلاحظ من الجدول (9) أن متغير صعوبة الفقرة (diff) قد احتل المركز الأول في الأهمية بين جميع المتغيرات، بينما كان متغير السلطة المشرفة (strm) والتعرض للتنمر (bullied) من أقل المتغيرات أهمية في هذا النموذج.

5. النموذج التنبؤي باستخدام نموذج الجوار الأقرب – كي (KNN):

قبل البدء ببناء النموذج التنبؤي لنموذج الجوار الأقرب -كي (KNN) تم تحديد المعالم الفائقة لهذا النموذج، وهي القيمة k والذي يُطلق عليه "عدد الجيران" لتحديد عدد نقاط البيانات التي يُراد ترشيحها لتصنيف نقطة البيانات الجديدة، ويُعتبر العدد المختار من الجيران المحددين، أمرًا بالغ الأهمية في تحديد النتائج، لذلك عادة ما يُوصى باختبار العديد من قيم k للعثور على أفضل ملاءمة كما يُنصح بتجنب ضبط قيمة k على مستوى منخفض جدًا أو مرتفع جدًا. (Theobald, 2017)

ضُبطت المعالم الفائقة للنموذج بتجريب أعدادًا مختلفة من الجيران (k) باستخدام برمجية R من خلال مكتبة (caret) وتم تحديد قيم ممكنة لعدد الجيران (K) التي نريد أن نختبرها وهي (3,5,7,9) ويبين شكل (12) قيم (K) والدقة لكل قيمة.



الشكل (12): تمثيل بياني لتحديد قيم ممكنة لعدد الجيران (k) لنموذج الجوار الأقرب – كي (K-NN)

وكما هو مبين من الأداء بعد تمثيله في شكل (12) لكل قيمة من قيم (K) أن القيمة (k=9) كانت أعلى دقة، وهي أفضل قيمة من بين عدد الجيران المختارة، واستُخدامت المكتبات (caret) من برمجية (R) لبناء النموذج التنبؤي لخوارزمية الجوار الأقرب – كي ، وتم حساب أهمية المتغيرات باستخدام معامل متوسط انخفاض جيني، وكانت النتائج كما هي مبينة في الجدول (10).

الجدول(10): أهمية المتغيرات في نموذج الجوار الأقرب – كي (K-NN)

Variable	MeanDecreaseGini
صعوبة الفقرة (diff)	554.8
مستوى الأسرة الاقتصادي والاجتماعي (ses)	149.5
استخدام التطبيقات الرقمية خارج المدرسة (usingapp)	116.6
مرات استخدام التطبيقات الرقمية في حصص الرياضيات (digitalmath)	96.6
امتلاك المعلمين للمهارات الرقمية (teacherskill)	77.27
السلطة المشرفة (strm)	74.85
توافر أجهزة رقمية موصولة بالإنترنت في المدرسة (digschool)	74.47
التعرض للتنمر (bullied)	49.87
الجنس (gender)	34.54

يُلاحظ من الجدول (10) أن متغير صعوبة الفقرة (diff) كان المتغير الأكثر أهمية في هذا النموذج حيث بلغ معامل متوسط انخفاض جيني (554.8)، وكان أقل المتغيرات أهمية في هذا النموذج هو الجنس (gender) حيث بلغ متوسط انخفاض معامل جيني له (34.54).

النتائج المتعلقة بالسؤال الثالث:

أي النماذج هو الأكثر دقة (فاعلية) في التنبؤ بأداء الطلبة في اختبار بيزا في الرياضيات للعام 2022؟

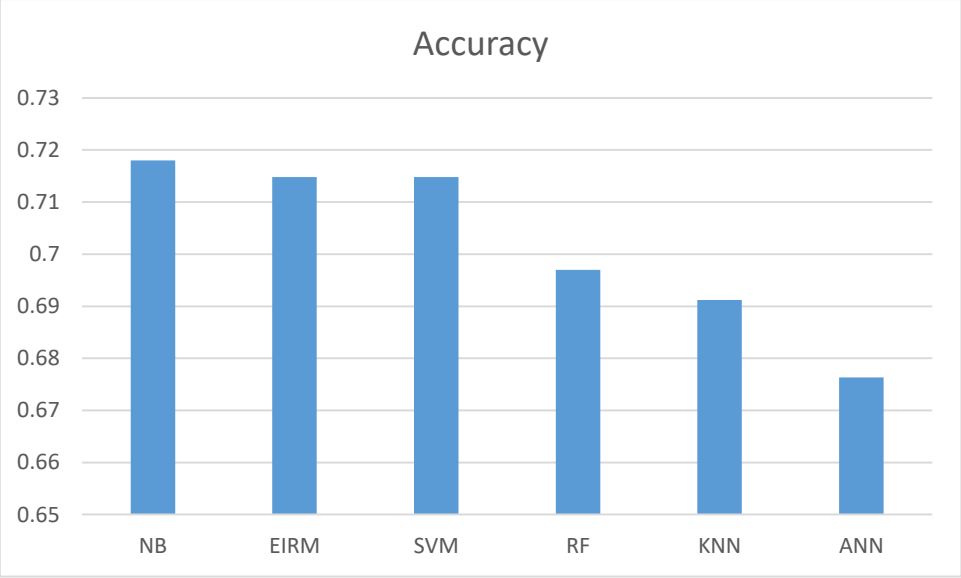
استُخدمت برمجية R لحساب مؤشرات الصدق التقاطعي بتكرار كي = 10 يبين الجدول (11) متوسط هذه المؤشرات لنموذج استجابة الفقرة التفسيرية (EIRM) ونماذج تعلّم الآلة الخمسة التي استُخدمت في هذه الدراسة.

الجدول(11): متوسط مؤشرات الصدق التقاطعي بتكرار- كي = 10 (10 Fold Cross Validation) لنموذج استجابة الفقرة التفسيرية ونماذج التعلّم الآلي الخمسة

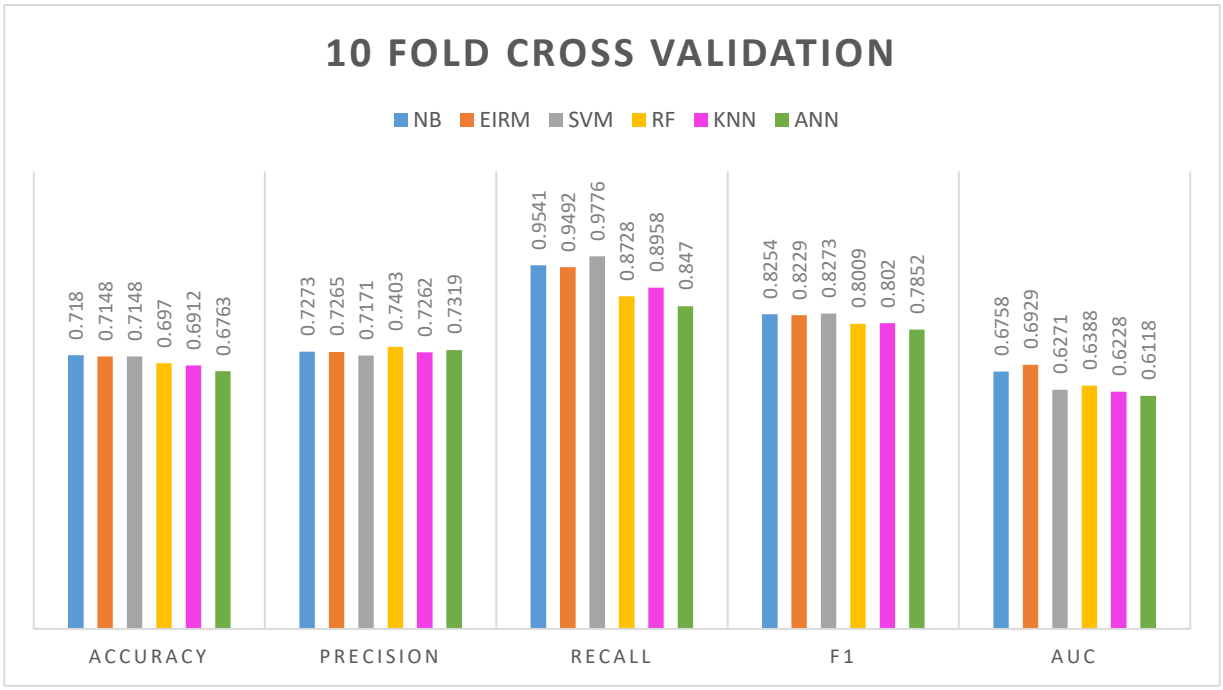
Model	Accuracy	Precision	Recall	F1	AUC
NB	0.718	0.727	0.954	0.825	0.676
EIRM	0.715	0.727	0.949	0.823	0.693
SVM	0.715	0.717	0.978	0.827	0.627
RF	0.697	0.740	0.873	0.801	0.639
KNN	0.691	0.726	0.896	0.802	0.623
ANN	0.677	0.732	0.847	0.785	0.611

ويبين جدول (11) متوسط مؤشرات الصدق التقاطعي عبر الطبقات العشرة مرتبة حسب مؤشر الدقة (Accuracy)، حيث أنه من أسهل المؤشرات التي يمكن فهمها من بين مؤشرات الصدق التقاطعي، ويمكن للباحث أن يعتقد أن النموذج الأفضل هو النموذج الذي يتمتع بمستوى عالٍ من الدقة، والدقة هي مؤشر ممتاز عندما تكون قيم معدلات النتائج الإيجابية والسلبية الكاذبة متساوية تقريبًا في مجموعة البيانات، ولكن يتعين على الباحث أن ينظر إلى معايير أخرى لتقييم أداء النموذج .(Baldi et al., 2000)

ويمثل الشكل (13) تمثيلًا مرئيًا لمؤشر الدقة (Accuracy) للنماذج الستة، كما يبين الشكل (14) التمثيل البياني لمتوسط مؤشرات الصدق التقاطعي بتكرار-كي = 10 للنماذج الستة.



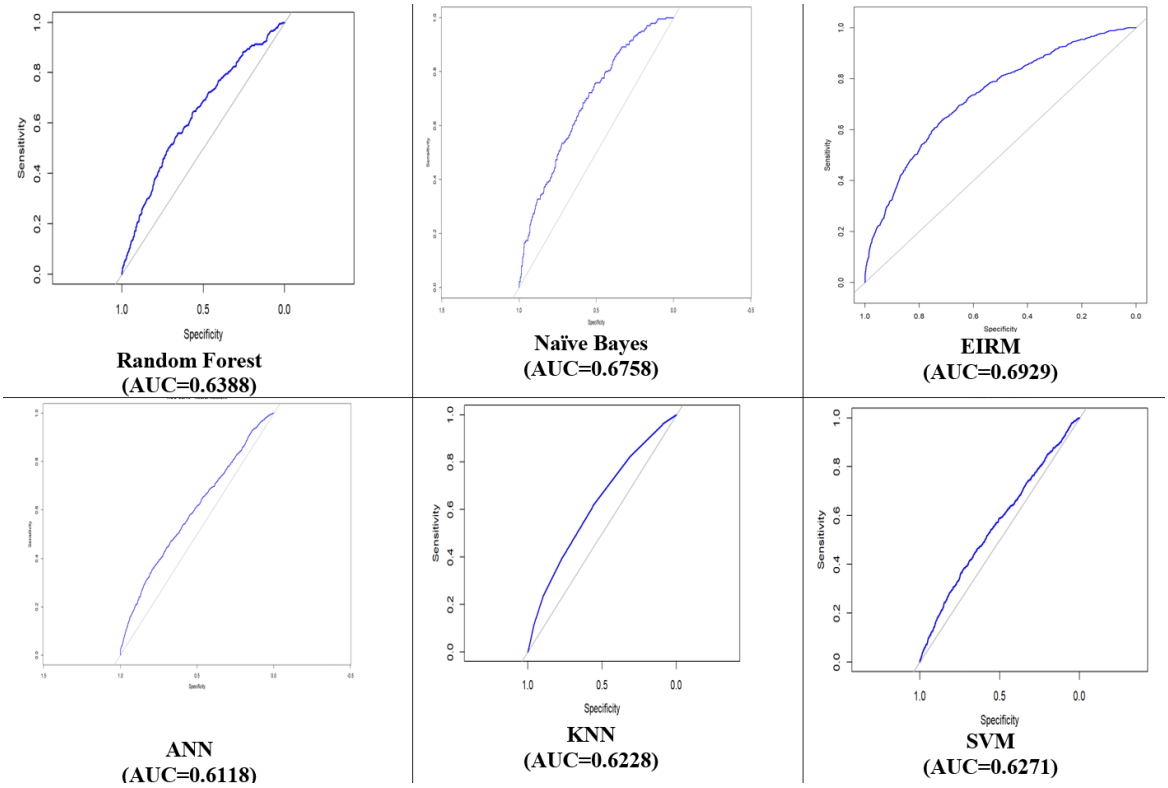
الشكل (13): التمثيل البياني لمؤشر الدقة (Accuracy) للنماذج الستة



الشكل (14): التمثيل البياني لمتوسط مؤشرات الصدق التقاطعي بتكرار-كي = 10 للنماذج الستة

اعتمادًا على جدول (11) وشكل (13) فإن نموذج نايف بييز (NB) سجل أعلى مؤشر دقة (Accuracy) بين النماذج الستة المستخدمة للتنبؤ بنتائج الطلبة في اختبار بيزا 2022 في الأردن بدقة بلغت (0.718)، وهذا يدلُّ على أن النموذج جيد في التعرف على الحالات الإيجابية وفي تحقيق تنبؤات صحيحة، يليها بانخفاض طفيف جدًا نموذج نظرية استجابة الفقرة التفسيرية (EIRM) وآلة دعم المتجه (SVM) واللذان كانا بنفس درجة الدقة حيث بلغت (0.715) ، ويُلاحظ من الجدول (11) والشكل (14) أن باقي المؤشرات لنموذج نايف بييز كانت من المؤشرات العالية حيث أن كل مؤشر يركز على زاوية معينة من الأداء للنموذج، وكان ترتيب نموذج نايف بييز الثالث من حيث مؤشر الإحكام وبلغ (0.727) وهو الذي يركز فقط على نسبة التنبؤات الصحيحة الإيجابية، وكان ترتيبه الثاني من حيث مؤشر الاستدعاء وبلغ (0.954) حيث يقيس هذا المؤشر نسبة القيم الفعلية الإيجابية التي تم التنبؤ بها بشكل صحيح، وأما مؤشر F1 فقد بلغ (0.825) وترتيبه الثاني بين النماذج والذي يجمع بين الدقة والاستدعاء ويأخذ بالاعتبار كل من النتائج السلبية الخاطئة والإيجابية حيث أن نموذج نايف بييز يفترض استقلالية الخصائص (Witten & Frank, 2000) وهذا الافتراض يبسّط النموذج ويقلّل من التعقيدات الحسابية مما يجعل النموذج أكثر كفاءة في التعامل مع بيانات كبيانات بيزا، لذلك فهو يتطلب بساطة في العمليات الحسابية مقارنة بباقي النماذج مما يجعله أقل تعقيدا وأكثر كفاءة من الناحية الحسابية، ونظرًا لبساطته يمكن لهذا النموذج التعامل بفاعلية مع مجموعات البيانات التي تحتوي على توزيع غير متوازن للفئات مثل بيانات بيزا المتناثرة والغير متوازنة، كما أنه مقارنة بنموذج استجابة الفقرة التفسيرية فهو أقل تعقيدًا مما يجعله أكثر قدرة على التعميم، ويتطلب هذا النموذج افتراضات أقل تقييدًا من نموذج استجابة الفقرة التفسيرية، فهو يعتمد فقط على الاحتمالات الشرطية وافتراض استقلالية الميزات مما يجعله أكثر مرونة. (Youssef, 2022)

ويبين الشكل (15) التمثيل البياني لمؤشر المساحة تحت منحنى AUC (ROC) للنماذج الستة الداخلة في الدراسة.



الشكل (15): االتمثيل البياني لمؤشر لمساحة تحت منحنى AUC (ROC) للنماذج الستة

يُلاحظ من الشكل (15) أن نموذج استجابة الفقرة التفسيرية كان من أفضل النماذج في قدرته على التمييز بين استجابات الطلبة في اختبار بيزا 2022 حيث بلغ مؤشر (AUC) له (0.6929)، وتفوّق على نماذج التعلّم الآلي، وكان نموذج نايف بييز الأفضل من بين نماذج التعلّم الآلي في التمييز بين استجابات الطلبة حيث بلغ مؤشر (AUC) له (0.6758)، إلا أنه لم يتفوق على نموذج استجابة الفقرة التفسيرية في هذا المؤشر في اختبار بيزا 2022، وتوفر نماذج استجابة الفقرة التفسيرية عادة نتائج أكثر وضوحًا من حيث تأثير المتغيرات العشوائية أكثر من نماذج التعلّم الآلي كالشبكات العصبية والغابات العشوائية.(Linden, 2016)، وتتفوق نظرية استجابة الفقرة التفسيرية في قدرتها على التعامل مع البنية الهرمية للبيانات. (De Boek & Wilson, 2004) حيث تعامل نموذج استجابة الفقرة التفسيرية (EIRM) مع بيانات بيزا ذات البنية الهرمية ووفرت تفسيرات واضحة للمتغيرات المؤثرة على أداء الطلبة.

واحتل متغير صعوبة الفقرة (diff) المركز الأول في الأهمية في نماذج التعلّم الآلي الخمسة، وكان تأثيره معنويًا في نموذج استجابة الفقرة التفسيرية مما يشير إلى أن هذا المتغير له أكبر تأثير في التنبؤ باستجابات الطلبة على اختبار PISA 2022، بينما احتل متغير التعرض للتنمر(bullied) المتغير قبل الآخير في الأهمية في نموذج الغابة العشوائية ونايف بييز وآلة دعم المتجه والجوار الأقرب-كي وكان تأثيره غير معنوي في نموذج استجابة الفقرة التفسيرية، واحتل متغير السلطة المشرفة (strm) المركز الأخير في الأهمية في نموذج نايف بييز ونموذج آلة دعم المتجه والمركز قبل الأخير في نموذج الشبكات العصبية الاصطناعية وكان تأثيره غير معنوي في نموذج استجابة الفقرة التفسيرية مما يدلّ على أن متغيريّ التعرض للتنمر (bullied) والسلطة المشرفة (strm) لا يمكن اعتبارهما متغيرين مهمين في التنبؤ باستجابات الطلبة على اختبار بيزا 2022 في الرياضيات.

التوصيات:

في ضوء النتائج التي توصلت إليها الدراسة تمت التوصية بما يأتي:

- إجراء دراسات بحثية جديدة على متغيرات جديدة بهدف إثراء الأدب التربوي بمزيد من المعلومات للباحثين في مجال تطوير نماذج تنبؤية قوية لتحسين آداء الطلبة في اختبارات بيزا القادمة.
- توسيع نطاق تطبيق النماذج التنبؤية التي توصلت إليها الدراسة؛ ليشمل اختبارات أخرى أو دول أخرى لاختبار القدرة على تعميم هذه النماذج.
- اختبار نماذج جديدة من نماذج التعلّم الآلي لم تُطبق في الدراسة لاستكشاف نماذج تنبؤية أقوى.

المصادروالمراجع

المركز الوطني لتنمية الموارد البشرية. (2023). *مستوى أداء طلبة الأردن في دراسة البرنامج الدولي لتقييم الطلبة (Program for International Students Assessment PISA 2022*. عمان: الأردن.

خليلية، هـ. وسمار، ث. وسليط، ي. (2020). التنبؤ بأداء الطلاب بناء على ملف الطالب الأكاديمي. *مجلة أبحاث جامعة فلسطين التقنية*، 8(2)، 23-39.

REFERENCES

Ahmed, E. (2024). Student performance prediction using machine learning algorithms. *Applied Computational Intelligence and Soft Computing*, *2024*(1), 1–15. https://doi.org/10.1155/2024/4067721

Alpaydin, E. (2005). *Introduction to machine learning*. *The Knowledge Engineering Review*, *20*(4), 432–433. https://doi.org/10.1017/S0269888906220745

Anderson, J., Lin, H., Treagust, D., Ross, S., & Yore, L. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, *5*(4), 591–614. https://doi.org/10.1007/s10763-007-9090-y

Arnold, C., Biedebach, L., Kupfer, A., & Neunhoeffer, M. (2024). The role of hyperparameters in machine learning models and how to tune them. *Political Science Research and Methods*, *12*(4), 1–8. https://doi.org/10.1017/psrm.2023.61

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C., & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics*, *16*(5), 412–424.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). John Wiley & Sons.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bulut, O. (2020, December 14). Explanatory IRT models in R. https://okan.cloud/posts/2020-12-14-explanatory-irt-models-in-r/

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer. https://doi.org/10.1007/978-1-4757-3990-9

Efron, B. (2013). Bayes' theorem in the 21st century. *Science*, *340*(6137), 1177–1178. https://doi.org/10.1126/science.1236536

Gonzalez, O. (2021). Psychometric and machine learning approaches for diagnostic assessment and tests of individual classification. *Psychological Methods*, *26*(2), 236–254. https://doi.org/10.1037/met0000317

Gupta, R., Sharma, A., & Alam, T. (2024). Building predictive models with machine learning. In P. Singh, A. R. Mishra, & P. Garg (Eds.), *Data analytics and machine learning* (pp. 39–59). Springer. https://doi.org/10.1007/978-981-97-0448-4_3

Halder, R., Uddin, M., Uddin, M., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications. *Journal of Big Data*, *11*(1), 1–55. https://doi.org/10.1186/s40537-024-00973-y

Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. Sage Publications.

Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). Pearson Education.

Khor, E. (2019). Predictive models with machine learning algorithms to forecast students' performance. In *Proceedings of the 13th International Technology, Education and Development Conference* (pp. 2831–2837). https://doi.org/10.21125/inted.2019.0757

Kilimci, Z., & Ganiz, M. (2015). Evaluation of classification models for language processing. *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)* (pp. 1–8). https://doi.org/10.1109/INISTA.2015.7276787

Kim, Y., Gutierrez, N., & Petscher, Y. (2024). Decomposing variation in vocabulary and listening comprehension task performance in Spanish and English into person, ecological, and assessment differences for Spanish-English bilingual children in the United States. *Journal of Speech, Language, and Hearing Research*, *67*(10), 3733–3747. https://doi.org/10.1044/2024_JSLHR-23-00702

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Linden, W. (2016). *Handbook of item response theory volume one: Models*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315374512

Maass, W., Parsons, J., Purao, S., Storey, V., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information system research. *Journal of the Association for Information Systems, 19*(12), 1253–1273. http://dx.doi.org/10.17705/1jais.00526

McCulloch, C., & Searle, S. (2001). *Generalized, linear, and mixed models*. Wiley & Sons.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning* (Adaptive Computation and Machine Learning Series). MIT Press.

Moukhafi, M., El Yassini, K., & Seddik, B. (2020). Intrusions detection using optimized support vector machine. *International Journal of Advances in Applied Sciences (IJAAS), 9*(1), 62–66.

Nguyen, Q., Ly, H., Ho, L., Al-Ansari, N., Le, H., Tran, V., Prakash, I., & Pham, B. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering, 2021*, Article 4832864, 1–15. https://doi.org/10.1155/2021/4832864

OECD. (2023a). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD. (2023b). "Foreword," in *PISA 2022 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

OECD. (2023e). "The PISA target population, the PISA samples, and the definition of schools," in *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD. (2023f). *PISA 2022 technical report*. OECD Publishing. https://doi.org/10.1787/01820d6d-en

Osmanbegovic, E., & Suljic, M. (2012). Data mining approach for predicting student performance. *Economic Review - Journal of Economics and Business, 10*(1), 3–12.

Pachouly, S., & Bormance, D. (2025). Exploring the predictive power of explainable AI in student performance forecasting using educational data. In D. Goyal (Ed.), *Recent advances in sciences, engineering, information technology & management* (pp. 362–370). CRC Press.

Park, J., Dedja, K., Pliakos, K., Kim, J., Joo, S., Cornillie, F., Vens, C., & Noortgate, W. (2023). Comparing the prediction performance of item response theory and machine learning methods on item responses for educational assessments. *Behavior Research Methods, 55*(4), 2109–2124. https://doi.org/10.3758/s13428-022-01910-8

Pliakos, K., Joo, S., Park, J., Cornillie, F., Vens, C., & Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education, 137*, 91–103. https://doi.org/10.1016/j.compedu.2019.04.009

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330

Sugiyama, M. (2015). *Introduction to statistical machine learning*. Morgan Kaufmann.

Swaminathan, S., & Tantri, B. (2024). Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research, 27*(4), 4023–4031. https://doi.org/10.53555/AJBR.v27i4S.4345

Theobald, O. (2017). *Machine learning for absolute beginners*. Oliver Theobald.

Tomar, P., & Verma, S. (2021). Impact and role of AI technologies in teaching, learning, and research in higher education. In S. Verma & P. Tomar (Eds.), *Impact of AI technologies on teaching, learning, and research in higher education* (pp. 190–203). IGI Global. https://doi.org/10.4018/978-1-7998-4763-2.ch012

Vapnik, V. (1995). *The nature of statistical learning*. Springer. http://dx.doi.org/10.1007/978-1-4757-2440-0

Vijayalakshmi, V., & Venkatachalapathy, K. (2019). Comparison of predicting student's performance using machine learning algorithms. *International Journal of Intelligent Systems and Applications, 11*(12), 34–45. https://doi.org/10.5815/ijisa.2019.12.04

Wilson, M., De Boeck, P., & Carstensen, C. (2006). *Explanatory item response models: A brief introduction*. Hogrefe & Huber Publishers.

Witten, I., & Frank, E. (2000). *Data mining – Practical machine learning tools and techniques* (2nd ed.). Morgan Kaufmann.

Youssef, Y. (2022). Bayes theorem and real-life application. Cairo University, Faculty of Economic and Political Science, Socio-Computing Department.

Zhang, Z. (2016). Naïve Bayes classification in R. *Annals of Translational Medicine, 4*(12), 241. https://doi.org/10.21037/atm.2016.03.38