

بسم الله الرحمن الرحيم جامعة افريقيا العالمية



عمادة الدراسات العليا والبحث العلمى والنشر كلية إقرأ لدراسات الحاسوب

قسم تقانة المعلومات

تنقيب بيانات طلاب المرحلة الاساسية للتنبؤ بدرجاتهم في مادة الرياضيات

بحث تكميلي لنيل درجة الماجستير في تقانة المعلومات

اعداد الطالب: ابوبكر محمد مصطفى عبدالله اشراف:أ. د/سعد صبير 2020



اهدى هذا العمل للعلماء الذين جعلوا حياتهم لخدمة البحث العلمى والعلم، كما اهدى لكل المسلمين في مشارق الارض ومغاربها، ولكل اساتذتي الكرام ، وامي الغالية حفظها الله ولوالدى الذى ارشدنى للطريق الصحيح وغرس في نفسي حب القراءة وطلب العلم والى زوجتي التى تقسمت معي مشاق البحث والحياة ، كما اوجه شكري خاصة للاهل والاصدقاء محمد ، الهادى سلمان ، والي دفعة تقانة المعلومات .

واسأل الله تعالى ان ينال استحسانكم ويضيف قيمة حقيقية في المجال والمكتبات.

الشكر والتقدير

يسرني ان اوجة شكرى لكل من نصحني او ارشدني او وجمنى او ساهم معي في اعداد هذا البحث في كل مراحله ،واشكر على وجة الخصوص ألاستاذ الدكتور (سعدصبير)على مساندتي وارشادي بالنصح والتصحيح،كما اوجة شكري لكلية أقرا لدراسات الحاسوب اساتذة وباحثين وطلاب.

<u>الفهرس</u>

| ٠ب | الإهداء |
|----|----------------------------|
| ج | الشكر والتقدير |
| د | القهرس |
| | فهرس الجدوال |
| و | فهرس الاشكال |
| ز | المستخلص |
| 1 | الفصل الاول: |
| 1 | الإطار المنهجي للبحث، |
| 2 | 1 اساسيات البحث |
| 2 | 1.1المقدمة: |
| 4 | 1.2مشكلة البحث: |
| 4 | 1.3اهداف البحث: |
| 5 | 1.4أهمية البحث : |
| 8 | الفصل الثاني: ادبيات البحث |
| 9 | 2تنقيب البيانات |
| 9 | 2.1تعاريف تنقيب البيانات |
| 13 | 2.2أساسيات التعلم الآلي |
| 24 | 2.3الدارسات السابقة: |
| 39 | الفصل الثالث: |
| 39 | عرض و تحلیل البیانات |

| 40 | 3عرض وتحليل البيانات |
|----|---|
| 40 | 3.1المواد |
| 42 | 3.2التحليل |
| 52 | الفصل الرابع: النتائج |
| 53 | 4اننتائج |
| 53 | 4.1النتائج المتحصلة |
| 54 | 4.2مصفوفة الارتباك |
| 55 | 4.3التقيم |
| 56 | 4.4المناقشة والاستنتاجات |
| 58 | الفصل الخامس:الخاتمة |
| 60 | 5.1التوصيات |
| 61 | المراجع والمصادر |
| | فهرس الجدوال |
| 14 | جدول 2-1مثال لنتائج الطلاب |
| 18 | جدول 2-2يوضح شجرة القرار |
| 21 | جدول 2-3 مثال على تعديل المتغيرات |
| 21 | جدول 2-4 تقيم فعالية التنبؤ |
| 42 | جدول 3-1 يوضح معنى الرموز الموجودة في بيانات الطالب |
| 48 | جدول 3-2بيانات التدريب |
| 49 | جدول 3-3بيانات الاختبار |
| 53 | جدول 4-1يوضح نتائج بيانات الاختبار |
| 54 | جدول 2-4 confusion matrix 2 |

| 4-3 نتائج الخورازميات | جدول |
|---|------|
| س الاشكال | |
| 1-2 مثال لشجرة قرار | شکل |
| 2-2 مجموعة من البيانات المبعثرة | شكل |
| 2-3تقسيم المتغيرات | شکل |
| 2-4مخطط شجرة القرار | |
| 1-3 يعرض نتائج طلاب أمادة الرياضيات | شكل |
| 2-2يعرض نتائج الطلاب | شکل |
| 3-3 يعرض درجات الطلاب للامتحان نصف الفصلي | شكل |
| 4-3 طريقة عمل الخورزامية | شكل |

المستخلص

مع تطور علم الحاسوب توفرت الكثير من البيانات في مختلف التخصصات والمؤسسات التعلمية واحدة من النظم التي تحتاج الى تتقيب بياناتها،مما دفع الباحث للانشاء نموذج للتوقع درجات الطلاب فتمحورت اشكالية البحث في انخفاض درجات الطلاب في مادة الرياضيات وفقا لسجلات الطلاب في مدارس الاساسيبة لذلك كانت هنالك حوجة للايجاد طريقة يمكن من خلالها التنبؤ بدرجات الطلاب عن طريق المتغير التابع Result في مرحلة مبكره ، هدفت هذه الدراسة على تصنيف الطلاب لمجموعتين (نجاحين وراسبين) لتوقع درجات الطلاب في مرحلة مبكرة ،اتبع الباحث المنهج الوصفي التحليلي لمناسبتها لموضوع قيد الدراسة كما تم جمع العينات لعدد 1500 طالب في العام 2015-2018 قسمت العينة الى مجموعتين المجموعة الاولى تمثل مجموعة التدريب وعددهم 500طالب بينما المجموعة الثانية تمثل مجموعة الاختبار وعددهم 1000طالب،اعتمد الباحث على الدرجات المعروفة مسبقا لنتائج الطلاب بغرض تحليل المتغيرات المؤثرة في التغيرات المستقبلية بخوارزمية decision tree باداة orange لتصنيف البيانات للتنبؤ بدرجات الطلاب النهائية على أساس درجاتهم السابقة،استخدم الباحث خمسة خوارزمة tree للتنبؤ ثم مقارنة نتائجها مع خوارزميات اخرى وهي: NN ،Naive Bayes، Logistic Regression و كانت SVM وكانت decision tree هي الأفضل أداء بنسبة . %99.7

Abstract

With the development of computer science, a lot of data became available in various disciplines and educational institutions and one of the systems that need to extract its data, which prompted the researcher to create a model for predicting students 'grades, and therefore the research problem centered on this study about students' declining grades in mathematics according to students 'records in Basic schools, so there was a need to find a method by which to predict students' grades using the outcome dependent variable at an early stage. It aims to categorize students into two groups (pass and fail) to predict students' grades at an early stage. The researcher followed the descriptive and analytical approach to its relationship with the topic under study, as it collected a sample of 1500 samples in the year 2015–2018, dividing the sample into two groups. The first group represents the training group of 500 students, while the second group represents the test group of 1000 students. The researcher relied on the previously known scores of the students' results for the purpose of analyzing the variables that affect future changes in the decision tree algorithm using an orange data classification tool to predict the final grades of students based on their previous grades. The researcher used five tree algorithms to predict their results and then compare them with other algorithms: Naive Bayes and NN. Logistic Regression, SVM, and the bestperforming decision tree was 99.7%

الفصل الاول:

الإطار المنهجي للبحث،

1 اساسبات البحث

1.1 المقدمة:

وفرة البيانات في مصادر التخزين المختلفة والتطور في مجالات الحاسوب في التنوع ،التخزين وسهولة إتاحتها مع التوسع في حجم المعلومات بصورة استباقية لم تشهد من قبل ، ادي لدخول الحاسب في كافة المجالات وجعلت الناس تفكر في الاستفادة من هذا الكم الهائل من المعلومات للخروج بنتائج تفيد البشرية، وهذه البيانات متعددة الاشكال والانماط والاحجام بصورة لا يمكن حصرها لذلك ظهرت الحاجة لدراستها وتحليلها والاستفادة منها من خلال تفسير هذه المعلومات الغير مفيدة في ذاتها لمعرفه يستفاد منها. من هنا ظهر مصطلح تعدين او تتقيب البيانات التي تشبه كثيراً تعدين المعادن من باطن الارض حيث يتم اخراج الكنوز من الارض من فضاء لا حصر له من الشوائب والمواد الغير مفيدة.

الكم الكبير من البيانات الموجودة حاليا في قواعد بيانات الطلاب يفوق القدرة البشرية على تحليل واستخراج المعلومات الأكثر فائدة دون مساعدة من تقنيات التعلم الآلي[6].

التنقيب عبارة عن عملية غير اعتيادية لاستخراج المعلومات الضمنية وذات الفائدة التي لم تكون معروفة سابقا من كمية هائلة من المعلومات،كما هي عملية اكتشاف وتحليل عن طريق ادوات لكميات ضخمة من البيانات للاكتشاف الانماط ذات معاني، والانماط هي المعرفة التي لم تكن معروفة من قبل او مجهوله عن استخدمها يوفر للمؤسسات مختلفة الاختصاص القدرة على اكتشاف المعرفة ،والسلوك والتنبؤ المستقبلي من المعلومات الكبيرة الموجودة باشكال مختلفة للمساعدة في اتخاذ القرارات.

ان المؤسسة التعلميمية احد هذه المؤسسات التي تحتاج الى تتقيب بياناتها للاستفادة منها في جودة التعليم وتحسين مستوى الطلاب ويمكن تطبيق التعدين في المدراس لقياس فهم الطلاب في مادة معينة او معرفة مستوى الطلاب في جوانب معينة وكذلك توقع درجاتهم المستقبلية وميولهم من هذه العوامل حتى تساعد الادارة التعليمية في اتخاذ القرار المناسبة.

تحتاج النظم التعليمية إلى طرق مبتكرة لتحسين جودة التعليم لتحقيق أفضل النتائج وتقليل معدل الفشل[52].

ولقد تم الاستفادة من تقنيات تنقيب البيانات في الاوسط التعليمية بكثرة، وهنالك الكثير من التجارب الناجحة في هذا المجال.

إنشاء نموذج للتنبؤ بأداء الطلاب يوفر فائدة كبيرة في تحديد الطلاب المعرضين للرسوب في وقت مبكر ، وخاصة في الفصول الكبير. باستخدام الخورازميات كالشبكات العصبية وتصنيف أشجار صنع القرارات ، ومع تحليل العوامل التي تؤثر على نجاح الطلاب [25].

قد تكون القدرة على التنبؤ بعلامة الطلاب مفيدة في عدد كبير من الطرق المختلفة المرتبطة بالتعلم على مستوى الجامعة[22].

يمكن استخدام خوارزميات استخراج البيانات للمساعدة في اكتشاف المعرفة ذات الصلة الواردة في قواعد البيانات التي تم الحصول عليها من النظم التعليمية على شبكة الإنترنت[23].

كما يمكن استخدام النتائج على حد سواء لمساعدة المعلمين في إدارة فصلهم التصنيف الطلاب من أجل التنبؤ بدرجاتهم النهائية بناء على الميزات المستخرجة من البيانات المسجلة في الشبكة [24].

يقدم هذا البحث دراسة لدرجات طلاب الاساس وذلك من خلال تحليل التقارير المدرسية الإحصائية الموجودة في مدرسة الانقاذ الاساسية، وذلك باستخدام التتقيب في البيانات التي تستطيع إيجاد آلية قادرة على دراسة العوامل التي تلعب دور في تحصيل الطلاب ونتائجهم من أجل الربط وتحديد مدى العلاقة بينهما وأهميتها في التنبؤ المستقبلي لدرجات الطلاب.

سوف نذكر في البحث مجموعة من النماذج التي تم اختبارها لتشكل عينة عن الاختبارات والامتحانات التي يبني عليها نتائج البحث.

1.2 مشكلة البحث:

تكمن مشكلة البحث في انخفاض درجات الطلاب في مادة الرياضيات وفقا لسجلات الطلاب في المدرسة لذلك كانت هنالك حوجة لوجود نموذج يمكن من خلاله التنبؤ بدرجات الطلاب في مرحلة مبكرة للمساعدة في تحسين مستوى الطلاب

1.3 اهداف البحث:

يمكن تلخيص الهدف من البحث في اربع نقاط:

- ◄ تصنیف الطلاب لمجموعتین (ناجحین وراسبین)
- الكشف عن قواعد الارتباط لفهم السمات التي تؤدي لنجاح او رسوب
 الطالب
 - الكشف عن الحالات الشاذة
 - ◄ توقع درجات الطلاب في مرحلة مبكرة

1.4 أهمية البحث:

تكمن أهمية البحث في استخدام أدوات وتقنيات التعلم الألى في مجال التنبؤ المستقبلي لدرجات الطلاب.

ومن المتوقع أن تسمح هذة الدراسة بتقديم مؤشرات لمدراء المدارس ، كما يقدم البحث في الجانب المعلوماتي اقتراحات للخوارزميات الأفضل في مجال التنبؤ اعتمادا على مجموعة من التحليل.

1.5 اسباب اختيار الموضوع:

يمكن تلخيص اسباب اختيار هذة المشكلة للسببين:

الدافع الذاتي للعمل في هذه المشكلة

خدمة المجتمع في ايجاد طريقة معينة تسهم في العملية التربوية لتنبؤ
 بدرجات الطلاب

1.6 منهجية البحث:

سوف يتم الاعتماد على اساليب التنقيب والتعلم الالى للاكتشاف المعرفة من نتائج الطلاب لمدرسة الانقاذ بغرض التنبؤ بالتغيرات التى تحدث فى المستقبل ، كما تم جمع عينة من نتائج الطلاب من المدرسة فى ملف 'Excel مكونة من 1500عينة ، اقتراح الباحث تطبيق هذة الدراسة عن طريق خوارزمية orange ببرنامج decision tree البيانات الى فئات معينة من ضمن فئات معروف مسبقاً حتى تعطى نظرة مستقبلية.

كما سوف يتبع الباحث فى هذة الدراسة المنهج الوصفى التحليلي بحيث يتم توصيف الطلاب والعوامل المؤثرة للنجاح والرسوب وتحليها بغرض التعميم والتنبؤ المستقبلي لسعى لحل الاشكالية قيد الدراسة.

1.7 حدود البحث:

الموضوعية :التوقع للامتحانات الرياضيات لمدأرس الاساسية محلية جبل اولياء -وحدة الازهري

الزمانية :يوليو -2019:يونيو-2020

الحدود المكانية: وزارة التربية والتعليم ولاية الخرطوم-محلية جبل اولياء – وحدة الازهري

1.8 مجتمع العينة:

تم استخدام ثلاثة مجموعات مختلفة من البيانات لهذا البحث المجموعة البيانات الأولى عينة التي جمعت من مدرسة الانقاذ والعينة الثانية من مدرسة الميثاق القرانية الاساسية الخاصة والعينة الثالثة من مدرسة العفاف الاساسية بنات كل هذه المدارس تقع في وحدة الازهري الادارية محلية جبل اولياء ولاية الخرطوم تحتوي على معلومات حول 1500 طالبًا من مختلف الفصول 500طالب لكل مدرسة.

قسمت العينة الى مجموعتين المجموعة الاولى تمثل مجموعة التدريب وعددهم 500طالب بينما المجموعة الثانية تمثل مجموعة الاختبار وعددهم 1000طالب.

1.9 هيكل البحث:

يتكون البحث من خمسة فصول:-

الفصل الاول | الاطار المنهجي للبحث

الفصل الثاني الدبيات البحث ، الدراسات السابقة

الفصل الثالث | عرض وتحليل البيانات

الفصل الرابع | النتائج

الفصل الخامس | الخاتمة

1.10 الاضافة العلمية:

تكمن الاضافة في متغيرات جديدة لم تستخدم في البحوث السابقة مثل عدد مرات فشل الطالب ،مجموعة الدرجات الكلية والنصفية للطالب.

الفصل الثاني: ادبيات البحث الدراسات السابقة

2 تنقيب البيانات 2.1 تعاريف تنقيب البيانات

هي استخراج أنماط مهمة من قاعدة بيانات معينة وهي أداة لها قيمة لتحويل البيانات إلى حقائق يمكن الاستفادة منها واستخدامها من قبل متخذي القرار أو طريقة لاكتشافها المعلومات من كمية كبيرة من البيانات .

التتقيب عن البيانات هو عملية اكتشاف المعرفة المثيرة للاهتمام ، مثل الجمعيات والنماذج والتغييرات، والهياكل الحرجة، من كميات كبيرة من البيانات المخزنة في قواعد البيانات أومستودعات البيانات أو مستودعات المعلومات الأخرى وقد استخدمت على نطاق واسع في السنوات الأخيرة لتوافر كميات هائلة من البيانات في شكل إلكتروني ، والحوجة لتحويل البيانات إلى معلومات مفيدة.

التنقيب عبارة عن عملية غير اعتيادية للاستخراج المعلومات الضمنية وذات الفائدة التي لم تكون معروفة سابقا من كمية هائلة من المعلومات ،كما هي عملية اكتشاف وتحليل عن طريق ادوات لمجموعة ضخمة من البيانات للاكتشاف انماط ذات معانى ، والانماط هي المعرفة التي لم تكن معروفة من قبل او مجهوله.

النمط: هو تعبير في اللغة يصف مجموعة فرعية للبيانات.

تهدف DMإلى تطوير الأساليب التي تستخدم أنواع فريدة من نوعها وبالتالي ، فإن نجاح تطبيق DMالتقنيات في بيانات الأعمال تشجع اعتمادها في مختلف مجالات المعرفة [6].

2.1.1 أسباب تنقيب البيانات

1. يهدف التتقيب على اكتشاف انماط جديدة

2. يساعد على اكتشاف الاتجاهات والسلوكيات المستقبلية التي تساعد لبناء القرارات

2.1.2 أنواع تنقيب البيانات 2.1.2.1 التنقيب الوصفي

يعتمد على اعادة تنطيم البيانات والتنقيب في اعماقها واستخراج النماذج الموجوة فيها كتشابة الزبائن

2.1.2.2 التنقيب التنبؤي

يعمل على ايجاد افضل التنبؤات من المعطيات المعرفة مسبقا عن طريق بعض المتغيرات واهمها عملية استخراج معلومات قديمة لتوقع لما سيحدث في المستقبل.

2.1.2.3 مهام تنقيب البيانات

العنقدة

الهدف منها هو تحديد الاتجاهات داخل البيانات ويحاول هذا الاسلوب العثور على مجموعات من العناصر التي توجد اعادة معا ،وهي عملية تقسيم البيانات الى مجموعة من الاصناف اعتمادا على اشتراكها بالخواص المتشابة وهي تقسيم غير موجة للبيانات.

التسلسل

تستخدم التحليل القائم على الوقت للانتزاع معلومات مفيدة وهي مماثلة لتجميع في انها تستخدم لتحديد العناصر التي تحدث معا.

التصنيف

يستخدم التصنيف لحل كثير من المشكلات من خلال تحليل مجموعة من البيانات ووضوعها بشكل اصناف يمكن استخدامها في ما بعد لتصنيف البيانات المستقبلية مثل شراء السلع عبر الانترنت .

2.1.3 تنقيب البيانات في التعليم

ازدهر تعدين البيانات التعليمية (EDM) في النظم التعليمية مؤخرًا يتيح تحليل أداء الطالب والتنبؤ به حتى يمكن اتخاذ التدابير مقدمًا [52].

دخلت تكنولوجيا تتقيب البيانات في الكثير من المجالات مثل :التسويق ، كشف الاحتيال ،الطب ، الانتخابات ، الخدمات المصرفية ، المراقبة ، وغيرة من المجالات.

فالتعليم واحد من أهم المجالات التي يمكن تنقيب البيانات في الكثير من الموضوعاتها مثل :استطاع رأي الطلاب في مادة معينة، تقيم المحتوى التعليمي، المساعدة في اتخاذ القرار، توقع مستوى الطلابالخ

وقد تم تطبيق DMفى الاوسط التعليمية للأهداف البحثية مثل تحسين عملية التعلم وتوجيه الطلاب التعلم أو اكتسابفهم أعمق للظواهر التعليمية [6].

مثال لهذه التطبيقات في مجالات مثل الذكاء الاصطناعي ، التعلم الآلي ، تحليل السوق ، الإحصاء وقاعدة البيانات[15] .

الهدف الرئيسي لمؤسسات التعليم العالي هو توفير التعليم الجيد لطلابها والطريق الوحيد الى تحقيق أعلى مستوى من الجودة هو تحديد العوامل التي تؤثر على الأداء الأكاديمي ثم محاولة حلها [18].

توقع الطالب هو احد تطبيقات التعليم يتطلب العديد من المعاير، والتنبؤ الفعال للأداء الأكاديمي للطالب يتطلب نموذج يشمل جميع الأحوال الشخصية والاجتماعية والنفسية وغيرها من المتغيرات البيئية.

مؤسسات التعليم العام مثل المدارس هي واحدة من بيئات التعليم التي يكتسب فيها الطالب الكثير من المهارات التعليمة والتربوية، المهمة الأساسية لهذه المؤسسات هي تلقين الطلاب وتهيئة النفسية والتربوية بحيث يكون الطالب جاهزا لمراحل القادمة.

تحتاج المؤسسة التعليمية إلى معرفة مسبقة للطلاب المسجلين للتنبؤ بأدائهم الأكاديمي في المستقبل ، مما يساعد المؤسسة التعليمية على تحديد الطلاب الناجحين وتحسين مستوى الذين قد يحصلون على درجات أقل[51].

يتم تخزين حجم متزايد من البيانات حاليا في قواعد البيانات التعليمية التي تحتوي على معلومات مخفية مختلفة يمكن أن تساعد في تحسين أداء الطلاب وبالتالي يتم استخدام تعدين البيانات التعليمية لدراسة البيانات واستخراج المعلومات المخفية للعمليات اللاحقة .

ويمكن استخدام البيانات التعليمية في العديد من العمليات مثل تقدير معدل تسرب الطلاب ،والتنبؤ عن القيم غير الطبيعية في نتائج الطلاب ،والتنبؤ بأداء الطالب في التمارين والواجبات الخ.

يمكن استخدام تقنيات التنبؤ لمساعدة المدارس للتنبؤ لدرجات الطلاب في مادة معينة أو متوسط مجموعة المواد،إذا يشير ناتج التنبؤ هذا إلى أن الطالب سيحصل على تقدير نجاح أو رسوب في المادة ، ثم بذل جهود إضافية يمكن إجراؤه لتحسين الأداء الأكاديمي للطالب في حالة التنبؤ الرسوب .

لخص الباحث أهداف التنقيب عن البيانات التعليمية في التنبؤ: في سلوك الطالب في التعلم، واكتشاف أو تحسين النماذج، ودراسة اثر الدعم التعليمي والتقدم في المعرفة العلمية لتنفيذ أساليب استخراج البيانات وأدوات تحليل البيانات المتاحة في المؤسسات التعليمية، والمعروفة باسم البيانات التعليمية [19].

يتطلب تعدين بيانات المدارس المعرفة السابقة من مجموعات البيانات المتاحة ، يمكن استخدامها لتمثيل و إيصال المعلومات إلى مديري المدارس لمراقبة الظروف و اتخاذ إجراءات لحل المشاكل.

يستخدم التنقيب عن البيانات في مجال التعليم لحل المشاكل التي يمكن عن تحدث في اداء الطالب الذي يمثل مصدر قلق كبير في المدارس والمعاهد التعليمية قد تؤثر عدة عوامل على الأداء، تشير الدراسة عن المكونات الثلاثة المطلوبة للتنبؤ هي :المعلمات التي تؤثر على أداء الطالب ، طرق التعدين البيانات ،الثالث هو أداة استخراج البيانات .قد تكون هذه المعلمات نفسية ،شخصية والبيئية [2] .

2.2 أساسيات التعلم الآلي

2.2.1 تعریف

التعريف الشائع للتعلم الآلي هو قدرة الكمبيوتر على التعلم من التجربة ،

ان الحاسب يمكنه التعلم من خلال التجربة Eعن طريق بعض المهام Tفاذا كان قياس الاداء افضل Pفيحدث تحسين في التجربة E،عادة ما يتم تقديم الخبرة في شكل بيانات الإدخال [40].

يستخدم التعلم الآلي للكشف عن بيانات مخفية في بيئات البيانات غير المهيكلة ، يمكن للكمبيوتر العثور على تبعيات في البيانات معقدة للغاية

بالنسبة للإنسان ، أو يمكن استخدامها للعثور على التوقعات في البيانات المنظمة.

2.2.2 التحليلات التنبؤية

التحليلات التنبؤية هي توقع الأحداث والسلوكيات المستقبلية الموجودة في البيانات غير المرئية سابقًا ، باستخدام نموذج مبني على بيانات سابقة مماثلة [41].

التحليلات التنبؤية يمكن تطبيقها في مجموعة واسعة من المجالات مثل: التمويل، التعليم، الرعاية الصحية والقانون[42].

نتشابه فى طريقة التطبيق في كل هذه المجالات باستخدام البيانات التي تم جمعها سابقا ، تجد خوارزمية التعلم الآلي العلاقات بين الخصائص المختلفة للبيانات.

النموذج التتبؤية هو القادر على التنبؤ بإحدى خصائص البيانات المستقبلية بناءً على الخصائص الموجودة [43] .

يعرض الجدول2-أمثلة للبيانات حول الطلاب الذين نجحوا أو فشلوا في الاختبار مع بعض المعلومات ألاخرى عن الطلاب.

| Age | Gender | GPA | Absences | Passed |
|-----|--------|-----|----------|--------|
| 14 | F | 3.2 | 5 | 1 |
| 13 | M | 2.4 | 7 | 0 |
| 15 | М | 3 | 6 | 1 |

جدول 2-2مثال لنتائج الطلاب

الهدف من ذلك هو التتبؤ بما إذا كان الطالب قد نجاح في الامتحان أم لا من خلال النظر إلى الخصائص و المتغيرات الآخرى .

في هذه الحالة ، يسمى العمود"Passed" المتغير التابع ، وكل متغير آخر يسمى المتغير المستقل، اذا حصل الطالب على "1" في العمود"Passed" " يعنى أن الطالب فشل في امتحان .

من خلال تطبيق خوارزمية تعلم الآلة على هذه البيانات ، يمكن انشاء نموذج التتبؤ ، والذي يعطي قيمة المتغير التابع ، ويأخذ كل متغير آخر كمدخل.

تسمى عملية إنشاء نموذج تتبؤ من البيانات المعروفة سابقاً بالتدريب ، وتسمى هذه البيانات بيانات التدريب أو مجموعة التدريب بعد إنشاء النموذج ، فإنه يجب تطبيقه على مجموعة بيانات أخرى لاختبار فعاليتها البيانات المستخدمة لهذا الغرض تسمى بيانات الاختبار أو مجموعة الاختبار السبب في استخدام مجموعتين مختلفتين هو التأكد من أن النموذج يتميز بالمرونة الكافية ليتم استخدامه في مجموعات البيانات بخلاف المجموعة التي تم تصميمها.

خلاف ذلك ، قد تحدث مشكلة over fitting ، وهو عندما يكون نموذج دقيق باستخدام مجموعة البيانات الأصلية الخاصة به ، لكن أداءه ضعيف على مجموعات البيانات الأخرى[44] .

طريقة شائعة لتجنب over fitting هو تقسيم مجموعة إدخال البيانات في مجموعات التدريب والاختبار.

لتقييم نموذج بيانات الاختبار ، يتم استخدام النموذج للتنبؤ بالمتغير التابع في مجموعة الاختبار.

ثم ، القيم المتوقعة والقيم الفعلية للمتغير التابع . التقييم أكثر تعقيدًا من النظر إلى عدد التوقعات .الصحيحة هناك العديد من معايير التقييم المختلفة

.

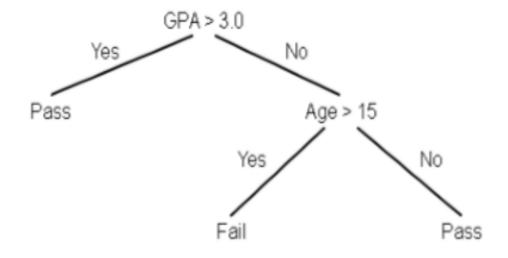
2.2.3 طرق مختارة

هناك العديد من الخوارزميات لإنشاء نموذج التنبؤ، في هذا البحث سوف نستخدم خوارزمية شجرة القرار DT وثم نقم بمقارنة نتائجها بخوارزميات مختلفة: svm, neural network, naïve Bayesليهم جميعًا نفس المهمة بشكل أساسي، وهي التنبؤ بمتغير تابع على متغيرات مستقلة، فهي تستند إلى طرق رياضية مختلفة.

2.2.3.1 أشجار القرار

أشجار القرار هي هياكل بيانية ،تتكون من قرارات حيث يقوم كل قرار محتمل بإنشاء عقدة جديدة ،قد تكن هذه العقدة قرار جديد مما أدى إلى رسم بياني يشبه الشجرة[45]. يوضح الشكل 2-دمثال على القرار

شجرة.



شكل 2-1 مثال لشجرة قرار

تستخدم هذه الشجرة للتنبؤ بما إذا كان الطالب قد نجاح فى الامتحان من خلال النظر في المعدل التراكمي "3.0 حPA"اذا كانت القيمة"نعم" يعتبرالطالب ناجحا و اذا كانت القيمة" لا " تنتقل الى شرط العمر إذا كان و 15<"العمر " يعتبرالطالب راسبا اذا كانت 15<"العمر " يعتبرالطالب ناجحا .

في التعلم الآلي ، تقسم أشجار القرار مجموعة البيانات في القيم المناسبة حتى ظهر هيكل الشجرة،تسمى هذه العملية التقسيم العودية[46] .

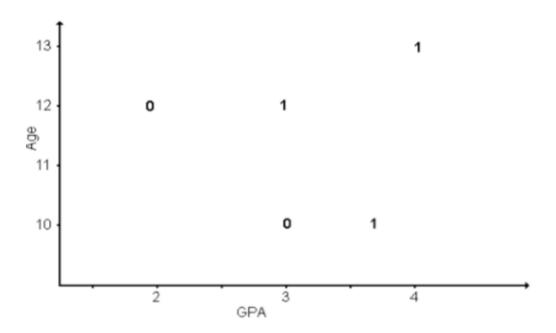
تحاول خوارزمية شجرة القرار العثور على أفضل طريقة لتقسيم البيانات بحيث تكون الأجزاء متجانسة قدر الإمكان .

يتم توضيح هذه العملية من خلال إنشاء شجرة قرار من بيانات الطلاب المبينة في الجدول2-4.

| • | GPA | Pass |
|----|-----|------|
| 10 | 3 | 0 |
| 10 | 3.6 | 1 |
| 12 | 2 | 0 |
| 12 | 3 | 1 |
| 13 | 4 | 1 |

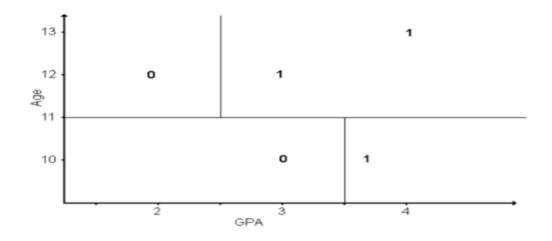
جدول 2-5يوضح شجرة القرار

الهدف هو النتبؤ بقيمة" النجاح "باستخدام القيم" العمر "و" المعدل التراكمي" ، تحتوي البيانات على متغيرين مستقلين فقط ، انظر الشكل(2-6).



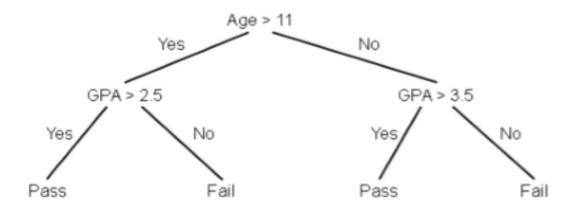
شكل 2-2 مجموعة من البيانات المبعثرة

مجموعة مبعثر من بيانات الطالب بين المحوريين X_e تمثل المتغيرات المستقلة، يمكن تقسيم المتغير التابع باستخدام خوارزمية شجرة القراركما هو مبين في الشكل 2-2.



شكل 2-4تقسيم المتغيرات

تحدد خوارزمية شجرة القرار مواقع وعدد ألاقسام .يمكن عرض البيانات الموجودة في المخطط المقسم على شكل شجرة قرار.



شكل 2-5مخطط شجرة القرار

2.2.4 ميزة الهندسة

في التعلم الآلي ، هندسة الميزات هي عملية اختيار المتغيرات أو إنشاء ميزات في مجموعة بيانات لتحسين نتائج التعلم الآلي [47].

كما يمكن أن يشمل الاختيار إزالة الميزات غير الضرورية أو الزائدة عن الحاجة و تتطلب إزالة المتغيرات غير الضرورية تقييم مدى أهمية المتغير ويمكن القيام به عن طريق إنشاء نموذج لاختبار ارتباط المتغير مع المتغير التابع.

يتضمن إنشاء ميزة تعديل المتغيرات وإنشاء جديد منها بدمج متغيرات مختلفة متعددة [48].

قد تحتوي بيانات الإدخال على متغيرات كثيرة ، بعضها لا يحسن أداء التنبؤ ، وبالتالي يجعل النموذج التنبؤي معقدة للغاية .في مثل هذا الحالة ، يجب إزالة المتغيرات غير الضرورية لجعل النموذج أكثر كفاءة يمكن تحديد المتغير المراد إزالته يدويًا باستخدام معرفة المجال أو ذلك يمكن القيام به تلقائيا 47] .

اختيار الميزة تم من خلال مراقبة إخراج نموذج الانحدار الخطي لإيجاد مقدارارتباط كل متغير مع المتغير التابع، الاستخدام الثاني للهندسة المميزة في الدراسة هو تعديل المتغيرات. هذه يمكن أن تشير إلى الجمع بين متغيرات متعددة لإنشاء متغير جديد، حساب المتغير بطريقة مختلفة بحيث يمكن استخدامها بشكل أفضل في التصنيف، أو تصنيف متغير بذلك أن لديها مجموعة محدودة من القيم الممكنة.

مثال على تعديل متغير يمكن يتم إعداده مع مجموعة بيانات الطالب التي تحتوي على اللغة الأصلية للطالب كواحدة من المتغيرات .يحتوي الجدول 2-7 على البيانات.

| ID | Age | level | Mather salary | Result |
|----|-----|-------|---------------|--------|
| 1 | 7 | 1 | 3500.00USD | 1 |
| 2 | 7 | 2 | 2700.00USD | 0 |

| 3 | 9 | 3 | 3500.00USD | 1 |
|---|---|---|------------|---|
| 4 | 8 | 2 | 2750.00USD | 1 |

جدول 2-8 مثال على تعديل المتغيرات

الجدول مثال على البيانات لهندسة الميزة.

في هذا المثال ، يحتوي المتغير " Mather salary " على قيم مالية ومع ذلك ، فإن هذه المتغير ليس له تاثير على اداء الطالب" .

في مثل هذه الحالة، يتم الاستغناء عن هذا المتغير بحيث يتم حذفه.

تعديل البيانات بهذه الطريقة لا يؤثر على أهمية المتغيرالتابعResult ، بينما يجعله أكثر بساطة و فائدة ونتائج هذا التعديل يتطلب المعرفه بالمجال في هذه الدراسة ، يتم تعديل الميزة عن طريق إنشاء ميزة جديدة للمتغير كدالة لمتغيرات مختلفة او عن طريق الاستغناء المتغيرات الغير مؤثرة.

2.2.5 طرق التقييم

من أجل تقييم فعالية نموذج التنبؤ ، يجب أن تحسب القيم المتوقعة مقارنة مع القيم الفعلية في حالة تساوى القيم بنسبة كبيرة يعني عن نموذج التنبؤ . جيد ، هناك معايير متعددة لفعالية التنبؤ.

يوضح الجدول2-9 النتائج المحتملة للتتبؤ بالقيم الثنائية.

| | Predicted as True | Predicted as False |
|----------------|-------------------|--------------------|
| Actually True | True Positive | False Negative |
| Actually False | False Positive | True Negative |

جدول 2-10 تقيم فعالية التنبؤ

تسمى المصفوفة التي تعرض نتائج التنبؤ المحتملة بمصفوفة الارتباك[49]

هناك معايير تقييم مختلفة يمكن الحصول عليها من هذه القيم .واحد هو الدقة[50].

قانون الدقة:

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

الدقة هي في الأساس نسبة التوقعات الصحيحة سواء عن كانت ايجابية او سلبية ومع ذلك ، فإن الدقة لها حدود في تقييم أداء التنبؤ .

على سبيل المثال ، مجموعة البيانات التي تحتوي على 100 طالب ، 90 منهم اجتازوا الاختبار ، ربما يعتبرتنبؤ خام (يُعرف باسم قاعدة الأغلبية) لا يستخدم في طرق التعلم الآلي ، ولكن بدلاً من ذلك تتوقع أن يمر كل طالب بالامتحان بدقة 90 ٪ يجب أن يكون الأداء أفضل من مجرد تخمين.

في هذه الدراسة ، يتم استخدام ثلاثة معايير أخرى .اثنان منهم الدقة والتذكر ،[50].

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

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

يتم استخدام الدقة والاستدعاء معًا لإجراء تقييم أفضل.

الفكرة الرئيسية فى ذلك هي التنبؤ بدقة النتائج إلايجابية ليست كافية .نموذج التنبؤ الجيد يجب لديه مزيج من التوقعات الإيجابية الناجحة .

المعيار الثالث الذي تستخدمه هذه الأطروحة يسمى F-measure ، وهو كذلك تم تعريفه على أنه[49]

$$F = 2.\frac{Precision \cdot Recall}{Precision + Recall}$$

F-measureهو وسيلة للحصول على قيمة واحدة تأخذ الدقة والتذكر فيها الحساب F-measure . هو معايير التقييم النهائية للمقارنات في هذه الرسالة.

2.3 الدارسات السابقة:

هناك الكثير من البحوث التى تم تتفيذه حول التتبؤ بالأداء الأكاديمي باستخدام تقنيات استخراج البيانات والتعلم الآلي.

الدراسة [1] هدف الباحث إلى تنفيذ العديد من التنبؤات التقنية في التنقيب عن البيانات لمساعدة المؤسسات التعليمية في التنبؤ بالمعدل التراكمي GPA لطلاب عند التخرج.

إذا كان الناتج المتوقع انخفاض مستوى الطلاب GPAعند التخرج، مما يمكن المؤسسات التعليمية من بذل مجهود إضافي لتحسين الأداء الأكاديمي ، استخدم الباحث المنهج الوصفي لاستخراج البيانات في شكل عددي ، كونت عينة الدراسة من سجلات الطلاب التي تحتوى على 127 طالب وطالبة من جامعة فيران التركية قسم الحاسوب من عام 2007 الى 2011 كان عليهم اجتياز 49 دورة مهنية وثقافية.

تستخدم الجامعة النظام المئوي 100% مقسم على درجات عملية 40% و 60% نظري ويرمز لمعاملات النجاح بال ((,60% نظري ويرمز لمعاملات النجاح بال ((,CD,1.5),(CD,1.5),(DD,1.),(BA,3.5),(BB,3.0),(CB,2.5),(CC,2.0),(CD,1.5)) تم إنشاء مصفوفة تحتوي على عدد الطلاب * عدد الدورات إي 127 *49 تظهر الصفوف أسماء الطلاب والأعمدة درجات الدورات للإغراض التنبؤ .

قسمت البيانات إلي مرحلتين في المرحلة الأولى تم استخدام المعدل التراكمي باستخدام خوارزميات الأولى للتنبؤ في المرحلة الثانية المعدل التراكمي الذي يتضمن مجموعه 38 درجة.

تم ضبط جميع المعاملات لكل طرق التنبؤ وفقا لهذه التجارب الواسعة النطاق SVM و SVM و ELM ويشير التحليل للنتائج عن تقنية NN وباستخدام خوارزميات NN و NN و ELM ويشير التحليل للنتائج عن تقنية 94.92 %) أكثر دقة بمعدل (97.98 %) و طريقة ELM الثاني الأكثر دقة معدل (94.92 %) تقييمها على أساس معيار معامل الارتباط. كما ذكرت الدراسة عن خوارزمية NN ذات المعدل الأقل للتنبؤ بمستوى دقة (93.76 %).

صنف الباحثون في الدراسة [2] العوامل المؤثرة على أداء الطلاب للتنبؤ لثلاثة مكونات هي: المعاملات التي تؤثر على أداء الطالب، طرق تعدين البيانات والثالث هو أداة استخراج البيانات. قد تكون هذه المعاملات نفسية وشخصية وبيئية.

تعمل هذه الدراسة للحفاظ أداء الطالب،ثم التنبؤ بأداء الطالب بواسطة تطبيق خوارزمية bays Naive و J48 كما تم تصنيف شجرة القرار بأداة WEKA.

تمكن الباحث من الحصول على المعرفة التي تصف أداء الطالب. هذه المعرفة التي سوف تساعد على تحسين جودة التعليم وأداء الطالب وخفض معدل الفشل.

في الدراسة [3]باكتشاف العلاقات بين العوامل الشخصية والاجتماعية للطلاب في الفصل الدراسي السابق باستخدام مهام استخراج البيانات، فيمكن التنبؤ بأدائها في الفصول القادمة.

ركز الباحث في الدراسة [6] على تحديد عوامل الفشل بين الطلاب وعرضها من خلال التتقيب عن البيانات التتبؤية باستخدام خوارزميات تصنيف القائمة.

أخذت الالدراسة، لمدرسة الثانوية باستخدام WEKA وتم استخدام المنهج التجريبي كما تم تطبيقها على خوارزميات تصنيف مختلفة مثل الإدراك المتعدد الطبقات ، REPTree و J48 ، SMO ، Naïve Bays ،تم إنشاء الإحصائيات بناءً على جميع خوارزميات التصنيف ومقارنة الكل كما يتم إجراء خمسة مصنفات للتنبؤ بالدقة ولإيجاد أفضل خوارزمية تصنيف أداء

بين الكل. في هذا الدراسة ، كما تم عرض نموذج تدفق المعرفة بين جميع المصنفات الخمسة.

هذه البحث يوضح أهمية استخدام خوارزميات التنقيب عن البيانات القائمة على التنبؤ والتصنيف في مجال التعليم، Multi-Layer Perception تعمل بشكل أفضل مع دقة 75 %، وبالتالي يثبت MLP يحتمل أن تكون أدقة خوارزمية لتصنف بالمقارنة مع جميع المصنفات الأخرى.

الدراسة [7] هدفت هذه الدراسة لتحديد الفجوات في طرق التنبؤ الحالية وتحديد المتغيرات المستخدمة في تحليل أداء الطلاب.ولدراسة طرق التنبؤ الحالية للتنبؤ بأداء الطلاب بالمنهج التجريبي لدراسة مدى ملائمة ألأسئلة مع أهداف هذه الدراسة. تتمتع Neural Network بأعلى دقة تنبؤ بنسبة أهداف هذه الدراسة. القرار بنسبة 91%، ثم شجرة القرار بنسبة 91%، المائدة القرار بنسبة 83%. كما كانت أقل نسبة لخوارزمية Naive Bayes بنسبة 76%.

قدم الباحثون دراالأربع،في تطوير طرق لاكتشاف المعرفة من البيانات التعليمية لتحسين أداء طلاب الدراسات العليا ، والتغلب على مشكلة الرسوب في المواد الدراسية حيث قام الباحثون باستخراج معلومات مفيدة من بيانات الطلاب التي تم جمعها من كلية العلوم والتكنولوجيا – خانيونس في الفترة ما بين 1993–2007، بتطبيق تقنيات استخراج البيانات لاكتشاف الارتباط والتصنيف والتجميع وكشف أنماط وقواعد جديدة مفيدة[8].

في كل من هذه المهام الأربع ، قدموا المعرفة المستخرجة ووصف أهميتها في المجال التعليمي.

استخدمت تقنيات استخراج المعرفة للبيانات التطبيقية وذلك لتصيف العلاقات بين الطالب والعوامل السلوكية لتطوير نموذج لتوقع أداء الطالب.

باستخدام خوارزمية دعم تصنيف آلة المتجهات (SSVM) وتقنيات التجميع [9].

فكانت نتائج هذا الدراسة نموذجًا للتنبؤ بالأداء الأكاديمي للطالب من خلال استخدام العوامل و المتغيرات السلوكية لتنبؤ كما توصلت هذه الدراسة إلى أن هنالك علاقة قوية بين الحالة العقلية للطالب و الأداء الأكاديمي.

تمت مقارنة أكثر من طريقة وتقنية لاستخراج البيانات لتوقع درجات الطلاب ، تم جمع البيانات من الدراسات الاستقصائية التي أجريت خلال الفصل الصيفي في الجامعة منتو زلا ، كلية الاقتصاد ، العام الدراسي 2010–2011 ، بين طلاب السنة الأولى والبيانات التي اتخذت أثناء التسجيل بواسطة إيدن عثمان بيكوفيتش وآخرون [10]تم تقييم النجاح في امتحان الصف وتأثير المتغيرات الاجتماعية والديموغرافية للطلاب ، والنتائج المحققة في المدرسة الثانوية ومن تم التحقيق في امتحان القبول ، والمواقف تجاه الدراسة التي يمكن أن يكون لها تأثير على النجاح.

أظهرت أن مصنف Naïve Bays يتفوق على شجرة القرار وأساليب الشبكة العصيبة.

يتم استخدام طريقة التصنيف Bayesian في قاعدة بيانات الطلاب للتنبؤ تقسيم الطلاب على أساس قاعدة بيانات العام السابق.

[11] أظهرت هذه الدراسة أن الأداء الأكاديمي للطلاب لا تعتمد دائما على جهودهم المبذولة في المواد (2005–10)، هناك عوامل أخرى لها أهمية كبيرة واثر على أداء الطلاب. وقد ساعدت الدراسة المقترحة من قبلهم في الحد من الفشل واتخاذ الإجراءات المناسبة في الوقت المناسب.

[12] تم تسجيل سلوك الطلاب المسجلين وأنشطتهم ومعالجتها وتحليلها في برنامج دورة محددة عبر 6 سنوات (2005–10) ، مع سمات متعددة تم جمعها من قاعدة بيانات الجامعة. ونتيجة لذلك، كانت هذه الدراسة قادرة على

التنبؤ، إلى حد ما بالدرجات النهائية للطلاب في مادة محددة، وكذلك، "مساعدة الطالب على تحسين أداء الطالب، وتحديد الطلاب في الوضع الحرج للحد من الفشل والمساعدة في اتخاذ القرارات المناسبة في الوقت المناسب ".

قام الباحثون [13] بإجراء بحث لتوقع أداء الطلاب نجاح أو رسوب باستخدام خورزامية التصنيف Naïve Bays لتحليل وتصنيف أداء الطلاب Bays خورزامية التصنيف الاحتمالات ، الذي يفترض أن جميع السمات المعطاة في مجموعة البيانات هي مستقلة عن بعضها البعض ، اعتمد البحث على عينة من البيانات الطلاب المسجلين في دبلوم الدراسات العليا بكلية تطبيقات الحاسوب(PGDCA) في جامعة D.R.M ، فايز آباد ، الهند. كان البحث قادراً على طريق استبيان لتنبؤ لدرجات الطلاب في عامهم القادم ، بناء على درجاتهم في العام السابق.

تم جمع البيانات عن طريق استبيان قدمت لجميع طلاب القسم ، والتي تشمل العديد من الخصائص والسمات كالسمات الشخصية والاجتماعية ، كما استخدمت بعض الأسئلة النفسية لمعرفة الحالة النفسية لطلاب وذالك لفهم العلاقات بين هذه العوامل وأداء الطالب والدرجات. وحدد الباحثون (4) أهداف رئيسية لهذه الدراسة على النحو التالي مرحلة الثانوية صدر بيانات المتغيرات التنبؤية (ب) تحديد العوامل المختلفة ،مما يؤثر على سلوك تعلم الطالب ، (ج) بناء نموذج التنبؤ باستخدام تقنيات استخراج بيانات و التصنيف على أساس المتغيرات التنبؤية المحددة ، و (د) التحقق من صحة تطوير نموذج لطلاب .

وجد الباحثون أن العامل المؤثرة في أداء الطالب هو درجته في المرحلة الثانوية والتي تساعدهم للحصول على أداء جيدا في دراسة البكالوريوس.

كذلك الوضع المعيشة، وسيلة التدريس، عادات الطالب الأخرى، دخل الأسرة السنوي، و وضع عائلة الطالب، تساهم بشكل كبير في الأداء التعليمي للطلاب وبالتالي فإنه يمكن التنبؤ بدرجة الطالب أو أدائه بشكل عام إذا تم جمع المعرفة الشخصية والاجتماعية عنه.

[5] قام الباحثون بمقارنة عدة خوارزميات لتصنيف الأداء على مجموعة البيانات التعليمية للطلاب ،تركز الدراسة بشكل أساسي على اختيار أفضل الخوارزميات، وتوفير معيار لكل واحد منهم.

اكتشف الباحثون إن أفضل طريقة لتصنيف مجموعة البيانات التي تم اختبارها هي شجرة القرار ، والتي تم اختيارها بناء على الدقة المنتجة باستخدام 10 أضعاف لتحقق من صحة.

قدمت هذه الدراسة طرق جيدة لتحديد أفضل خوارزمية لتصنيف مجموعات البيانات المحددة عن طريق اختبار خوارزميات وتقنيات متعددة قبل اختيار أي واحدة سوف تعمل بها في نهاية الدراسة ، فمن المستحسن للغاية لاختبار مجموعة البيانات مع المصنفات متعددة أولا ، ثم اختر الأكثر دقة وافضل طريقة لتصنيف لأي مجموعة بيانات.

[15]قم الباحثون بتحليل بيانات الطلاب المسجلين في السنة الأولى بكلية الهندسة بغرض التنبؤ . و تم الحصول على البيانات من المعلومات المقدمة من قبل الطلاب المقبولين في المعهد التي تتضمن الاسم الكامل والجنس ومعرف الطلب والنتائج في امتحانات الفصول، درجات امتحان القبول والفئة ونوع القبول.

طبقت خوارزمية شجرة القرارات D3او C4.5 للتنبؤ بنتائج الطلاب في الفصل الدراسي الأول .

كما قسم الباحثون التنفيذ إلى خمس مراحل. في المرحلة الأولى، جمع معلومات الطلاب الذين تم قبولهم وتشمل كل التفاصيل المقدمة إلى الكلية إثناء التسجيل.

في المرحلة الثانية: تنظيف البيانات التي تم جمعها كما تم إدخال المعلومات ذات الصلة في قاعدة بيانات.

المرحلة الثالثة هي المعنية بتطبيق خوارزميات أشجار القرار ID3 و C4.5.

وفي المرحلة الرابعة، بيانات الاختبار، أي معلومات عن الطلاب المسجلين في السنة الأولى، تم تطبيقه على أشجار القرار.

وتألفت المرحلة النهائية من تطوير الواجهات في شكل تطبيق ويب.

استخدم الباحثون برنامج Rapid Miner في العثور على معلومات مخفية من مجموعة بيانات التدريب.

ثم بعد ذلك تم تنفيذ أنماط تتبؤية مستفادة حديثًا للتتبؤ بأداء الطلاب وعمل تطبيق ويب للموظفين لاستخدامه للحصول على النتائج المتوقعة للقبول الطلاب تم الحصول على مستوى دقة 75.275

عرض الباحث في هذه الدراسة [16] نموذجًا لاستخراج البيانات التعليمية من أجل التنبؤ بأداء الطالب في مادة البرمجة، وذلك بتحديد المتغيرات التي تساعد الطالب في تحسين الأداء في مادة البرمجة لمساعدة المعلمين في تحسين العملية التعليمية.

استخدم الباحث في هذه الدراسة عوامل كثيرة مثل الخلفية الرياضية للطلاب، والاستعداد للبرمجة، مهارات حل المشكلات، الجنس، الخبرة السابقة، المدرسة

الثانوية، اختبار الرياضيات، الصف، برمجة الكمبيوتر السابقة، واستخدام التعلم الإلكتروني.

النموذج المقترح يتضمن ثلاث مراحل ؛ معالجة البيانات ، واختيار السمات المناسبة واختيار الخوارزمية الأفضل تناسبا مع غرض الدراسة وتم إجراء العديد من المحاولات لتصميم الاختبارات التي يمكن الاعتماد عليها في توقع درجات الطلاب في دورات البرمجة ، بحثت الدراسة العلاقة بين البرمجة واكتساب المهارات والتدابير المختلفة للفرد بما في ذلك: 1) الخلفية الرياضية. 2) مشكلة حل القدرات ؛ و 3) المهارات المعرفية

كون الباحث اختبار يحتوي على 20 أسئلة لطلاب قسم علوم الحاسوب في جامعة المنصورة على مدار ثلاث سنوات.

النتائج التي تم الحصول عليها وبيانات الطلاب حول المتغيرات الأخرى التي تم جمعها. هذه المتغيرات أو السمات الواردة مثل الجنس، درجة الطالب في العام الدراسي السابق درجة الرياضيات في المدرسة الثانوية، الكمبيوتر السابق، تجربة البرمجة، واستخدام التعلم الإلكتروني الذي هو أنتجت تحت إشراف المركز الوطني للتعليم الإلكتروني، مصر.

دورة التعلم الإلكتروني مفعلة على ثلاثة سنوات أكاديمية يتم تحليل البيانات التي تم جمعها عن طريق استخراج البيانات التقنيات والتنبؤ بالنتائج

[17] استخدم الباحث مجموعة من النماذج الرياضية المختلفة للتنبؤ بأداء الطلاب للتدخل في مرحلة المبكرة لتحسين مستوى الطلاب في المرحلة الجامعية ، استخدم الباحث مجموعة العوامل من المؤسسات الأكاديمية ، مثال لهذه العوامل النفسية والاقتصادية لتصنيف درجات الطلاب إلى نجاحين و راسبين باستخدام الشبكات العصبية متعددة الطبقات NN.

خوارزمية التعلم ليفنبرغ ماركوارت كانت قادرة على تحقيق أفضل أداء مع دقة تصنيف متوسط 83.7 % ، والحساسية

77.37 %، خصوصية 85.16 %، القيمة التنبؤية الإيجابية 94.04 %، و القيمة التنبؤية السلبية 85.16 %، وتمت مقارنة نموذج NN مع المصنفات الأخرى على وجه التحديد كأقرب جار ، شجرة القرار و Vector Machine على نفس مجموعة البيانات باستخدام نفس الميزات. وكانت النتائج إلى أن NN يتفوق على جميع المصنفات الأخرى من حيث الإجماليدقة التصنيف

[18] هدفت الدراسة لمعرفة ما إذا كان هناك أي أنماط في البيانات المتاحة (سجلات الطلاب والدورات التدريبية) قد تكون مفيدة للتنبؤ بأداء الطلاب.

شملت هذه الدراسة عينة مكونة من 150 طالبًا تم جمعهم من جامعة نجران بالمملكة العربية السعودية. تم التقاطها وترتيبها باستخدام حزمة إحصائية للعلوم الاجتماعية (SPSS) وأداة استخراج البيانات تستخدم نموذج البيانات القائمة على التعدين لتحديد أي من العوامل المعروفة التي يمكن أن تعطي مؤشرًا مبكرًا للأداء المتوقع. وظفت هذه الدراسة كلا ميزة تخفيض وتقنية التصنيف للحد من معدل الخطأ. النتائج التجريبية تكشف عن أهمية العلاقات بين كل من العمل العملي والواجبات في الدورة ومعدل النجاح. ولكن، من ناحية أخرى نجد عن توزيع عدد المهام المعينة له تأثير سلبي على الأداء الأكاديمي للطالب، كما نجد العامل الأكثر تأثيرًا هو حضور الطلاب في الفصل بالإضافة إلى الاختبار النهائي ودرجات امتحانات النصفية

كما يركز الباحثون [20]في البحث في مجال التعدين التربوي لنمذجة أداء الطالب بدلاً من أداء المدرسين.

واحدة من الأدوات الشائعة لتقييم أداء المعلمين استبيان عن طريق الطلاب استخدمت في هذه الدراسة أربعة تقنيات لتصنيف مختلفة ،خوارزميات شجرة القرار ، ناقلات الدعم ، الشبكات العصبية الاصطناعية ، والتحليل التمييزي لبناء نماذج المصنف.

قدم الباحثون في [21] نموذجًا للتنبؤ بأداء الطالب من خلال امتحان.الصف استخدم الباحث مجموعة المعلمات الموجهة للتنبؤ مثل المتغيرات الاجتماعية والديموغرافية ، النتائج المحققة من المدرسة الثانوية ،امتحان القبول .

[12] ركز الباحثون في هذه الدراسة على مهنة الاباء ونوع المدرسة. النتائج التحصلة عليها من الفرضية تكشف أن هذا النوع من المدرسة لا يؤثر على أداء الطلاب بينما تلعب مهنة الوالدين دورًا رئيسيًا في التنبؤ

[26] استخدم الباحث أدوات تتقيب البيانات التعليمية لتوقع المعدل التراكمي النهائي للطلاب بناءً على درجاتهم في الدورات السابقة، كما قم الباحثون بجمع نسخة من بيانات الطلاب التي تضم المعدل النهائي ودرجاتهم في جميع الدورات الدراسية.

بعد المعالجة المسبقة للبيانات ، تم تطبيق خوارزمية شجرة قرارات J48 لاكتشاف قواعد التصنيف لاستخراج معرفة مفيدة ، وتحديد الدورات الأكثر أهمية في الخطة الدراسية للطلاب بناءً على درجاتهم

قام الباحثون في الدراسة [27] ببناء نموذج تصنيف للتنبؤ بمسار الدراسة المناسب لطلاب المدارس. تتكون البيانات من 248 حالة التي تم جمعها من ست مدارس أساسية في مدينة المفرق في الأردن. وتم الحصول على مستوى دقة 87 ٪. ، باستخدام خوارزمية شجرة القرار

قام الباحثون في الدراسة [28] بتحليل بيانات الطلاب باستخدام خوارزمية تصنيف ID3 لتوقع درجات الطلاب في نهاية الفصل الدراسي. لقد قدمت هذه الدراسة بغرض الحصول على درجة الماجستير في تطبيقات الحاسب الآلي لعام 2007 حتى عام 2010 في جامعة VBS بورفانشال ، تهدف الدراسة لمساعدة المعلمين على إيجاد طرق للتحسين أداء الطلاب. تم جمع البيانات من 50 طالبًا ،ثم تم استخراج مجموعة من القواعد لتحليلها.

الدراسة [29] ركزت على تحسين أداء الطلاب باستخدام تقنيات استخراج البيانات حيث جمعت البيانات من السجلات الشخصية و الأكاديمية للطلاب لتحديد خصائصهم واكتشاف القيم الشاذة من نظام إدارة قواعد البيانات بجامعة غزة الإسلامية من 151 حالة.

لتنفيذ تقنيات استخراج البيانات، وهي: القواعد، التصنيف، التجميع والكشف الخارجي وكشفت النتائج الكثير من المعلومات المفيدة من قواعد و نماذج التصنيف.

تم الحصول على معارف مفيدة لتحسين أداء الطلاب في دورة قواعد البيانات.

تهدف الدراسة [30] لتوقع درجات الطلاب باستخدام بيانات التسجيل من جامعة الغربية في ولاية فرجينيا كونت العينة من 112،390 حالة. . قارن الباحثون النتائج المختلفة وجدوا أن خوارزمية 48J و هي الأفضل.

في [31] ، جامعة الاقتصاد الوطني والعالمي (UNWE) في بلغاريا تم تحليل بيانات تسجيل الطلاب كان الهدف لهده الدراسة هو التنبؤ بأداء الطالب على أساس خصائص الطلاب ما قبل الجامعة.

قام الباحث بأجراء التطبيق بعدة خوارزميات وأظهرت النتيجة أن شجرة القرار JRip) و k-NN.

وفي الدراسة حاول الباحثون تطبيق مجموعة مختلفة من تقنيات التصنيف لمجموعة البيانات للمقارنة أدائها واختيار أفضل الخوارزميات لتكون متكامل في ألأداة (E-Learning Web Miner).

هذه الأداة تهدف إلى مساعدة المعلمين على اكتشاف أداء طلابهم.حيث جمعت البيانات من مادة "مقدمة إلى أساليب الوسائط المتعددة "، في جامعة كانتا بريا لثلاث سنوات أكاديمية من 2007 إلى 2010.

واستخدم الباحثون طرق مختلفة لتصنيف البيانات حيث ووجدوا أن أداء ودقة التقنيات تعتمد على نوع السمات وحجم مجموعة البيانات[32].

تم تنفيذ مجموعة من خوارزميات تنقيب البيانات المختلفة حيث هدفت إلى تصنيف الطلاب حسب العلامات النهائية في مجموعات مختلفة اعتمادا على الأنشطة المنجزة على شبكة الإنترنت.

وتشمل هذه الأنشطة: عدد المهام المنجزة ، وعدد من الاختبارات التي اتخذت ، وعدد الاختبارات التي المستخدم في الاختبارات ، تتكون مجموعة البيانات من 438 طالبًا من جامعة قرطبة في 7دورات .

حيث قاموا بتقييم ألأداء بخوار زميات تستند إلى نوع السمات: الرقمية.

وجدوا أن خوارزمية شجرة القرار CART و C4.5 هي الأفضل[33].

استخدم الباحثون التصنيف للتنبؤ بدرجات الطلاب على أساس المتغيرات الاجتماعية والديموغرافية مثل (العمر والجنس ،العرق التعليم ، وحالة العمل ، والإعاقة) ومتغيرات البيئة (برنامج الدورة ،كتلة الدورة التدريبية).

مجموعة البيانات تحتوي على بيانات الطلاب في دورة نظم المعلومات في كلية الفنون التطبيقية بنيوزيلندا.

استخدمت في هذه الدراسة أربعة أشجار للتصنيف البيانات وهي: CHAID، OUEST هي الأفضل أداء بنسبة لتصنيف 60.5٪ [34].

نطاق هذه الدراسة هي التحقيق في دقة تقنيات استخراج البيانات.

الخطوة الأولى من الدراسة هي جمع بيانات الطالب جمعت 300 من سجلات طلاب الدراسات العليا في دورة علوم الكمبيوتر ، من مؤسسة تعليمية خاصة.

والخطوة الثانية هي تنظيف البيانات واختيار السمات ذات الصلة.

الخطوة الثالثة تم بناء الخوارزميات وتقييم أدائها ، NaiveBayesSimple، الخطوة الثالثة تم بناء الخوارزميات وتقييم أدائها ، REPTree ، J48 ، SMO ، Multilayer Perception أن Multilayer Perception أكثر دقة من الخوارزميات الأخرى[35].

حاول الباحثون التنبؤ بالفشل في الفصول الأساسية لمادتي (الرياضيات والبرتغالية) لطلاب المدارس الثانوية من منطقة الينتيخو البرتغال عن طريق استخدام 29 من المتغيرات التنبؤية.

استخدمت أربعة خوارزميات لتعدين البيانات مثل شجرة القرار (DT)، عشوائية الغابات (RF)، الشبكة العصبية (NN) ودعم المتجهات الآلة (SVM) على مجموعة من 788 طالبًا، الذي ظهر في عام 2006 الفحص.

حصل الباحثون على أن DT أكثر دقة تتبؤية بمستوى 93 % وكان خوارزميات 91NN لمجموعة البيانات من فئتين (النجاح / الرسوب) [36].

طبقت خمس خوارزميات لتنبؤ بأداء طلاب علوم الكمبيوتر في الجامعة اليونانية الهيلينية المفتوحة وهي أشجار القرار ، التعلم القائم على بيرسبترون ، شبكات بايزية ، التعلم القائم على المثيل.

كونت البيانات من مجموعه 365 سجل للطلاب التي تضم عدد من المتغيرات مثل الجنس والعمر والحالة الاجتماعية، و سمة الأداء " علامة مهمة يتم استخدامها كمدخلات ثنائية لتصنيف (النجاح / الفشل) ".

تم استخدام تقنيات لتحديد أكثر المتغيرات المؤثرة في أداء الطالب تمكن الباحثون من الحصول على أفضل خوارزمية هي سذاجة بايز بمستوى دقة تتبؤية عالية (74٪)[37] .

قام الباحثون بتطوير نظام يمكنهم من التنبؤ بأداء الطلاب من السنوات السابقة باستخدام مفاهيم تقنيات استخراج البيانات و التصنيف ، وتحليل المعلومات التي تحتوي على بيانات الطلاب، مثل النوع، علامات، اختبارات المجلس، علامات امتحان القبول، النتائج في السنة الأولى من الدفعة السابقة عن الطلاب.

من خلال تطبيق خوارزمية ID3 و C4.5-، لقد تمكن الباحثون من توقع الأداء العام والفردي للطلاب المقبولين حديثً [51].

من خلال الدراسات المذكورة أعلاه ، تحصلنا إلى أن أداء الطالب يمكن يعتمد على عوامل مختلفة مثل العوامل الأكاديمية والاجتماعية و الاقتصادية وغيرها من العوامل البيئية.

وان الاختلاف في الدقة التنبؤية يمكن أن يرتبط مع طبيعة مجموعة بيانات الطالب وحجمها.

كما ركزت معظم الدراسات السابقة على استخدام التصنيف للتنبؤ على أساس بيانات التسجيل والدرجات في المواد، على نفس المنوال سوف يكون خط التركيز في هذا البحث مع اضافة متغيرات جديدة لم تستخدم في البحوث السابقة مثل عدد مرات فشل الطالب ،مجموعة الدرجات الكلية والنصفية للطالب ، سنقوم في هذا البحث بتصنيف البيانات ببناء شجرة القرار للتنبؤ بدرجات الطلاب النهائية على أساس درجاتهم السابقة

الفصل الثالث:

عرض وتحليل البيانات

3 عرض وتحليل البيانات3.1 المواد

تم استخدام ثلاثة مجموعات مختلفة من البيانات لهذا البحث المجموعة الأولى جمعت من مدرسة الانقاذ والعينة الثانية من مدرسة الميثاق القرانية الاساسية الخاصة الثالثة من مدرسة العفاف الاساسية بنات كل هذه المدارس تقع في وحدة الازهري الادارية محلية جبل اولياء ولاية الخرطوم تحتوي على معلومات حول 1500 طالبًا من مختلف الفصول ، تحتوي 23 متغيرًا تم اختيارهم من قبل الباحث كما في الجدول 1.3.1.

كونت هذه البيانات في شكل جدول بيانات Microsoft Excel بواسطة الباحث ،تحتوي على تفاصيل كل طالب مثل الاسم الكامل والعمر والجنس والمستوى ودرجات المواد التي تم الحصول عليها فيها امتحانات مادة الرياضيات للفصول الاول اساس الى والثامن اساس ، والدرجات التي تم الحصول عليها في مادة الحاسوب والجغرافيا ، درجات مادة الرياضيات التى تم الحصول عليها في امتحانات السابقة ، ، إلخ سهولة إجراء .

| Туре | Description | Column | الرقم |
|------|--|--------|-------|
| | | | |
| ٠٠ | . 1 (* :11 11 1 | Λαο | 1 |
| رفمي | يرمز لعمر الطالب الذي يتراوح ما بين | Age | 1 |
| | 5الي16سنة في مرحلة الاساس | | |
| حرفي | النوع :عبارة عن متغير ثنائي يحمل قيمتين | Sex | 2 |
| | (ذکر ، انثي) | | |
| حرفي | مكان المدرسة: يقصد به مكان الذى يوجد فيه | Place | 3 |
| | المدرسة | | |

| | | T | 1 |
|------------|---|-------------|----|
| حرفي | الدولة: الجنسية الاصلية لطالب | Nationality | 4 |
| ، حرفي | مكان الولادة: يقصد به مكان ولادة الطالب | place-pr | 5 |
| 1) رقمي | المستوى ترتيب حسب الفصول (2,38 | Level | 6 |
| بات رقمي | عدد مرات فشل الطالب في مادة الرياضي | num- | 7 |
| | (1,2,3) | Failures | |
| مادة رقمي | درجة الطالب في الامتحان النصقي لـ | m-exam | 8 |
| | الرياضيات يتراوح ما بين (1 الى30) | | |
| مادة رقمي | درجة الطالب في امتحان السنة السابقة في | exam-last | 9 |
| | الرياضيات يتراوح ما بين (1 الى30) | year | |
| مواد رقمي | مجموعة درجة الامتحانات النصفيةل | totalm- | 10 |
| | السابقة يتراوح ما بين (1 الى280) | exam | |
| مابقة رقمي | مجموعة درجة الامتحانات الكلية لمواد ال | per-total | 11 |
| | يتراوح ما بين (1 الى280) | exam | |
| بين رقمي | درجة الطالب في مادة الجغرافيا يتراوح ما | gh-exam | 12 |
| | (1 الى30) | | |
| بين رقمي | درجة الطالب في مادة الحاسوب يتراوح مـ | cs-exam | 13 |
| | (1 الى30) | | |
| ين رقمي | تقيم الطالب في الواجبات يتراوح ما | Accentment | 14 |
| | (100كى) | | |
| 0 الى رقمي | عدد مرا ت الغياب يتراوح ما بين (| Absences | 15 |
| | (100 | | |
| ثنائي حرفي | الحالة الصحية لطالب عبارة عن متغير | Health | 16 |
| | يحمل قيمتين (جيد ، غير جيد) | | |
| تغير حرفي | مشاركة الطالب في الفصل عبارة عن م | Shar | 17 |
| | ثنائی یحمل قیمتین(yes, no) | | |

| d-pe 18 |
|-----------|
| |
| edu 19 |
| |
| -edu 20 |
| |
| size-f 21 |
| f-jop 22 |
| |
| f-jop 23 |
| |
| Reslut 24 |
| |

جدول 3-1 يوضح معنى الرموز الموجودة في بيانات الطالب

3.2 التحليل

كانت الخطوة الأولى هي تطبيق أساليب التعلم الآلي على البيانات الأولية.

3.2.1 مراحل التحليل

قسم الباحث مرحلة التحليل الى ثالثة مراحل:

3.2.1.1 المرحلة الاولى تقسيم البيانات

كانت الخطوة التالية هي تقسيم البيانات إلى مجموعات التدريب والاختبار. تم استخدام مجموعة التدريب (500 عينة من البيانات) لبناء التنبؤتم استخدام الطراز ومجموعة الاختبار (1000 من البيانات) لاختبار النموذج.

3.2.1.2 المرحلة الثانية معالجة البيانات

نجد عن سجل الطالب تتوفر فية الكثير من المعلومات والبيانات الاكاديمية المتعلقة بالطالب ، في هذا البحث تم تحديد بعض البيانات للتعامل معها وقد تمت معالجتها وفقا الخطوات التالية:

3.2.1.2.1 تهيئة البيانات

معد جمع البيانات هناك سمات غير مفيدة وهى معدد مع البيانات هناك سمات غير مفيدة وهى ،Place، f-jop، size-f، m-edu،f-edu Absences, Health, found-pe، Shart،Nationality تمت إزالتها.على سبيل المثال ، فإن حجم الاسرة ليس لدية اي صلة في التنبؤ بالأداء المستقبلي للطالب.

وكما تم اكمال البيانات المفقودة في بعض الحقول بقيم مناسبة وتعديل بعض القيم النصية الى حرفية مثل النوع ذكر وانثى الى (f،m).

3.2.1.2.2 اختيار السمات

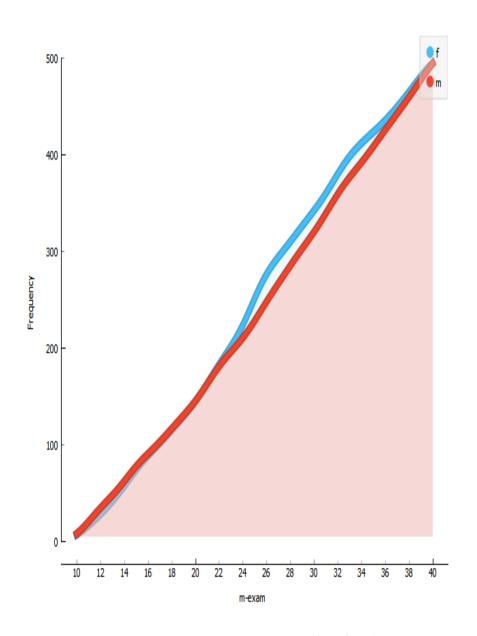
قام الباحث باختيار السمات ذات الفائدة والتاثير في المتغير التابع وهي

Age, sex, , level,m-exam, persoivse-exam, total-m-exam, per-total exam, gh-exam, cs-exam, Result.

حدد الباحث اهم الصفات الموثرة على نجاح الطلاب وذلك باجراء العمليات على كل صفة محددة لاكتشاف اكثر المتغيرات تاثرا المتغير التابع وهى الامتحان الفصلي لمادة الرياضيات ومجموعة الدرجات للامتحانات الكلية والنصفية ودرجة الطالب في مادتي الحاسوب والجغرافيا.

المخططات التالية تبين العلاقة بين المتغيرات وتفسير اسباب تكرار القيم بشكل رسومي ، تم تمثيل القيم عن طريق عينة من بيانات التدريب.

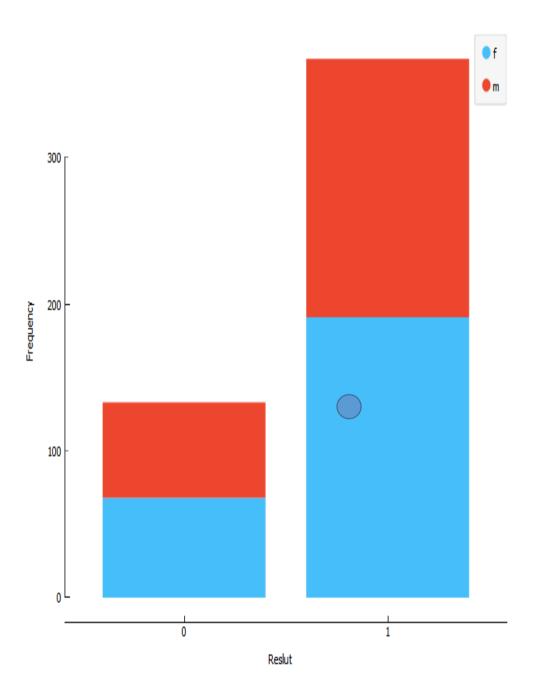
اظهرت الدراسة عن هنالك تتسبأ طردي للجنسين في امتحان الرياضيات النصفي كما هو مبيت في شكل 3-1.



شكل 3-2 يعرض نتائج طلاب لمادة الرياضيات

المفتاح = : ذكر المفتاح

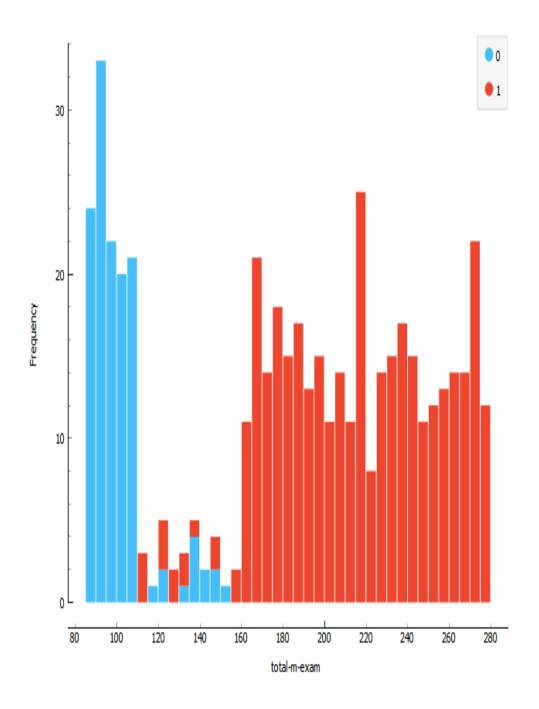
توصل الباحث على عن الاناث اكثر تفوقا من الطلاب في مادة الرياضيات كما تساويا في عدد مرات الرسوب كما هو موضح في الشكل2-3.



شكل 3-3يعرض نتائج الطلاب

المفتاح = : ذكر ، : انثى

كما توصل الباحث على عن الطلاب الناجحين اكثر عددا من الراسبين هو موضح في الشكل3-3.



شكل 3-4 يعرض درجات الطلاب للامتحان نصف الفصلي

المفتاح 🛑 :نجاح 🔃: رسوب

3.2.1.3 المرحلة الثالثة التطبيق الخوارزمية

اعتمد الباحث على استخدام شجرة القرارات في التنبؤ باداء الطلاب ومقارنتها ببعض طرق التعلم الالى للخروج بافضل النتائج، تم إنشاء نماذج التنبؤ باستخدام اداة التعلم الالى orange وذلك لانها اداة ساهلة الاستخدام كما توفر واجهة رسومية .

باستخدام طريقة اشجار القرار ،يوضح الجدول 3-2 نتائج االطلاب في مادة الرياضيات.

| | Age | sex | level | m-exam | Perv_exam | Total-m-exam | Per-m- exam | gh-exam | cs_exam | Result |
|----|-----|-----|-------|--------|-----------|--------------|----------------|---------|---------|--------|
| 1 | 9 | F | 3 | 22 | 24 | 219 | 185 | 30 | 29 | 1 |
| 2 | 7 | М | 1 | 11 | 7 | 101 | 90 | 23 | 30 | 0 |
| 3 | 5 | F | 1 | 21 | 28 | 166 | 175 | 28 | 8 | 1 |
| 4 | 12 | M | 7 | 21 | 23 | 169 | 193 | 29 | 27 | 1 |
| 5 | 13 | М | 8 | 31 | 32 | 237 | 268 | 27 | 25 | 1 |
| 6 | 13 | F | 7 | 20 | 16 | 159 | 105 | 22 | 20 | 1 |
| 7 | 14 | M | 8 | 15 | 7 | 88 | 89 | 13 | 12 | 0 |
| 8 | 11 | M | 6 | 38 | 30 | 265 | 201 | 30 | 27 | 1 |
| 9 | 6 | M | 1 | 36 | 32 | 260 | 242 | 29 | 26 | 1 |
| 10 | 7 | F | 2 | 24 | 24 | 215 | 218 | 25 | 29 | 1 |

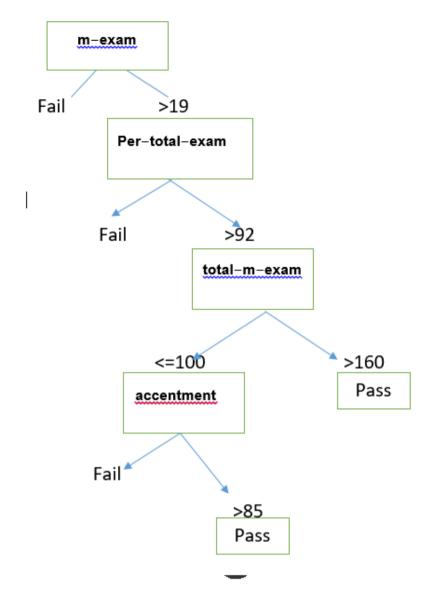
جدول 3-3بيانات التدريب

يتضمن الجدول بيانات التدريب في بشكل أساسي على تحديد بيانات الإدخال ،المتغير التابع ، والمتغيرات المستقلة .بعد إنشاء النموذج ، يتم تطبيقه إلى مجموعة بيانات الاختبار في الفصل الرابع.

| | Age | sex | level | m- exam | Perv_exam | Total-m- exam | Per-m- exam | gh-exam | cs_exam | Result |
|----|-----|-----|-------|------------|-----------|------------------|----------------|---------|---------|--------|
| 1 | 6 | F | 1 | 19 | 18 | 137 | 86 | 20 | 27 | 3 |
| 2 | 9 | E | 3 | 25 | 22 | 201 | 214 | 30 | 28 | ? |
| 3 | 7 | F | 1 | 27 | 25 | 204 | 174 | 26 | 23 | ? |
| 4 | 13 | F | 7 | 26 | 24 | 164 | 185 | 30 | 29 | ? |
| 5 | 6 | F | 1 | 18 | 17 | 98 | 98 | 23 | 20 | ? |
| 6 | 12 | М | 7 | 33 | 33 | 237 | 266 | 29 | 28 | ? |
| 7 | 14 | F | 8 | 33 | 39 | 262 | 234 | 29 | 29 | ? |
| 8 | 14 | F | 8 | 40 | 39 | 254 | 248 | 27 | 24 | ? |
| 9 | 9 | М | 4 | 11 | 7 | 95 | 99 | 18 | 23 | ? |
| 10 | 13 | M | 8 | 37 | 37 | 242 | 249 | 26 | 28 | ? |

جدول 3-4بيانات الاختبار

يمكن ان نتوقع فى المستقبل الحصول على نتائج تماثل بيانات التدريب بعد استخدام الخوارزمية الموضحة فى الشكل3-5 للمتغير التابع Result.



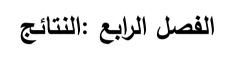
شكل 3-6 طريقة عمل الخورزامية

تم بناء شجرة القرار بتكرار التقسيم لتتيصف البيانات ، وفقا للاختيار افضل الميزات المتوقعة.

اختارت هذه الخوارزمية المتغير m-exam لتقسيم البيانات لعقدتين وفقا شروط فاذا كانت القيمة اصغر من 19يعتبر الطالب راسبا واذا كانت per-total- يتم تقسيم البيانات بواسطة متغير اخر وهو per-total-exam اذا كان exam الطالب والا سيتم per-total-exam اذا كان اصغر من 92 يرسب الطالب والا سيتم تقسيم البيانات مرة اخرى total-m-exam اذا كان اصغر من 160 يعتبر

الطالب راسب وسوف تقسم البيانات مرة اخرى عن طرق الصفة accentment فاذا كان accentmentكبر من 85 يعتبر الطالب ناجحا والافطالب راسبا.

بعد تطبيق الخوارزمية المبينة في الشكل تمكن الباحث من الحصول الى النتائج الموضحة في الجدول3-5في الفصل الرابع.



4 النتائج4.1 النتائج المتحصلة

تمكن الباحث من الحصول الى نتائج التوقع للمتغير التابع Result فى الجدول 1-4 العمود التعمود التعمود الموضحة فى الشكل 1-1الفصل الثالث.

| | Age | sex | level | m- exam | Perv_exam | Total-m- exam | Per-m- exam | gh-exam | cs_exam | Result | Tree Result |
|----|-----|-----|-------|------------|-----------|------------------|----------------|---------|---------|--------|----------------|
| 1 | 6 | F | 1 | 19 | 18 | 137 | 86 | 20 | 27 | ? | 0 |
| 2 | 9 | F | 3 | 25 | 22 | 201 | 214 | 30 | 28 | ? | 1 |
| 3 | 7 | F | 1 | 27 | 25 | 204 | 174 | 26 | 23 | ? | 1 |
| 4 | 13 | F | 7 | 26 | 24 | 164 | 185 | 30 | 29 | ? | 1 |
| 5 | 6 | F | 1 | 18 | 17 | 98 | 98 | 23 | 20 | ? | 0 |
| 6 | 12 | M | 7 | 33 | 33 | 237 | 266 | 29 | 28 | ? | 1 |
| 7 | 14 | F | 8 | 33 | 39 | 262 | 234 | 29 | 29 | ? | 1 |
| 8 | 14 | F | 8 | 40 | 39 | 254 | 248 | 27 | 24 | ? | 1 |
| 9 | 9 | М | 4 | 11 | 7 | 95 | 99 | 18 | 23 | ? | 0 |
| 10 | 13 | М | 8 | 37 | 37 | 242 | 249 | 26 | 28 | ? | 1 |

جدول 4-2يوضح نتائج بيانات الاختبار

للتاكد من صحة التوقع قام الباحث بحساب مصفوفة الارتباك التي تحتوي على بيانات حول القيم المتوقعة والقيم الفعلية .

4.2 مصفوفة الارتباك

ويبين الجدول 4-3 مصفوفة الارتباك لنموذج التنبؤ بخوارزمية شجرة القرار .

Confusion matrix for Tree

| | | Predicted | | |
|--------|--------|-----------|--------|----------|
| | | 0 | 1 | ~ |
| | | Т | F | Σ |
| Actual | 0 T | 313 TP | 1 FN | 314 |
| | 1 F | 2 FP | 684 TN | 686 |
| | \sum | 315 | 685 | 1000 |

جدول 4-4 confusion matrix

تمكن الباحث من الحصول على مستوى دقة من مصفوفة الارتباك بنسبة 99.7 %. عن طريق DT خوارزمية.

تقييم النتائج هي مقارنة طرق التعلم الآلي المختارة في شروط أدائها التنبؤ يوضح الجدول 4-5 نتائج التنبؤ ل5 طرق تعلم الآلة لمجموعة البيانات

| model | Accuracy | F-measure | Precision | Recall |
|------------------------|----------|-----------|-----------|--------|
| Tree | 0.997 | 0.997 | 0.997 | 0.997 |
| Svm | 0.969 | 0.969 | 0.970 | 0.969 |
| Neural network | 0.971 | 0.971 | 0.971 | 0.971 |
| Naïve bayes | 0.971 | 0.971 | 0.972 | 0.971 |
| Logistic Regression | 0.974 | 0.974 | 0.974 | 0.974 |

جدول 4-6 نتائج الخورازميات

التقيم

4.3

تمكن الباحث من الوصول على عن خوارزمية tree وهي الأفضل لتصنيف البيانات بنسبة أداء 99.7٪، ثم تليها خوارزمية Neural Network ، Naive Bayes، 97.4 بنسبة أداء Svm اقل أداء بنسبة 96.9٪.

4.4 المناقشة والاستنتاجات

يعتمد نجاح التعلم الآلي في التنبؤ بأداء الطلاب على استخدام البيانات وخوارزميات التعلم الآلي واختيار آلة التعلم الصحيحة.

لتحقيق أفضل النتائج فإن الخوارزمية وحدها لا يمكن أن توفر ذلك،فان ميزة الهندسة وعملية تعديل البيانات هما أيضا عوامل مهم في الحصول على أفضل نتائج التنبؤ.

تمكن الباحث من تحسين نتائج التنبؤ عن طريق اختيار الأسلوب والهندسة الميزة مع 5 طرق مختلفة التعلم الآلي Svm ،Neural Network ، Naive Bayes و مقارنة نتائجها باستخدام أربعة اساليب للتقييم.

اظهرت نتائج الدراسة أوجه التشابه واختلاف بسيط بين الطرق المختلفة ،

نجد عن هذه الدراسة حصلت على اعلى مستوى دقة مما كانت عليه الدراسات السابقة ، يعود الفرق في استخدام واختيار المتغيرات التابعة .

كما استخدم الباحث متغيرات إضافية تشير إلى درجات الامتحانات السابقة والفصلية والنصقية ودرجات الحاسوب والجغرافيا كانت أكثر ارتباطًا بنجاح الطالب ،هذا يدل على أهمية البيانات واختيار السمات .

نتائج هذه الدراسة تشير إلى أن ميزة الهندسة توفر المزيد من التحسن لنتائج التنبؤ من اختيار الطريقة.

و على الرغم من عن ميزة الهندسة قد تم في قدرة محدودة ، جعلت فرقا أكبر في أداء التتبؤ .علاوة على ذلك، تم تحقيق أكبر قفزة في التحسين في الأشجار القرار ، حيث يتم تطبيق الاختيار وتعديل الميزة على البيانات .

توصل الباحث عند محاولة تحسين التنبؤ بأداء الطالب ، يعد تعديل بيانات الإدخال عاملاً مهمًا إلى جانب اختيار الطريقة الصحيحة للبيانات.

على الرغم من أن ميزة الهندسة كانت أكثر فعالية من اختيار الطريقة ، فإن مزيج من كلا النهجين قدمت أفضل النتائج.

السابق ذكره يدل على أن استخدام التعلم الآلي هو وسيلة فعالة للتنبؤ بالطالب أداء.

الفصل الخامس:الخاتمة

5 الخاتمة:

تم استفاء جميع اهداف البحث في الفقرات التاليبة:

- \checkmark تصنیف الطلاب لمجموعتین (ناجحین وراسبین)، تمکن الباحث من تصنیف البیانات الی مجوعتین عن طریق المتغیر التابع result ورمز لجموعة الطلاب الناجحین ب1 والراسبین 0
- ✓ الكشف عن قواعد الارتباط لفهم السمات التي تؤدي لنجاح او رسوب الطالب ،قام باختيار افضل السمات التي توثر على المتغير التابع مما اسهم في خروج افضل النتائج
- ✓ الكشف عن الحالات الشاذة ،وذلك عن طريق مصفوفة الارتباك التي تحسب القيم الفعلية والقيم المتوقعة لنجاح والرسوب
- ✓ توقع درجات الطلاب في مرحلة مبكرة، من خلال بيانات التدريب تم توقع قيم المتعير التابع result في مجموعة بيانات الاختبار مما يجعل النموذج قادرا على التنبؤ لدرجات الفصول القادمة

من خلال انشاء نموذج لتوقع درجات الطلاب تمكن الباحث من الحصول على عدة نتائج وهي:

- ✓ وجود طريقة لمساعدة المدرسة في تقيم اداء الطالب من خلال درجاته السابقة
 - ✓ المساهمة في رفع درجات الطلاب في مادة الرياضيات
 - ✓ المساهمة في ايجاد افضل الطرق لتحسين مستوى الطالب
 - ✓ توفير بيانات محوسبة تساعد المدرسة في كثير من العمليات

5.1 التوصيات

يحتوي هذا البحث على بعض القيود التي يجب مراعاتها، لم يكن هنالك وصول مباشر وسهل للبيانات تم جمع البيانات من سجلات المدرسة الاستغناء عن المتغيرات الغير مؤثرة لجعل النموذج أكثر كفاءة.

الورقية كنتائج الشهرية والتقارير الموجودة ، من ثم قام الباحث بتحويل هذه البيانات الورقية الى قواعد بيانات مما كلف الباحث الكثير من الوقت بالإضافة إلى ذلك ، مجموعات البيانات كانت كبيرة ، تحتوي على أكثر من ألف سجل .

مجال آخر يمكن أن تحسنه الأبحاث المستقبلية هو تنوع طرق التعلم الآلي

الطرق الستخدمة في هذا البحث decision tree ، البحث sym و Logistic Regression ، Neural Network نوصى باستخدام طرق أخرى ، مثل التجميع واقرب جار مما يساعد لفهم الاشكالية قيد الدراسة بصورة أفضل.

المجال الأخير الذي يمكن تحسينه هو عملية إنشاءو تعديل الميزةمن البيانات و عن طريق إنشاء ميزة جديدة للمتغير كدالة لمتغيرات مختلفة .

مع الأخذ في الاعتبار أنه كلما زاد متغير مخصص ، كلما زاد صعوبة تفسيرالعلاقة بينه وبين المتغير التابع.

المراجع والمصادر

| [1] | تيكين ، أحمد،2014، التنبؤ المبكر بمعدلات درجة الطلاب عند التخرج نهج استخراج البيانات،المجلة الأوراسية للبحوث التربوية 54 ، |
|------|---|
| | .226-207 |
| [2] | غورميت كور، ويليامجيت سينغ،2016 ، التتبؤ بأداء الطلاب باستخدام أداة WEKA، مجلة دولية للعلوم الهندسية ، يناير 2016 ، |
| | المجلد. \ISSN: 2229-6317 |
| [3] | امجد ابو الساع،2016 ، تعدين البيانات التربوية وأداء الطلاب تتبؤ، (IJACSA) المجلة الدولية لعلوم الحاسب الآلي المتقدمة |
| | والتطبيقات ،المجلد. 7 ، رقم 5 ، 2016 |
| [4] | سميرة محمد،2018، تطبيق تقنيات التنقيب في البيانات لتقيم اداء طلاب، بحث لاستكمال متطلبات نيل درجة البكالوريوس في علوم |
| | الحاسب الالي، جامعة سبها كلية العلوم قسم الحاسوب |
| [5] | أ.د عبدالرحمن احمد ،مناهج البحث العلمي ،2006،جامعة السودان المفتوحة ،الطبعة الاولى 2006 |
| [6] | Parneet Kaura, Manpreet Singhb, Gurpreet Singh Josanc," Classification and prediction based data |
| | mining algorithms to predict slow learners in education sector", International Conference on |
| | Recent Trends in Computing 2015(ICRTC-2015) |
| [7] | Amirah Mohamed Shahiria,*,Wahidah Husaina, Nur'aini Abdul Rashida، "A Review on Predicting |
| | Student's Performance using Data Mining Techniques", The Third Information Systems International |
| | Conference, Procedia Computer Science 72 (2015) 414 – 42 |
| [8] | غورميت كور 1 ، ويليامجيت سينغ ،النتبو الستخدام أداةWEKA ، مجلة دولية للعلوم الهندسية، المجلد. 6913–2229 (الط |
| | /SSN: 2320-0332 |
| [9] | سيمبرينج ، م. زارليس ، ديدي هارتاما ، رامليانا إس ، إلفي واني ، " الأداء من خلال تطبيق تنبؤ الطالب الأكاديمي تقنيات استخراج |
| | البيانات "، المؤتمر الدولي حول الإدارة والذكاء الاصطناعي ، 2011 |
| [10] | إريك إيدي ، مارك هـ. شوالتر ، "تأثير جودة المدرسة على أداء الطلاب منهج الانحدار "رسائل الاقتصاد 58 (1998) ، 19 |
| [11] | هانوماثابا ، مارغريت ماري، "تعدين البيانات التربوية والتنبؤ بصعوبات التعلم"، مجلة آسيا والمحيط الهادئ للبحوث ، 2013. |
| [12] | راجيسواري ، سوشيتا بوركار ، شاول "التنبؤ بالأداء الأكاديمي للطلاب باستخدام التعدين البيانات "، المجلة الدولية لعلوم الكمبيوتر |
| | والحوسبة المنتقلة ، IJCSMC ، المجلد. 2 ، العدد. 7 بيوليو 2013. |
| [13] | كارتيك إن شاه ، سرينيفاسولو كوثورو ، وس. فيراموثو ، "تجميع الطلاب بناءً على الأداء الأكاديمي السابق "، المجلة الدولية للبحوث |
| | الهندسية والتطبيقات (IJERA) ، مايو 2013. |
| [14] | لوبو ، "تعدين البيانات في النظام التعليمي باستخدام WEKA" ، دولي مؤتمر حول اتجاهات النكنولوجيا الناشئة (2011 (ICETT |
| [15] | نصير أحمد ، "تعدين المعلومات في بيانات تقييم أداء الطلاب" ، مجلة العلوم الهندسية ، 2012. |
| [16] | براشانت ، "التنبؤ بطلاب الدراسات العليا لدرجة الماجستير بناءً على أدائها السابق باستخدام شجرة القرار في بيئة Weka "،المجلة |

| الدولية لتطبيقات الكمبيوتر (0975 - 8887) ، المجلد 74 - 11 ، يوليو 2013 |
|---|
| SaharAl-Sud," Predicting students final degree classification, Education and Information Technologies |
| 24:2357–2369 https://doi.org/10.1007/s10639-019-09873-8, 2019 |
| Nawal Ali Yassein1, Rasha Gaffer M Helali," Predicting Student Academic Performance in KSA using |
| Data Mining Techniques", Journal of Information Technology & Software Engineering KSA Yassein et |
| al., J Inform Tech Softw Eng 2017, 7:5 DOI: 10.4172/2165-7866.1000213 |
| Baker, Ryan SJD, "The state of educational data mining in A review and future visions", JEDM ,2009 |
| Agaoglu , "Predicting instructor performance using data mining techniques in higher education", 2016. |
| Osman begovic E, Mirza " Data mining approach for predicting student performance", J Econ Bus 10: |
| 3-12,2012 |
| Ramesh, Parkavi, Ramar"Predicting student performance: A statistical and data mining approach",IJC |
| 35–39,2013 |
| Agathe M, Yacef K "Educational data mining: A case study", AIED, pp: 467-474. |
| Minaei-Bidgoli "Predicting student performance: an application of data mining methods with an education |
| web-based system", Frontiers in education, 2003 |
| Zekic-Susac" Neuron networks and trees of decision-making for prediction of eficiency in studies",2009 |
| Baker, Ryan SJD, "The state of educational data mining in A review and future visions", JEDM ,2009 |
| Al-Radaideh, , "A classification model for predicting the suitable study track for school students " |
| International Journal of Research and Reviews in Applied Sciences, vol. 8, 2001 |
| Baradwaj and S. Pal, "Mining educational data to analyze students' performance " International |
| Journal of Advanced Computer Science and Applications, vol. 2, 2011 . |
| El-Halees, "Mining student data to analyze learning behavior: A case study," presented at the |
| International Arab Conference of Information Technology (ACIT2008), Tunisia, 2008. |
| A. Nandeshwar and S. Chaudhari," Enrollment Prediction Models Using Data Mining, Available |
| http://nandeshwar.info/wpcontent/uploads/2008/11/DMWVU_Project.pdf ,2008 |
| Kabakchieva, "Predicting student performance by using data mining methods for classification," |
| Cybernetics and Information Technologies, vol. 13, 2013. |
| Garc´ıa-Saiz and M. Zorrilla, "Comparing classification methods for predicting distance students' |
| performance," The Journal of Machine Learning Research, 2011. |
| Romero,"Data mining algorithms to classify students," ,presented at the the 1st International |
| Conference on Educational Data Mining, 2008. |
| Kovačić, "Early prediction of student success: Mining students enrolment data," presented at the |
| |

| | Informing Science & IT Education Conference, 2010. |
|------|--|
| [35] | Ramesh, "Performance Analysis of Data Mining Techniques for Placement Chance Prediction", Interna |
| | Journal of Scientific & Engineering Research Volume2,Issue8,August-2011 ISSN 2229-5518 |
| [36] | P. Cortez, and A. Silva, "Using Data Mining To Predict Secondary |
| | School Student Performancel", In EUROSIS, A. Brito and J. Teixeira |
| | ·2008pp.5-12.,2008 |
| [37] | Kotsiantis, "Prediction of Student's Performance in Distance Learning Using Machine Learning |
| | Techniques "«Applied Artificial Intelligence, Vol. 18, No. 5, 2004, pp. 411-426. 22: 314-327,2004 |
| [38] | Kalpesh Adhatrao، "PREDICTING STUDENTS' PERFORMANCE USIN |
| | ID3 AND C4.5 CLASSIFICATION ALGORITHMS", International Journal of Data Mining & Knowledge |
| | Management Process (IJDKP) Vol.3, No.5, September 2013 |
| [39] | Aysha Ashraf " A Comparative Study of Predicting Student's Performance by use of Data Mining |
| | Techniques", American Scientific Research Journal for Engineering, Technology, and Sciences |
| | (ASRJETS) ISSN (Print) 2313-4410, ISSN (Online) 2313-4402,2018 |
| [40] | Mitchell, "Machine Learning, McGraw-Hill, 1997 |
| [41] | Nyce , CPCU "Predictive analytics white paper", American Institute for CPCU. Insurance Institute of |
| | America, 9–10,2007 |
| [42] | Sas," Predictive Analytics: What it is and why it matters", SAS. |
| | https://www.sas.com/en_us/insights/analytics/predictive-analytics.html. Retrieved April24, 2017 |
| [43] | Wayne, Eckerson "Predictive analytics. Extending the Value of Your DataWarehousing Investment, TDW |
| | Practices Report 1, 1-36,2007. |
| [44] | Nitish Srivastava, Geoffrey Hinton," Dropout: A simple way to prevent neural networks from |
| | overfitting", The Journal of Machine Learning Research 15(1), 1929-1958,2014 |
| [45] | J. Ross Quinlan," Simplifying decision trees", International Journal of Man-Machine Studies 27(3), |
| | 234,1987 |
| [46] | Carolin Strobl, "An introduction to recursive partitioning: rationale, application, and characteristic |
| | classification and regression trees, bagging, and random forests", Psychological Methods 14(4), |
| | 348,2009 |
| [47] | Pedro, Domingos, "A few useful things to know about machine learning", Communications of the ACM |
| | 55(10), 78-87,2012 |

| [48] | Roman | Kern | "Feature | Engineering, | Knowledge | Discovery | and | Data |
|------|--------------|--------------|-----------------|----------------------|------------------|------------------|------------|--------|
| | Mining",http | ://kti.tugra | z.at/staff/deni | s/courses/kddm1/f | featureengineeri | ng.pdf. Retrieve | d May 3,2 | 017. |
| [49] | Tom Fawo | ett," An int | roduction to R | OC analysis", Patt | tern Recognition | Letters 27(8), 8 | 61-874,20 | 005 |
| [50] | J. Ross (| Quinlan," Si | mplifying dec | ision trees", Intern | national Journal | of Man-Machin | ne Studies | 27(3), |
| | 221-234,1 | 987 | | | | | | |
| | | | | | | | | |
| | | | | | | | | |