ينتمي هذا المشروع البحثي إلى المجال المعرفي المختص بتنقيب البيانات وعلى وجه التحديد تنقيب البيانات في مجال التعليم . بدأنا بجمع البيانات من موقع KDD CUP ويقوم هذا الموقع بمسابقة سنوية لاستخراج البيانات واكتشاف المعرفة التي تنظمها ACM المجموعة ذات الاهتمام الخاص لاكتشاف المعرفة وتنقيب البيانات ، المنظمة المهنية الرائدة لمنقبين للبيانات ،حيث كانت البيانات عبارة عن بيانات طلبة من موقع edx و Coursera ، وتم تجميع بيانات من ٣٩ دورة وكانت الدورات من مختلف المجالات وكان عدد المستخدمين ١١٢٤٨٨ مستخدم وكان عدد الأنشطة التي تم تسجيلها مثل مشاهدة الفيديو وحل الواجبات المقدم من الدورة .. الخ حوالي ٨ مليون نشاط تم استخراج الميزة يدويا بإستخدام مكتبة Pandas وتم استخدام مكبتة Scikit learn وهي احدى مكتبات لغة برمجة بايثون لتطبيق لوغاريتميات تعلم الآلة ، ثم تم تطبيق ٦ لوغاريتميات من لوغاريتميات المستخدمة في التصنيف من مجال تعلم الآلة المراقب وهم : ١- الإنحدار اللوجستي ٢- Gradient Boosting٣- Support Vector Machine ٤- Naïve Bayes ٥- Random forrest٦- Decision Tree وتم أيضا استخدام نوعين من الشبكة العصبية وهم Multilayer Perceptron (MLP) وشبكة عصبية تم إنشاؤها يدويا وتعتبر من احدى الشبكات التي تسمى بتعلم العميق ، والهدف من هذا المشروع هو تنبؤ إنسحاب الطلبة من الدورات المفتوحة على شبكة الإنترنت، وفي نهاية المشروع البحثي سنعرض ونناقش النتائج التي تم الحصول عليها باستخدام أدوات فعالة لتصور البيانات وتحليلها و مقارنتها مع قريناتها من الطرق السابق استخدامها في المحتوى العلمي على مستوى دقة كل لوغاريتم تم استخدامها في هذا المشروع ، وأخيراً، من الناحية العلمية التجريبية، من عدة مقارنات علمية ونظرا لدقة التصنيف وعوامل أخرى نرى مدى فاعلية نماذج التي تمت تجربتها في هذا المشروع وإثبات أن النهج المقترح لدينا وهو نموذج من الشبكة العصبية بدقة تصل ل ٩٥% ذا دقة تصنيفية متفوقة مقارنتا مع غيره من الخوارزميات والنُهُجٌ المستخدمة تقليدياً في البحوث العلمية ذات الصلة.

Massive Open Online Courses (MOOCs) have produced creativity, enthusiasm, involvement and various discussions from different contributors across the educational field. The enrollment in the MOOCs technology does not demand specific criteria, such as acquiring high intelligence within this area. Rather, this form of technology facilitates the enrollment in enormous online classes with the absence of any actual physical attendance. However, and despite such facilitation, the MOOCs does not get high levels of completion rates. Whilst considerable number of people apply for an online course, just few of them who accomplish the entire course. This dissertation intends to analyze MOOCs from a distinctive dataset that was provided from KDDCUP Competition 2015. The employment of machine learning mechanisms considers a key aspect for such analysis in order to categories contributors as either dropouts or completers. The engagement with different modalities of mechanisms, including Logistic Regression, Support Vector Machines (SVM), Gradient Boosting, Naïve Bayes, Random Forrest and Decision Trees are examples of such techniques. In addition, the utilization of deep learning models, such as the Deep Learning sample constructed from scratch and the Multi-Layer Perceptron can assist the analysis. Moreover, the dissertation aims to confirm to what extent various domains and systems could be analyze using a general process. Finally, an evaluation of the characteristics has been used in the analysis will be demonstrated