



جامعة الموصل  
كلية علوم الحاسوب والرياضيات

استخدام خوارزمية تعزيز التدرج لتعلم الآلة في توقع  
النقر على الاعلانات ضمن صفحات موقع التواصل  
الاجتماعي (فيس بوك)

شهد شرف الدين يحيى

رسالة الدبلوم العالي

قسم علوم الحاسوب

بإشراف

المدرس

د. بان شريف مصطفى

## ملخص البحث

التعلم الآلي هو مجال الدراسة التي تسخر مبادئ علوم الحاسوب والإحصاءات لإنشاء نماذج إحصائية. وعادة ما تستخدم هذه النماذج للقيام بشيئين: التنبؤ أي القيام بتوقعات حول المستقبل استنادا إلى بيانات عن الماضي والاستدلال أي اكتشاف أنماط في البيانات. وهناك العديد من خوارزميات التعلم الآلي التي تصنف على أنها خاضعة للإشراف وغير خاضعة للإشراف.

تعد صناعة الإعلانات عبر الانترنت من الصناعات الواعدة في الوقت الحاضر إذ تبلغ قيمتها مليارات الدولارات. إذ إن المُعلن يقوم بالدفع مقابل ردود المستخدمين القابلة للقياس، لذا فإن أنظمة التنبؤ بالنقر على إعلانات الفيس بوك هي من الأنظمة الأساسية لشركات الإعلانات وهي مهمة تعلم آلي صعبة. إذ تعتمد في بنائها على خوارزميات التعلم الآلي.

في هذا البحث تم تدريب إحدى خوارزميات التي تعمل كمصنف ثنائي على مجموعة البيانات والتي أخذت من موقع "Kaggle" للمسابقات العالمية. في هذه المجموعة من البيانات، يعرض كل مثال (سطر نصي) إعلانا معروضا بالنقر على التصنيف، إما نعم (+) أو لا (-). والهدف هو توقع احتمال النقر على إعلان جديد. ويطلق على مشكلة تقدير نسبة النقر إلى الظهور ب (Click Through Rate).

بُني النموذج بالاعتماد على خوارزمية (Boosting Gradient Decision Tree). وقد تم دراسة الميزات المؤثرة في بناء النموذج ووضعهم في مجاميع لاختيار النموذج الاكفأ بالاعتماد على احد مقاييس الكفاءة المعتمدة (منحنى ROC).

**University Of Mosul**  
**College Of Computer Science And Mathematics**



**Use the Gradient Boosting Algorithm to  
learn the machine in anticipation of  
clicking on ads within the social  
networking site pages  
(Facebook)**

**Shahad Sharaf Aldeen Yehya**

**Higher Diploma Thesis**

**Department of Computer Science**

**Supervised By**

**Lecturer**

**Dr.Ban Sharief Mustafa**

## **Abstract**

Machine learning is the field of study which uses the principles of computer sciences and statistics to build statistic models. Usually, these models are used to do two things: Predict any prediction about the future based on past and Induce any data pattren. There are many automated learning algorithms that are classified as supervised and unsupervised.

Internet advertising is one of the most promising industries today ,it worths billions of dollars. As the advertiser companies are paying based on measurable user responses , these companies are mainly depended on the prediction systems of clicking on Facebook ads which is a difficult learning task and needs a machine learning algorithm to be introduced.

In this work, we training binary classification on the dataset Criteo Kaggle CTR competition. In this data set, each example displays an advertise by clicking on (1+) or not (1-). The target is to predict whether the new ad will be clicked. This is the standard click through rate estimation problem.

The model is built depending on Boosting Gradient Decision Tree algorithm. Features affecting model building are analyzed and putted in groups for the selection of the more efficient model which examined using one of the accredited efficiency measures (ROC curve).