التنبؤ بالعلاوة السعرية باستخدام خوارزمية أشجار التصنيف والانحدار في مزادات eBay

مفرح حمير محمد

بإشراف د :

محمود كامل

ملخص

اجتذب استخدام تقنيات استخراج البيانات في مجال المزادات اهتمامًا كبيرًا من مجتمع البحث. في المزادات ، يحاول المستخدمون تحقيق أعلى ربح وتجنب الخسارة قدر الإمكان. لذلك ، يمكن تنفيذ تقنيات استخراج البيانات في مجال المزاد لتطوير طريقة ذكية يمكن استخدامها من قبل المستخدمين في المزادات عبر الإنترنت. ومع ذلك ، فإن تحديد العوامل التي تؤثر على نتيجة المزاد ، وخاصة السعر الأولي ، أمر بالغ الأهمية. بالإضافة إلى ذلك ، يجب إنشاء نظام ذكي يعتمد على البيانات النظيفة لضمان دقة النتائج. في هذا العمل ، نقترح نظام ذكي (مصنف) للتنبؤ بسعر المزاد الأولي. يستخدم النظام المقترح طريقة التنعيم المزدوج لتنظيف البيانات من حيث المعالجة المسبقة. يتم تنفيذ هذا النظام على مجموعة بيانات تم جمعها من موقع أي باي وتنظيفها باستخدام منهجية التنعيم المزدوج المقترحة. في مرحلة التدريب ، يتم استخدام تقنية الكارت لبناء المصنف. بالمقارنة مع التقنيات المماثلة ، يُظهر النظام المقترح أداءً أفضل من حيث الدقة والمتانة مقابل البيانات المزعجة ، كما هو محدد باستخدام منحنيات روك.

Price premiums prediction using classification and regression trees (CART) algorithm in eBay auctions

Mofareah Humeer Mohammed

Supervised By

Dr. Mahmoud Kamel

Abstract

The use of data mining techniques in the field of auctions has attracted considerable interest from the research community. In auctions, the users try to achieve the highest gain and avoid loss as much as possible. Therefore, data mining techniques can be implemented in the auctioning domain to develop an intelligent method that can be used by the users in online auctions. However, determining the factors that affect the result of an auction, especially the initial price, is critical. In addition, the intelligent system must be established based on clean data to ensure the accuracy of the results. In this work, we propose an intelligent system (classifier) to predict the initial price of auctions. The proposed system uses the double smoothing method (DSM) for data cleaning in terms of preprocessing. This system is implemented on a data set collected from the eBay website and cleaned using the proposed DSM. In the training phase, the CART technique is employed for the classifier construction. Compared to similar techniques, the proposed system exhibits a better performance in terms of the accuracy and robustness against noisy data, as determined using ROC curves.