



وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

التحقق من عدد الأشجار الأمثل في الغابة العشوائية

رسالة مقدمة

إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة دبلوم عالي في
علوم الحاسوب

من قبل

نور رافع عبد الباقي عبد الرحمن

بإشراف

د. محمد جاجان يونس

استاذ مساعد

الخلاصة (Abstract)

تُعد الغابة العشوائية تقنية فعالة من الناحية الحسابية يمكنها العمل بسرعة عبر مجموعات البيانات الكبيرة. حيث تم استخدامها في العديد من المشاريع البحثية الحديثة والتطبيقات الواقعية في مجالات متنوعة. ومع ذلك، فإن الدراسات ذات الصلة لا تقدم تقريباً أي توجيهات عن عدد الأشجار التي يجب استخدامها لتكوين هذا النوع من خوارزميات تعلم الآلة.

تحلل هذه الدراسة ما إذا كان هناك عدد مثالي من الأشجار داخل الغابة العشوائية، أي الحد الذي لا يؤدي منه زيادة عدد الأشجار إلى تحقيق مكاسب في الأداء (من حيث دقة النتائج والوقت المستغرق) وسيؤدي فقط إلى زيادة التكلفة الحسابية.

تم في هذه الدراسة بعد تنفيذ الخوارزمية المقترحة استنتاج وتوضيح أن مع زيادة عدد الأشجار في الغابة العشوائية قد لا يعني دائماً أن أداء الغابة سيكون أفضل من الغابات السابقة (عدد أقل من الأشجار). كما تم استنتاج أن عدد أشجار الغابة العشوائية ٢٠ ممكن أن يعتبر حجم أشجار مثالي من حيث الدقة والسرعة بالنسبة لقاعدة البيانات التي تم اعتمادها في هذه الدراسة.

**Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Computer Science**



Investigating Optimal Trees Number in Random Forest

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Higher Diploma
in
Computer Science**

**By
Noor Rafea Abed Al Baki Abed Al Rahman**

**Supervised by
Dr. Mohammed Chachan Younis
Assistant Professor**

2022 A.D.

1443 A.H.

Abstract

A random forest algorithm is a computationally efficient approach that can work quickly across big data sets. This method has been used in numerous state-of-the-art studies as well as real-time applications in a variety of fields. However, relevant research projects provide almost no guidance on the number of trees that should be utilized to compose this type of machine learning technique.

This study is mainly concerned with detecting the optimal number of trees to be used within the related random forest, i.e. the extent to which the growth of the number of trees does not affect the performance improvement (in terms of accuracy of results and time taken). This will only lead to an increase in arithmetic costs.

In addition to that, in this study, after performing the recommendation algorithm, it could be concluded that the increase in the number of trees in the random forest does not always mean that the increase in the performance of the forest would be better if it is compared to the previous forests (fewer number of trees). It was also concluded that the number of random forest trees 20 could be considered an ideal tree size in terms of accuracy and speed regarding the database that was adopted in this study.