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

# تحسين تقدير كلفة البرمجيات بأستخدام التعلم التجميعي

رسالة مقدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير علوم في
البرمجيات

من قبل

نداء ثامر قاسم سلطان

بإشراف

أ. د. إبراهيم أحمد صالح أحمد

١٤٤٦ هـ ٢٠٢٥

## المستخلص

يعدّ تقدير تكلفة تطوير البرمجيات أحد التحديات الرئيسة التي تواجه فرق التصميم والتطوير. مع ازدياد تعقيد الأنظمة البرمجية وتعدد التقنيات المستخدمة أصبح التخمين الدقيق للكلفة يواجه تحديات إدارية مختلفة تتطلب اتخاذ قرارات صحيحة من مديري المشاريع لضمان نجاح المشروع والتخطيط الفعّال للموارد.

الهدف من هذه الرسالة هو تحسين دقة تقديرات التكلفة، ولتحقيق هذه الغاية، تم استخدام خوارزميات التعلم التجميعي Ensemble Learning مع إطار العمل التقليدي لنموذج COCOMO، إذ تم تطبيق نموذج COCOMO كمرجع لتقدير التكلفة النهائية للمشروع، هو نموذج رياضي يعتمد على معادلة حسابية لتقدير التكلفة بواسطة ضرب الجهد المبذول بتكلفة الساعة البرمجية Cost Per Hour لتحويل الجهد إلى تقدير مالي ،تكون قيمة الساعة البرمجية غير ثابتة يحددها المستخدم وتعتمد على المتغيرات الاقتصادية والرواتب. وتم استخدام ثلاث مجموعات من بيانات المشاريع الحقيقية المختلفة من حيث الحجم وعدد الخصائص وتم جمعها من قواعد بيانات معروفة هي بيانات الصين ، وكيتشنهام و ISBSG لتقييم دقة النموذج المقترح ،وتمت مقارنة النتائج للمجاميع المختلفة كذلك مقارنتها مع نتائج مشاريع أخرى.

تم استخدام ثلاثة نماذج تجميعية رئيسة لتخمين الجهد المبذول لإنجاز المشروع البرمجي وهي: التكديس (Stacking)، التعزيز (Boosting)، والتعبئة (Bagging) ويقاس(شخص _ساعة) او (شخص _شهر) كمرحلة اولى من بناء أداة تخمين الكلفة، المرحلة الثانية تم فيها استخدام نموذج الكلفة البنائية Cost Constructive Model(COCOMO) لتخمين الكلفة باستخدام الجهد الناتج من نماذج التعلم التجميعي. تم تركيب نماذج التعلم التجميعي الثلاث باستخدام الخوارزميات التالية :الغابة العشوائية Random Forest، الانحدار الخطيLinear Regression ،الانحدار اللوجستي Logistic Regression، التعزيز التكيفي Adaptive Boosting، تعزيز التدريج Gradient Boosting، وتعزيز التدريج الفائق Extreme Gradient Boosting كمتعلمين أساسيين لبناء نماذج التجميع للحصول على نماذج أكثر مرونة ودقة للتنبؤ بكلفة المشاريع البرمجية.

تم تدريب النماذج باستخدام بيانات التدريب بنسبة 80% من مجموع البيانات، واختبارها على بيانات الاختبار بنسبة 20% من البيانات. تم قياس أداء النماذج باستخدام مقاييس التقييم التالية: MMRE، R-squared، PRED(25)، MDMRE .

حققت نتائج تنفيذ النماذج على المجموعات الثلاث من البيانات مستويات متفاوتة من الدقة في التنبؤ، إذ تفوقت النماذج الثلاثة :التكديس ،التعزيز، التعبئة في دقة التنبؤ بنسبة تتراوح بين (97% –98%) عند تطبيقها على مجموعة بيانات ISBSG ، أما عند تطبيق النماذج على مجموعة بيانات الصين كانت دقة التنبؤ جيدة ،أعطى نموذج التكديس دقة تنبؤ 98% أما نموذج التعبئة فأعطى دقة تنبؤ 94% ، ونموذج التعزيز أعطى دقة تنبؤ 86% .في حين كانت نتائج تطبيق النماذج على بيانات كيتشنهام هي الأقل دقة بين النتائج ، فأعطت دقة تنبؤ تتراوح بين (22% – 36%) .إذ تفوقت نماذج التعلم التجميعي الثلاثة التكديس والتعبئة والتعزيز من حيث دقة التنبؤ عند تطبيقها على مجموعة بيانات ISBSG ،لذا تم اعتمادها للتنبؤ بالجهد في المرحلة الأولى من بناء الأداة المقترحة. في المرحلة الأتية بعد تقدير جهد العمل بواسطة نماذج التجميع ، تم استخدام نموذج COCOMO لتخمين الكلفة النهائية للمشروع .

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

# Enhancing Software Cost Estimation Using Ensemble Learning

**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 Master of Science
in
Software**

**By**

## Nedaa Thamer Qasim Sultan

**Supervised by**

## Prof. Dr. Ibrahim Ahmed Saleh Ahmed

---

**2025 A.D.** **1446 A.H**

# Abstract

Cost estimating for software development is one of main challenges facing design and evolution teams. With increasing complexity of software systems and multiplicity of technologies used, accurate cost estimation faces various administrative challenges that require making correct decisions by project managers to ensure project success and effective resource planning.

The goal of this thesis is to improve accuracy of cost estimates, to achieve this goal, Ensemble Learning algorithms were used with traditional framework of the COCOMO model. COCOMO model was applied as a reference to estimate final cost of the project. It is mathematical model where the mathematical equation used to estimate the cost by multiplying effort by cost per hour , to convert effort into a financial estimate. The value of programming hour is not fixed and determined by user and depends on economic variables and salaries. Three sets of real projects data were used of different size and number of features; they were collected from well-known databases namely China, Kitchenham and ISBSG data, to evaluate accuracy of proposed model. The results were compared for the different groups and were also compared with the results of other projects.

Three main Ensemble models are used to estimate effort involved in completing software project namely Stacking, Boosting, and Bagging which is measured in (person_hour) or (person_month) as first stage of building cost estimation tool, second stage in which cost constructive model (COCOMO) used to estimate cost using effort resulting from Ensemble learning models. The three ensemble learning models were constructed using the following algorithms: Random Forest, Linear Regression, Logistic Regression, Adaptive Boosting, Gradient Boosting and Extreme Gradient Boosting as basic learners to build

Ensemble models to obtain more flexible and accurate models to predict cost of software projects .

The models were trained using training data for 80% of the total data, and tested on testing data for 20% of data. Model performance was measured using the following Evaluation Metrics: MMRE, R-squared, PRED (25), MDMRE.

The results of implementing the models on the three data sets achieved varying levels of accuracy in prediction, where three models: Stacking, Boosting, and Bagging outperformed accuracy of the prediction by (97%-98%) when applied to the ISBSG data set, but when applying the models to China dataset, the prediction accuracy was good, stacking model gave a prediction accuracy of 98%, while bagging model gave a prediction accuracy of 94%, The boosting model gave a prediction accuracy of 86%, while results of applying models to Kitchenham data are least accurate among results, giving a prediction accuracy ranging by (22% - 36%). The three Ensemble learning models of stacking, bagging and boosting outperformed in prediction accuracy when applied to the ISBSG dataset, so they were depended on to predict the effort in the first stage of building the proposed tool. In next stage, after estimating work effort by Ensemble models, the COCOMO model used to estimate final cost of project.