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

تصميم أداة لتخمين كلفة مشاريع البرمجيات بالاعتماد على تقنيات الذكاء الاصطناعي

رسالة مُقدَّمة
إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
وهي جزء من متطلبات نيل شهادة ماجستير علوم في
البرمجيات

من قبل
محمد ماهر حسن داؤود العباسي

إشراف
أ.م.د جمال صلاح الدين سيد مجيد النعيمي

1444 هـ
2023 م

الملخص

تعاني الأساليب التقليدية لتقدير تكلفة البرمجيات من التحيز والعبء الزائد وللتغلب على مثل هذه المشاكل في الأساليب التقليدية، فقد ظهرت التقنيات القائمة على التعلم الآلي كحل فعال لمشكلة تقدير تكلفة البرمجيات. أُقترح أنموذج تجميع بالتراص في هذه الدراسة كنهج فعال لتقدير تكلفة تطوير البرمجيات. إذ تم بناء هذا الأنموذج باستعمال تقنية تعلم المجموعة. وتم إنشاء أنموذج مركب من ثلاثة مستويات. ففي المستوى الأول تم استعمال ست خوارزميات تعلم آلي مختلفة لغرض تخمين جهد التطوير. وفي المستوى الثاني تم استعمال خوارزمية الغابة العشوائية لغرض توليد تخمين الجهد النهائي بالاعتماد على تخمينات المستوى الأول. وفي المستوى الثالث والأخير تم استعمال طريقة الانحدار الخطي لغرض تخمين الكلفة بالاعتماد على الجهد المخمن الذي حُصِلَ عليه من طريقة الغابة العشوائية والتي تمثل متعلم المستوى الثاني. تم تدريب الأنموذج المقترح باستعمال مجاميع بيانات ISBSG الإصدار الثاني لسنة 2021 وهي مجاميع بيانات عالية الجودة تتكون من بيانات مشاريع مجموعة من شركات متنوعة موزعة حول مختلف دول العالم، أذ تحتوي على بيانات 10,600 مشروع برمجي لكل منها 252 خاصية.

يجمع الأنموذج المركب المقترح بين نتائج العديد من المُتعلمات التي تشمل: انحدار المتجه الداعم SVR، الانحدار البايزي Bayesian Regression، الانحدار الخطي Linear Regression، خوارزمية الجار الأقرب KNN والشبكات العصبية متعددة الطبقات MLP. تم دمج نتائج المُتعلمات الأساسية باستعمال Random Forest بوصفها المتعلم الفوقي. وتم ضبط المعلمات المختلفة لطرائق التعلم الآلي المطبقة باستعمال Particle Swarm Optimization (PSO). كشفت التجارب ان الانموذج المركب المقترح يتفوق على الطرق الفردية، وان التقنيات المختلفة لمعالجة البيانات بصورة مسبقة يمكن ان يكون لها أثر كبير في دقة التنبؤ. تم استعمال مجموعة بيانات ISBSG في بناء الأنموذج مع ثلثين من البيانات للتدريب وثلث للاختبار والتحقق من الصحة. لتقييم أداء أنموذج التنبؤ تم استعمال متوسط حجم الخطأ النسبي MMRE و Pred (25) وR-Squared. حقق انموذج تخمين الجهد دقة بلغت 98.63% حسب مقياس Pred(25)، ونسبة خطأ 0.04 حسب مقياس MMRE، و 1.0 حسب مقياس R2. فيما حقق الأنموذج المقترح لتقدير تكلفة البرمجيات دقة Pred(25) بنسبة بلغت 90٪، ونسبة خطأ 0.14 حسب مقياس MMRE، و 0.93 حسب مقياس R2.

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

Design a Tool to Estimate the Cost of Software Projects based on Artificial Intelligence Techniques

**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
Mohammed Maher Hasan Daud Al-Abbasi

Supervised by
**Asst. Prof. Dr. Jamal Salahaldeen Majeed
Alneamy**

2023A.D. **1444A.H.**

Abstract

Conventional methods of software cost estimation suffer from subjectivity and overhead. To overcome these limitations in the traditional approaches, Machine learning-based techniques have been emerged as an effective solution for the problem of software cost estimation. A Stacking ensemble model has been proposed in this study as an effective approach for software development cost estimation. This model was built using ensemble learning technique. A three-level ensemble model has been created. In the first level, six different machine learning algorithms were used for the purpose of estimating the development effort. In the second level, the random forest algorithm was used to generate the final effort estimate based on the first level estimates. In the third and final level, the linear regression method was used for the purpose of estimating the cost, depending on the estimated effort obtained from the random forest method, which represents the second level learner. The proposed model was trained using ISBSG data sets, second edition of 2021, which are high-quality data sets consisting of project data from a variety of companies distributed around the world, which contains data of 10,600 software projects, each with 252 characteristics.

The proposed composite model combines the results of several learners including: Support Vector Regression (SVR), Bayesian Regression, Linear Regression, KNN and Multilayer Neural Networks (MLP). The results of the basic learners were combined using Random Forest as the meta-learner. The different parameters of the applied machine learning methods were tuned using Particle Swarm Optimization (PSO). Experiments revealed that the proposed composite model is superior to the individual methods, and the different techniques of pre-processing for the data could have a significant impact on the accuracy of the prediction. The ISBSG dataset was used to build the model with two-thirds of the data for training and one-third for testing and validation. To evaluate the performance of the prediction model, MMRE, Pred (25) and R-Squared were used. The effort estimation model achieved an accuracy of 98.63% according to the Pred(25) scale, an error rate of 0.04 according to the MMRE scale, and 1.0 according to the R2 scale. The proposed software cost estimation model achieved Pred(25) accuracy of 90%, an error rate of 0.14 according to the MMRE scale, and 0.93 according to the R2 scale.