وزارة التعليم العالي والبحث العلمي
جامعـــــة المـــوصل
كلية علـوم الحاسوب والرياضيات
قســـم البرمـــجيات

تصنيف حـالة متطلبات البرمـجيات بالأعتماد على خوارزميات التقنيات الذكائية

رسـالة مقـدمة
الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهاده دبلوم عالي في
البرمجيات

من قبل
سمـاء عمـاد شـيت رسول
بإشـراف
أ.د. إبراهيـم أحمد صالح

١٤٤٦ هـ ٢٠٢٤ م

المستخلص

يعد تصنيف متطلبات البرمجيات من أهم التحديات في دورة حياة البرمجيات ،التي تساعد مصممي البرمجيات على تحليل المتطلبات البرمجية ووضع المواصفات الأساسية للمشروع ولتجنب الأخطاء. فهذه المتطلبات قد تؤدي إلى نجاح المشروع أو فشله ،أو حتى تهديد أمن المستخدمين. لذا، فإن عملية تصنيف البرمجيات تلعب دوراً حاسماً في نجاح المشروع والتحقق من صحة إحتياجات المستخدم الأكثر أهمية.

تهدف هذه الرسالة إلى تصنيف المتطلبات البرمجية بالأعتماد على خوارزميات التعلم الآلي. وتم إستخدام مجموعة بيانات تحتوي على سمات متطلبات البرمجيات المتعلقة بأنواع المتطلبات الوظيفية والغير وظيفية ، اذ تم إجراء معالجة أولية للبيانات تشمل أولاً تنظيف البيانات من الأعمدة التي لا تؤثر على عملية التصنيف ،ومن ثم إجراء معالجة نصية لكون مجموعة البيانات (dataset) عبارة عن مجموعة نصية عن طريق حذف المقاطع النصية الزائدة مثل (ing ,an) ،ثم حذف الكلمات التي تتكون من حرفين ،أما لجعل البيانات مناسبة كبيانات إدخال إلى خوارزميات التعلم الآلي فقد تم إعتماد تقنيتين لتحويل النص إلى أعداد ،الأولى حقيبة الكلمات (Bag of Words:BoW) والأخرى تقنية تردد المصطلح العكسي (Term Frequency-Inverse Document Frequenc:TF-IDF حيث الأولى تحول النص إلى أرقام صحيحة والأخرى تحولها إلى أرقام نسبية ،ومن ثم إستخلاص الخواص بالأعتماد على تقنية إحصائية تحليل التباين (Analysis of variance ANOVA) لتصبح البيانات جاهزة كإدخالات الى خوارزميات التعلم الآلي ،وقد إعتمد في هذه الرسالة خمس خوارزميات هي الأنحدار اللوجستي وآلة دعم المتجهات ، وخوارزمية التعزيز التكيفي وخوارزمية التعبئة وخوارزمية التكديس.

وتم تطبيق الخوارزميات أعلاهُ وإستخدام مقاييس الأداء المُعتمدة وهي الدقة والضبط والأستدعاء ومقياس (F1) لتقييم أداء هذه الخوارزميات، إذ أظهرت النتائج أن أفضل خوارزمية هي خوارزمية التكديس بدقة بلغت (0.95)، كذلك من الملاحظ أن إعتماد تقنية تردد المصطلح العكسي حققت نتائج أفضل من تقنية حقيبة الكلمات بصورة عامة وفي أغلب الخوارزميات المستخدمة حيث بلغت خوارزمية التكديس في الدقة(0.95) ،والضبط(0.95) ،ومعدل الأستدعاء(0.94) ،ودرجة F1(0.94) وهي من أفضل النتائج مقارنة بالبحوث التي إستخدمت نفس قاعدة البيانات ولكن لم تصل إلى نتائج جيدة.

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

Software Requirements Status Classification Based on Intelligent Techniques Algorithms

**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

Sama Emad Sheet Rasool

Supervised by

Ibrahim Ahmed Salih

2024 A.D. **1446 A.H.**

Abstract

Software requirements classification is one of the most important challenges in the software life cycle, which helps software designers analyze software requirements and develop basic project specifications and avoid errors. These requirements may lead to the success or failure of the project, or even threaten the security of users. Therefore, the software classification process plays a crucial role in the success of the project and validating the most important user needs.

This thesis aims to classify software requirements based on machine learning algorithms. A dataset containing software requirements features related to functional and non-functional requirements types was used. The data was preprocessed, including first cleaning the data from columns that do not affect the classification process, then text processing, since the dataset is a text set by deleting redundant text segments such as (ing, an), then deleting words consisting of two letters. To make the data suitable as input data to machine learning algorithms, two techniques were adopted to convert text to numbers: the first is Bag of Words (BoW) and the other is Term Frequency-Inverse Document Frequency (TF-IDF), where the first converts text to integers and the other converts it to relative numbers. Then, the properties were extracted based on the statistical technique of Analysis of Variance (ANOVA) so that the data becomes ready as inputs to machine learning algorithms. Five algorithms were adopted in this thesis: Logistic Regression, Support Vector Machine, and Adaptive Boosting Algorithm. The packing algorithm and the stacking algorithm.

The above algorithms were applied and the approved performance measures were used, namely accuracy, precision, recall and (F1) measure to evaluate the performance of these algorithms. The results showed that the best algorithm is the stacking algorithm with an accuracy of (0.95). It is also noted that adopting the inverse term frequency technique achieved better results than the bag of words technique in general and in most of the algorithms used, as the stacking algorithm reached (0.95) accuracy, (0.95) precision, (0.94) recall rate, and (0.94) F1 score, which is one of the best results compared to the research that used the same database but did not reach good results.