كشف الغموض في مستندات المتطلبات المكتوبة باللغة العربية باستخدام خوارزميات التعلم الآلي

إعداد
بيان السوارية

المشرف
د.أحمد الذنيبات

المشرف المشارك
د. بلال الحواشين

جامعة الزيتونة الأردنية , 2023

الملخص

يلعب تحديد الغموض في مستندات المتطلبات العربية دورًا مهمًا في هندسة البرمجيات . وذلك لأن جودة المتطلبات تؤثر بشكل مباشر على النجاح العام لمشاريع تطوير البرمجيات. تقليديا ، استخدم المهندسون الطرق اليدوية لتقييم جودة المتطلبات ، مما يؤدي إلى عملية مستهلكة للوقت وذاتية وعرضة للأخطاء. تستكشف هذه الرسالة استخدام خوارزميات التعلم الآلي لأتمتة تقييم المتطلبات المعبر عنها باللغة الطبيعية. تهدف الأطروحة إلى مقارنة خوارزميات التعلم الآلي المختلفة وفقًا لقدراتها في تصنيف المتطلبات المكتوبة باللغة العربية كشجرة قرار ، وآلة ناقلات الدعم (SVM) ، وأقرب جيران (KNN) ، والانحدار اللوجستي (LR)، والغابات العشوائية (RF) ، و XGBoost ، ونموذج .AraBERT يتم تقييم أداء هذه الخوارزميات بناءً على دقتها(F1 ، الاسترجاع ، الدقة (كما تم استخدام اثنين من المشتقات المختلفة لمقارنة المصطلحات الكاملة والمصطلحات المشتقة. كشفت النتائج أن Random Forest تفوقت على جميع المشتقات ، حيث حققت دقة قدرها 0.95 دون استخدام ساق ، و 0.99 مع ISRI Stemmer ، و 0.97 مع .Arabic Light Stemmer تسلط هذه النتائج الضوء على المتانة والتطبيق العملي لخوارزمية .Random Forest أظهرت خوارزمية K-Nearest Neighbours (KNN)أيضًا أداءً قويًا ، حيث حققت دقة قدرها 0.96 بدون جذع ، و 0.97 مع ISRI Stemmer ، و 0.97 مع Arabic Light Stemmerونتائج AraBERT في جميع الحالات العشوائية ، كان

متوسط الدقة 0.978 ، في حين أن متوسط الاسترجاع 0.98 ومتوسط درجة F1 هو 0.978. يهدف هذا البحث إلى التغلب على قيود التقييم اليدوي من خلال الاستفادة من خوارزميات التعلم الآلي ونموذج التعلم العميق لأتمتة تقييم جودة المتطلبات. الهدف هو تعزيز كفاءة وموضوعية عملية التقييم.

الكلمات المفتاحية: هندسة برمجيات ، متطلبات عربية تصنيف متطلب غامض ، متطلبات لا لبس فيها ، تعلم الآلة