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

أسلوب ممنهج لتتبع المتطلبات باستخدام التعلم الآلي

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

من قبل

شفق عزام نوري محمود

بإشراف

أ.د. نجلاء أكرم يونس ذنون

الخلاصة

تُعد عملية تتبع المتطلبات عنصـــرًا حيويًا في هندســـة البرمجيات لضـــمان توافق المنتج النهائي مع المواصفات الأولية وتسهيل إدارة التغيير. يهدف هذا العمل إلى تطوير وتحسين عملية تتبع المتطلبات من خلال أتمتة إنشـاء روابط التتبع بين "المتطلبات النصـية" و"الشـيفرة البرمجية" باسـتخدام تقنيات متقدمة في معالجة اللغات الطبيعية والتعلم الآلي.

تضـــمن العمل تمثيل المتطلبات والشـــيفرة البرمجية من مجموعة بيانات eTour بعد ترجمتها في فضـــاء دلالي باســـتخدام تقنيات التوجيه النحوي Vectorization مثل TF-IDF وWord2Vec، بهدف تحسين دقة المطابقة الدلالية بينهما. كما تم التركيز على تحسين جودة بيانات التدريب من خلال استكشاف تأثير إضافة وحذف سمات مختلفة وتطبيق استراتيجيات متنوعة لموازنة البيانات لمعالجة مشكلة عدم التوازن بين الروابط الصالحة وغير الصالحة، حيث أظهرت تقنية التقليل Undersampling باسـتخدام Tomek Links أفضـل النتائج. تم تقييم أداء خمس خوارزميات تعلم آلي شائعة وهي: الغابة العشوائية Random Forest، الانحدار اللوجستي Logistic Regression، شجرة القرار Decision Tree، الجيران الأقرب K-Nearest Neighbors، ومصنف تعزيز التدرج Gradient Boosting Classifier باســـتخدام مقاييس الدقة Precision، والاســـترجاع Recall، ومقياس F1-Score.

أظهرت نتائج التجارب النهائية تفوق نموذج الغابة العشـــوائية (Random Forest) بشكل عام، حيث حقق أعلى توازن بين المقاييس محققًا أعلى درجة F1-Score بلغت 0.59، مع قيم جيدة للدقة (0.60) والاستدعاء (0.59). يشير هذا إلى قدرته المتوازنة على تقليل الإيجابيات الخاطئة مع الحفاظ على قدرة جيدة في اكتشـــاف الحالات الإيجابية الفعلية. وعلى الرغم من تحقيق الانحدار اللوجستي (Logistic Regression) لأعلى استرجاع(0.89) ، إلا أن دقته المنخفضة جدًا (0.20) حدّت من فعاليته الكلية (F1=0.32) للتطبيقات التي تتطلب دقة عالية. أظهرت شـــجرة القرار (Decision Tree)أداءً متوســـطًا .(F1=0.45) بينما قدم مصـــنف تعزيز التدرج (Gradient Boosting Classifier)أعلى دقة (0.69) لكن مع اســـتدعاء منخفض (0.31) أدى إلى F1-Scoreمتوسـط (0.43). وكان أداء الجيران الأقرب (K-Nearest Neighbors) محدودًا نسـبيًا (F1=0.36)، خاصة بسبب انخفاض الاستدعاء.(0.29)

تجدر الإشـارة إلى أن أعلى درجة F1-Score المحققة في هذا العمل (0.59 بواسـطة الغابة العشـوائية) قد تتفوق على النتائج المبلغ عنها في بعض الدراسـات السـابقة التي اسـتخدمت نفس مجموعة البيانات أو مقاربات مشابهة.

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

A Systematic Approach to Requirements Traceability Using Machine 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

Shafaq Azzam Noori Mahmood

Supervised by

Prof. Dr. Najla Akram Youniss Thanoon

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

Abstract

Requirements traceability is a vital component in software engineering to ensure the final product aligns with initial specifications and to facilitate change management. However, traditional manual methods suffer from challenges related to efficiency, accuracy, and scalability, especially with increasing project complexity and the volume of unstructured textual data. This work aims to develop and improve the requirements traceability process by automating the creation of traceability links between "textual requirements" and "source code" using advanced techniques in Natural Language Processing (NLP) and Machine Learning (ML).

The methodology involved representing requirements and source code (from the eTour dataset, after translation) in a semantic space using vectorization techniques such as TF-IDF and Word2Vec, aiming to enhance the semantic matching accuracy between them. Emphasis was also placed on improving the quality of training data by exploring the impact of adding and removing different features and applying various data balancing strategies to address the imbalance between valid and invalid links. Undersampling using Tomek Links yielded the best results.

The performance of five common machine learning algorithms was evaluated: Random Forest, Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), and Gradient Boosting Classifier (GBC), using Precision, Recall, and F1-Score metrics.

The final experimental results demonstrated the overall superiority of the Random Forest model, which achieved the best balance among the metrics, attaining the highest F1-Score of 0.59, along with good Precision (0.60) and Recall (0.59). This indicates its balanced ability to reduce false positives while maintaining a good capability for detecting actual positive

cases. Although Logistic Regression achieved the highest Recall (0.89), its very low Precision (0.20) limited its overall effectiveness $(F1=0.32)$ for applications requiring high precision. The Decision Tree showed moderate performance $(F1=0.45)$. While the Gradient Boosting Classifier provided the highest Precision (0.69), it suffered from low Recall (0.31), resulting in a moderate F1−Score (0.43). The performance of K−Nearest Neighbors was relatively limited $(F1=0.36)$, especially due to its low Recall.(0.29)

It is noteworthy that the highest F1−Score achieved in this work $(0.59$ by Random Forest$)$ may surpass results reported in some previous studies using the same dataset or similar approaches.