Chapter 16
Classification of Product Backlog Items in Agile Software Development Using Machine Learning

Nirubikaa Ravikumar
Sabaragamuwa University of Sri Lanka, Sri Lanka

Ashansa Kithmini Wijeratne
Sabaragamuwa University of Sri Lanka, Sri Lanka

Banujan Kuhaneswaran
https://orcid.org/0000-0002-0265-2198
Sabaragamuwa University of Sri Lanka, Sri Lanka

B. T. G. S. Kumara
https://orcid.org/0000-0003-3941-2275
Sabaragamuwa University of Sri Lanka, Sri Lanka

Adeeba Saleem
Sabaragamuwa University of Sri Lanka, Sri Lanka

G. A. C. A. Herath
https://orcid.org/0000-0001-9610-9964
Sabaragamuwa University of Sri Lanka, Sri Lanka

ABSTRACT

In agile software development, product backlog items (PBI) are used to capture the user requirements prior to the product implementation. Many types of requirements can be observed within a software project. Proper classification of PBI can positively impact the software development process. PBI can be classified into three categories: user stories, foundational stories, and spikes. After the extreme literature survey, no research was held on classifying the PBI into the categories mentioned above. This paper proposed a machine learning (ML) based approach to classify the PBI into three categories. 4,721 PBI were collected from different software projects and manually labelled into the three classes mentioned above. Then the PBI were cleaned using different pre-processing techniques. Classification models were constructed using ML techniques. The performance of each ML model was evaluated using accuracy, precision, recall, and F1 score. Support vector machine (SVM) outperformed other ML models by providing 88% accuracy.

INTRODUCTION

Software Engineering (Sommerville, 2004, 2020) is a systematic engineering approach and an engineering discipline concerned with all aspects of software development and maintenance of software products. It helps to adopt a systematic approach to developing products (Glass, Vessey, & Ramesh, 2002). Rather than individual programming, software engineering is intended to support professional software development.

The Software Development Life cycle (SDLC) (Ragunath, Velmourougan, Davachelvan, Kayalvizhi, & Ravimohan, 2010) is a systematic process that supports the production of high-quality software at a lower cost. The primary goal of this SDLC is to produce the best software that meets the customer's expectations in the shortest possible time. The traditional waterfall (Adenowo & Adenowo, 2013; Balaji & Murugaiyan, 2012) model is more rigid and based entirely on following a set of steps that keeps the team moving forward. It doesn't have any space for unexpected changes or revisions in this form. A sudden change in requirements in the project will result in the need for rework. The waterfall model has less end-user or client involvement within a project. Adding opinions and clarifying what clients want as the project moves is impossible in this waterfall model (Alshamrani & Bahattab, 2015).

The Agile methodology (Sharma, Sarkar, & Gupta, 2012) was proposed to respond to the limitations of the traditional waterfall model. It is an iterative project management and software development methodology that enables teams to deliver value to customers [6] quickly. It provides customer involvement in decision-making, leading to excellent customer retention (Balaji & Murugaiyan, 2012). The Agile methodology was very flexible, and even late changes in requirements were welcomed.

In Agile development, Product Backlog Items (PBI) (Sedano, Ralph, & Péraire, 2019) are a prioritised set of requirements to be implemented in product development. It's a decision-making tool that aids in estimating, refining, and prioritising everything that might be done in the future. Many individuals consider the backlog a to-do list, and they define it as a list of tasks that must be completed to get the product into the market. In detail, we can say the PBI is a prioritised list of all potential product features, all fundamental elements that must be in place for the attributes to be implemented, and all critical research activities and studies that must take place to achieve accurate decisions about how to approach feature development and requirements. To increase the quality of the product, we need to consider the features in PBI.

We can classify the items in PBI as (i) user stories (Amna & Poels, 2022; Raharjana, Siahaan, & Fatichah, 2021): simple statements which explain the need or requirements of the product and are written from a user perspective (ii) foundational stories: a foundational or infrastructure needs that we have to fulfil the user stories, and (iii) spikes (Al Hashimi, Altaleb, & Gravell, 2020; Al Hashimi & Gravell, 2020): a most important feature that includes researches and studies to achieve a better understanding on user needs and development procedures.

Classification in ML is a method of grouping a given data set into classes. It can be done with both structured and unstructured data. Classifying the given PBI is the primary research objective. Since three types of PBI will be considered in this research, authors have to use multiclass classification models such as Artificial Neural Network (ANN), Naive Bayes, Support Vector Machine (SVM), Logistic Regression, and Decision Tree.

The structure of this paper is divided into six sections. In Section II, provides the background of the study. Section III explains the related works so far held in the proposed approach and how our works differ from the existing studies. Section IV presents the motivation of the research. Section V describes

Related Content

The Border: Between Hybridization and Separation
Elvira Martini, Raffaele De Luca Picioneand Antonio Ciaschi (2022). *Handbook of Research on Applying Emerging Technologies Across Multiple Disciplines (pp. 391-408).*
www.irma-international.org/chapter/the-border/301330

Digitalization of the Financial Sector: New Opportunities and Challenges During the COVID-19 Crisis
Karima Toumi Sayari (2023). *Revolutionizing Financial Services and Markets Through FinTech and Blockchain (pp. 78-98).*
www.irma-international.org/chapter/digitalization-of-the-financial-sector/326986

A Discussion on the Relationship Between Information and Communication Technologies (ICT) and Entrepreneurship
Mehmet Erylmaz (2021). *Research Anthology on Digital Transformation, Organizational Change, and the Impact of Remote Work (pp. 1889-1900).*
www.irma-international.org/chapter/a-discussion-on-the-relationship-between-information-and-communication-technologies-ict-and-entrepreneurship/270381

An Analysis of How Cognitive Cities Better Address Sustainability and Equity Concerns in Society as Compared to Smart Cities
Kevin Richardson (2023). *Applied Research Approaches to Technology, Healthcare, and Business (pp. 157-170).*
www.irma-international.org/chapter/an-analysis-of-how-cognitive-cities-better-address-sustainability-and-equity-concerns-in-society-as-compared-to-smart-cities/331647

New Trends and Tools for Customer Relationship: Challenges in Digital Transformation
Ana Limaand Jorge Pacheco (2021). *Research Anthology on Digital Transformation, Organizational Change, and the Impact of Remote Work (pp. 1916-1934).*
www.irma-international.org/chapter/new-trends-and-tools-for-customer-relationship/270384