I. Introduction

The initial and one the most crucial stage in developing software is requirement engineering. The output of this stage is Software Requirement Specification (SRS) [1]. This document describes all system requirements that must be met by the developed software [2]. Several processes including identification and collection of requirements are required to design artifacts in the SRS. These processes are categorized into Requirement Elicitation, that can be done by performing direct communications such as focus group and interviews, or indirect communication through online questionnaires to parties interested in the project [3], [4]. During requirement elicitation, natural language is might be used as the obtained results can be from discussions and interviews, confirming the applicability of Natural Language Processing (NLP) for the data analysis [5].

Requirement specification consists of requirements that a system must meet or fulfil, contracts and conformance standards based on user needs and expectations. Therefore, it is highly recommended for a system analyst or developer to have the same understanding and perspective on requirement specifications so that the software development results follow user needs [6]. In several instances, there are differences in perspective between clients and developers related to software requirements specifications, leading to incorrect requirement specifications. Therefore, a study is required to help validate the software requirements. As both requirement specification and elicitation can use NLP, there is an opportunity of using text analysis or text mining method to automate the validation process.

This research aims to propose a method to validate the requirements provided in the SRS document using text mining analysis. The validation method utilizes the interview transcripts from the requirement elicitation process as inputs. The method then compares the similarity and suitability of the transcripts with the provided functional and non-functional requirements included in the SRS document. Our method also describes the text pre-processing methods prior to text analysis. To evaluate our method, this research uses the SRS documents from an application called Baker's Corner. Baker's Corner is an Android-based pastry product marketing and sales application that is available in Google Play. The selection of the Baker's Corner application was based on the ease of accessing direct interviews with stakeholders and it being an application previously developed by the author. Therefore, the SRS documents of the Baker's Corner application are used as a case study and can be beneficial for general application development purposes. The research aims to mitigate discrepancies in user and developer perceptions that can lead to requirement misalignment. The contributions of our study are as follows:

- Proposing a method to validate functional and non-functional requirements using text mining analysis, including pre-processing and text similarity, by comparing them against the requirement elicitation artefacts.

- Evaluating our proposed method comprehensively by performing reliability testing with Cohen's Kappa.

- Generate a similarity value from the text mining analysis process to compare it with the results of reliability testing for validating the requirement specification.

- Providing discussions related to factors that might affect the low similarity score and some software requirement improvement suggestions.