

Abstract

Steam is a platform that provides services for purchasing video games digitally. It also offers features for users to share video game reviews with each other. These reviews reflecting the players' experiences and opinions. However, these reviews do not reveal the main issues reported by the reviewers or their changing perceptions over time. Contradictions between the reviews users write and the labels they provide can also occur. Therefore, the reviews written by users will be classified through sentiment analysis. This study conducts a sentiment analysis on Disco Elysium video game reviews from Steam using K-Nearest Neighbor (KNN) method and Chi-Square feature selection. There are five scenarios in this study, first is to compare the impact of using stemming process in the preprocessing stage, second is to measure the effect of applying Grid Search for hyperparameter tuning, third is to compare the effects of data splitting with a ratio of 80:20 and 50:50, fourth is to compare the results of Euclidean Distance and Manhattan Distance in KNN classification, and fifth is to compare the result of using KNN method and SVM method. The result indicates that using stemming in the preprocessing stage, not applying Grid Search hyperparameter tuning, data splitting with a ratio of 80:20, and KNN with Euclidean Distance provides the best performance. The best performance results achieved are 79.76% precision, 75.24% recall, and 75.21% F1-Score.

Keywords: sentiment analysis, k-nearest neighbor, chi-square, Steam, Disco Elysium