

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Currently, games are one of the easiest entertainments to find, because there are already many game sales platforms available, such as Steam Store, Epic Game Store, Blizzard, PlayStation Store and others. Not only do they sell games digitally they also provide features to download and play them directly. The Steam Store also provides a means for the community to discuss, profile to show off the number of games and playtime, chat features, and exchange and sell digital items or trade items [1]. With so many features, steam is able to invite many users, even in 2020 Steam's active users reached 120 million monthly active users with peak concurrent users reaching 24.8 million users [2].

On Steam game store page, there are features that can help users such as overall user reviews, customer reviews, and tags that have a function to navigate user. User reviews and customer reviews which are important for evaluating game quality [4]. On Steam Store, every review is sorted based on the number of upvotes and comments, when a new review is posted it will appear at the very end, but if a review get a lot of upvotes, then the review will appear at the top of the list. This is a plausible theory, but in practice it takes time and it is rare for new reviews to appear at the top of the list. The result is many useful reviews that are not visible to users, as most user only see them in a cursory manner. In addition, there is also a high probability that there is a bad review that should be at the bottom of the list but could enter the top ranking because of getting fake votes and comments.

Data obtained from Kaggle [3], namely data from 2010 to 2019. Data will be processed using the Naïve Bayes and K-Means algorithm, each algorithm will be matched with original algorithm from Python library, after that both algorithm will be matched again to determine which algorithm is more accurate for predicting sentiment of Steam reviews.

## **1.2 Problem Formulation**

1. Is every review that has the most upvotes not a fake review?
2. Does a long review mean it's a useful review?
3. Is a review from player with high hour of played useful?

## **1.3 Scope**

The data to be used is review data for each game on Steam Store starting from 2010 until 2019.

## **1.4 Objective**

The purpose of this project is to provide insight to customers those reviews can be trusted.

