I. Introduction

Opinion is one of the factors that influence human behavior [1]. Commentary on a film is a text that often describes someone's opinion about a film [2]. Comments can play an important role in the success of a film because comments are a way used to express feelings, opinions, and ratings of a film [3]. In [4], the usage of the Naïve Bayes method can make it difficult to group an opinion into two aspects, positive and negative sentiment. In [5], they used information gain as a feature selection by removing unrelated features because most of the irrelevant features are removed.

The information contained in the film can be concluded as the quality of a film experience from the audience, but the rating of 'inappropriate' with the context sentence makes the rating of film 'not recommended'. **This issue supports this research where a film should be classified by sentiment. This paper uses the Naïve Bayes method for classification, TF-IDF as the feature extraction, and Information Gain as the feature selection. However, the uses of Naïve Bayes to classify the word are conditionally independent of each other [10]. This research uses TF-IDF and Information Gain to prevent this problem**. In this research, TF-IDF, Information Gain, and Naïve Bayes can result in precision, recall, and f1-score above 82%.

This research identifies the effects against precision, recall, dan f1-score values using data preprocessing, subtraction from the number of feature used or term, the threshold against precision, recall, and f1-score values, and cross-validation against precision, recall, and f1-score. This research also has a limitation that becomes a constraint, the large number of datasets. Hence, this research does not use all the terms or words of feature extraction due to the limitations on the number of available laptops.

The methodology used in this research is sentiment analysis on the dataset movie review from IMDb Movie Review using the Naïve Bayes method as the classification, the TF-IDF method as the feature extraction, and Information Gain method as the feature selection and to obtain higher precision, recall and f1-score values that are optimized using multiple scenarios.

The contents of the paper are divided five sections. Sections I presents introduction. Sections II presents related work. Sections III presents methodology. Sections IV presents experimental results. And sections V presents conclusion and future works of all experiments.