PREPARING MOVIE REVIEW TEXT DATA FOR SENTIMENT ANALYSIS IN AZERBAIJANI

Nijat Zeynalov

University of Tartu, The School of Economics and Business Administration

zenijat16631@sabah.edu.az

ABSTRACT: Sentiment analysis is the interpretation and classification of emotions (positive, negative and neutral) within text data using text analysis techniques. The main objective of this paper is to retrieve and pre-process movie review text data in the Azerbaijani. Although, most of data on social media is text-based, it is not possible to directly apply deep learning processes to this raw data and text data preparation is different for each problem. Preparation starts with simple steps, like loading data, but quickly gets difficult with cleaning tasks, filtering only the relevant data which are very specific to the data we are working with. [14] One of the biggest obstacles to this study was the lack of resources that prevents us to produce better quality text data in the Azerbaijani language. Some processes are required to prepare the data before the algorithms are implemented and in this paper, you will discover how to prepare movie review text data for sentiment analysis in Azerbaijani.

KEYWORDS : sentiment analysis, movie review data, data preprocessing, machine learning, deep learning, classification

1. INTRODUCTION

Movie review is an article which expresses the opinions of its readers, allowing us to grasp the general concept of the film and determine whether to watch it or not, by reviewing the film, either positive or negative. A film review can affect the whole team that is working on the film.

A study illustrates that in some cases, the success or the failure of a movie depends on its reviews [2]. Therefore, a vital challenge is to be able to classify movies reviews to capture, retrieve, quantify and analyze watchers more effectively [3]. The grading of film reviews as favorable or negative feedback is linked with word events from the text reports, and the environment used such terms in a positive or negative way previously.