Image Caption Validation for Public Complaints on Social Media

Naufal Scudetto[1], Ade Romadhony[2]

[1,2,3]School of Computing, Telkom University, Bandung
[1]naufalscudetti@students.telkomuniversity.ac.id, [2]aderomadhony@telkomuniversity.ac.id,

Abstrak

Image Caption Validation adalah tugas memvalidasi apakah sebuah caption cocok dengan gambar yang diberikan. Hasil dari tugas ini dapat mengurangi kesusahan dengan memvalidasi keluhan yang disampaikan dari masyarakat setempat. Dalam tulisan ini, kami memperkenalkan satu set data khusus keluhan masyarakat Indonesia yang terdiri dari keluhan banjir dan membuang sampah sembarangan dari Twitter. Kami menggunakan model VGG16 fine-tuned untuk klasifikasi gambar dan Model IndoBERT (Model Bert Indonesia) untuk klasifikasi teks. Kedua model ini digabungkan menjadi sistem yang memvalidasi apakah keterangan yang diberikan mendeskripsikan gambar secara akurat. Hasil eksperimen menunjukkan bahwa VGG16 fine-tuned memiliki skor akurasi 93% , Model IndoBERT memiliki akurasi 89%, dan Model Prediksi Keluaran Validasi memiliki akurasi 56%.

Kata kunci : image caption validation, klasifikasi gambar, klasifikasi teks, VGG16 fine-tuned , indoBERT, Model Prediksi Keluaran Validasi

Abstract

Image Caption Validation is the task of validating whether a caption matches the image provided. The result of this task can reduce distress by validating complaints conveyed from the local community. In this paper, we introduce a custom data set of Indonesian public complaints that consists of floods and littering complaints from Twitter. We used a fine-tuned VGG16 model for image classification and an IndoBERT Model (Indonesian Bert Model) for text classification. Both of these models were combined to into a system that validates whether a given caption accurately describes the image. The experimental result showed that the fine-tuned VGG16 had an accuracy score of 93% ,the IndoBERT Model had an accuracy of 89%, and the Validation Output Prediction Model had an accuracy of 56%.

Keywords : image caption validation, image classification, text classification, fine-tuned VGG16, indoBERT, validation output prediction model

1. Introduction

There have been quite a few interests in computer vision tasks, one of those is the evaluation of image captioning. The task at hand is considered very difficult because the outputs of the image captions are in the form of natural language and must mirror the con- tent of the given image (1). A major challenge for establishing a structural connection between captions and images is that the links between predicates and the corresponding object regions are often ambiguous (2). Social media can provide researchers with a data set richer in diverse demographic background characteristics versus the conventional homogeneous groups mostly used in university settings where students are often the sole participant group for scholarly research (3).

Image caption validation on social media is important because it can validate the representation of community distress online (4). The goal would be to improve community distress by validating the relevance of the post and address the actual complaint. By validating the image caption, many people can benefit from this to highlight and focus on important and urgent issues without the trouble of actually having to validate the complaint manually. This also inherently reduces noise coming in from these complaints by removing irrelevant ones.

This paper addresses the task of Image Caption Validation to a data set that is mostly composed of complaints from Twitter. We build a model that can extract the aspects of floods (banjir), littering (sampah) and non complaints (bukan keluhan). These complaints are the most frequent ones that are voiced by the people of Indonesia through social media (Twitter).

Fig. 1 shows an example of a complaint where an image and caption are highly relatable, which in this case is about flooding(banjir).