

## ABSTRAK

Purbalingga merupakan sebuah wilayah yang terletak di Provinsi Jawa Tengah. Purbalingga menawarkan keindahan alam dan tempat wisata yang menarik. Banyak wisatawan mengabadikan momen kunjungannya dalam bentuk foto/gambar yang kemudian diunggah di media sosial. Namun, sebuah gambar dapat mengandung banyak informasi dan setiap individu dapat menafsirkannya secara berbeda. Tanpa adanya keterangan atau *caption* pada gambar, manusia akan kesulitan mengurai informasi tersebut. *Image captioning* mampu mengatasi tantangan ini dengan menghasilkan deskripsi teks otomatis pada gambar. Selain itu, *text-to-speech* digunakan untuk meningkatkan aksesibilitas bagi penyandang tunanetra dalam memahami keterangan gambar. Penelitian ini bertujuan mengembangkan model *image captioning* pada gambar objek wisata di Purbalingga menggunakan arsitektur *transformer* serta ResNet50. Arsitektur *transformer* menggunakan *attention mechanism* untuk mempelajari konteks dan hubungan antara *input* dan *output*, sementara ResNet50 merupakan jaringan konvolusional yang handal dalam ekstraksi fitur gambar. Evaluasi model menggunakan metrik BLEU, yang membandingkan kalimat prediksi dengan kalimat referensi, menunjukkan hasil terbaik BLEU- $\{1, 2, 3, 4\} = \{0.672, 0.559, 0.489, 0.437\}$ . Eksperimen dengan berbagai kombinasi *hyperparameter* menunjukkan penambahan *embedding* dan *layer* meningkatkan waktu pelatihan dan menurunkan skor BLEU, sementara perubahan jumlah *head* tidak terlalu memengaruhi hasil. Model terbaik diimplementasikan dalam aplikasi berbasis web menggunakan metode SDLC *waterfall*, *framework Flask*, dan basis data *MySQL*. Aplikasi ini memungkinkan pengguna mengunggah gambar objek wisata, mendapatkan deskripsi otomatis dalam bahasa Indonesia, dan mendengarkan *caption*-nya yang dibacakan secara audio menggunakan fitur *text-to-speech* berbasis *Web Speech API*. Hasil pengujian dengan metode *blackbox* menunjukkan hasil valid pada setiap pengujian, menandakan aplikasi berjalan sesuai kebutuhan dan layak digunakan.

**Kata kunci:** *image captioning, objek wisata, text-to-speech, transformer*

## **ABSTRACT**

*Purbalingga is a region located in Central Java Province, offering interesting natural beauty and tourist destinations. Many tourists capture their moments in photos, which are then uploaded to social media. However, a picture can contain a lot of information, and each individual may interpret it differently. Without captions, people may struggle to extract this information. Image captioning addresses this challenge by automatically generating text descriptions for images. Additionally, text-to-speech is used to enhance accessibility for the visually impaired in understanding image descriptions. This research aims to develop an image captioning model for images of tourist attractions in Purbalingga using transformer architecture and ResNet50. The transformer architecture employs an attention mechanism to learn the context and relationships between inputs and outputs, while ResNet50 is a robust convolutional network for image feature extraction. Model evaluation using BLEU metrics, which compare generated sentences to reference sentences, shows the best results as  $BLEU-\{1, 2, 3, 4\} = \{0.672, 0.559, 0.489, 0.437\}$ . Experiments indicate that increasing embeddings and layers extends training time and lowers BLEU scores, while changing the number of heads has minimal impact on results. The best model is implemented in a web-based application using the SDLC waterfall method, Flask framework, and MySQL database. This application allows users to upload tourist attraction images, receive automatic descriptions in Indonesian, and listen to the captions read aloud using the Web Speech API-based text-to-speech feature. Blackbox testing results show valid outcomes for all tests, indicating that the application operates as required and is suitable for use.*

**Keywords:** *image captioning, text-to-speech, tourist attractions, transformer*