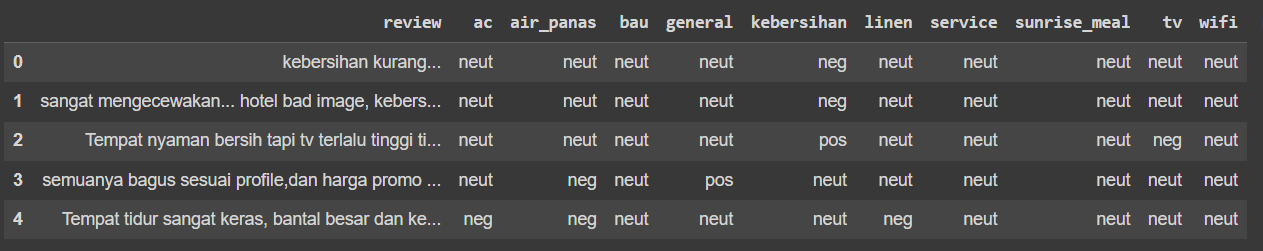
1. Data Exploration

Dataset Hoasa merupakan dataset yang berisi mengenai review dari pelanggan mengenai aspek yang ada di hotel. Didalam dataset tersebut terdapat 12 atribut yang terdiri dari 11 aspek penilaian hotel dan review pelanggan. Nilai dari setiap aspek penilaian diambil review yang diberikan oleh pelanggan hotel. Penilaian terhadap aspek tersebut terdapat 4 nilai yaitu positive, negative, neutral, dan positive-negative. Dari dataset tersebut akan diimplementasikan aspect based sentiment analysis. Berikut merupakan informasi yang ada di dataset hoasa.



1. Methodology

Dalam kegiatan ini digunakan metode BERT yang merupakan salah satu metode klasifikasi transformer yang biasanya digunakan untuk melakukan text classification.

1. Data preprocess

Hal yang pertama dilakukan adalah dengan melakukan Pra-proses dataset hoasa yang akan digunakan sebagai input ke model BERT. Karena dalam dataset terdapat banyak aspek maka hal yang dilakukan adalah menggabungkan semua aspek menjadi satu kolom dan menggabungkan penilaian aspek dengan aspek itu sendiri. Hasilnya bisa dilihat dibawah ini.



1. Data Preparation

Hal selanjutnya dilakukan adalah menyiapkan dataset untuk bisa diproses oleh model. Oleh karena itu karena atribut label yang ada di dataset yang awalnya berbentuk string diubah menjadi integer dengan menggunakan fungsi MultiLabelBinarizer() dari kelas sklearn. Setelah itu untuk atribut review perlu dilakukan tokenize dan encode data teks sesuai dengan format BERT.

1. Membangun Model

Metode yang digunakan untuk melakukan klasifikasi teks adalah BERT. Model yang dibangun disesuaikan dengan bahasa yang ada di dataset. Bahasa yang ada di dataset ini merupakan bahasa indonesia. Dalam membangun model menggunakan bantuan library python.

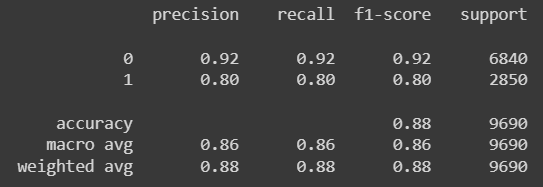
1. Melakukan Training Model

Setelah membangun model adalah melakukan train model dengan menggunakan bantuan pytorch. Pytorch melakukan training berulang kali dengan menerapkan perulangan untuk epochs, batches, melakukan backpropagation dan menghitung loss. Parameter yang diinputkan hanya jumlah epoch, jenis optimizer, dan learning rate. Dengan menggunakan pytorch untuk melatih model dapat menemukan model terbaik dari iterasi yang dilakukan ketika melakukan training model sehingga tidak perlu melakukan hyperparameter tuning.

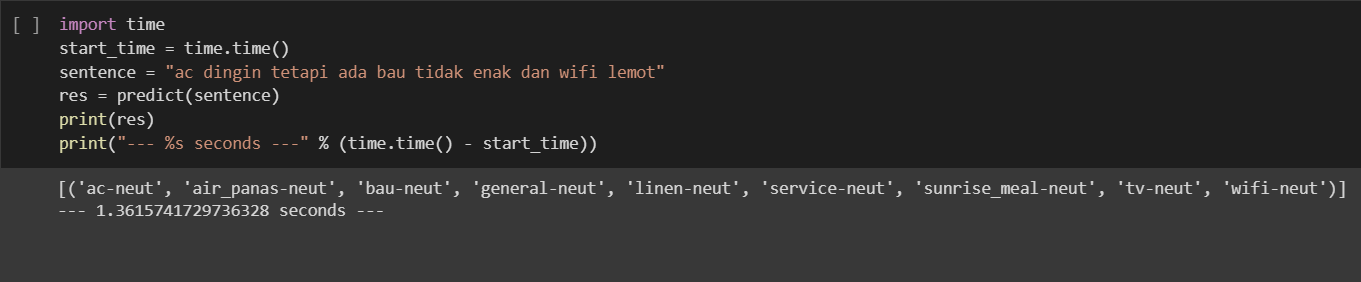
1. Melakukan Evaluasi

Setelah melakukan training model hal yang dilakukan melakukan evaluasi dari hasil yang didapat dengan membandingkan hasil prediksi dengan data aktual dengan cara menggunakan fungsi classification report.

1. Result and Analysis

Dari hasil yang didapat akurasi dari metode BERT ketika melakukan klasifikasi teks berdasarkan aspek yang ada di dataset adalah 0.88 atau 88%. Hasil ini selalu sama ketika dilakukan training model berulang kali. Sementara itu macro average dari precision, recall dan f1 score adalah 86% dan weighted average dari ketiga metrik tersebut adalah 88%.

Selain itu ketika dilakukannya inference hasil yang didapat adalah diperlukan 1 detik untuk melakukan melakukan inference per kalimat. Berikut merupakan waktu running dari proses inference



Hasil prediksi yang diberikan dari hasil training yang dilakukan memberikan hasil neutral yang cukup banyak di setiap aspek. Alasan dari hal tersebut karena dengan banyaknya nilai neutral di setiap aspek di dataset maka ketika dilakukan inference hasil yang dikeluarkan hanya mengeluarkan nilai netral untuk setiap aspek dan nilai dari semua aspek dijadikan satu label.