

Identifikasi Kesamaan Pertanyaan pada Soal Bahasa Indonesia Menggunakan Metode Recurrent Neural

Muhammad Iqbal¹, Hasmawati², Ade Romadhony³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹muhammadiqbaaall@students.telkomuniversity.ac.id, ²hasmawati@telkomuniversity.ac.id,

³aderomadhony@telkomuniversity.ac.id,

Abstrak

Dalam sebuah forum tanya jawab, identifikasi kesamaan pertanyaan digunakan untuk melihat kemiripan antara dua buah pertanyaan. proses ini dilakukan untuk membandingkan apakah pertanyaan yang di *input* oleh *user* memiliki kesamaan pada pertanyaan yang ada pada *database* untuk meningkatkan performa sebuah sistem pada forum tanya jawab *online*. Pada saat ini, identifikasi kesamaan pertanyaan banyak dilakukan dalam bahasa asing. Tujuan dari penelitian ini adalah untuk mengidentifikasi kesamaan pertanyaan dalam bahasa Indonesian dan mengevaluasi keefektifan metode yang digunakan. Data yang digunakan adalah data public dengan pasangan pertanyaan yang berlabel 0 dan 1, dimana label 0 untuk pasangan pertanyaan yang berbeda, dan label 1 untuk pasangan pertanyaan yang sama. Metode yang digunakan adalah Recurrent Neural Network (RNN) dengan pendekatan Manhattan Distance untuk menghitung jarak kesamaan antara dua pasangan pertanyaan. pasangan pertanyaan yang berlabel tersebut dijadikan dua inputan untuk diidentifikasi kesamaan nya menggunakan metode RNN kemudian jarak kesamaan antaran dua pasangan pertanyaan tersebut dihitung menggunakan formula Manhattan Distance. Kami mengevaluasi model menggunakan tiga optimizer yang berbeda yaitu RMSprop, Adagrad, dan Adam. Hasil terbaik yang diperoleh adalah dengan menggunakan optimizer Adam dengan rasio split-data 80:20 dan hasil keseluruhan adalah akurasi sebesar 76%, presisi 74%, recall 98,8%, dan F1-Score sebesar 85,1%.

Kata kunci : Kesamaan pertanyaan, Manhattan Distance, Pengukuran Semantik, RNN.
