Pengembangan Sistem Penilaian Esai Otomatis (SIMPLE-O) pada Ujian Esai Bahasa Indonesia Menggunakan Stacked Bidirectional Gated Recurrent Unit dengan Manhattan Distance dan Cosine Similarity = The Development of an Automatic Essay Scoring System (SIMPLE-O) for Indonesian Language Essay Exams Using Stacked Bidirectional Gated Recurrent Unit with Manhattan Distance and Cosine Similarity

Sidauruk, Febriana Pasonang, author

Abstrak

Skripsi ini membahas mengenai Pengembangan Sistem Penilaian Esai Otomatis (SIMPLE-O) dengan menggunakan Stacked Bidirectional GRU dengan Manhattan Distance dan Cosine Similarity yang diterapkan untuk menilai jawaban esai bahasa Indonesia. Data yang digunakan pada sistem terdiri dari jawaban esai pelajar dan kunci jawaban dari pengajar. Sistem akan melalui tahapan pre-processing, word embedding, kemudian proses training, dan terakhir proses testing. Data sebelumnya diolah untuk dilakukan training terlebih dahulu dengan memberikan tujuh skenario pengujian agar memberikan selisih dan error yang rendah. Kedua jawaban akan diuji menggunakan dengan variasi hyperparameter sesuai dengan hasil terbaik dari seluruh skenario pengujian, kemudian diukur kemiripan hasil keduanya menggunakan dua jenis metrics yaitu, Manhattan Distance dan Cosine Similarity. Model menggunakan Cosine Similarity menghasilkan rata-rata nilai selisih 1.935 untuk fase training dan 8 untuk fase testing. Sedangkan Manhattan Distance menghasilkan selisih 1.887 untuk fase training dan 9.039 untuk fase testing.
..... This thesis discusses the design of an Automatic Essay Scoring System (SIMPLE-O) using Stacked Bidirectional GRU with Manhattan Distance and Cosine Similarity for Indonesian essay grading. The system utilizes a dataset consisting of student essay answers and corresponding teacher's answer key. The system goes through several stages including pre-processing, word embedding, training, and testing processes. The data is pre-processed and then trained using seven different testing scenarios to achieve low difference and low-error results. The system is evaluated using various hyperparameter settings based on the best results obtained from all testing scenarios. The similarity between the generated scores and the reference scores is measured using two metrics: Manhattan Distance and Cosine Similarity. The Cosine Similarity-based model achieved an average difference of 1.935 during the training phase and 8 during the testing phase. On the other hand, the Manhattan Distance-based model achieved an difference of 1.887 during the training phase and 9.039 during the testing phase.