PENERAPAN SPEECH RECOGNITION MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM) UNTUK PRESENTASI DINAMIS

LATAR BELAKANG

Presentasi merupakan metode komunikasi dengan menyampaikan informasi, ide, dan gagasan oleh seseorang kepada sekelompok orang/audiens yang banyak. Penyampaian materi ini dapat dilakukan dengan bantuan software presentasi, seperti power point. Namun, untuk kontrol pengoperasian software presentasi ini memerlukan bantuan perangkat pendukung lain seperti mouse, keyboard, atau remote. Penggunaan perangkat tambahan ini tidak jarang menimbulkan distraksi bagi pembicara karena harus terus menerus mengoperasikan perangkat selama presentasi, sehingga fokus penyampaian materi akan terbagi. Selain itu, terdapat kemungkinan bahwa perangkat tambahan ini mengalami malfungsi yang dapat mengganggu berjalannya proses presentasi. Dengan demikian, dibuatlah sistem kontrol software presentasi dengan menerapkan speech recognition sebagai voice command untuk mengoperasikan presentasi secara dinamis.

DATA PENELITIAN

Dataset yang diperoleh memuat file suara yang berisi beragam kata ucapan dengan durasi 1 detik. Data dikumpulkan oleh Warden (2018) dengan total 34 variasi kata berbahasa inggris berjumlah 103.807 ucapan. Penelitian ini mengaambil 8 kosa kata untuk dijadikan sebagai kata perintah.

Tabel 1. Data Kata Perintah

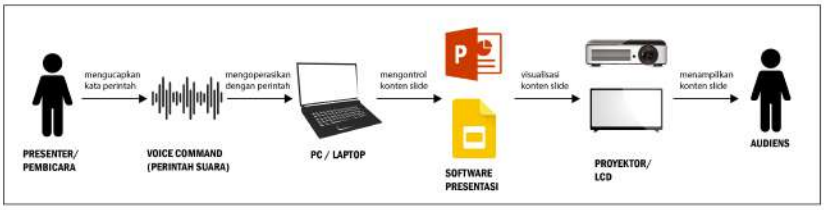
Kata	Jumlah Ucapan	Perintah
Down	3.917	Bergeser ke bawah
Go	3.880	Nyalakan media player
Left	3.801	Pindah slide ke kiri (sebelum)
Off	3.754	Matikan mode presentasi
On	3.845	Nyalakan mode presentasi
Stop	3.872	Hentikan audio stream dan terminate program
Up	3.723	Bergeser ke atas

HASIL EKSPERIMEN

Tabel 2. Evaluasi Model

Arsitektur Model	Training		Validation		Testing	
	Loss	Acc	Loss	Acc	Loss	Acc
128 - 64 - 8	0.3466	0.8844	0.3136	0.8954	0.3188	0.8896
128 - dropout(0.5) - 64 - 8	0.4438	0.8616	0.3117	0.8793	0.3889	0.8691
256 - 128 - 8	0.1850	0.9375	0.2286	0.9243	0.2528	0.9167
256 - dropout(0.5) - 128 - 8	0.1704	0.9418	0.1752	0.9434	0.1883	0.9398

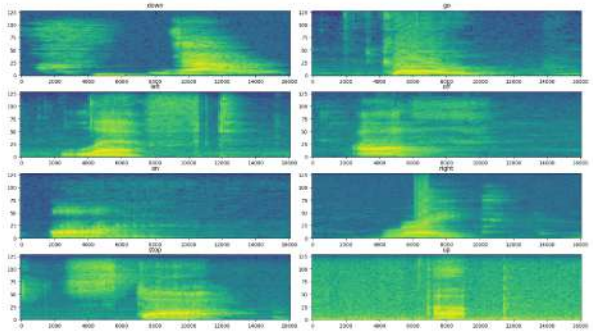
METODE PENELITIAN



Gambar 1.
Alur Sistem Speech Recognition

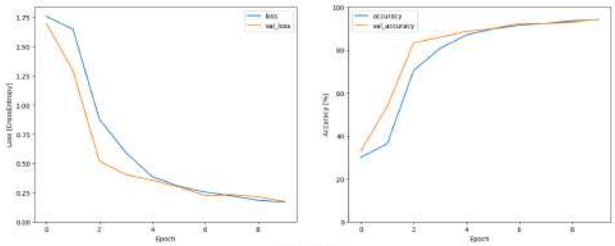
IMPLEMENTASI SISTEM

Model dengan arsitektur LSTM 256 - dropout(0.5) - LSTM 128 - 8, sistem mampu mendeteksi kata 8 kata perintah dengan baik. Audio streaming menghasilkan input berupa sinyal suara yang kemudian diubah menjadi spectogram. Spectogram sepanjang 16000 sample rate (1 detik) ini akan diprediksi kata perintah yang muncul. Threshold 0.7 digunakan sebagai batas confidence minimum terdeteksinya suatau kata perintah

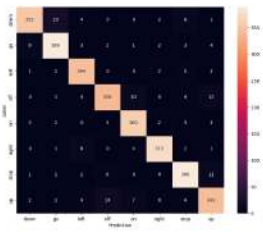


Gambar 2.
Audio Spectrogram

Gambar 3.
Speech Recognition



Gambar 4.
Fitting Model Terbaik



Gambar 5.
Confusion Matrix Model Terbaik

KESIMPULAN

Dari 4 model yang diujikan, model denan arsitektur layer LSTM256 - dropout(0.5) - LSTM128 - 8, memiliki hasil performa yang lebih unggul. Model ini menghasilkan loss 0.1704 loss dan 94.18% untuk training, 0.1752 loss dan 94.34% untuk validation, serta 0.1883 loss dan 93.98% untuk testing. Penambahan dropout layer berpengaruh terhadap peningkatan performa model. Dengan demikian, LSTM dapat tepat diterapkan dengan baik untuk data sequential seperti audio.