Laporan Proyek
"Voice Activity Detection Using Support Vector Machine"

Anggota Kelompok:
- Benny Strata Wijaya - 2540128682
- Bryan Fendis - 2502003112
- Nisrina Marwah -2502011354

1. Judul
   Voice Activity Detection on Audio with Support Vector Machine

2. Deskripsi Proyek
   Proyek yang akan dilakukan adalah mengenai *voice activity detection* (VAD) yang akan memeriksa apakan terdapat suara manusia/pembicaraan pada suatu audio. Hal tersebut dilakukan dengan membuat model VAD dengan input berupa dataset audio dan output adalah bagian dari audio yang memiliki *speech*. VAD tersebut akan menggunakan *support* vector *machine* (SVM) untuk pengklasifikasian dan *log filterbank energy* dan *log energy* sebagai fiturnya. Audio akan dibagi menjadi beberapa frame dan berdasarkan fitur yang diperoleh, frame tersebut akan diklasifikasi sebagai *speech* atau *non-speech* yang dapat diketahui kebenarannya dari anotasi atau label audio pada data tersebut. Kemudian, hal tersebut akan diaplikasikan untuk pelatihan model sehingga model dapat mempelajari pola dan metrik evaluasi. Metrik evaluasi seperti, akurasi akan digunakan untuk mengukur performa dari model dalam mendeteksi *speech* atau *voice acitivty* pada audio.

3. Dataset dan penjelasannya
   Dataset yang akan digunakan pada proyek adalah Speech activity detection datasets dari Kaggle. Dataset tersebut memiliki 719 file audio beserta dengan anotasinya yang bersumber dari TIMIT, PTDB-TUG, dan Noizeus. Data audio terdiri atas suara lak-laki dan perempuan dari TIMIT dan PTDB, kemudian terdapat file Noizeus yang merupakan data audio dengan noises yang terbagi menjadi 7 kategori yaitu, ocehan, mobil, tidak ada noises, restoran, stasiun, jalan, dan kereta.
   Link: https://www.kaggle.com/datasets/lazyrac00n/speech-activity-detection-datasets

4. Metode
   - Persiapan data
     Mengimport data berupa audio dan anotasi

   - Ekstraksi fitur
     Mengekstraksi fitur dari audio tersebut dimana fitur yang akan digunakan adalah log filterbank energies dan log energy. Hal tersebut dilakukan dengan mengekstraksi fitur energi log filterbank dan energi log dari setiap file audio, lalu membagi label yang sesuai ke frame. Pembagian frame dilakukan setiap 30 milisekon dari durasi audio. Kemudian, menentukan label ke setiap frame dengan *speech* atau *non* speech berdasarkan threshold. Fitur dan label tersebut kemudian disimpan dalam variabel dataset untuk digunakan pada processing dan model

   - Penentuan variabel X dan y

Variabel X akan menyimpan fitur dari variabel dataset, sedangkan variabel y akan menyimpan label.

- Pembagian data menjadi data latih dan test
  Data akan dibagi menjadi data latih:test sebesar 70:30
- Preprocessing data
  Preprocessing data digunakan dengan menggunakan standard scaler dan dilakukan kepada data latih dan test.

- Pelatihan model
  Dengan menggunakan model SVM dari sklearn, pelatihan dilakukan dengan menggunakan data latih. Model akan mengklasifikasi setiap frame pada audio dari data dan mengklasifikasikan apakah masing-masing frame merupakan *speech* atau *non-speech*. Lalu, anotasi akan menjadi referensi label dalam mempelajari pola dan karakteristik dari bagian *speech* dan *non-speech*.

- Testing
  Melakukan prediksi menggunakan data test pada model yang sudah dilatih.

- Evaluasi Model
  Melakukan evaluasi model dari hasil prediksi dengan metrik evaluasi akurasi, presisi, recall, dan f1-score.

5. Fitur
   Fitur yang akan digunakan sebagai input model adalah log filterbank energy dan log energy agar model dapat mempelajari pola dan karakteristik dari *speech*.

6. Model
   Dengan menggunakan model Support Vector Machine (SVM) untuk melatih klasifikasi antara *speech* dan *non-speech* dari fitur log filterbank energy dan log energy.

7. Rencana Evaluasi Metrik
   Metrik yang akan digunakan untuk mengevaluasi performa adalah akurasi, precision, recall, dan F1-score, dan confusion matrix.

8. Hasil
   Hasil model dinilai berdasarkan evaluasi yang dilakukan pada testing dengan nilai yang evaluasi metrik pada Tabel 1 dan nilai confusion matrix pada Tabel 2 atau “Fig. 1”.

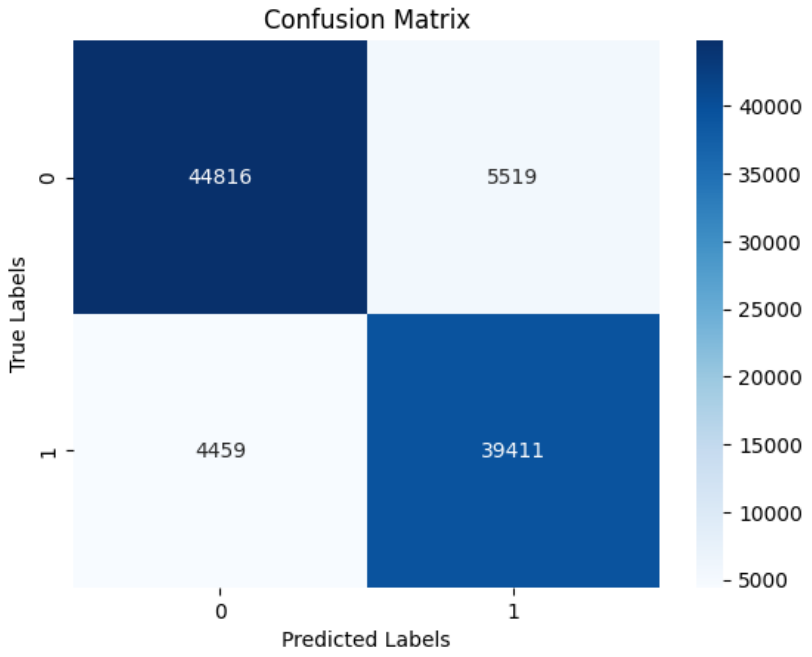
Tabel 1. Hasil evaluasi metrics dari data test

Evaluation Metrics	**Value (%)**
Accuracy	89.41
Precision	87.72
Recall	89.84
F1 score	88.76

Tabel 2. Hasil confusion matrix dari data test

Confusion Matrix			
True Label	0	44816	5519
	1	4459	39411
		0	1
		Predicted Label	

Fig. 1. Gambar confusion matrix dari data test



Dari hasil yang diperoleh dapat dikatakan bahwa model memperoleh performa yang baik dalam mendeteksi bagian *speech* dan bukan pada audio. Hal tersebut dapa dilihat dari akurasi dengan nilai 89.41% menunjukan performa yang tinggi dari model dalam mengklasifikasi dengan benar bagian *speech* dan *non-speech*. Presisi dengan nilai 87.72% menunjukan akurasi dalam pengklasifikasian bagian *speech* dan nilai recall 89.84% menunjukan kesuksesan model dalam mengidentifikasi bagian *speech* yang sebenarnya pada dataset. Kemudian, nilai F1-score 87.76% menunjukan performa model yang seimbang dari presisi dan recall.

9. Konklusi

Makalah ini fokus dalam implementasi VAD menggunakan support vector machine sebagai algoritma klasifikasi. Penelitian dilakukan dengan dataset yang diambil dari kaggle. Metode ini dipilih karena kemampuan untuk menangani blablablablablabla. Dengan mengekstrak fitur dari sinyal audio dan melakukan training pada svm model, untuk mendeteksi *speech* segments pada audio. Setelah dilakukan eksperimen dapat disimpulkan bahwa metode SVM ini cukup akurat, mendapatkan akurasi yang cukup tinggi 89,41%. Jika dibandingkan dengan BiLSTM yang memiliki akurasi 88.35% metode SVM memperoleh tingkat akurasi yang sedikit lebih tinggi. Voice activity detection ini dapat diimplementasikan di berbagai hal dalam aplikasi

multimedia seperti video conferencing, streaming service, and voice-controlled device. Memungkinkan untuk audio processing dan resource allocation yang efisien menjadikan fokus ke *speech* segmentnya saja. Penelitian dan eksperimen lebih lanjut dapat menjelajahi efektivitas dan performa VAD berbasis SVM pada berbagai skenario dan aplikasi dunia nyata, yang berpotensi menghasilkan kemajuan dalam teknologi pemrosesan suara dan komunikasi.

Link code:
https://colab.research.google.com/drive/1sxGLynabVNaLId-CNycX7Z6R0TP0JIxQ?usp=sharing
Link ppt:
https://www.canva.com/design/DAFltipfGD0/z-xLszFuWDq4MxTWFMnwzg/edit?utm_content=DAFltipfGD0&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton