

## **Ringkasan Kuliah Tamu Dr.Ir.Kuntjoro Pinardi**

### **Automatic Speech Recognition (ASR)**

ASR merupakan pengembangan dari teknologi Natural Language Processing (NLP) untuk mendeteksi suara manusia menjadi bahasa yang dimengerti oleh computer. ASR memiliki beberapa fungsi yang berguna untuk manusia yaitu

- Speech to text (pengubahan suara menjadi teks)
- Speaker ID (kemampuan mengenali identitas pembicara)
- Paralinguistic Aspect (kemampuan mengenali suara dengan intonasi, timing, kualitas audio)

ASR membutuhkan proses untuk mengenali suara manusia. Suara manusia setiap orang adalah hal yang unik karena setiap manusia perbedaan pita suara sehingga perlu analisis karakter suara untuk suku kata tertentu dengan arti yang sama tetapi dengan kualitas suara yang berbeda. Proses pengenalan suara berasal dari pemecahan kalimat menjadi suku kata kemudian dipecah menjadi pengucapan dan dideteksi spektrum frekuensi.

Dari spektrum frekuensi suara yang dideteksi perlu adanya pemisahan frame width sebesar 25 ms dan melakukan komputasi setiap frame shift 10 ms. Proses pengolahan spektrum frekuensi ini menggunakan diskrit fourier transform (DFT) dan menggunakan teorema bayes. Proses pengolahan suara ini tidak bisa 100 persen akurat karena berdasarkan teorema bayes yang menggunakan probabilitas sehingga computer perlu memprediksi kata-kata selanjutnya berdasarkan kata-kata sebelumnya yang sudah diucapkan dan dideteksi microphone.

Dalam implementasi ASR tentu ada beberapa kendala yang terus muncul seiring dengan perkembangan zaman.

- Keterbatasan sumber daya untuk ASR seperti device ASR yang semakin kecil menimbulkan masalah kurangnya memory dan storage.
- Variasi suara yang dideteksi ASR seperti suku kata yang sama diucapkan oleh intonasi yang berbeda atau dicampur dengan noise.
- Ragam Bahasa yang dideteksi ASR karena perbedaan pengucapan untuk masing-masing suku kata dalam Bahasa dan perlu adanya corpus (database pengucapan suku kata dalam Bahasa tertentu) untuk proses pembelajaran ASR dalam pendeteksian Bahasa

Proses ASR dalam mendeteksi suara manusia menggunakan 2 pendekatan yaitu acoustic model dan language model. Dalam mengembangkan acoustic model dan language model memerlukan training data menggunakan corpus Bahasa yang diambil dari sampling masyarakat.