



# RUNAS VALODAS NOTEIKŠANA

## Kas tas ir?

Runas valodas noteikšanas uzdevums ir noteikt audio ieraksta valodu.



## Mērķis

Izpētīt runājamās valodas noteikšanas metodes, koncentrējoties uz 3 Latvijā runājamām valodām: latviešu, krievu, angļu.

## Kopsavilkums

Tika aplūkota literatūra par 3 valodas noteikšanas metodēm: mašīnmācīšanās, dziļa mācīšanās un runas dziļas reprezentācijas. Tika secināts:

- Mašīnmācīšanās ir nepieciešamas labas gan lingvistiskas, gan mašīnmācīšanās zināšanas, bet precizitāte šai metodei atpaliek no jaunākām.
- Dziļas mācīšanās metodē ir nepieciešamas lielas apmācību datu kopas, lai sasniegtu labu veikspēju.
- Runas dziļas reprezentācijas labi darbojas runas atpazīšanas uzdevumos, īpaši zemu resursu gadījumos. Neprasa eksperta līmeņa zināšanas.

## Piedāvātais risinājums

Turpmākā darba ietvaros plānoti vairāki eksperimenti. Plānots izmantot runas dziļas reprezentācijas metodi un wav2vec algoritmu. Plānots iegūt 6 stundu datu kopas katrai valodai.

