**Thành viên:**

Nguyễn Thành Nam – 17021305

Nguyễn Trung Hiếu – 17021246

**Báo cáo bài tập môn xử lý tiếng nói**

**Dataset**: ~100 file wav cho các từ “bệnh nhân”, “chúng ta”, “có thể”, “người”, “Việt Nam”. Chia tập dữ liệu thành tập train và test theo tỉ lệ 70-30.

**Thư viện sử dụng**: MultinomialHMM của hmmlearn.

**Kmeans**: 14 cluster cho kết quả tốt nhất.

**Các mô hình hmm sử dụng kĩ thuật subdividing phones(3 trạng thái cho mỗi âm vị) và 1 skip connection**:

* “bệnh nhân”: có tổng 6 âm vị không tính dấu cách nên sử dụng mô hình có 18 trạng thái:

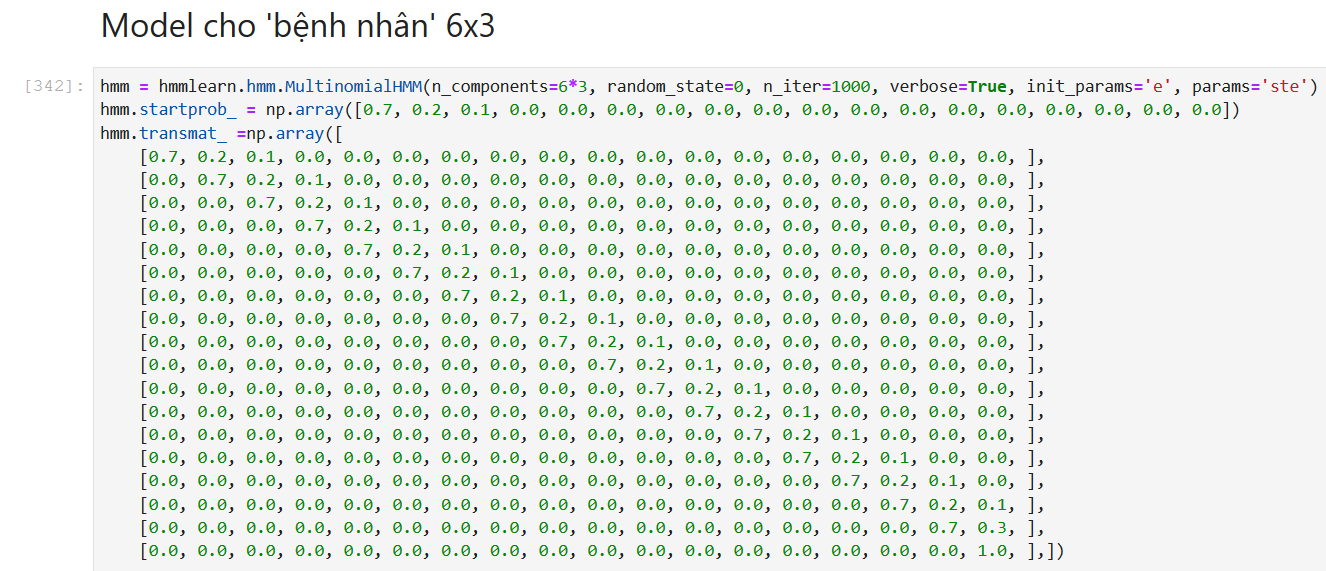


Figure : khởi tạo cho hmm "bệnh nhân".

* “chúng ta”: có 5 âm vị không trính dấu cách, mô hình có 15 trạng thái:

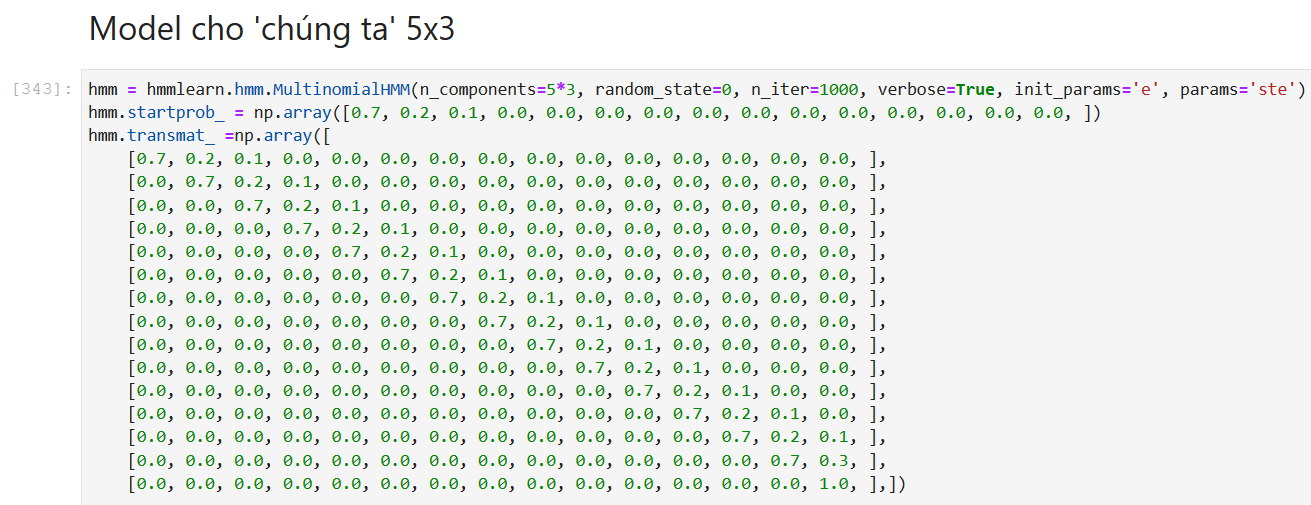


Figure : khởi tạo cho hmm "chúng ta".

* “có thể”: có 4 âm vị không tính dấy cách, mô hình có 12 trạng thái:

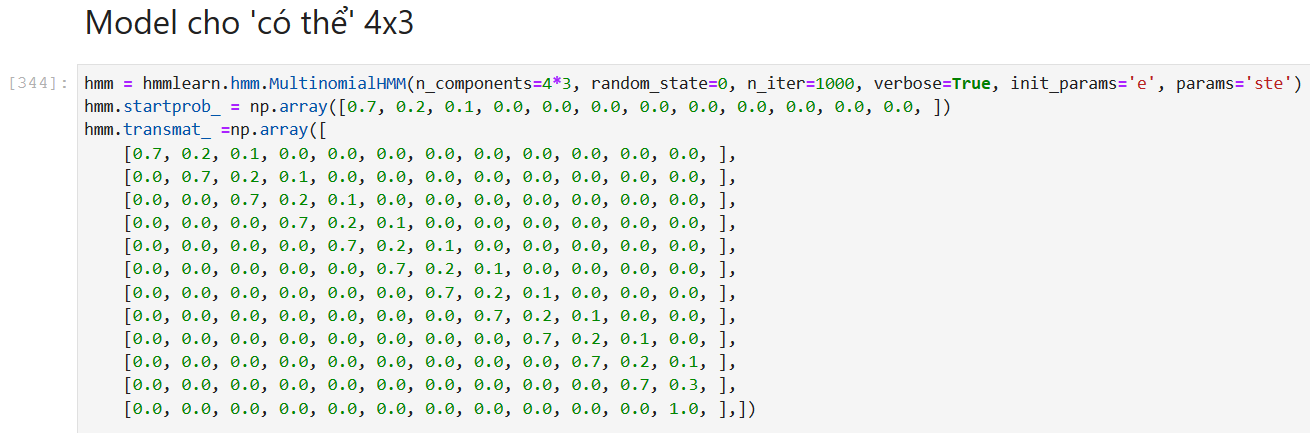


Figure : khởi tạo cho hmm "có thể".

* “người”: có 3 âm vị, mô hình có 9 trạng thái:

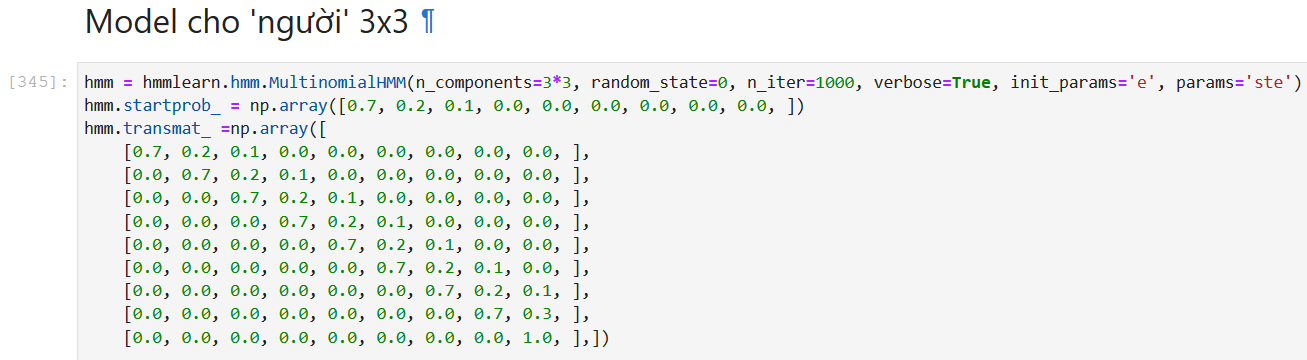


Figure : khởi tạo cho hmm "người".

* “Việt Nam”: có 6 âm vị không tính dấu cách, mô hình có 18 trạng thái:

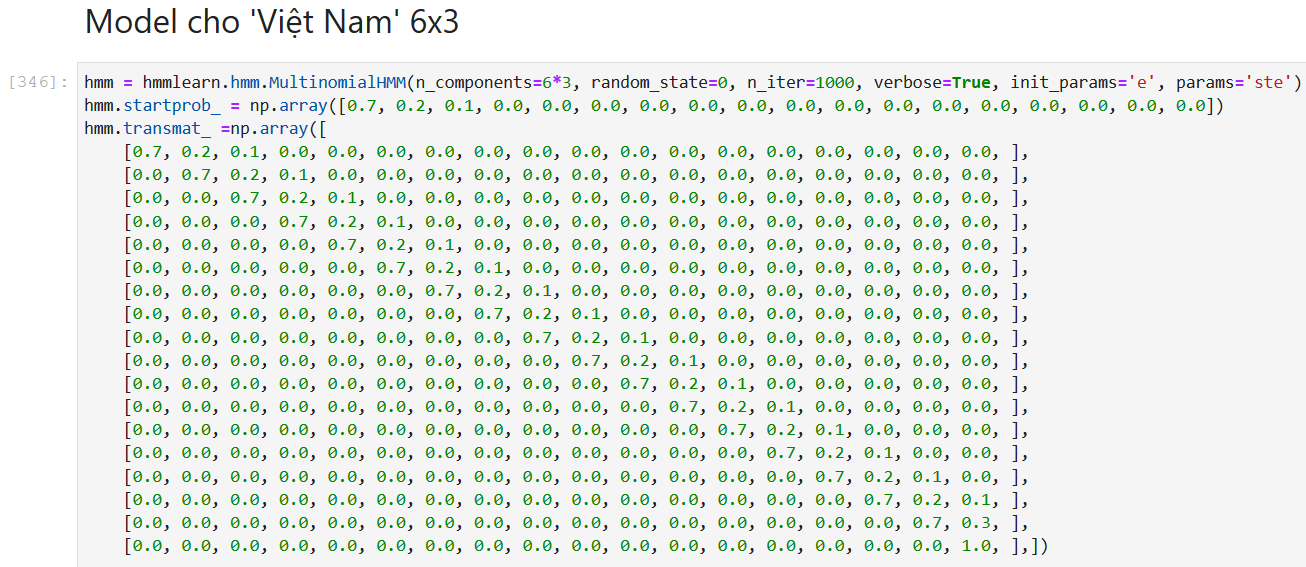


Figure : khởi tạo cho hmm "Việt Nam".

**Kết quả trên tập test:**

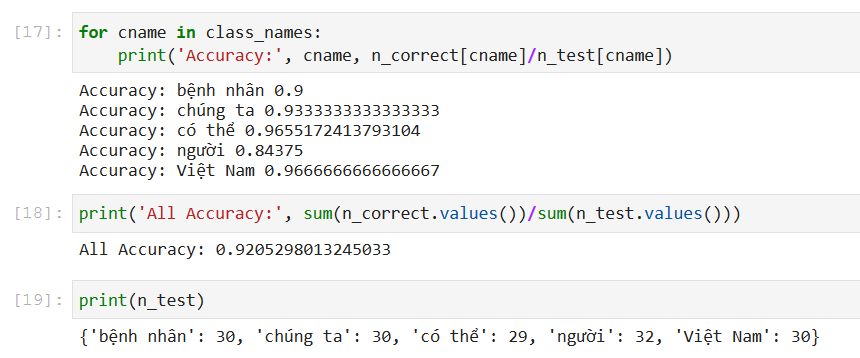


Figure : độ chính xác trên tập test.

**Kết quả trên tập các file wav tự phát âm các từ qua micro:**

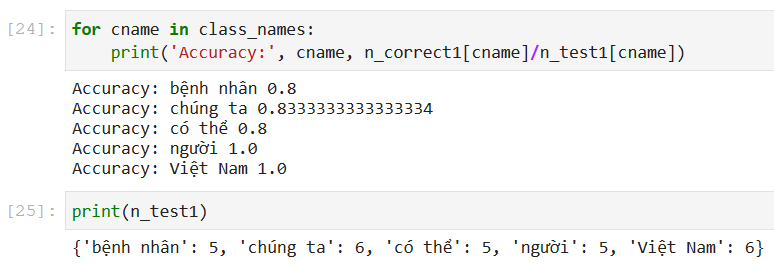


Figure : độ chính xác trên tập tự phát âm qua micro.

**Code & bộ tham số của các mô hình**: https://github.com/not-nam-or-am-i/voice-processing-hmm/blob/master/hmms.ipynb