Personally Identifiable Information Protection Process

Preparation data → **Speech to Text** → **PII Detection** → **Speech Masking**

Preparation data

Creating conversations between customer and call center in text format

↓

Creating the conversations (audio files type) by Siri's voice

Speech to Text

Converting audio files type for speech to text model

↓

Using Google Cloud Speech to convert speech to text

↓

Storing texts which are converted and audio time in dictionary format

PII Detection

NER Tagger Process

Using Stanford NER Tagger for named entities recognizer

Using NLTK for named entities recognizer

Using spaCy for named entities recognizer

↓

Comparing between 3 models for appropriately named entities recognizer

↓

Creating rule-based conditions for detecting the numbers which are personally identifiable informations

↓

Storing words which are named entities in the data frame

Speech Masking

Matching named entities and words of the conversations

↓

Matching audio time of words and then masking the speech in that time