Sign Language Translation

Aziza M
Saumya S
Srik G
Vinit T

Agenda

Problem

The Idea
The Problem Statement

Data

The Sourcing
The Processing

Models and Results

The Models Tried
The Results

What's Next

The Integration
The Improvements

What is Sign Language

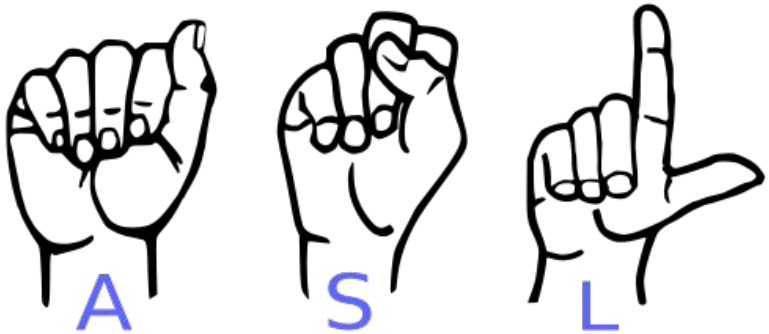
Specifically, American Sign Language

Characteristics

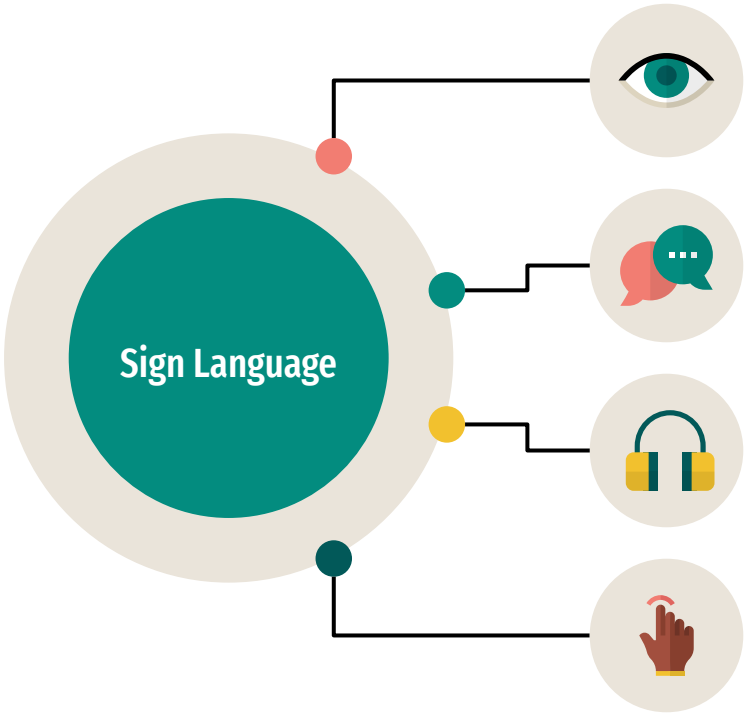
1. **Visual**
   Movement of the face, torso, and hands

2. **American “and Global”**
   USA, Anglophone Canada. Parts of West Africa and South-East Asia

3. **Loan words**
   English loanwords are often borrowed through finger spellings



Why Deal With ASL?



Common

0.25 million - 0.50 million ASL users in USA

Complex

Difficult for untrained spoken language users

Impactful

Employment gap of 22.5% for hearing impaired

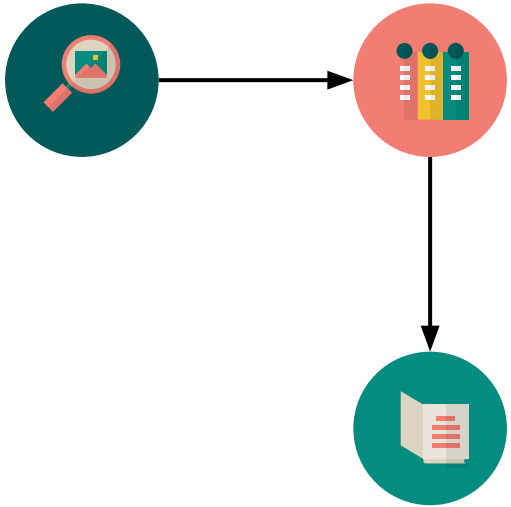
Mental Health

0.6 million deaf individuals require mental health attention

Problem Statement

Hand Gesture Images

Images of ASL gestures



Columnar Data

Table with Columns
Extracted from the Images

Spoken Language

Predicting Word for a given Image

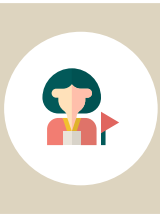
All Things, Data, Considered

Source

Kaggle Sign Language Recognition

Observations

10000 Rows
(5146 ASL + 4853 Non-ASL)



Dimensions

428 Columns of Hand positions, directions etc.

Labels

18 words + 1 <UNK>

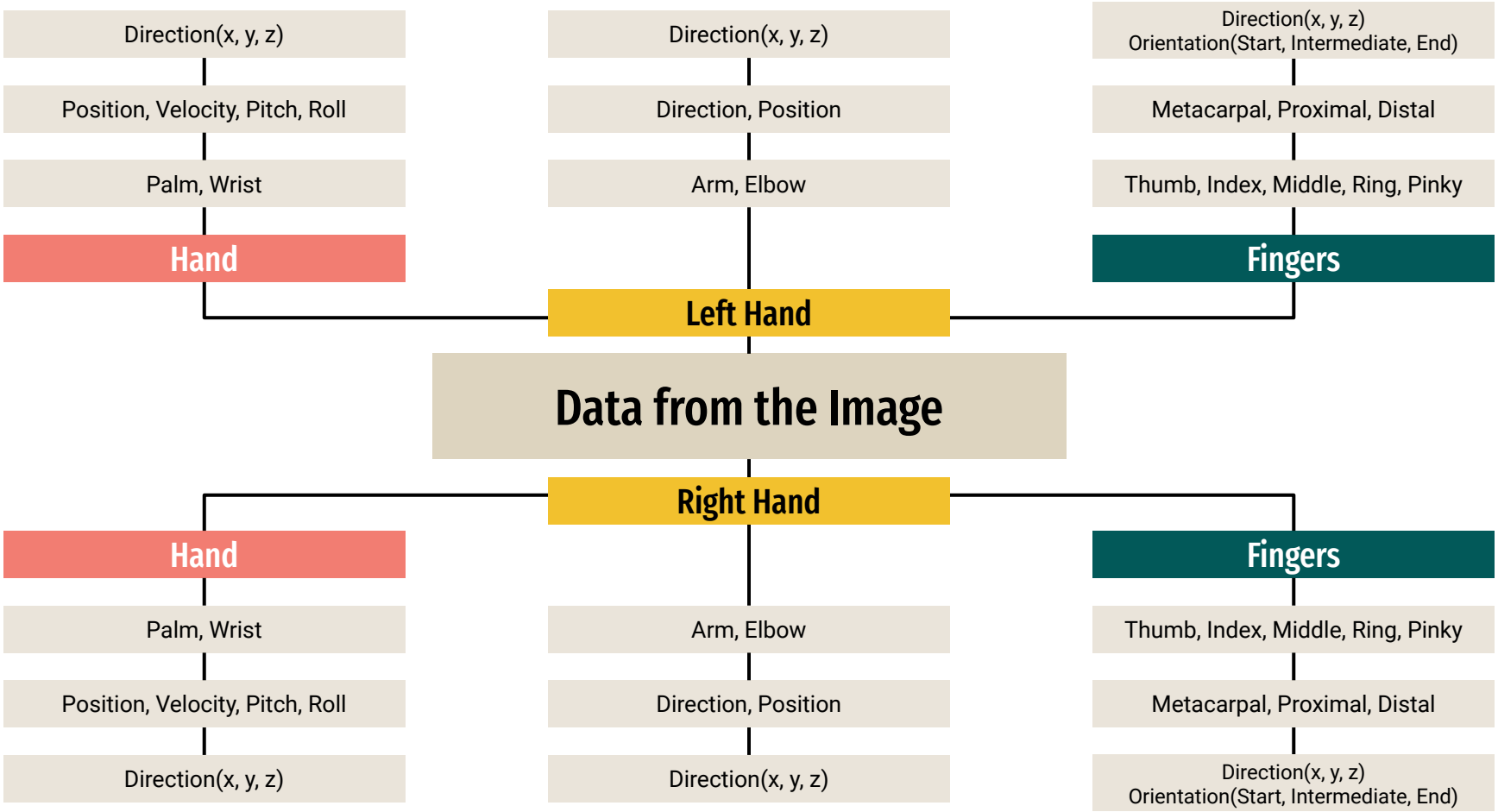
EDA

Multicollinearity

Processing

Normalization

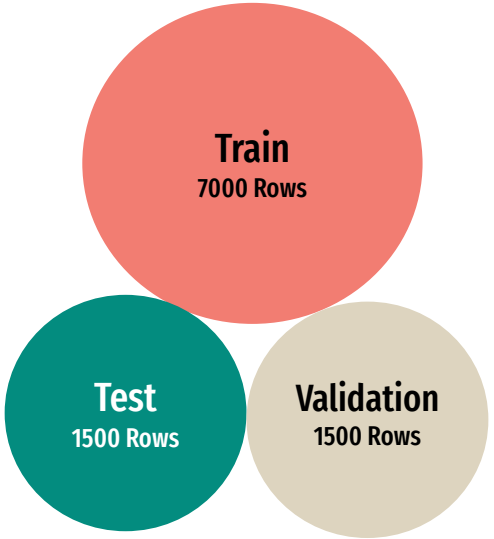
PCA


Direction(x, y, z)
Position, Velocity, Pitch, Roll
Palm, Wrist
Hand
Direction(x, y, z)
Direction, Position
Arm, Elbow
Direction(x, y, z)
Orientation(Start, Intermediate, End)
Metacarpal, Proximal, Distal
Thumb, Index, Middle, Ring, Pinky
Fingers
Left Hand
Data from the Image
Right Hand
Hand
Palm, Wrist
Position, Velocity, Pitch, Roll
Direction(x, y, z)
Arm, Elbow
Direction, Position
Direction(x, y, z)
Fingers
Thumb, Index, Middle, Ring, Pinky
Metacarpal, Proximal, Distal
Direction(x, y, z)
Orientation(Start, Intermediate, End)

Labels and Stratification

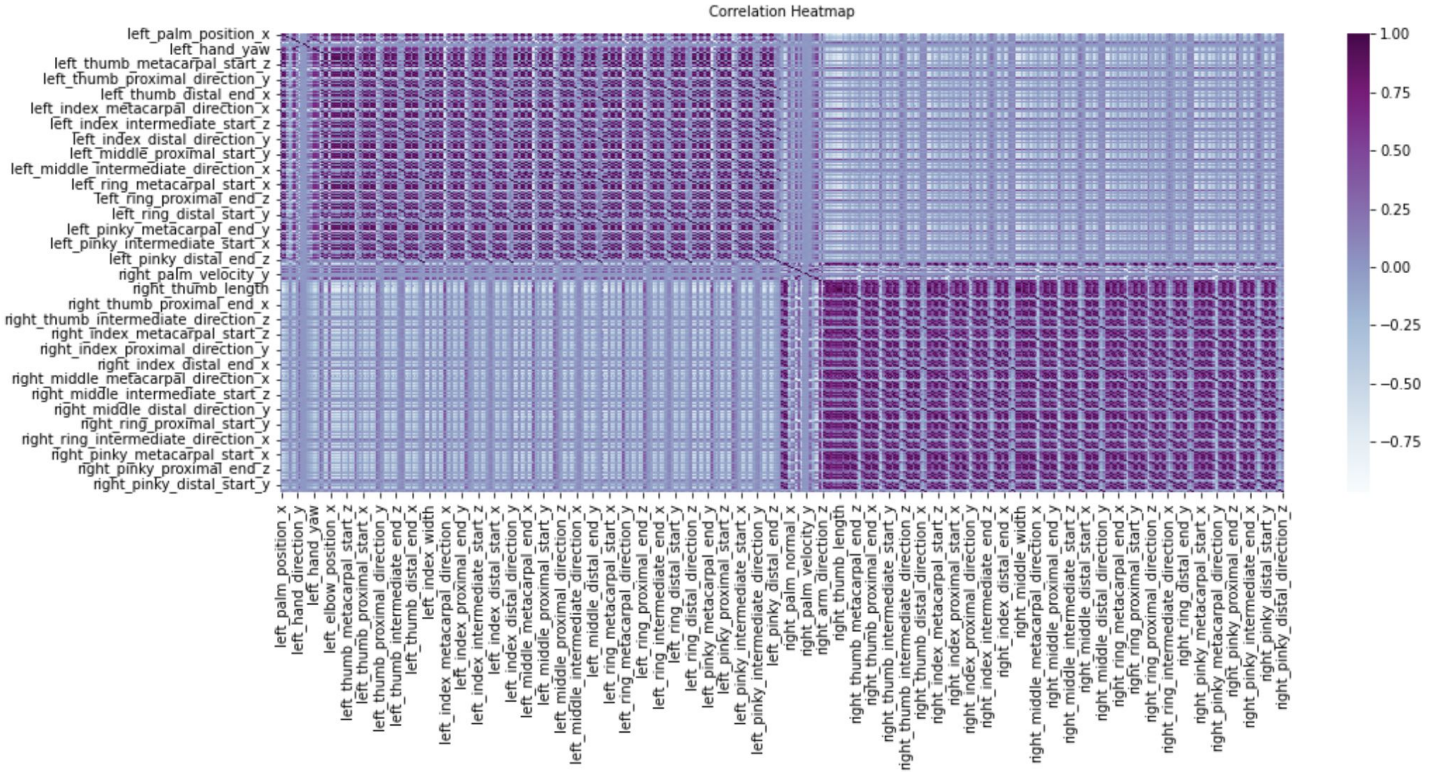
Labels	Frequencies
<UNK>	4854
AIRPORT	381
BAD	271
BUS	346
CAR	300
DRINK	394
EXCUSE_ME	291
FOOD	312
GOOD	256
HELLO	228
I	279
NAME	250
PLANE	200
RESTAURANT	260
SORRY	234
TAXI	250
THANKS	253
TIME	299
YOU	341
Grand Total	**9999**

Frequencies of the 19 Labels



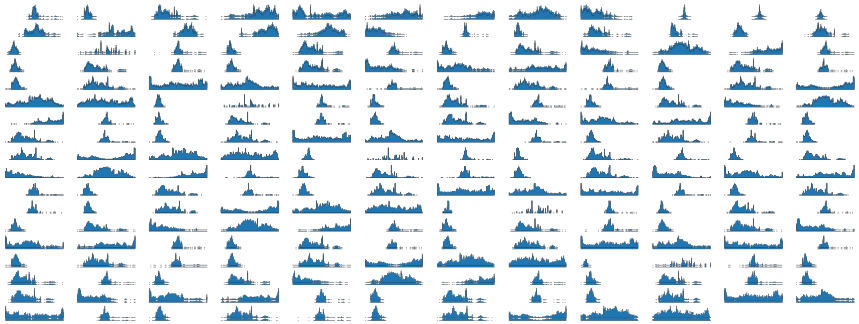
Stratified Sampled Train, Test and Validation

MultiCollinearity



Scaling Before PCA - Normalization

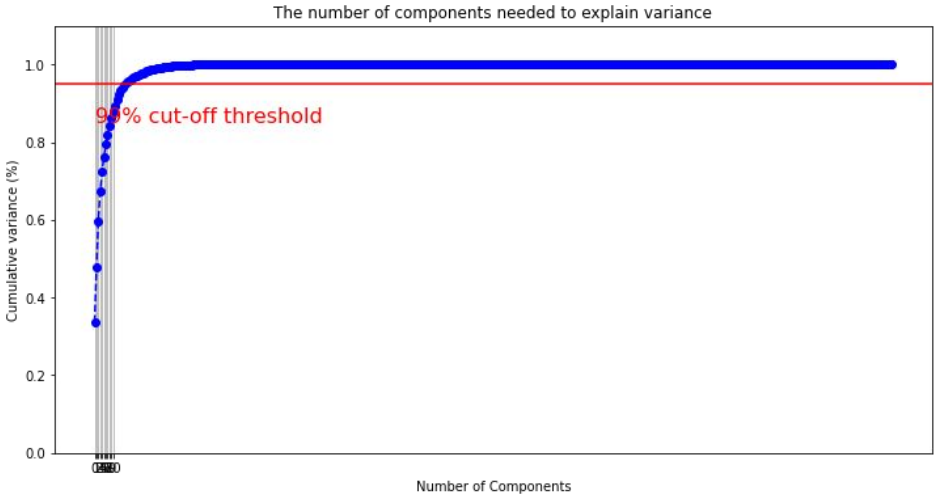
Pre-Normalization



Post-Normalization

No Change in the Distribution of the columns

PCA



Details

Principal Components
We had 85 PCAs
ie, sqrt(7000), our train size

1

Explaining Variance
We ended up with 18 PCs based on the variance explained

2

On Test and Validation
We performed PCA on the Train and fit it to Test and Validation Datasets

3

Accuracy & F1-Score Calculation for Multi-Class Classification

Using an Example

		Predictions			
		0	**1**	**2**	***Total***
Actuals	**0**	20	10	1	*31*
	1	5	30	1	*36*
	2	2	6	5	*13*
	Total	*27*	*46*	*7*	*80*

Accuracy

= 20+30+5 / 80

= 55/80

= 0.5625

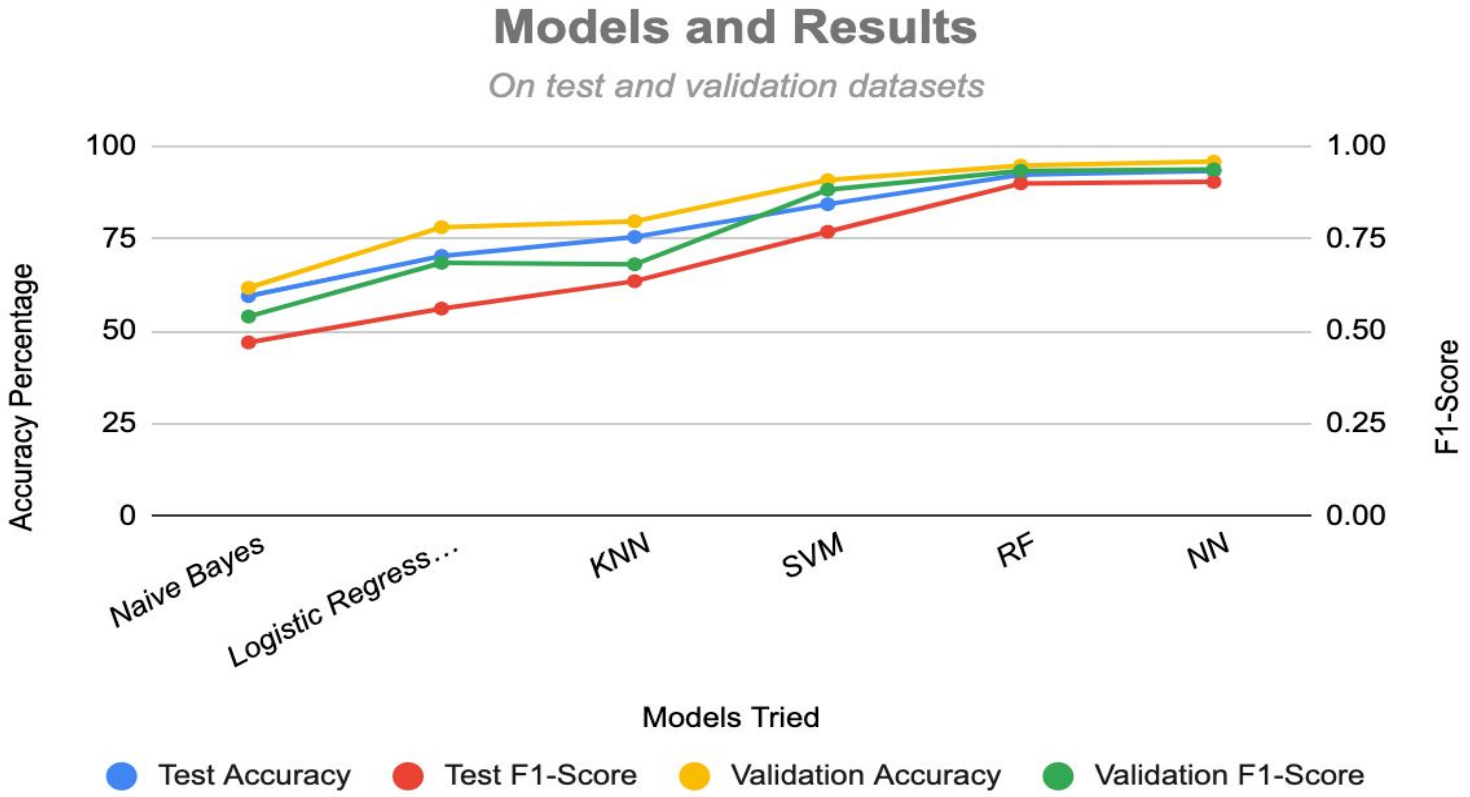
F1-Score

Label	Precision	Recall	F1-Score
0	20/27 = 0.7407	20/31 = 0.6452	0.6897
1	30/46 = 0.6522	30/36 = 0.8333	0.7317
2	5/7 = 0.7143	5/13 = 0.3846	0.5000

= 0.6897 + 0.7317 + 0.5000 / 3

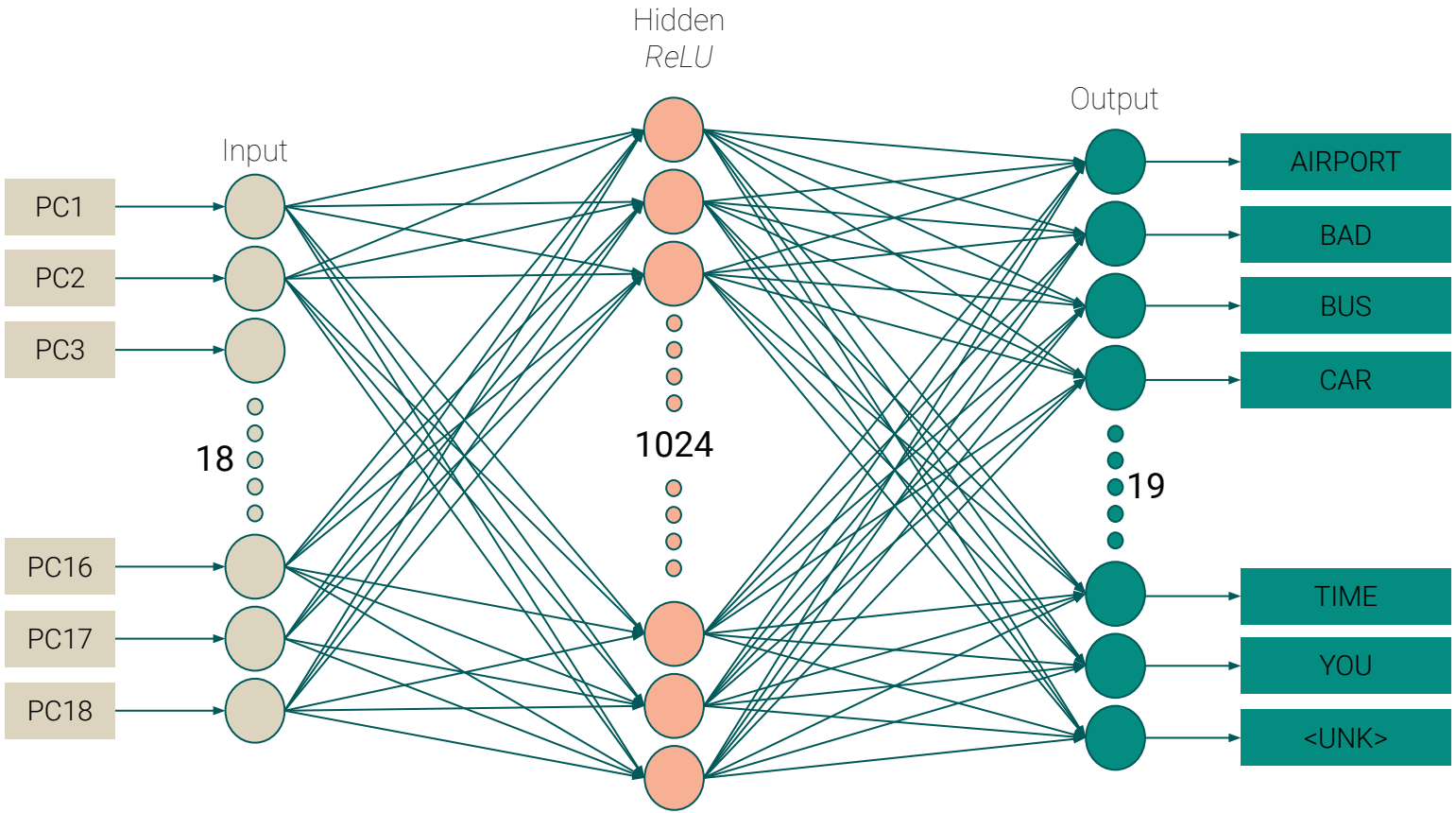
= 1.9214/3

= 0.6405



Colab Worksheet with the processing and models

NN Model Architecture



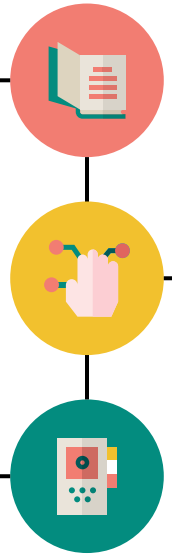
NN Confusion Matrix

		PREDICTS																	
		AIRPORT	BAD	BUS	CAR	DRINK	EXCUSE_ME	FOOD	GOOD	HELLO	I	NAME	PLANE	RESTAURANT	SORRY	TAXI	THANKS	TIME	YOU	<UNK>
ACTUALS	AIRPORT	48	0	0	0	0	0	1	0	0	0	0	1	2	0	1	0	1	1	2
	BAD	0	38	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	BUS	0	0	48	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	CAR	0	0	0	41	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0
	DRINK	0	1	0	0	54	0	0	0	2	0	0	1	0	0	0	0	0	0	1
	EXCUSE_ME	0	0	0	1	0	42	0	0	0	0	0	0	0	0	0	0	0	0	1
	FOOD	1	0	0	0	0	0	45	1	0	0	0	0	0	0	0	0	0	0	0
	GOOD	0	0	0	0	0	2	0	36	0	0	0	0	0	0	0	0	0	0	0
	HELLO	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	1
	I	0	0	0	0	2	0	0	0	0	37	0	0	0	0	1	0	0	0	2
	NAME	1	0	0	0	0	0	1	0	0	0	34	0	2	0	0	0	0	0	0
	PLANE	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	1	0	0
	RESTAURANT	0	0	0	0	0	0	1	0	0	0	0	0	34	1	0	0	0	0	3
	SORRY	0	0	0	0	2	0	0	0	3	1	0	0	4	25	0	0	0	0	0
	TAXI	1	0	0	0	0	0	0	0	1	1	0	0	0	0	31	0	0	0	4
	THANKS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	37	0	0	1
	TIME	0	0	0	0	0	0	0	0	0	0	2	1	0	0	3	0	38	0	1
	YOU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	1
	<UNK>	3	0	0	2	0	0	2	0	15	0	0	0	2	0	2	0	0	2	700

Road Ahead

Sentence Based System

- Currently, Image -> Word
- Can be, Images -> Sentences



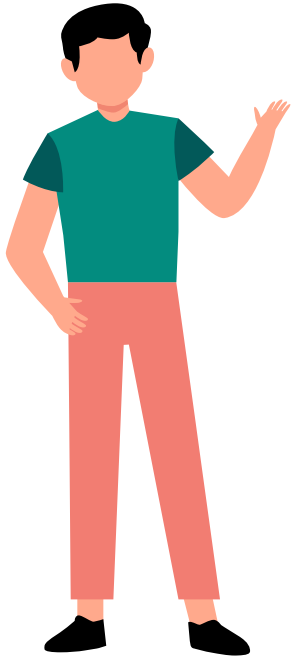
Sensor Integration

- Currently, implementation is only model
- Can be, integrated with capturing sensor for seamless translation

Image -> Audio

- Currently, no audio
- Can use off-the-shelf speech synthesizers to translate in real-time

Thank You



And after we are done with the presentation.....