MIXING NATURAL AND SYNTHETIC (MALE) DATA

DATASET

```
Total number of utterances : 25
-- Check Last Transaction : 3
-- CheckBalance : 9
-- Withdraw Money : 1
-- Send Money : 11
-- Deposit : 1
```

Only Natural Data, for both Train and Test
Best results (5,1,1)

```
For Ngram : 1
-- For threshold 0 : 0.5652173913043478
-- For threshold 1 : 0.5217391304347826
-- For threshold 2 : 0.5652173913043478
-- For threshold 3 : 0.4782608695652174
-- For threshold 4 : 0.6086956521739131
-- For threshold 5 : 0.6956521739130435
-- For threshold 6 : 0.6521739130434783

For Ngram : 2
-- For threshold 0 : 0.4782608695652174
-- For threshold 1 : 0.6086956521739131
-- For threshold 2 : 0.4782608695652174
-- For threshold 3 : 0.4782608695652174
-- For threshold 4 : 0.21739130434782608
-- For threshold 5 : 0.13043478260869565
-- For threshold 6 : 0.08695652173913043

For Ngram : 3
-- For threshold 0 : 0.17391304347826086
-- For threshold 1 : 0.30434782608695654
-- For threshold 2 : 0.08695652173913043
-- For threshold 3 : 0.08695652173913043
-- For threshold 4 : 0.08695652173913043
-- For threshold 5 : 0.08695652173913043
-- For threshold 6 : 0.08695652173913043
```

```
Uni- Accuracy: 0.6956521739130435
Combo Accuracy: 0.8260869565217391
```

Only Synthetic Data, for both Train and Test
Best Result (2, 4, 2)

```
For Ngram : 1
-- For threshold 0 : 0.782608695652174
-- For threshold 1 : 0.6521739130434783
-- For threshold 2 : 0.782608695652174
-- For threshold 3 : 0.7391304347826086
-- For threshold 4 : 0.782608695652174
-- For threshold 5 : 0.7391304347826086
-- For threshold 6 : 0.6956521739130435

For Ngram : 2
-- For threshold 0 : 0.782608695652174
-- For threshold 1 : 0.782608695652174
-- For threshold 2 : 0.8260869565217391
-- For threshold 3 : 0.9130434782608695
-- For threshold 4 : 0.9565217391304348
-- For threshold 5 : 0.8695652173913043
-- For threshold 6 : 0.6521739130434783

For Ngram : 3
-- For threshold 0 : 0.6956521739130435
-- For threshold 1 : 0.6521739130434783
-- For threshold 2 : 0.7391304347826086
-- For threshold 3 : 0.6956521739130435
-- For threshold 4 : 0.6086956521739131
-- For threshold 5 : 0.30434782608695654
-- For threshold 6 : 0.13043478260869565
```

```
Uni- Accuracy: 0.782608695652174
Combo Accuracy: 0.8695652173913043
```

MIXED DATA : SYNTH + NAT , for both train/test
Best Result(2, 1, 1)

```
Ngram : 1
--Threshold 0 : Accuracy =  0.6304347826086957
--Threshold 1 : Accuracy =  0.6521739130434783
--Threshold 2 : Accuracy =  0.717391304347826
--Threshold 3 : Accuracy =  0.717391304347826
--Threshold 4 : Accuracy =  0.5

Ngram : 2
--Threshold 0 : Accuracy =  0.7826086956521740
--Threshold 1 : Accuracy =  0.8478260869565217
--Threshold 2 : Accuracy =  0.8043478260869565
--Threshold 3 : Accuracy =  0.7391304347826086
--Threshold 4 : Accuracy =  0.7391304347826086

Ngram : 3
--Threshold 0 : Accuracy =  0.5652173913043478
--Threshold 1 : Accuracy =  0.5869565217391305
--Threshold 2 : Accuracy =  0.4347826086956522
--Threshold 3 : Accuracy =  0.4347826086956522
--Threshold 4 : Accuracy =  0.3695652173913043
```

```
Uni- Accuracy: 0.717391304347826
Combo Accuracy: 0.8478260869565217
```

MIXED DATA : SYNTH + NAT , tested with NAT
Best Result(2, 1, 1)

```
Ngram : 1
--Threshold 0 : Accuracy =  0.3913043478260870
--Threshold 1 : Accuracy =  0.5217391304347826
--Threshold 2 : Accuracy =  0.6086956521739131
--Threshold 3 : Accuracy =  0.5652173913043478
--Threshold 4 : Accuracy =  0.3913043478260870

Ngram : 2
--Threshold 0 : Accuracy =  0.6086956521739131
--Threshold 1 : Accuracy =  0.7826086956521740
--Threshold 2 : Accuracy =  0.7391304347826086
--Threshold 3 : Accuracy =  0.6086956521739131
--Threshold 4 : Accuracy =  0.6086956521739131

Ngram : 3
--Threshold 0 : Accuracy =  0.30434782608695654
--Threshold 1 : Accuracy =  0.5217391304347826
--Threshold 2 : Accuracy =  0.21739130434782608
--Threshold 3 : Accuracy =  0.17391304347826086
--Threshold 4 : Accuracy =  0.17391304347826086
```

```
Uni- Accuracy: 0.6086956521739131
Combo Accuracy: 0.7826086956521740
```

MIXED DATA : SYNTH + NAT , tested with SYNTH
Best Result(0, 0/2, 0)

```
Ngram : 1
--Threshold 0 : Accuracy =  0.8695652173913043
--Threshold 1 : Accuracy =  0.8260869565217391
--Threshold 2 : Accuracy =  0.8260869565217391
--Threshold 3 : Accuracy =  0.8260869565217391
--Threshold 4 : Accuracy =  0.6521739130434783

Ngram : 2
--Threshold 0 : Accuracy =  0.9130434782608695
--Threshold 1 : Accuracy =  0.8695652173913043
--Threshold 2 : Accuracy =  0.9130434782608695
--Threshold 3 : Accuracy =  0.8695652173913043
--Threshold 4 : Accuracy =  0.8695652173913043

Ngram : 3
--Threshold 0 : Accuracy =  0.782608695652174
--Threshold 1 : Accuracy =  0.6956521739130435
--Threshold 2 : Accuracy =  0.7391304347826086
--Threshold 3 : Accuracy =  0.6956521739130435
--Threshold 4 : Accuracy =  0.6086956521739131
```

(0,0,0)

```
Uni- Accuracy: 0.8695652173913043
Combo Accuracy: 0.9130434782608695
```

(0,2,0)

```
Uni- Accuracy: 0.8695652173913043
Combo Accuracy: 0.9565217391304348
```

DATA OVERVIEW

DATA

```
Total number of utterances : 25
-- Check Last Transaction : 3
-- CheckBalance : 9
-- Withdraw Money : 1
-- Send Money : 11
-- Deposit : 1
```

Natural

```
Total phones in the dataset :  640
Average number of phones per dialogue : 25.6
Number of unique phones :  38
```

Synthesized Male

```
Total phones in the dataset :  727
Average number of phones per dialogue : 29.08
Number of unique phones :  37
```

Synthesized Female 1

```
Total phones in the dataset :  740
Average number of phones per dialogue : 29.6
Number of unique phones :  40
```

Synthesized Female 2

```
Total phones in the dataset :  673
Average number of phones per dialogue : 26.92
Number of unique phones :  36
```

Voice Samples?

COMMON PHONES WITH NATURAL SOURCE

DATA SOURCE	MEAN SIMILARITY	STD SIMILARITY
Natural + Male Synthesized	0.61	0.139
Natural + Female1 Synth	0.59	0.106
Natural + Female2 Synth	0.62	0.117

COMMON PHONES WITHIN SYNTHESIZED SOURCES

DATA SOURCE	MEAN SIMILARITY	STD SIMILARITY
Male Synth + Female1 Synth	0.68	0.0851
Male Synth + Female2 Synth	0.76	0.098
Female1 Synth + Female2 Synth	0.71	0.088

Natural (5,1,1)

```
For Ngram : 1
-- For threshold 0 : 0.5652173913043478
-- For threshold 1 : 0.5217391304347826
-- For threshold 2 : 0.5652173913043478
-- For threshold 3 : 0.4782608695652174
-- For threshold 4 : 0.6086956521739131
-- For threshold 5 : 0.6956521739130435
-- For threshold 6 : 0.6521739130434783

For Ngram : 2
-- For threshold 0 : 0.4782608695652174
-- For threshold 1 : 0.6086956521739131
-- For threshold 2 : 0.4782608695652174
-- For threshold 3 : 0.4782608695652174
-- For threshold 4 : 0.21739130434782608
-- For threshold 5 : 0.13043478260869565
-- For threshold 6 : 0.08695652173913043

For Ngram : 3
-- For threshold 0 : 0.17391304347826086
-- For threshold 1 : 0.30434782608695654
-- For threshold 2 : 0.08695652173913043
-- For threshold 3 : 0.08695652173913043
-- For threshold 4 : 0.08695652173913043
-- For threshold 5 : 0.08695652173913043
-- For threshold 6 : 0.08695652173913043

Uni- Accuracy: 0.6956521739130435
Combo Accuracy: 0.8260869565217391
```

Synth Male (2,4,2)

```
For Ngram : 1
-- For threshold 0 : 0.782608695652174
-- For threshold 1 : 0.6521739130434783
-- For threshold 2 : 0.782608695652174
-- For threshold 3 : 0.7391304347826086
-- For threshold 4 : 0.782608695652174
-- For threshold 5 : 0.7391304347826086
-- For threshold 6 : 0.6956521739130435

For Ngram : 2
-- For threshold 0 : 0.782608695652174
-- For threshold 1 : 0.782608695652174
-- For threshold 2 : 0.8260869565217391
-- For threshold 3 : 0.9130434782608695
-- For threshold 4 : 0.9565217391304348
-- For threshold 5 : 0.8695652173913043
-- For threshold 6 : 0.6521739

For Ngram : 3
-- For threshold 0 : 0.6956521739130435
-- For threshold 1 : 0.6521739130434783
-- For threshold 2 : 0.7391304347826086
-- For threshold 3 : 0.6956521739130435
-- For threshold 4 : 0.6086956521739131
-- For threshold 5 : 0.30434782608695654
-- For threshold 6 : 0.13043478260869565

Uni- Accuracy: 0.782608695652174
Combo Accuracy: 0.8695652173913043
```

Synth Female1 (0,2,1)

```
For Ngram : 1
-- For threshold 0 : 0.782608695652174
-- For threshold 1 : 0.782608695652174
-- For threshold 2 : 0.6956521739130435
-- For threshold 3 : 0.6086956521739131
-- For threshold 4 : 0.6521739130434783
-- For threshold 5 : 0.6521739130434783
-- For threshold 6 : 0.7391304347826086

For Ngram : 2
-- For threshold 0 : 0.8260869565217391
-- For threshold 1 : 0.782608695652174
-- For threshold 2 : 0.8260869565217391
-- For threshold 3 : 0.7391304347826086
-- For threshold 4 : 0.5652173913043478
-- For threshold 5 : 0.391304347826087
-- For threshold 6 : 0.173913043

For Ngram : 3
-- For threshold 0 : 0.5652173913043478
-- For threshold 1 : 0.7391304347826086
-- For threshold 2 : 0.391304347826087
-- For threshold 3 : 0.17391304347826086
-- For threshold 4 : 0.08695652173913043
-- For threshold 5 : 0.08695652173913043
-- For threshold 6 : 0.08695652173913043

Uni- Accuracy: 0.782608695652174
Combo Accuracy: 0.8260869565217391
```

Synth Female2 (2,2,1)

```
For Ngram : 1
-- For threshold 0 : 0.6521739130434783
-- For threshold 1 : 0.6086956521739131
-- For threshold 2 : 0.6956521739130435
-- For threshold 3 : 0.6521739130434783
-- For threshold 4 : 0.6521739130434783
-- For threshold 5 : 0.6086956521739131
-- For threshold 6 : 0.5217391304347826

For Ngram : 2
-- For threshold 0 : 0.6086956521739131
-- For threshold 1 : 0.7391304347826086
-- For threshold 2 : 0.8260869565217391
-- For threshold 3 : 0.7391304347826086
-- For threshold 4 : 0.5652173913043478
-- For threshold 5 : 0.5217391304347826
-- For threshold 6 : 0.43478260869565216

For Ngram : 3
-- For threshold 0 : 0.5217391304347826
-- For threshold 1 : 0.6086956521739131
-- For threshold 2 : 0.5652173913043478
-- For threshold 3 : 0.4782608695652174
-- For threshold 4 : 0.21739130434782608
-- For threshold 5 : 0.08695652173913043
-- For threshold 6 : 0.08695652173913043

Uni- Accuracy: 0.6956521739130435
Combo Accuracy: 0.782608695652174
```

TRAINING ON SYNTHETIC DATA

TRAINED ON MALE SYNTHETIC DATA
TESTED ON NATURAL DATA

```
Ngram : 1
--Threshold 0 : Accuracy =  0.36
--Threshold 1 : Accuracy =  0.52
--Threshold 2 : Accuracy =  0.64
--Threshold 3 : Accuracy =  0.52
--Threshold 4 : Accuracy =  0.48

Ngram : 2
--Threshold 0 : Accuracy =  0.44
--Threshold 1 : Accuracy =  0.52
--Threshold 2 : Accuracy =  0.44
--Threshold 3 : Accuracy =  0.36
--Threshold 4 : Accuracy =  0.28

Ngram : 3
--Threshold 0 : Accuracy =  0.32
--Threshold 1 : Accuracy =  0.2
--Threshold 2 : Accuracy =  0.12
--Threshold 3 : Accuracy =  0.08
--Threshold 4 : Accuracy =  0.08
```

(2,1,0)

```
Uni- Accuracy: 0.64
Combo Accuracy: 0.64
```

TRAINED ON ALL FEMALE SYNTHETIC DATA
TESTED ON NATURAL DATA

```
Ngram : 1
--Threshold 0 : Accuracy =  0.4
--Threshold 1 : Accuracy =  0.48
--Threshold 2 : Accuracy =  0.48
--Threshold 3 : Accuracy =  0.48
--Threshold 4 : Accuracy =  0.44

Ngram : 2
--Threshold 0 : Accuracy =  0.4
--Threshold 1 : Accuracy =  0.72
--Threshold 2 : Accuracy =  0.6
--Threshold 3 : Accuracy =  0.6
--Threshold 4 : Accuracy =  0.52

Ngram : 3
--Threshold 0 : Accuracy =  0.28
--Threshold 1 : Accuracy =  0.16
--Threshold 2 : Accuracy =  0.12
--Threshold 3 : Accuracy =  0.12
--Threshold 4 : Accuracy =  0.04
```

(2,1,0)

```
Uni- Accuracy: 0.48
Combo Accuracy: 0.72
```

TRAINED ON ALL SYNTHETIC DATA
TESTED ON NATURAL DATA

```
Ngram : 1
--Threshold 0 : Accuracy =  0.4
--Threshold 1 : Accuracy =  0.44
--Threshold 2 : Accuracy =  0.56
--Threshold 3 : Accuracy =  0.56
--Threshold 4 : Accuracy =  0.48

Ngram : 2
--Threshold 0 : Accuracy =  0.48
--Threshold 1 : Accuracy =  0.72
--Threshold 2 : Accuracy =  0.64
--Threshold 3 : Accuracy =  0.6
--Threshold 4 : Accuracy =  0.6

Ngram : 3
--Threshold 0 : Accuracy =  0.4
--Threshold 1 : Accuracy =  0.32
--Threshold 2 : Accuracy =  0.24
--Threshold 3 : Accuracy =  0.28
--Threshold 4 : Accuracy =  0.2
```

(3,1,0)

```
Uni- Accuracy: 0.56
Combo Accuracy: 0.76
```

MIXING ALL SYNTHETIC DATA AND NATURAL DATA

APPENDIX :

SYNTHETIC DATA

Only Synthetic MALE Data, for both Train and Test
Best Result (2, 4, 2)

```
For Ngram : 1
-- For threshold 0 : 0.782608695652174
-- For threshold 1 : 0.6521739130434783
-- For threshold 2 : 0.782608695652174
-- For threshold 3 : 0.7391304347826086
-- For threshold 4 : 0.782608695652174
-- For threshold 5 : 0.7391304347826086
-- For threshold 6 : 0.6956521739130435

For Ngram : 2
-- For threshold 0 : 0.782608695652174
-- For threshold 1 : 0.782608695652174
-- For threshold 2 : 0.8260869565217391
-- For threshold 3 : 0.9130434782608695
-- For threshold 4 : 0.9565217391304348
-- For threshold 5 : 0.8695652173913043
-- For threshold 6 : 0.6521739130434783

For Ngram : 3
-- For threshold 0 : 0.6956521739130435
-- For threshold 1 : 0.6521739130434783
-- For threshold 2 : 0.7391304347826086
-- For threshold 3 : 0.6956521739130435
-- For threshold 4 : 0.6086956521739131
-- For threshold 5 : 0.30434782608695654
-- For threshold 6 : 0.13043478260869565
```

```
Uni- Accuracy: 0.782608695652174
Combo Accuracy: 0.8695652173913043
```

Only Synthetic FEMALE 1 Data, for both Train and Test
Best Result (0/1, 0/2, 1)

```
For Ngram : 1
-- For threshold 0 : 0.782608695652174
-- For threshold 1 : 0.782608695652174
-- For threshold 2 : 0.6956521739130435
-- For threshold 3 : 0.6086956521739131
-- For threshold 4 : 0.6521739130434783
-- For threshold 5 : 0.6521739130434783
-- For threshold 6 : 0.7391304347826086

For Ngram : 2
-- For threshold 0 : 0.8260869565217391
-- For threshold 1 : 0.782608695652174
-- For threshold 2 : 0.8260869565217391
-- For threshold 3 : 0.7391304347826086
-- For threshold 4 : 0.5652173913043478
-- For threshold 5 : 0.391304347826087
-- For threshold 6 : 0.17391304347826086

For Ngram : 3
-- For threshold 0 : 0.5652173913043478
-- For threshold 1 : 0.7391304347826086
-- For threshold 2 : 0.391304347826087
-- For threshold 3 : 0.17391304347826086
-- For threshold 4 : 0.08695652173913043
-- For threshold 5 : 0.08695652173913043
-- For threshold 6 : 0.08695652173913043
```

(0,2,1)

```
Uni- Accuracy: 0.782608695652174
Combo Accuracy: 0.8260869565217391
```

Only Synthetic FEMALE 2 Data, for both Train and Test
Best Result (2, 2, 1)

```
For Ngram : 1
-- For threshold 0 : 0.6521739130434783
-- For threshold 1 : 0.6086956521739131
-- For threshold 2 : 0.6956521739130435
-- For threshold 3 : 0.6521739130434783
-- For threshold 4 : 0.6521739130434783
-- For threshold 5 : 0.6086956521739131
-- For threshold 6 : 0.5217391304347826

For Ngram : 2
-- For threshold 0 : 0.6086956521739131
-- For threshold 1 : 0.7391304347826086
-- For threshold 2 : 0.8260869565217391
-- For threshold 3 : 0.7391304347826086
-- For threshold 4 : 0.5652173913043478
-- For threshold 5 : 0.5217391304347826
-- For threshold 6 : 0.43478260869565216

For Ngram : 3
-- For threshold 0 : 0.5217391304347826
-- For threshold 1 : 0.6086956521739131
-- For threshold 2 : 0.5652173913043478
-- For threshold 3 : 0.47826086956521740
-- For threshold 4 : 0.21739130434782608
-- For threshold 5 : 0.08695652173913043
-- For threshold 6 : 0.08695652173913043
```

```
Uni- Accuracy: 0.6956521739130435
Combo Accuracy: 0.7826086956521740
```