Audio Language Classifier:
Model Metrics

Kirsten Regier

1 Gender Classifier

1.1 Model Architecture

Two different models were trained and evaluated for the gender classifier. The structure of each model is shown in Table 1. Model 1 has one hidden layer with 128 nodes, while Model 2 has two hidden layers with 128 and 64 nodes, respectively.

Table 1: Gender Classifier Model Architectures.

Layer	Model 1	Model 2
Input	(32, 10, 128)	(32, 10, 128)
Dense	128 nodes	128 nodes
Dropout	50%	50 %
Dense		64 nodes
Dropout		50%
Flatten	(32, 1280)	(32, 640)
Output	(32, 1)	(32, 1)

1.2 Model Performance Comparison

Table 2: Gender Classifiers - Metrics Comparison

	Model 1	Model 2
Loss	0.061640	0.076962
Accuracy	0.981419	0.975507
Precision	0.982818	0.960199
Recall	0.979452	0.991438

1.3 Confusion Matrices

Table 3: Confusion Matrices for the Gender Classifier Models

Model 1		Predicted		Recall
		F	M	
Actual	F	590	10	0.9833
	M	12	572	0.9795
Precision		0.9801	0.9828	
Accuracy				0.9814

Model 2		Predicted		Recall
		F	M	
Actual	F	576	24	0.9650
	M	5	579	0.9914
Precision		0.9914	0.9602	
Accuracy				0.9755

2 Language Classifier

2.1 Model Architecture

Three models were trained and evaluated for the language classifier. The structure of each model is shown in Table 4. Model 1 has one hidden layer with 12 nodes, Model 2 has one hidden layer with 128 nodes, and Model 3 has two hidden layers with 128 and 64 nodes, respectively.

Table 4: Language Classifier Model Architectures.

Layer	Model 1	Model 2	Model 3
Input	(32, 10, 128)	(32, 10, 128)	(32, 10, 128)
Dense	12 nodes	128 nodes	128 nodes
Dropout	50%	50%	50 %
Dense			64 nodes
Dropout			50%
Flatten	(32, 120)	(32, 1280)	(32, 640)
Output	(32, 11)	(32, 11)	(32, 11)

2.2 Model Performance Comparison

A naive classifier that always predicted the majority class (Arabic) would have an accuracy of 15%. All three of the models improved upon this baseline accuracy rate, as shown in the second row of Table 5.

Table 5: Language Classifiers - Metrics Comparison

	Model 1	Model 2	Model 3
Loss	2.25485	2.323	2.20684
Accuracy	0.232955	0.238636	0.25
Precision	0.625	0.5	0.605263
Recall	0.0142045	0.0738636	0.0653
F1 Macro	0.16999	0.195512	0.191103
F1 Weighted	0.20067	0.228178	0.224565

2.3 Confusion Matrix - Model 3

The confusion matrix for the Model 3 predictions is shown in Table 6, with the precision and recall rates by class in Table 7.

Table 6: Confusion matrix for Model 3 predictions. The bold numbers on the diagonal represent correct predictions.

lang	R	A	T	K	G	D	S	F	E	P	M	Segments
Russian	**0**	2	1	4	0	1	7	4	1	1	7	28
Arabic	0	**24**	0	4	1	7	2	2	4	0	11	55
Turkish	0	1	**1**	4	0	3	4	0	2	1	5	21
Korean	1	10	0	**6**	0	3	2	2	4	2	1	31
German	0	0	0	1	**0**	0	5	1	5	2	1	15
Dutch	0	4	0	0	0	**12**	3	1	3	1	1	25
Spanish	1	2	0	6	2	1	**12**	3	8	0	7	42
French	1	3	0	4	1	2	10	**5**	1	1	4	32
English	0	1	0	3	0	5	2	2	**13**	0	4	30
Portuguese	0	3	0	3	1	0	7	5	3	**0**	6	28
Mandarin	0	2	0	8	2	4	9	1	2	2	**15**	45
Total	3	52	2	43	7	38	63	26	46	10	62	**352**

Table 7: Model 3 Metrics by Language

Language	Precision	Recall	F1-score	Samples
Russian	0.00	0.00	0.00	28
Arabic	0.46	0.44	0.45	55
Turkish	0.50	0.05	0.09	21
Korean	0.14	0.19	0.16	31
German	0.00	0.00	0.00	15
Dutch	0.32	0.48	0.38	25
Spanish	0.19	0.29	0.23	42
French	0.19	0.16	0.17	32
English	0.28	0.43	0.34	30
Portuguese	0.00	0.00	0.00	28
Mandarin	0.24	0.33	0.28	45
accuracy			0.25	352
macro avg	0.21	0.22	0.19	352
weighted avg	0.23	0.25	0.22	352