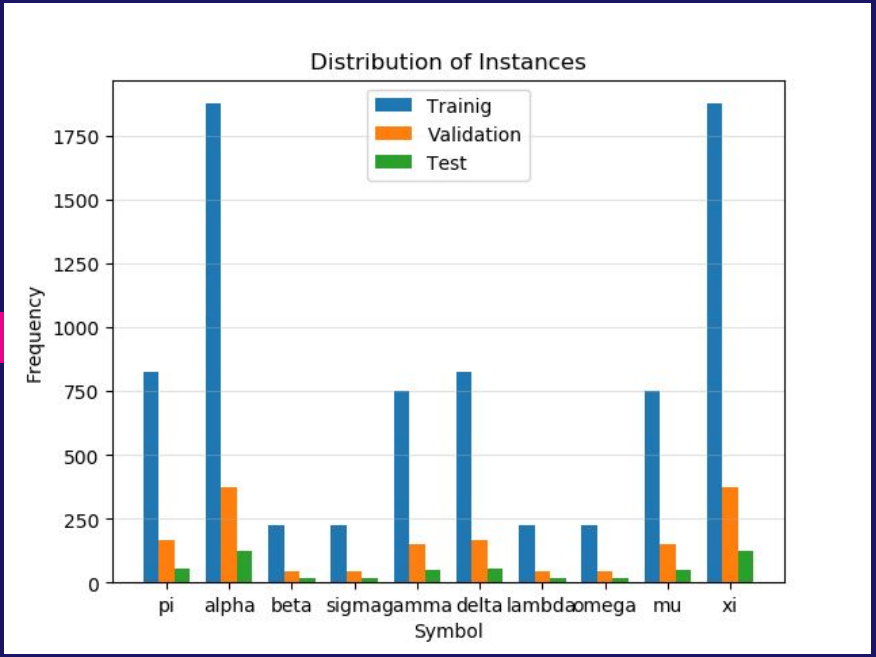
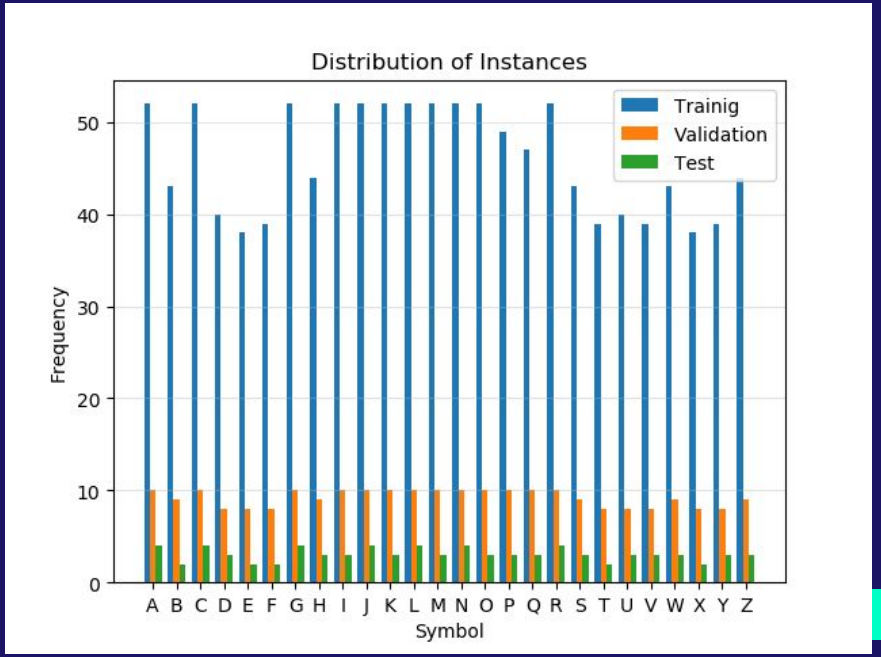
Assignment 1: Experimenting with Machine Learning

Team Coders For Hire:
Ihsaan Malek, 40024975 & Olivier Racette 40017231

01

OVERVIEW





Datasets

What we notice as a recurring theme with the models, is for Dataset 2, alpha, pi, gamma, delta, mu, and xi are generally the characters most accurately predicted.

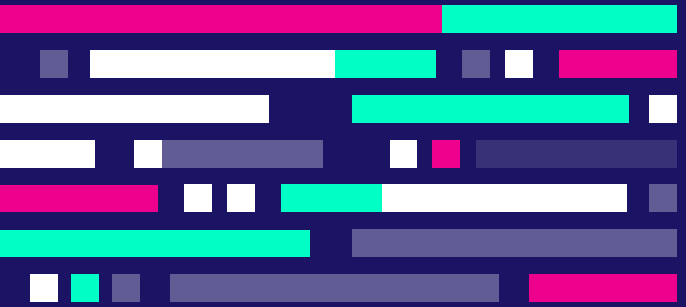
With Dataset 1, we see a more evenly distribution of character instances. This suggests that there should not be any sampling issues when it comes characters having more samples. However, the number of samples per letter is much less than Dataset 2

Summary of Model Performances

	DATA SET 1	*DATA SET 2*
MODEL	**F1-SCORE**	
GNB	0.69	0.68
Base-DT	0 .48	0.79
Best-DT	0.44	0.73
PER	0.77	0.83
Base-MLP	0.32	0.83
Best-MLP	0.83	0.87

GNB:Gaussian Naive Bayes Classifier

- Basic Model with default parameters
- Has Trouble predicting complex letters

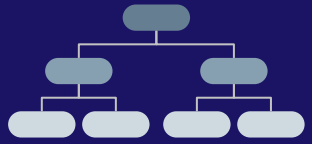


- Basic Model with default parameters
- For Dataset 1: Generally High level of precision; there are many predictions with a value of 1.0

Data Set 1		***Data Set 2***	
Letters/Class	F1-Score	**Letters/Class**	F1-Score
I,L,O,T,U	1 > F1 > 0.86	ω	1 > F1 > 0.86
C,Q,S,W,Y	0.86 > F1 >0.7	α, ξ	0.86 > F1 >0.7
A,B,D,F,G,J,M,N ,P,R,V,X,Z	0.7 > F1 > =0.5	π,σ,δ	0.7 > F1 > =0.5
E,H,K	F1 < 0.5	β,γ,λ,μ	F1 < 0.5

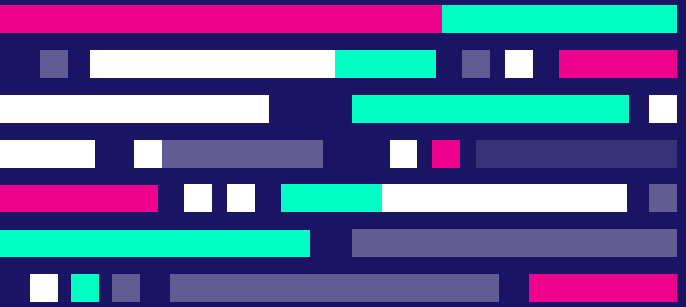
Weighted Average F1-SCORE

DataSet 1	DataSet 2
0.69	0.68



BASE-DT: DECISION TREE

- Using default parameters and Entropy as decision criterion

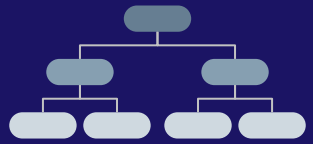


- Model is poor at predicting the Alphabet, too many characters
- Better at predicting a smaller training set

Data Set 1		***Data Set 2***	
Letters/Class	F1-Score	**Letters/Class**	F1-Score
	1 > F1 > 0.86	α	1 > F1 > 0.86
L,O,S,U	0.86 > F1 >0.7	π, δ, ω, μ, ξ,	0.86 > F1 >0.7
A,B,C,E,H,I,J,N, R,V,X,Y	0.7 > F1 > =0.5	β, σ, γ, λ	0.7 > F1 > =0.5
D,F,G,K,M,PQ,T ,W,Z	F1 < 0.5		F1 < 0.5

Weighted Average F1-SCORE

DataSet 1	DataSet 2
0.48	0.79



BEST DECISION TREE, GridSeachCV

- Dataset1: Criterion: Entropy, Max_depth: None, Min_sample_split: 5, Min_inpurity_decrease: 0, class_weight: None
- Dataset2: Criterion: Gini, Max_depth: None, Min_sample_split: 2, Min_inpurity_decrease: 0, class_weight: Balanced

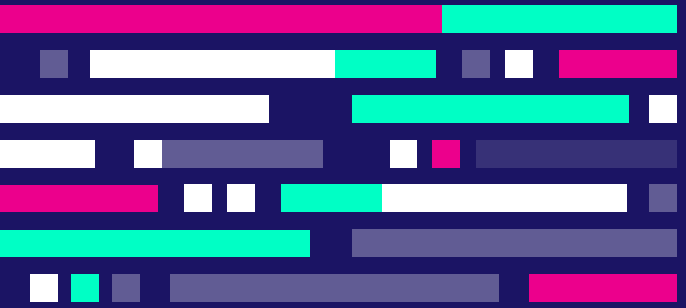
Data Set 1		***Data Set 2***	
Letters/Class	F1-Score	**Letters/Class**	F1-Score
	1 > F1 > 0.86	ω	1 > F1 > 0.86
L,	0.86 > F1 >0.7	α , ξ	0.86 > F1 >0.7
A,C,F,H,I,J,N,O, Q,R,S,U,V,Y	0.7 > F1 > =0.5	π,σ,δ	0.7 > F1 > =0.5
B,D,E,G,K,M,P, T,W,X,Z	F1 < 0.5	β,γ,λ,μ	F1 < 0.5

Weighted Average F1-SCORE

DataSet 1	DataSet 2
0.44	0.73

PER : PERCEPTRON

- Generally, high level of precision and high recall values
- Better at predicting letter with curvature

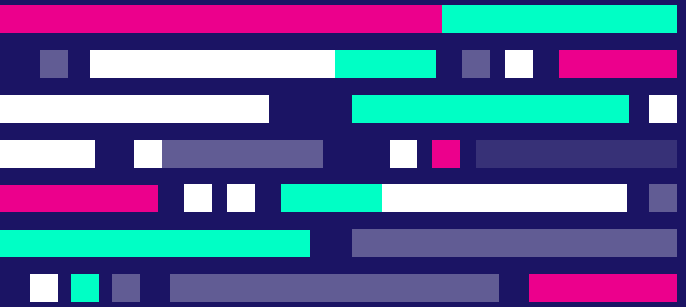


Weighted Average F1-SCORE	DataSet 1	DataSet 2
	0.77	0.83

Data Set 1		***Data Set 2***	
Letters/Class	F1-Score	**Letters/Class**	F1-Score
B,D,E,I,L,N,O,U, V,W,Y	1 > F1 > 0.86	α, σ, ω,	1 > F1 > 0.86
C,G,,J,P,Q,S,	0.86 > F1 >0.7	π, δ, μ, ξ,	0.86 > F1 >0.7
A,R,,X,	0.7 > F1 > =0.5	β, γ,	0.7 > F1 > =0.5
F,H,K,M,T,	F1 < 0.5	λ	F1 < 0.5

BASE-MLP: Multi-Layered Perceptron, 1 hidden Layer of 100 Neurons

- Model is poor at predicting values of a larger data set



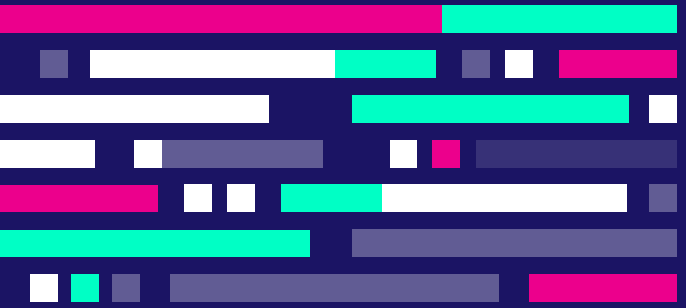
Data Set 1		***Data Set 2***	
Letters/Class	F1-Score	**Letters/Class**	F1-Score
	1 > F1 > 0.86	α, ω,	1 > F1 > 0.86
G,W	0.86 > F1 >0.7	π, σ, δ, μ, ξ,	0.86 > F1 >0.7
I,J,K,M,N,Q,	0.7 > F1 > =0.5	β, γ, λ	0.7 > F1 > =0.5
A,B,C,D,E,F,H,L,O,P,R,S,T,U,V,X,Y,Z	F1 < 0.5		F1 < 0.5

Weighted Average F1-SCORE

DataSet 1	DataSet 2
0.32	0.83

Best-MLP: Multi-Layered Perceptron

- Network architectures tested: [30, 50] and [10,10,10]
- Dataset1: activation: identity, hidden_layer_sizes: [30, 50], solver: adam
- Dataset2: activation: tanh, hidden_layer_sizes: [30, 50], solver: sgd



Weighted Average F1-SCORE	DataSet 1	DataSet 2
	0.83	0.87

Data Set 1		***Data Set 2***	
Letters/Class	F1-Score	**Letters/Class**	F1-Score
C,D,E,I,L,Q,U ,V,W,Y,Y	1 > F1 > 0.86	α, δ, ω, μ, ξ,	1 > F1 > 0.86
A,B,H,N,O,P, R	0.86 > F1 >0.7	π, σ, λ	0.86 > F1 >0.7
F,G,J,S,T,X,Z	0.7 > F1 > =0.5	β, γ,	0.7 > F1 > =0.5
K,M	F1 < 0.5		F1 < 0.5

Contributions

Description of contributions and responsibilities

Main.py
Olivier

GNB,Base-MLP, Best-MPL
Olivier

Base-DT, Best-DT, PER
Ihsaan

PPT Slides
Ihsaan, Olivier

EXERCISE

```
<!DOCTYPE html>
<html>
<body>

<canvas id="myCanvas"
width="250" height="25"
style="border:4px solid #00ffc5;">
</canvas>

</body>
</html>
```