Setup

```
In [1]: import sys, os
        current_directory = os.getcwd()
        root_directory = os.path.abspath(os.path.join(current_directory, os.pardir))
        sys.path.append(root_directory)
```

Import utils

1- Using single K Model

Predicting: 100%|

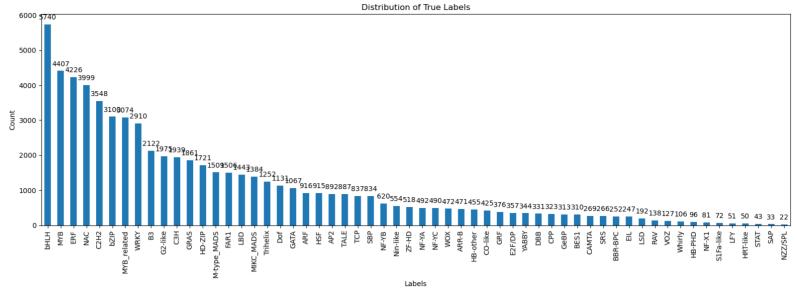
In [9]: **from** pretrained.predictor **import** SingleKModel, MultiKModel, OneTestKModel, BatchSingleKModel import metrics

Predict

In [3]: kmodel = SingleKModel(kmer_size=2) kmodel.load("../data/testset-full/k2/testset.csv", format="csv", type='kmer_file') genboard = kmodel.predict()

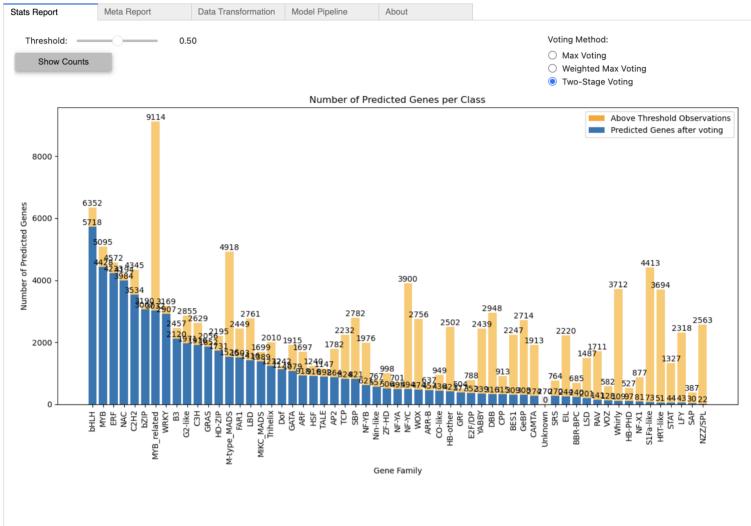
58/58 [01:50<00:00, 1.90s/it]

In [4]: metrics.plot_testset(true_label_df_path="../data/testset-full/k2/true_labels.csv", class_mapping_df



In [5]: genboard.display()

Tab(children=(VBox(children=(VBox(children=(VBox(children=(FloatSlider(value=0.5, continuous_update =False, des...



genboard.prediction.shape

Confusion report for validation

Out[6]: (64091, 58)

In [7]: **import** json import pandas as pd

voting_threshold=0.5,

binary_class_threshold=0.5,

```
true_label = pd.read_csv('../data/testset-full/k2/true_labels.csv')['true_label'].values
        with open('../data/testset-full/k2/class_mapping.json', 'r') as json_file:
            class_mapping = json.load(json_file)
        class_mapping['Unknown'] = 0
In [8]: genboard.show_eval_metric(
            true_label=true_label,
            class_mapping_rules=class_mapping,
            voting_method="Two-Stage Voting", #Two-Stage Voting
```

Gene

ARF

В3

Family

Accuracy

0.99

0.99

Precision

0.54

0.85

kmer_size=2

Recall

1.00

0.98

0.98

F1 Score

0.70

0.91

components=['confusion_matrix', 'general_accuracy', 'accuracy_per_family'] **Overall Accuracy Score**

Recall

0.97

0.96

F1 Score

0.65

0.82

AP2 0.99 0.48 ARR-B 1.00 0.71

Accuracy per Gene Family

Accuracy Precision

Gene

Family

0.99 0.98 0.97 0.99 1.00	0.36 0.79 0.14 0.35 0.90	0.99 0.97 1.00 0.99	0.53 0.87 0.25	BES1 C3H CO-like	0.97 0.99	0.14 0.71	0.99 0.96	0.24 0.82
0.97 0.99 1.00	0.14 0.35	1.00						
0.99	0.35		0.25	CO-like				
1.00		0.99	I	CO-like	0.99	0.44	0.99	0.61
	0.90		0.52	DBB	0.96	0.11	0.99	0.20
0.97		0.99	0.94	E2F/DP	0.99	0.44	0.97	0.61
	0.11	0.98	0.20	ERF	0.99	0.91	0.98	0.95
0.98	0.60	0.98	0.75	G2-like	0.98	0.67	0.97	0.80
0.99	0.55	0.98	0.70	GRAS	1.00	0.88	0.98	0.93
1.00	0.74	1.00	0.85	GeBP	0.96	0.11	0.96	0.20
0.99	0.18	1.00	0.31	HB-other	0.97	0.17	0.91	0.28
0.99	0.78	0.99	0.87	HRT-like	0.94	0.01	0.96	0.03
0.99	0.73	0.99	0.84	LBD	0.98	0.51	0.98	0.67
0.96	0.02	0.96	0.04	LSD	0.98	0.13	0.99	0.23
0.94	0.29	0.95	0.45	MIKC_MADS	0.99	0.81	1.00	0.89
0.99	0.86	0.99	0.92	MYB_related	0.90	0.31	0.93	0.47
				NF-X1	0.99	0.09	1.00	0.17
				NF-YB	0.98	0.31	0.97	0.47
				NZZ/SPL	0.96	0.01	1.00	0.02
				RAV	0.98	0.08	1.00	0.15
				SAP	0.99	0.09	1.00	0.16
				SRS	0.99	0.34	0.99	0.51
				TALE	1.00	0.77	0.99	0.86
				Trihelix	0.99	0.61	0.97	0.75
				WOX	0.96	0.17	0.99	0.29
				Whirly	0.94	0.03	0.98	0.05
				ZF-HD	0.99	0.50	0.97	0.66
				bZIP	1.00	0.95	0.98	0.96
0.00	0.00	0.00		Confusion N	/latrix			
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1 0	Confusion Matrix 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
	0.99 1.00 0.99 0.99 0.99	0.99 0.55 1.00 0.74 0.99 0.78 0.99 0.73 0.96 0.02 0.99 0.86 0.99 0.94 1.00 0.70 0.95 0.12 1.00 0.72 0.93 0.02 0.97 0.29 0.98 0.03 0.99 0.91 0.99 0.91 0.99 0.88	0.99 0.55 0.98 1.00 0.74 1.00 0.99 0.18 1.00 0.99 0.78 0.99 0.99 0.73 0.99 0.96 0.02 0.96 0.99 0.86 0.99 0.99 0.94 0.98 1.00 0.70 0.99 0.95 0.12 0.98 1.00 0.72 0.99 0.93 0.02 1.00 0.97 0.29 0.98 0.98 0.37 0.98 0.99 0.91 0.99 0.99 0.88 0.98	0.99 0.55 0.98 0.70 1.00 0.74 1.00 0.85 0.99 0.18 1.00 0.31 0.99 0.78 0.99 0.87 0.99 0.73 0.99 0.84 0.96 0.02 0.96 0.04 0.94 0.29 0.95 0.45 0.99 0.86 0.99 0.92 0.99 0.94 0.98 0.96 1.00 0.70 0.99 0.82 0.95 0.12 0.98 0.22 1.00 0.72 0.99 0.83 0.93 0.02 1.00 0.03 0.97 0.29 0.98 0.45 0.98 0.03 1.00 0.06 0.98 0.37 0.98 0.54 0.99 0.91 0.99 0.94 0.99 0.88 0.98 0.93	0.99 0.55 0.98 0.70 GRAS 1.00 0.74 1.00 0.85 GeBP 0.99 0.18 1.00 0.31 HB-other 0.99 0.78 0.99 0.87 HRT-like 0.99 0.73 0.99 0.84 LBD 0.96 0.02 0.96 0.04 LSD 0.94 0.29 0.95 0.45 MIKC_MADS 0.99 0.86 0.99 0.92 NF-X1 0.99 0.94 0.98 0.96 NF-YB NF-YB NF-YB NZZ/SPL RAV 0.95 0.12 0.98 0.22 RAV 0.95 0.12 0.98 0.22 RAV 0.95 0.12 0.98 0.22 RAV 0.93 0.02 1.00 0.03 SRS 0.97 0.29 0.98 0.45 WOX 0.99 0.91 0.99 0.94 Whirly <td>0.99 0.55 0.98 0.70 GRAS 1.00 1.00 0.74 1.00 0.85 GeBP 0.96 0.99 0.18 1.00 0.31 HB-other 0.97 0.99 0.73 0.99 0.84 LBD 0.98 0.96 0.02 0.96 0.04 LSD 0.98 0.99 0.86 0.99 0.92 0.99 0.94 0.98 0.96 1.00 0.70 0.99 0.82 0.95 0.12 0.98 0.22 1.00 0.72 0.99 0.83 0.93 0.02 1.00 0.03 0.97 0.29 0.98 0.45 0.98 0.03 1.00 0.06 0.99 0.91 0.99 0.94 0.99 0.91 0.99 0.94 0.99 0.91 0.99 0.94 0.99 0.88 0.98 0.93 0.99 0.88 0.98 0.93 Confusion Matrix Confusion Matrix Confusion Matrix</td> <td>0.99 0.55 0.98 0.70 GRAS 1.00 0.88 1.00 0.74 1.00 0.85 GeBP 0.96 0.11 0.99 0.18 1.00 0.31 HB-other 0.97 0.17 0.99 0.73 0.99 0.84 LBD 0.94 0.01 0.96 0.02 0.96 0.04 LSD 0.98 0.51 0.94 0.29 0.95 0.45 MIKC_MADS 0.99 0.81 0.99 0.86 0.99 0.92 MIKC_MADS 0.99 0.81 0.99 0.86 0.99 0.92 MIKC_MADS 0.99 0.81 0.99 0.84 0.98 0.96 NF-YB 0.99 0.09 0.95 0.12 0.98 0.22 RAV 0.98 0.03 0.97 0.29 0.98 0.45 TALE 1.00 0.77 0.98 0.37 0.99 0.24 WOX</td> <td>1.00</td>	0.99 0.55 0.98 0.70 GRAS 1.00 1.00 0.74 1.00 0.85 GeBP 0.96 0.99 0.18 1.00 0.31 HB-other 0.97 0.99 0.73 0.99 0.84 LBD 0.98 0.96 0.02 0.96 0.04 LSD 0.98 0.99 0.86 0.99 0.92 0.99 0.94 0.98 0.96 1.00 0.70 0.99 0.82 0.95 0.12 0.98 0.22 1.00 0.72 0.99 0.83 0.93 0.02 1.00 0.03 0.97 0.29 0.98 0.45 0.98 0.03 1.00 0.06 0.99 0.91 0.99 0.94 0.99 0.91 0.99 0.94 0.99 0.91 0.99 0.94 0.99 0.88 0.98 0.93 0.99 0.88 0.98 0.93 Confusion Matrix Confusion Matrix Confusion Matrix	0.99 0.55 0.98 0.70 GRAS 1.00 0.88 1.00 0.74 1.00 0.85 GeBP 0.96 0.11 0.99 0.18 1.00 0.31 HB-other 0.97 0.17 0.99 0.73 0.99 0.84 LBD 0.94 0.01 0.96 0.02 0.96 0.04 LSD 0.98 0.51 0.94 0.29 0.95 0.45 MIKC_MADS 0.99 0.81 0.99 0.86 0.99 0.92 MIKC_MADS 0.99 0.81 0.99 0.86 0.99 0.92 MIKC_MADS 0.99 0.81 0.99 0.84 0.98 0.96 NF-YB 0.99 0.09 0.95 0.12 0.98 0.22 RAV 0.98 0.03 0.97 0.29 0.98 0.45 TALE 1.00 0.77 0.98 0.37 0.99 0.24 WOX	1.00

