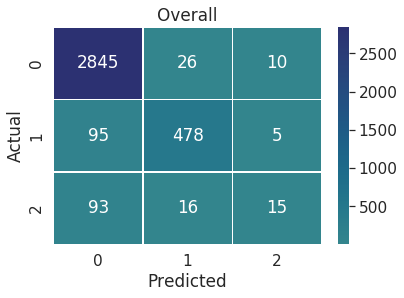
**March 30, 2024: CHECK FOR DUPLICATE DATA, REMOVE DUPLICATES AND RUN THE MODEL AFTER REMOVING DUPLICATE DATA FOR 100 TRIALS**

100%|██████████| 100/100 [18:58<00:00, 11.38s/trial, best loss: 0.039574482791014276]



precision recall f1-score support

0 0.94 0.99 0.96 2881

1 0.92 0.83 0.87 578

2 0.50 0.12 0.19 124

accuracy 0.93 3583

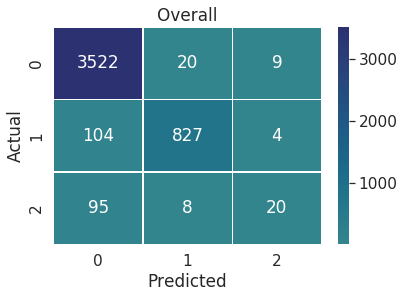
macro avg 0.79 0.65 0.68 3583

weighted avg 0.92 0.93 0.92 3583

**FEB 26, 2024: NO SIGNIFICANT DIFFERENCE IN THE ACCURACY AND F1 SCORE FOR 100 AND 500 TRIALS. SO, STICKING WITH 100 TRIALS IS PREFERABLE BECAUSE IT TAKES ONE-FOURTH OF TIME TO RUN. ALSO, MORE NUMBER OF 1’S ARE PREDICTED CORRECTLY IN 100 TRIALS**

* **Used UNC parameter values, changed ‘year’ to int and ‘jim\_crow’ to int, added Morgan/Taylor labelled data to re-train the model**
* **Used “SMOTEN” and stopwords removal**

100%|██████████| 100/100 [33:14<00:00, 19.94s/trial, best loss: 0.03014936236119825]



precision recall f1-score support

0 0.95 0.99 0.97 3551

1 0.97 0.88 0.92 935

2 0.61 0.16 0.26 123

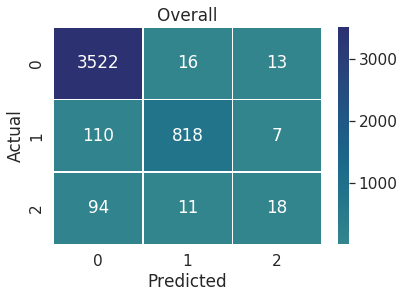
**accuracy 0.95** 4609

macro avg 0.84 0.68 0.72 4609

weighted avg 0.94 0.95 0.94 4609

**RUN THE MODEL FOR 500 TRIALS ON FEB 26, 2024**

100%|██████████| 500/500 [2:23:29<00:00, 17.22s/trial, best loss: 0.029639438994228606]



precision recall f1-score support

0 0.95 0.99 0.97 3551

1 0.97 0.87 0.92 935

2 0.47 0.15 0.22 123

accuracy **0.95** 4609

macro avg 0.80 0.67 0.70 4609

weighted avg 0.94 0.95 0.94 4609