data/raw\_data/drugsComTest\_raw.tsv

Started model training...

--------------------------------

Data Cleansing...

List of conditions: ['OTHERS', 'ADHD', 'Birth Control', 'Bipolar Disorde', 'Depression', 'Obesity', 'Insomnia', 'Pain', 'Anxiety', 'Acne', 'Weight Loss']

features path: data/processed\_data/spacy\_features\_for\_prediction.pkl

Saved spacy features to data/processed\_data/spacy\_features\_for\_prediction.csv

--------------------------------

--------------------------------

Building features & load models...

TfidfVectorizer(max\_df=0.5, ngram\_range=(1, 2))

--------------------------------

--------------------------------

Predict & Evaluate results...

Test accuracy: 0.848584607372689

Classification Report:

precision recall f1-score support

ADHD 0.85 0.92 0.88 1126

Acne 0.89 0.96 0.92 1847

Anxiety 0.57 0.82 0.67 1908

Bipolar Disorde 0.76 0.87 0.81 1380

Birth Control 0.87 0.99 0.93 9648

Depression 0.61 0.89 0.73 3095

Insomnia 0.73 0.89 0.80 1231

OTHERS 0.97 0.78 0.87 28994

Obesity 0.79 0.83 0.81 1189

Pain 0.58 0.88 0.70 2100

Weight Loss 0.80 0.85 0.83 1248

accuracy 0.85 53766

macro avg 0.77 0.88 0.81 53766

weighted avg 0.88 0.85 0.85 53766

Confusion Matrix:

Confusion matrix, without normalization

Confusion matrix saved to file path: reports/figures/validation\_confusion\_matrix\_for\_prediction.png