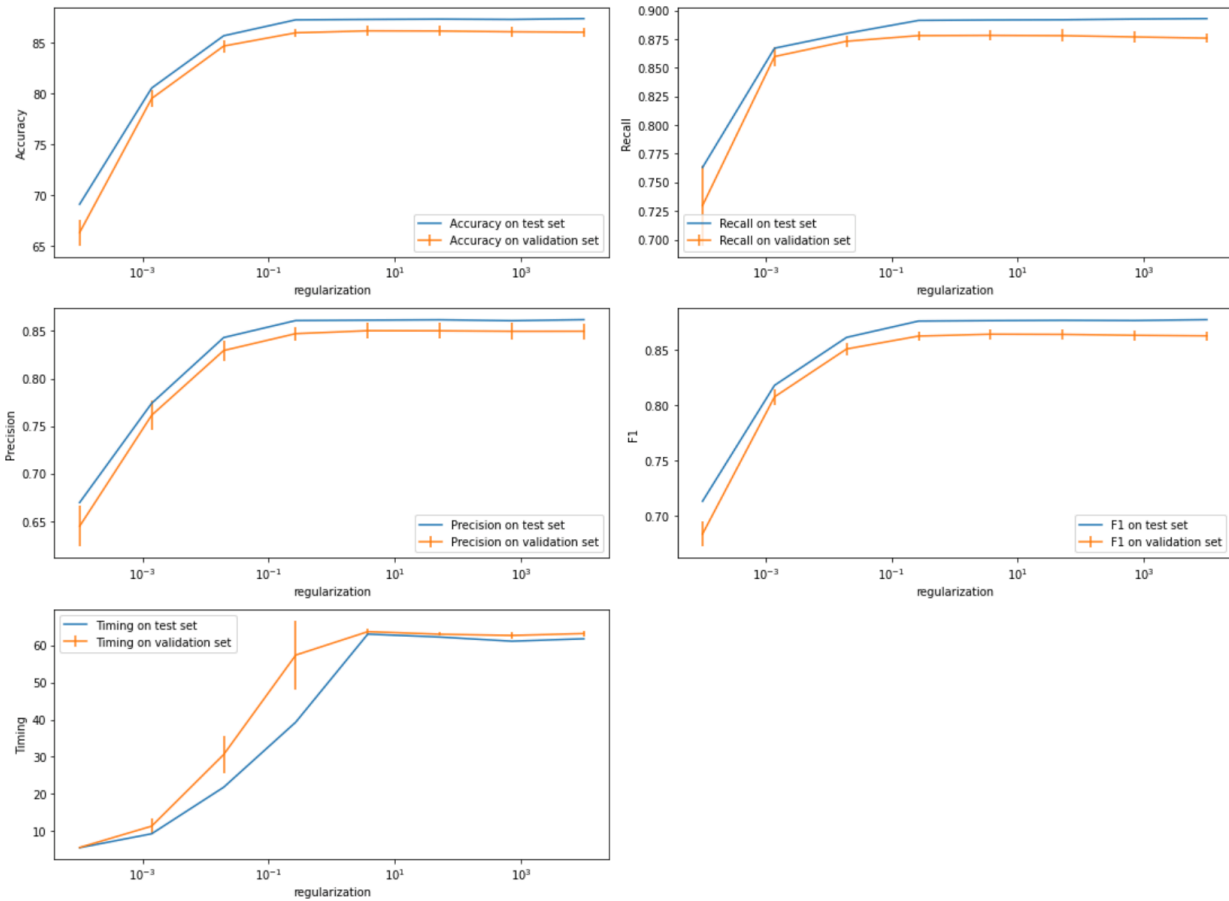


1 IMDB Movie Reviews Dataset

1.1 Logistic Regression

Test 1: Varying regularization

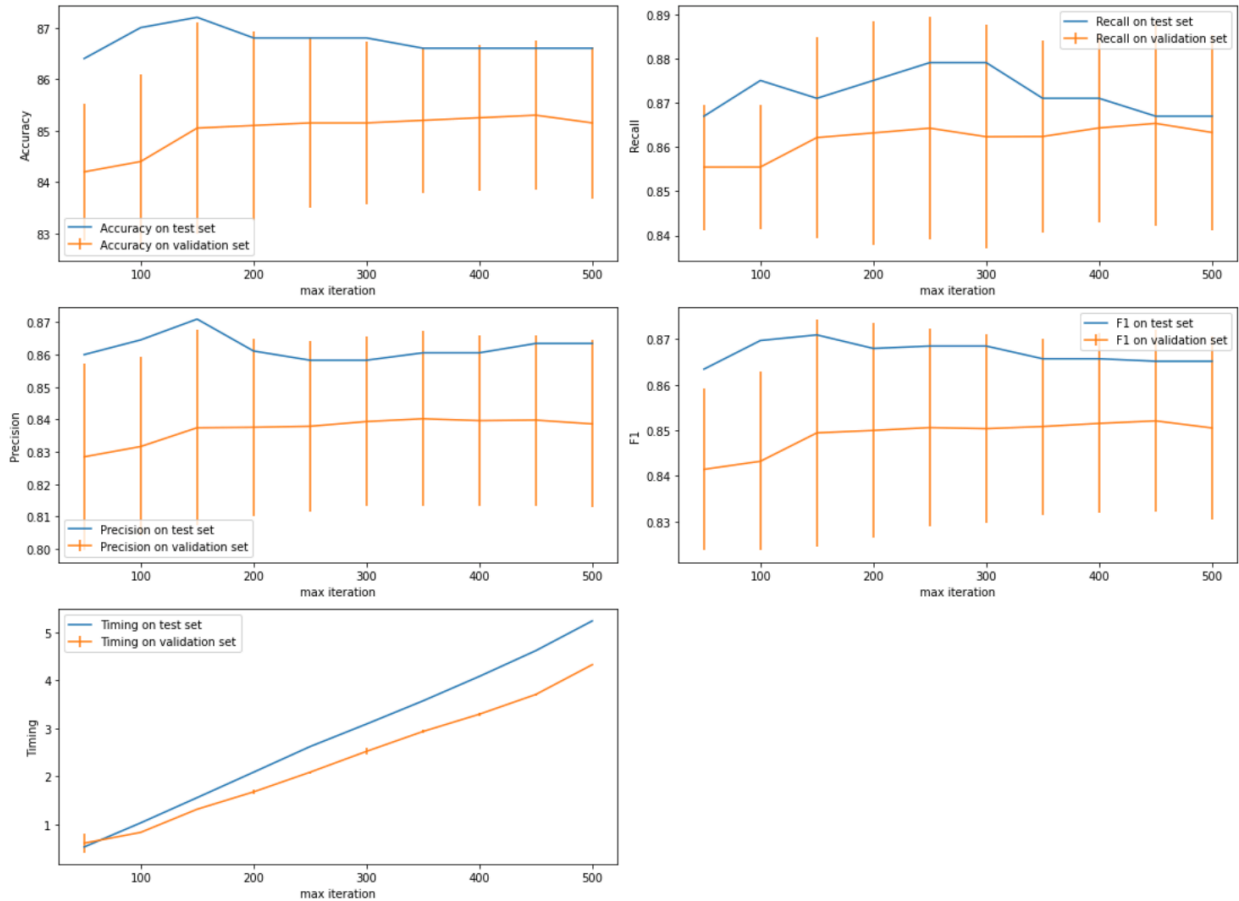
- Gradient Descent: LBFGS
- Max iterations: 100
- Ngrams: 2
- Vectorizer: Count



Test 2 of 11

Test 2: Varying max iterations

- Gradient Descent: LBFGS
- Regularization: 100
- Ngrams: 2
- Vectorizer: Count



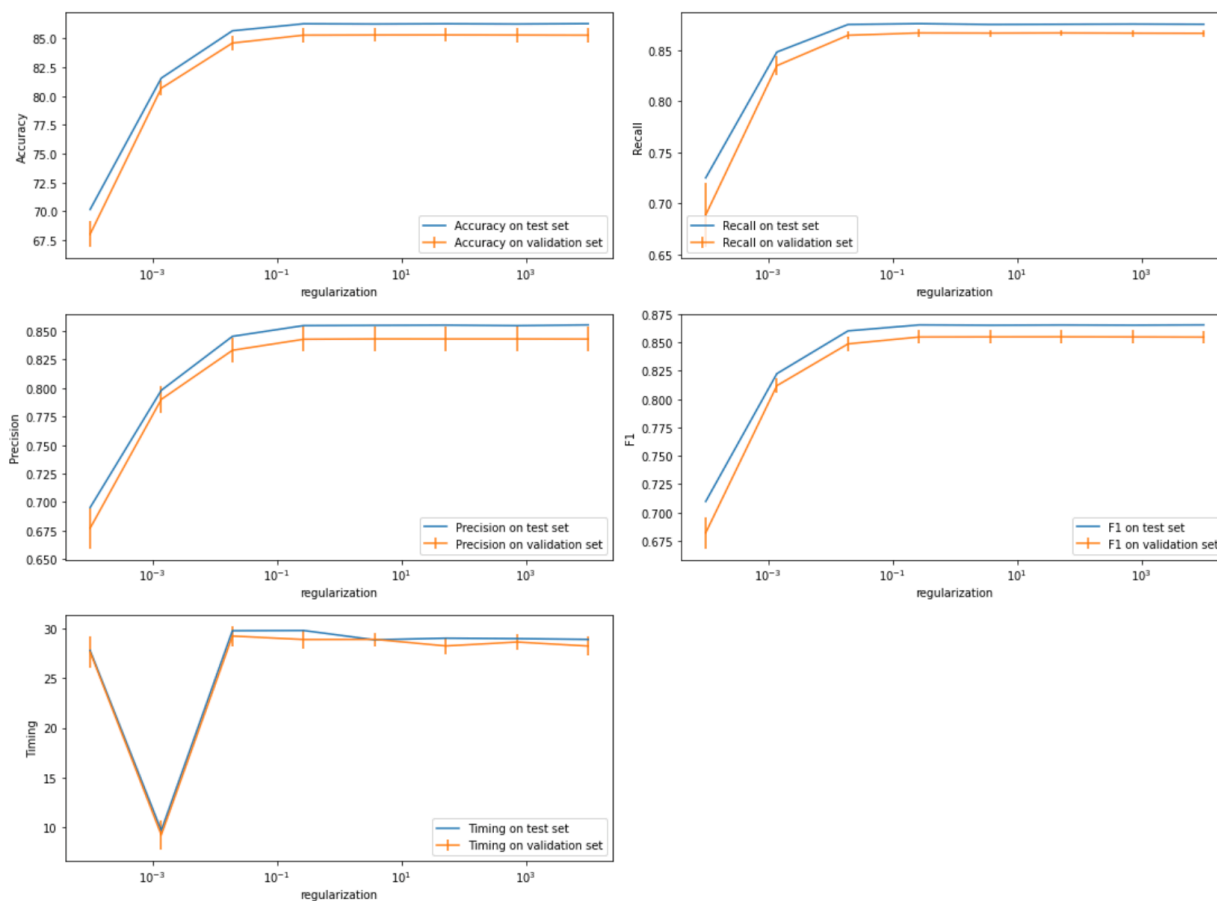
Highest average accuracy on validation set: 85.300 with max iteration of 450.000

Accuracy on test set: 85.300 with max iteration of 450.000

Test 3 of 11

Test 3: Varying Regularization

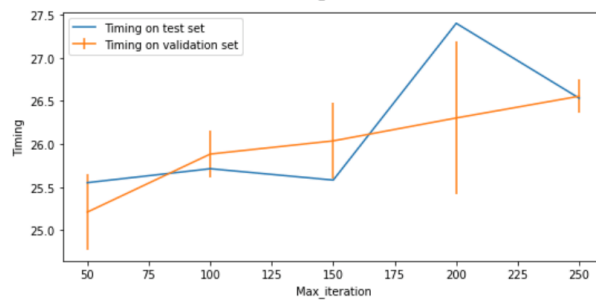
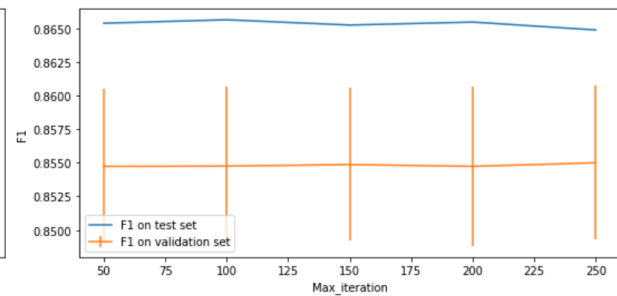
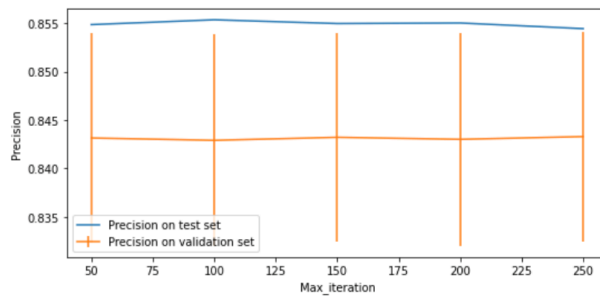
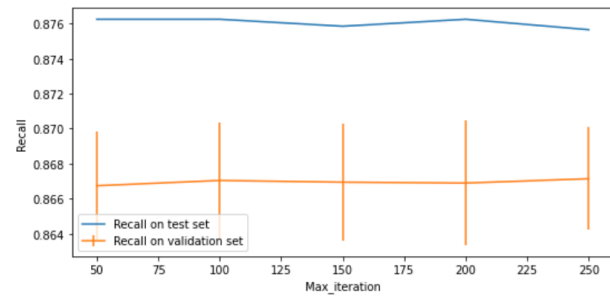
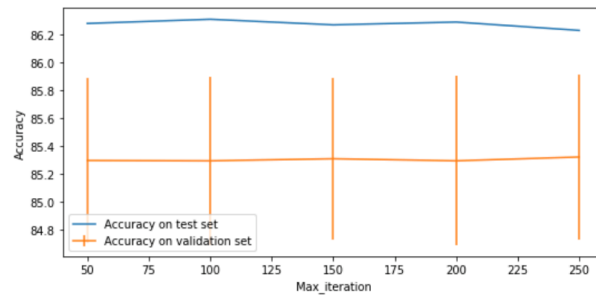
- Gradient Descent: SAGA
- Max iterations: 100
- Ngrams: 2
- Vectorizer: Count



Test 4 of 11

Test 4: Varying Max Iterations

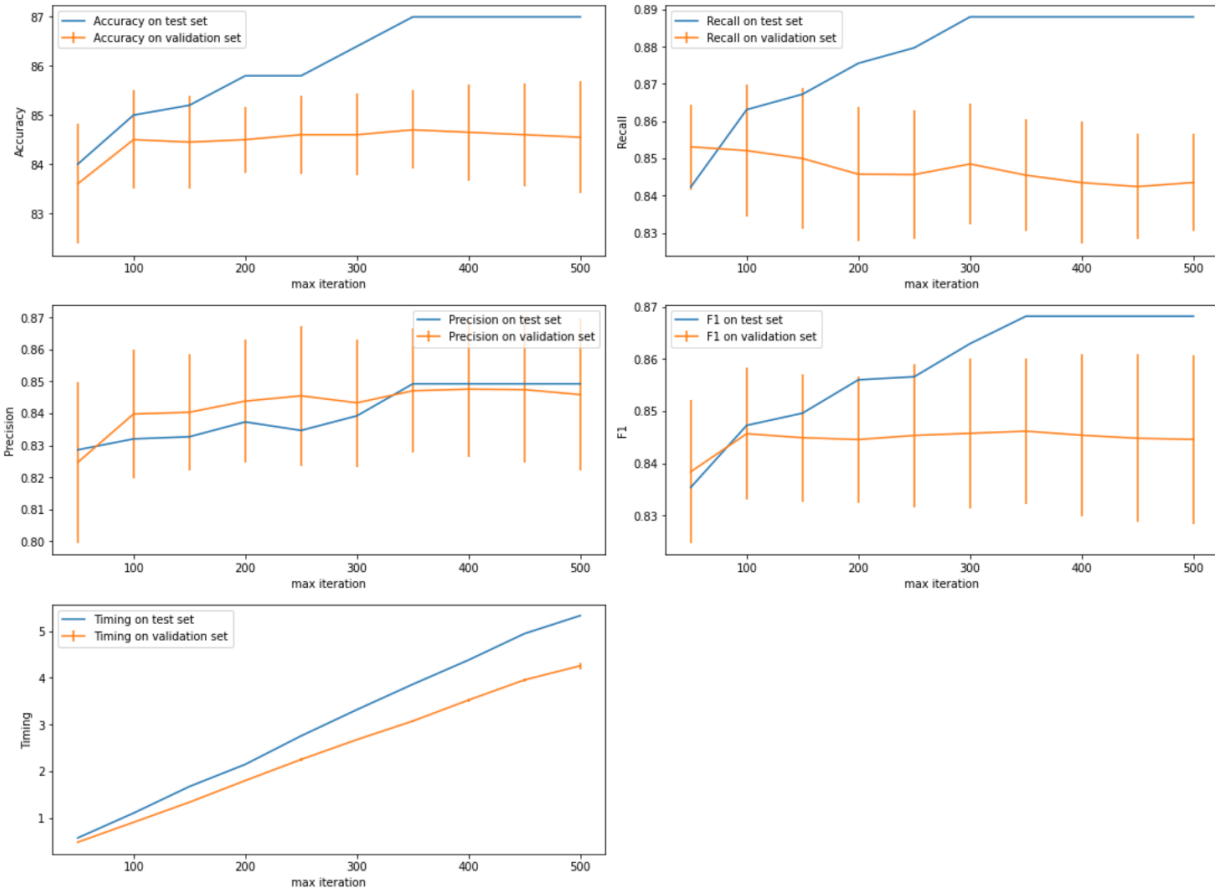
- Gradient Descent: LBFGS:
- Regularization: 1
- Ngrams: 2
- Vectorizer: Count



Test 5 of 11

Test 5: Varying max iterations

- Gradient Descent: SAGA
- C: 65
- Ngrams: 1
- Vectorizer: Count



Highest average accuracy on validation set: 84.700 with max iteration of 350.000

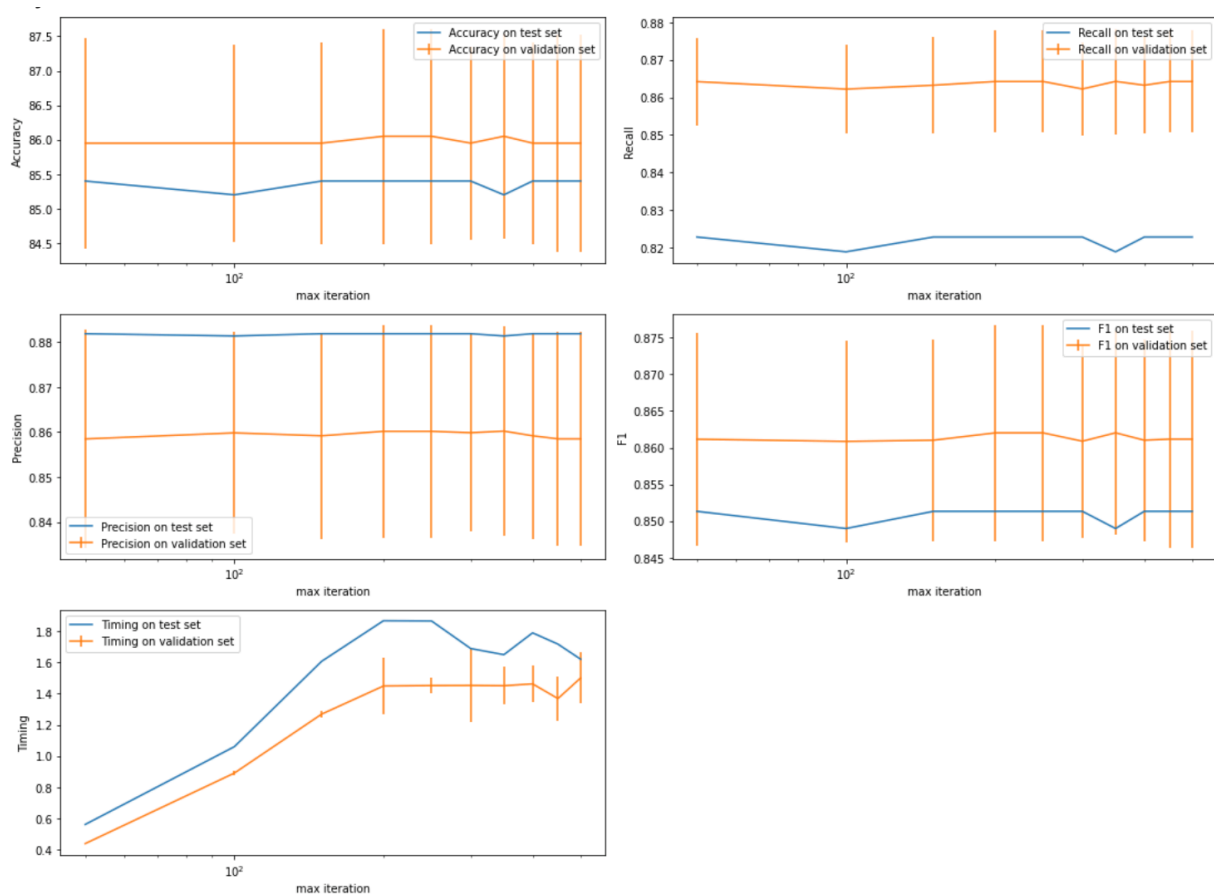
Accuracy on test set: 84.700 with max iteration of 350.000

Accuracy: 87.000

Test 6 of 11

Test 6: Varying max iterations

- Gradient Descent: SAGA
- C: 65
- Ngrams: 1
- Vectorizer: TF-IDF



Highest average accuracy on validation set: 86.050 with max iteration of 200.000

Accuracy on test set: 86.050 with max iteration of 200.000

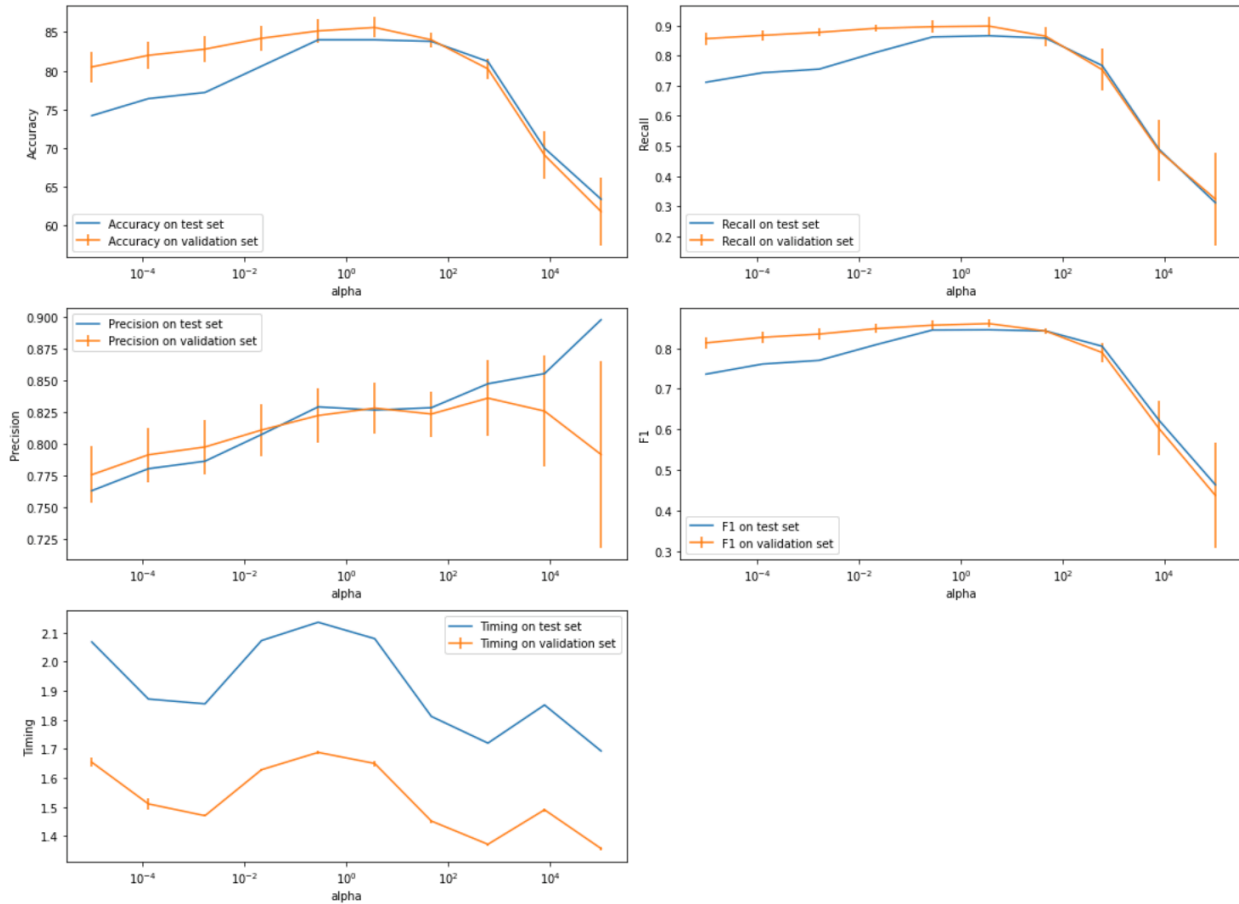
Accuracy: 85.400

1.2 Multinomial Naive Bayes

Test 7 of 11

Test 7: Varying smoothing factor

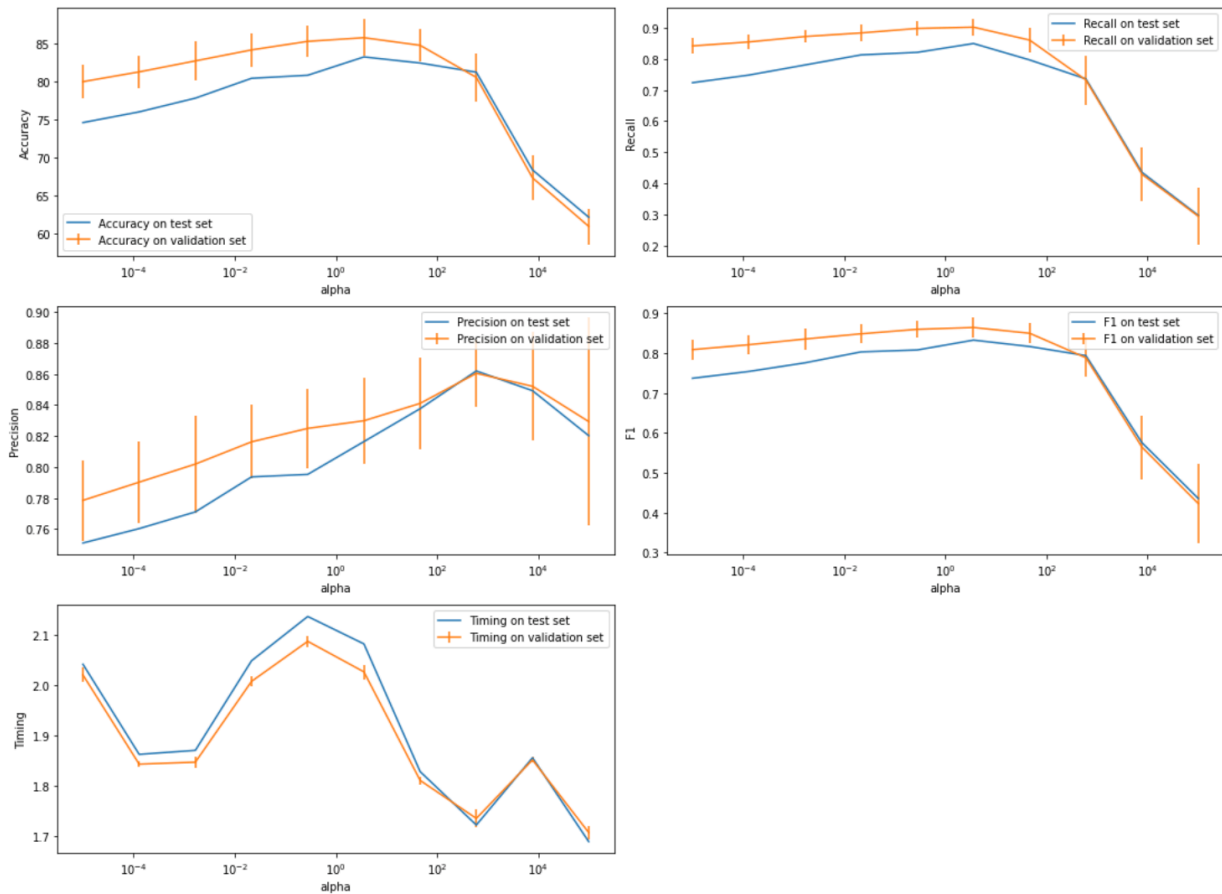
- Ngrams: 2
- Vectorizer: Count



Test 8: Varying smoothing factor

Test 8 of 11

- Ngrams: 1
- Vectorizer: Count



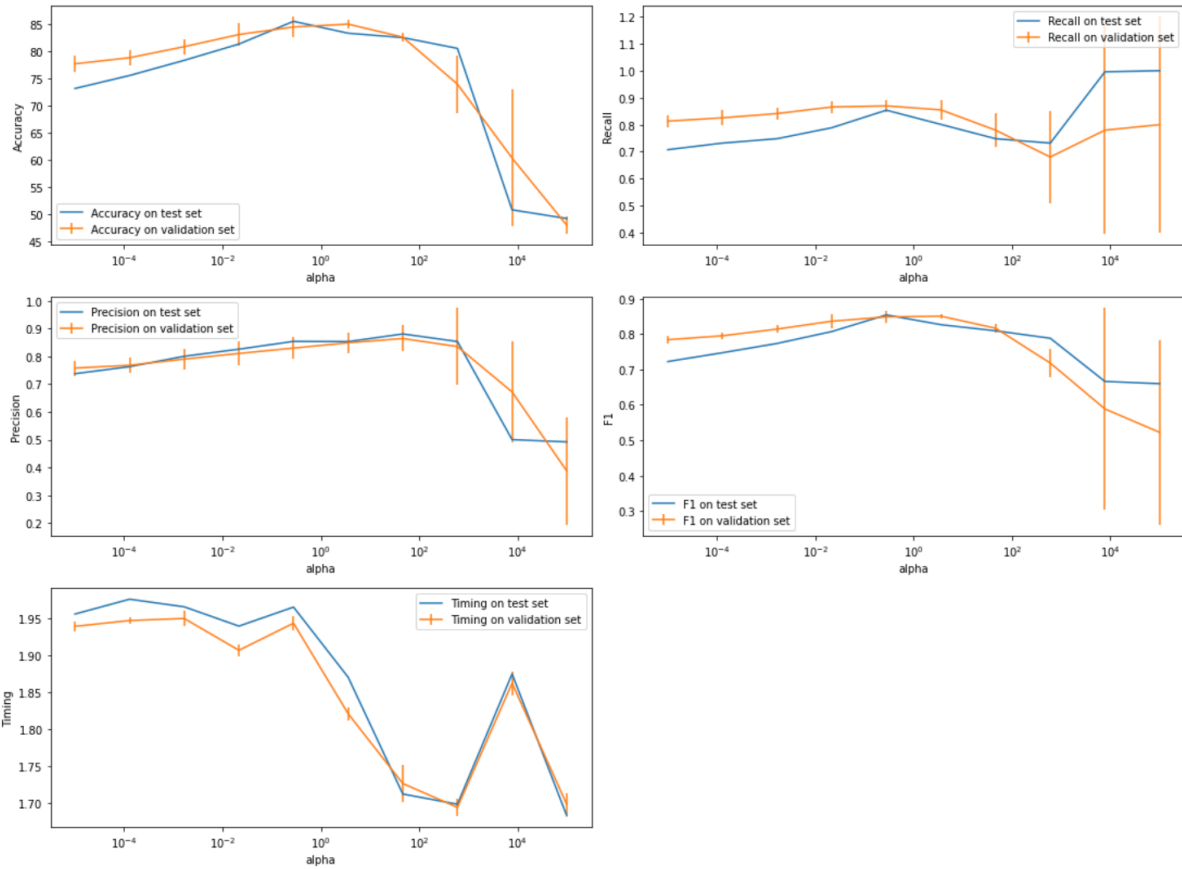
Highest average accuracy on validation set: 85.720 with alpha of 3.594

Accuracy on test set: 85.720 with alpha of 3.594

Test 9: Varying smoothing factor

Test 9 of 11

- Ngrams: 1
- Vectorizer: TF-IDF



Highest average accuracy on validation set: 85.080 with alpha of 3.594

Accuracy on test set: 85.080 with alpha of 3.594

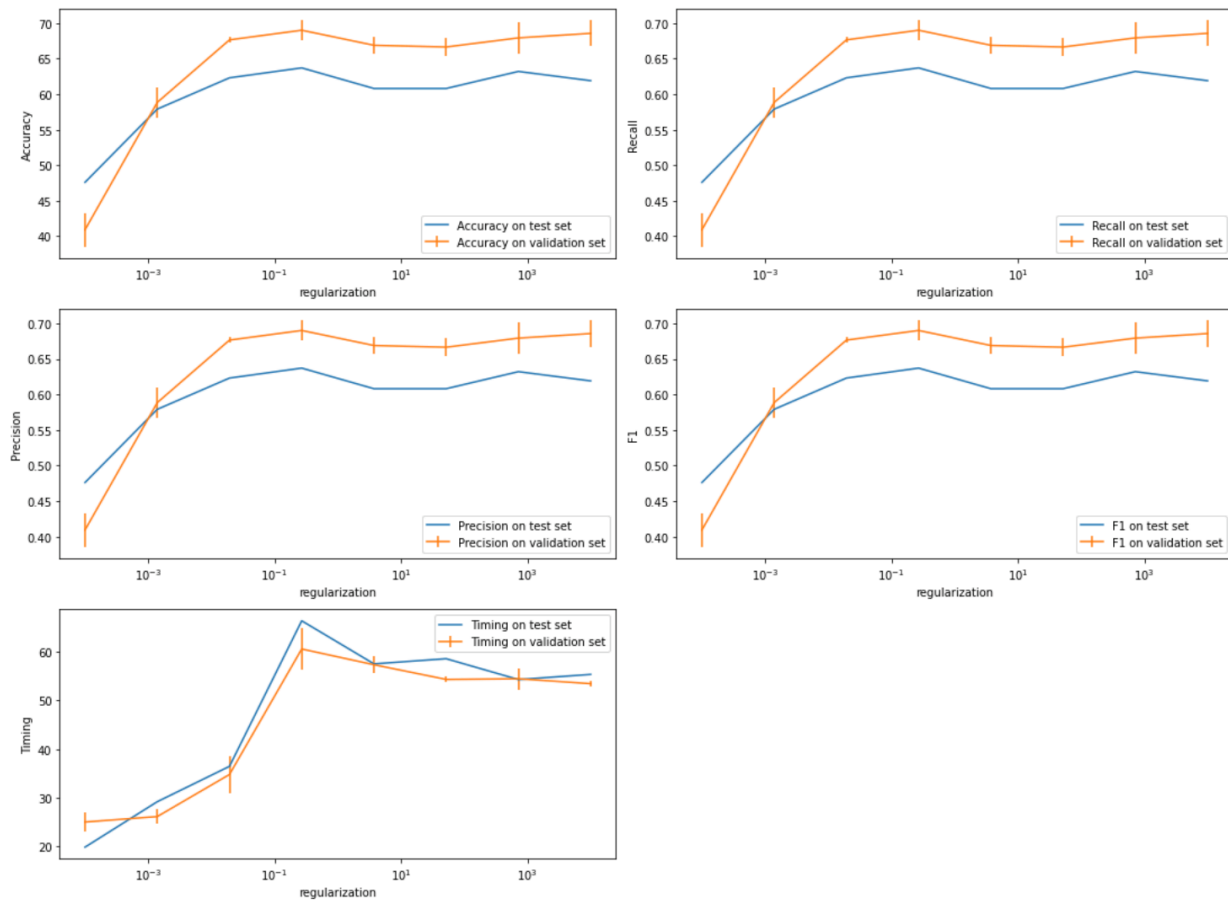
Accuracy: 83.400

2 20 News Dataset

1.1 Logistic Regression

Test 10: Varying smoothing factor

- Ngrams : 1
- Vectorizer: Count



Highest average accuracy on validation set: 69.000 with regularization of 0.268

Accuracy on test set: 69.000 with regularization of 0.268

Accuracy: 63.700

2.2 Multinomial Naive Bayes

Test 10: Varying smoothing factor

- Ngrams : 1
- Vectorizer: Count

