

Models	AGNews	Yelp Bin	Yelp Full	DBPe
Char shallow-and-wide CNN	90.7	94.4	60.3	98.0
Char-DenseNet $N_b = (4 - 4 - 4 - 4)$ Global Average-Pooling	90.4	94.2	61.1	97.7
Char-DenseNet $N_b = (10 - 10 - 4 - 4)$ Global Average-Pooling	90.6	94.9	62.1	98.2
Char-DenseNet $N_b = (4 - 4 - 4 - 4)$ Local Max-Pooling	90.5	95.0	63.6	98.5
Char-DenseNet $N_b = (10 - 10 - 4 - 4)$ Local Max-Pooling	92.1	95.0	64.1	98.5
Word shallow-and-wide CNN	92.2	<b>95.9</b>	<b>64.9</b>	<b>98.7</b>
Word-DenseNet $N_b = (4 - 4 - 4 - 4)$ Global Average-Pooling	91.7	95.8	64.5	<b>98.7</b>
Word-DenseNet $N_b = (10 - 10 - 4 - 4)$ Global Average-Pooling	91.4	95.5	63.6	98.6
Word-DenseNet $N_b = (4 - 4 - 4 - 4)$ Local Max-Pooling	90.9	95.4	63.0	98.0
Word-DenseNet $N_b = (10 - 10 - 4 - 4)$ Local Max-Pooling	88.8	95.0	62.2	97.3
bag of words	88.8	92.2	58.0	96.6
ngrams	92.0	95.6	56.3	98.6
ngrams TFIDF	92.4	95.4	54.8	<b>98.7</b>
fastText	<b>92.5</b>	95.7	63.9	98.6
char-CNN	87.2	94.7	62.0	98.3
char-CRNN	91.4	94.5	61.8	98.6
very deep char-CNN	91.3	95.7	64.7	<b>98.7</b>
Naive Bayes	90.0	86.0	51.4	96.0
Kneser-Ney Bayes	89.3	81.8	41.7	95.4
MLP Naive Bayes	89.9	73.6	40.4	87.2
Discriminative LSTM	92.1	92.6	59.6	<b>98.7</b>
Generative LSTM-independent comp.	90.7	90.0	51.9	94.8
Generative LSTM-shared comp.	90.6	88.2	52.7	95.4