Anay Karnik (2001CS05), Dhruv Chitkara (2001CS23), Harshit Singh (2001CS29)
AI lab 6 report

We build the decision tree model and train on the given dataset as suggested. Test results in the form of a classification report:

	precision	recall	f1-score	support
ABBR	1.00	0.67	0.80	9
DESC	0.84	0.97	0.90	138
ENTY	0.80	0.74	0.77	94
HUM	0.78	0.88	0.83	65
LOC	0.85	0.77	0.81	81
NUM	0.95	0.85	0.90	113
accuracy			0.85	500
macro avg	0.87	0.81	0.83	500
weighted avg	0.85	0.85	0.85	500

We achieve an F1 score of **85.**

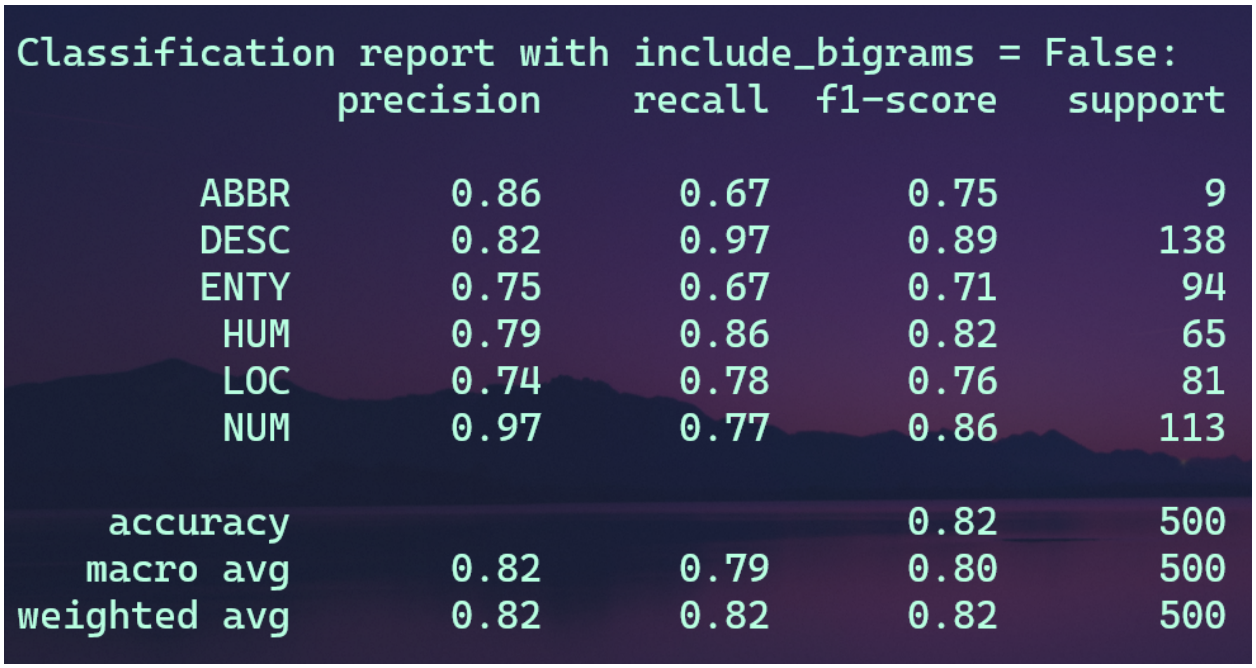
Now, we conduct ablation studies:

1. Results without including unigrams in features:

Classification report with include_unigrams = False:

	precision	recall	f1-score	support
ABBR	0.86	0.67	0.75	9
DESC	0.78	0.96	0.86	138
ENTY	0.56	0.48	0.51	94
HUM	0.57	0.74	0.64	65
LOC	0.76	0.69	0.72	81
NUM	0.89	0.66	0.76	113
accuracy			0.72	500
macro avg	0.74	0.70	0.71	500
weighted avg	0.73	0.72	0.72	500

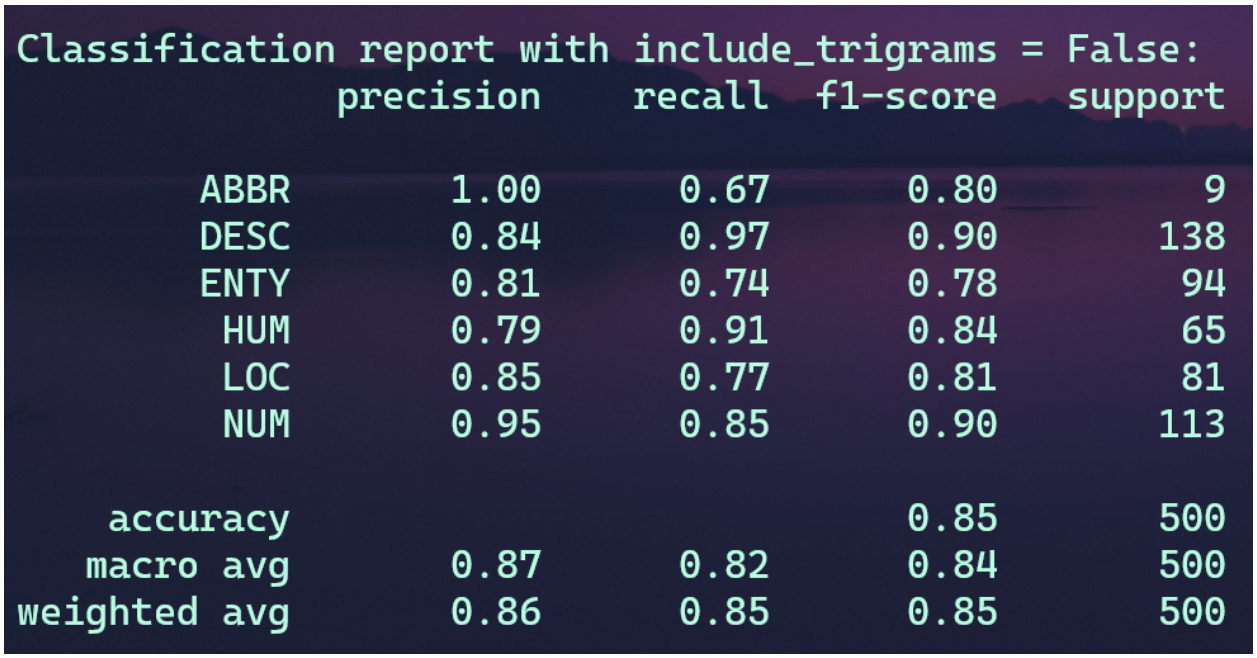
2. Results without including bigrams in features:



Classification report with include_bigrams = False:

	precision	recall	f1-score	support
ABBR	0.86	0.67	0.75	9
DESC	0.82	0.97	0.89	138
ENTY	0.75	0.67	0.71	94
HUM	0.79	0.86	0.82	65
LOC	0.74	0.78	0.76	81
NUM	0.97	0.77	0.86	113
accuracy			0.82	500
macro avg	0.82	0.79	0.80	500
weighted avg	0.82	0.82	0.82	500

3. Results without including trigrams in features:



Classification report with include_trigrams = False:

	precision	recall	f1-score	support
ABBR	1.00	0.67	0.80	9
DESC	0.84	0.97	0.90	138
ENTY	0.81	0.74	0.78	94
HUM	0.79	0.91	0.84	65
LOC	0.85	0.77	0.81	81
NUM	0.95	0.85	0.90	113
accuracy			0.85	500
macro avg	0.87	0.82	0.84	500
weighted avg	0.86	0.85	0.85	500

4. Results without including length of sentence in features:

Classification report with include_sentence_length = False:

	precision	recall	f1-score	support
ABBR	0.86	0.67	0.75	9
DESC	0.80	0.99	0.88	138
ENTY	0.77	0.68	0.72	94
HUM	0.85	0.88	0.86	65
LOC	0.89	0.79	0.84	81
NUM	0.94	0.82	0.88	113
accuracy			0.84	500
macro avg	0.85	0.81	0.82	500
weighted avg	0.85	0.84	0.84	500

5. Results without including pos tags in features:

Classification report with include_pos_features = False:

	precision	recall	f1-score	support
ABBR	0.88	0.78	0.82	9
DESC	0.82	0.97	0.89	138
ENTY	0.71	0.71	0.71	94
HUM	0.78	0.86	0.82	65
LOC	0.91	0.75	0.82	81
NUM	0.97	0.81	0.88	113
accuracy			0.83	500
macro avg	0.84	0.82	0.82	500
weighted avg	0.84	0.83	0.83	500

These results suggest that the features are important in the following order:

1. Unigrams 0.85 -> 0.72
2. Bigrams 0.85 -> 0.82
3. POS tags 0.85 -> 0.83
4. Sentence length 0.85 -> 0.84
5. Trigrams 0.85 -> 0.85

Thus, we find that **unigrams** are the most important among the features considered. Note that these results may vary based on particulars of implementation.