

REPORT - ASSIGNMENT 4

Aishik Mandal(18EC10074)

- **Feature Description :**

- Word: Lowercase version of the current word
- Suffixes: 2,3-length suffixes of the word
- Prefixes: 2,3-length prefixes of the word
- Is uppercase? (Bool)
- Is Title? (Bool)
- Is Digit? (Bool)
- EOS (End of Sentence) or BOS (Beginning of Sentence)? (Bool)
- Preceding Word: Features [a, d, e, f] for the previous word in the sentence
- Succeeding Word: Features [a, d, e, f] for the next word in the sentence

- **Hyperparameters tuned (if any) :**

Hyperparameters tuned using Sklearn function for Randomized Search using 3-fold cross-validation on the training Dataset.

	Parameters(C1,C2)	Test(f1)
Default	(0.1,0.1)	0.8405
Finetuned	(0.027825,0.00954)	0.84493

#		Train	Test
1.	10 most common transition features	VERB -> AUX 4.220089 PROPN -> PROPN 2.738212 ADJ -> NOUN 2.690111 AUX -> AUX 2.546099 PRON -> ADP 2.478454 PROPN -> ADP 2.461452 NOUN -> ADP 2.109384 NUM -> NOUN 2.103555 AUX -> SCONJ 2.073803 DET -> NOUN 1.988450	VERB -> AUX 4.922226 AUX -> AUX 3.955383 PROPN -> PROPN 3.141663 PROPN -> ADP 2.948683 NUM -> NUM 2.435607 AUX -> PUNCT 2.220667 PROPN -> PUNCT 2.026026 PROPN -> CCONJ 1.975470 NUM -> NOUN 1.946505 AUX -> PRON 1.886845
2.	10 least common transition features	PROPN -> PART -1.434184 ADV -> AUX -1.498448 PROPN -> DET -1.508396 DET -> CCONJ -1.577239 NUM -> PRON -1.675126 PROPN -> AUX -1.775343 CCONJ -> AUX -1.967885	VERB -> PRON -0.789593 PROPN -> ADJ -1.211473 VERB -> PROPN -1.256408 ADP -> PUNCT -1.482762 AUX -> VERB -1.496361 NOUN -> DET -1.516454 ADJ -> ADP -1.564752

		ADJ -> ADP -2.096025 ADJ -> PRON -2.187434 DET -> ADP -2.605487	NOUN -> NOUN -1.572386 SCONJ -> NOUN -1.643669 DET -> PROPN -1.729365
3.	Precision (per tag)	PART 0.937 CCONJ 0.960 SCONJ 1.000 ADJ 0.434 ADP 0.912 ADV 0.333 VERB 0.664 DET 0.741 NOUN 0.609 PRON 0.743 PROPN 0.398 NUM 0.876 PUNCT 0.993 AUX 0.760	PART 0.968 CCONJ 1.000 SCONJ 0.750 ADJ 0.667 ADP 0.947 ADV 0.600 VERB 0.849 DET 0.800 COMMA 0.000 NOUN 0.772 PRON 0.794 PROPN 0.655 NUM 0.917 PUNCT 1.000 AUX 0.936
4.	Recall (per tag)	PART 0.724 CCONJ 0.807 SCONJ 0.426 ADJ 0.442 ADP 0.947 ADV 0.216 VERB 0.681 DET 0.606 NOUN 0.704 PRON 0.645 PROPN 0.349 NUM 0.743 PUNCT 0.996 AUX 0.896	PART 0.909 CCONJ 1.000 SCONJ 1.000 ADJ 0.766 ADP 0.977 ADV 0.429 VERB 0.798 DET 0.889 COMMA 0.000 NOUN 0.866 PRON 0.831 PROPN 0.497 NUM 0.880 PUNCT 0.830 AUX 0.942
5.	F-Score (per tag)	PART 0.817 CCONJ 0.877 SCONJ 0.598 ADJ 0.438 ADP 0.929 ADV 0.262 VERB 0.672 DET 0.667 NOUN 0.653 PRON 0.691 PROPN 0.372 NUM 0.804 PUNCT 0.995 AUX 0.823	PART 0.937 CCONJ 1.000 SCONJ 0.857 ADJ 0.713 ADP 0.962 ADV 0.500 VERB 0.823 DET 0.842 COMMA 0.000 NOUN 0.817 PRON 0.812 PROPN 0.565 NUM 0.898 PUNCT 0.907 AUX 0.939
6.	Overall Accuracy	0.84493	0.71514