REPORT - ASSIGNMENT 4
Name: Debjoy Saha, Roll No. 18EC30010

- **Feature Description :**
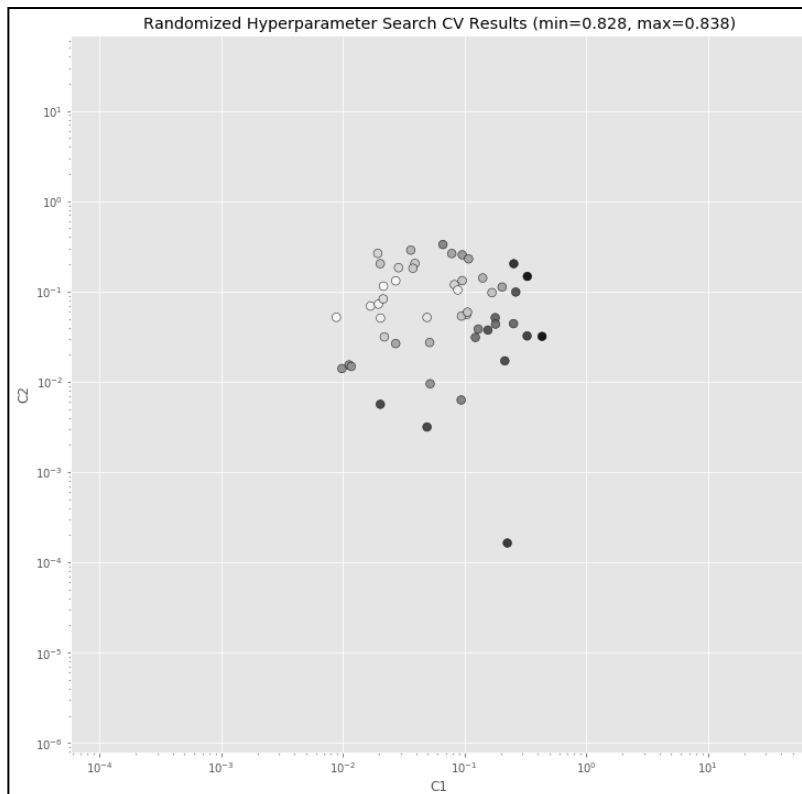   a. Index: Position of the word in the sentence
    b. Word: Lowercase version of the current word
    c. Suffixes: 1, 2, 3-length suffixes of the word
    d. Prefixes: 1, 2, 3-length prefixes of the word
    e. Is Uppercase? (bool)
    f. Is Title? (bool)
    g. Is Digit? (bool)
    h. EOS(End of a sentence) or BOS(Beginning of a sentence)? (bool)
    i. Preceding Word: Features [b, e, f, g] for the previous word in the sentence
    j. Succeeding Word: Features [b, e, f, g] for the next word in the sentence

- **Hyperparameters tuned:**
    Hyperparameter tuned using Sklearn function for Randomized Search using 3-fold cross-validation on the training dataset.

	Parameters(C1, C2)	Train(acc, f1)	Test(acc, f1)
Default	(0.1, 0.1)	(0.982, 0.982)	(0.851, 0.849)
Finetuned	(0.009, 0.05)	(0.988, 0.988)	(0.843, 0.839)

Parameter Search Space: (_CV score colour: White = 0.82(Min), Black = 0.84(Max))_

Features and Performance

#		Train		Test	
1.	10 most common transition features	VERB -> AUX 4.03 PROPN -> PROPN 3.10 AUX -> AUX 2.59 PART -> NUM 2.23 DET -> NOUN 2.06 ADJ -> NOUN 2.011 VERB -> SCONJ 1.99 PROPN -> ADP 1.88 ADP -> ADP 1.80 PRON -> ADP 1.78		VERB -> AUX 3.04 AUX -> AUX 2.68 PROPN -> PROPN 2.34 PROPN -> ADP 2.05 NUM -> NUM 1.85 AUX -> PUNCT 1.82 VERB -> PUNCT 1.74 PART -> NUM 1.57 ADV -> DET 1.57 PROPN -> CCONJ 1.48	
2.	10 least common transition features	NUM -> PRON -1.90 DET -> CCONJ -2.01 ADV -> ADV -2.27 CCONJ -> PART -2.58 CCONJ -> AUX -2.74 DET -> ADP -2.88 ADJ -> PRON -2.96 ADV -> AUX -3.01 ADJ -> ADP -3.02 ADP -> COMMA -3.55		AUX -> VERB -0.88 DET -> ADP -1.01 NOUN -> DET -1.22 VERB -> PROPN -1.23 DET -> PROPN -1.35 ADP -> PUNCT -1.36 PROPN -> AUX -1.39 NOUN -> NOUN -1.42 PROPN -> ADJ -1.67 ADJ -> ADP -1.82	
3.	Precision (per tag)	X PART CCONJ SCONJ ADJ ADP ADV VERB DET COMMA NOUN PRON PROPN NUM PUNCT AUX	1.000 1.000 0.993 1.000 0.991 0.989 0.910 0.995 0.950 1.000 0.992 0.988 0.991 1.000 1.000 0.971	X PART CCONJ SCONJ ADJ ADP ADV VERB DET COMMA NOUN PRON PROPN NUM PUNCT AUX	0.000 1.000 1.000 0.750 0.731 0.942 0.529 0.825 0.821 0.000 0.773 0.930 0.596 0.840 1.000 0.910
4.	Recall (per tag)	X PART CCONJ SCONJ ADJ	1.000 1.00 0.993 1.000 0.991	X PART CCONJ SCONJ ADJ	0.000 1.000 1.000 1.000 0.723

		ADP	0.989	ADP	0.967
		ADV	0.910	ADV	0.429
		VERB	0.995	VERB	0.859
		DET	0.950	DET	0.889
		COMMA	1.000	COMMA	0.000
		NOUN	0.992	NOUN	0.873
		PRON	0.988	PRON	0.815
		PROPN	0.991	PROPN	0.472
		NUM	1.000	NUM	0.840
		PUNCT	1.000	PUNCT	0.836
		AUX	0.971	AUX	0.957
5.	F-Score (per tag)	X	1.000	X	0.000
		PART	1.000	PART	1.000
		CCONJ	0.997	CCONJ	1.000
		SCONJ	1.000	SCONJ	0.857
		ADJ	0.991	ADJ	0.727
		ADP	0.991	ADP	0.954
		ADV	0.914	ADV	0.474
		VERB	0.977	VERB	0.842
		DET	0.966	DET	0.853
		COMMA	1.000	COMMA	0.000
		NOUN	0.992	NOUN	0.820
		PRON	0.978	PRON	0.869
		PROPN	0.989	PROPN	0.527
		NUM	1.000	NUM	0.840
		PUNCT	1.000	PUNCT	0.911
		AUX	0.984	AUX	0.933
6.	Overall Accuracy	0.988		0.843	