

(M) A Splitting Disagreement (1/1) [Solution]

Q1. Match each algorithm in the table (“Alg A” to “Alg C”) to its name (“Baseline”, “Pavan’s algorithm”, or “Arun’s algorithm”): **Baseline = A** **Pavan = C** **Arun = B**

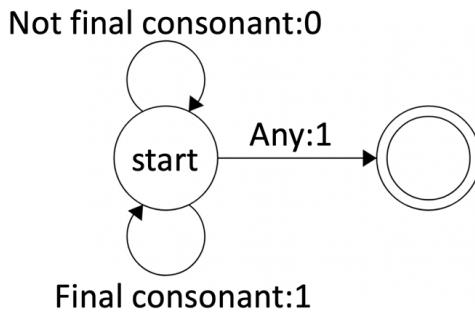
Q2. The output is still the same: [0,0,1,0,0,1,0,0,0,0,0,1]

But now the target is: [0,0,1,0,0,1,0,0,0,1,0,0,1]

So we have: TP: 3, FP: 0, FN: 1.

And thus **F1-score of Arun’s algorithm = 0.86**

Q3. (a) Pavan’s algorithm:



(b) Arun’s algorithm: This one is tricky. It needs to handle the scenario where the last V in one V-FC-IC-V template is the first V in another. It also needs to handle cases where the last V in a V-FC-IC-V template is the end of the input; or where V-FC-IC-V is started but not finished, with the sequence ending partway through. Abbreviations used here: Punct = punctuation; FC = final consonant. IC = initial consonant. V = vowel.

