PREDICTION MODELS BASED ON MAX-STEMS

(or harnessing imbalanced data)

Episode Four: Advanced Examinations

Ahmet Furkan EMREHAN

(matahmet@gmail.com)

PREDICTION MODELS BASED ON MAX-STEMS

- Episode One: One-Word Based
- Episode Two: A Combinatorial Approach
- Episode Three: Effect of Hyperparameters
- Episode Four: Advanced Examinations

INTRODUCTION

- In this chapter, I examine some cases:
- A) How do Accuracy Rates of Categories change, if the sample is changed ?
- B) Let’s compare all methods for each category
- C) Can a Trigonometric Approach contribute to hyperparametic model ?

A) How do Accuracy Rates of Categories change, if the sample is changed ?

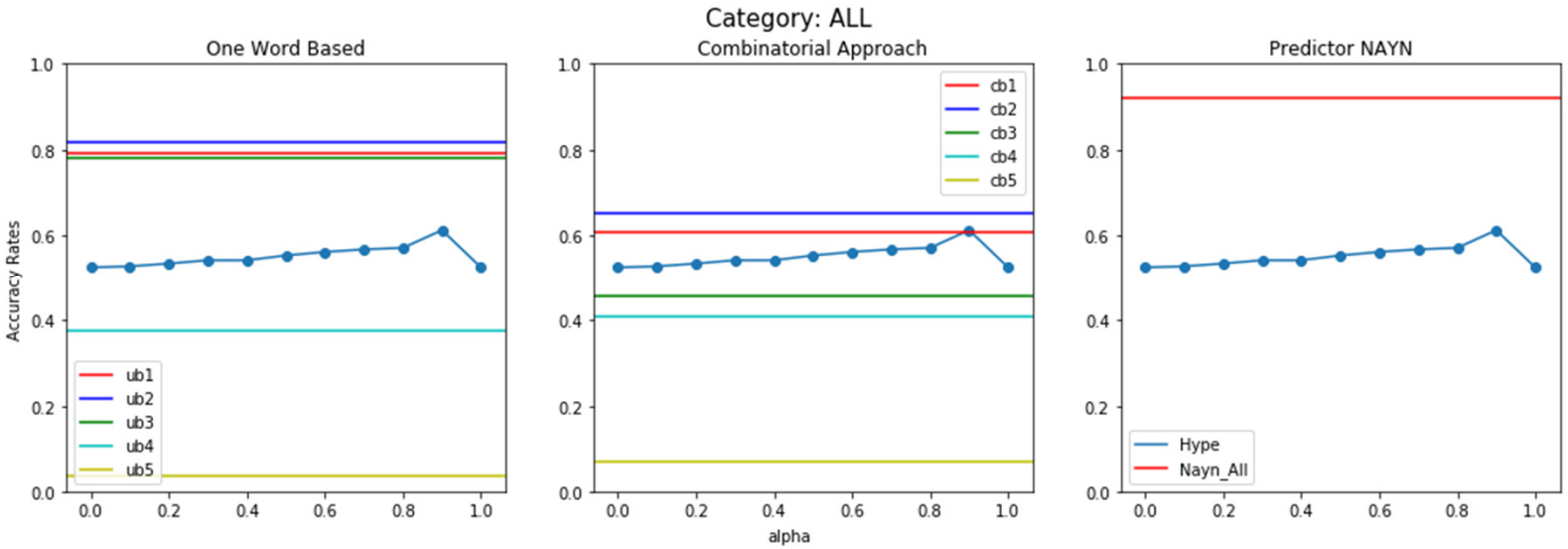
(Based on Predict Model 1 (On Word))

Accuracy Rates of Cateagories				
Sample	DÜNYA	SPOR	SANAT	Teknoloji
1	0.791	0.167	0.029	0.013
2	0.798	0.166	0.025	0.012
3	0.783	0.181	0.023	0.013
4	0.794	0.169	0.024	0.013
5	0.792	0.172	0.024	0.012
6	0.793	0.175	0.024	0.009
7	0.797	0.172	0.022	0.010
8	0.809	0.159	0.022	0.010
9	0.793	0.175	0.024	0.009
10	0.797	0.172	0.022	0.010

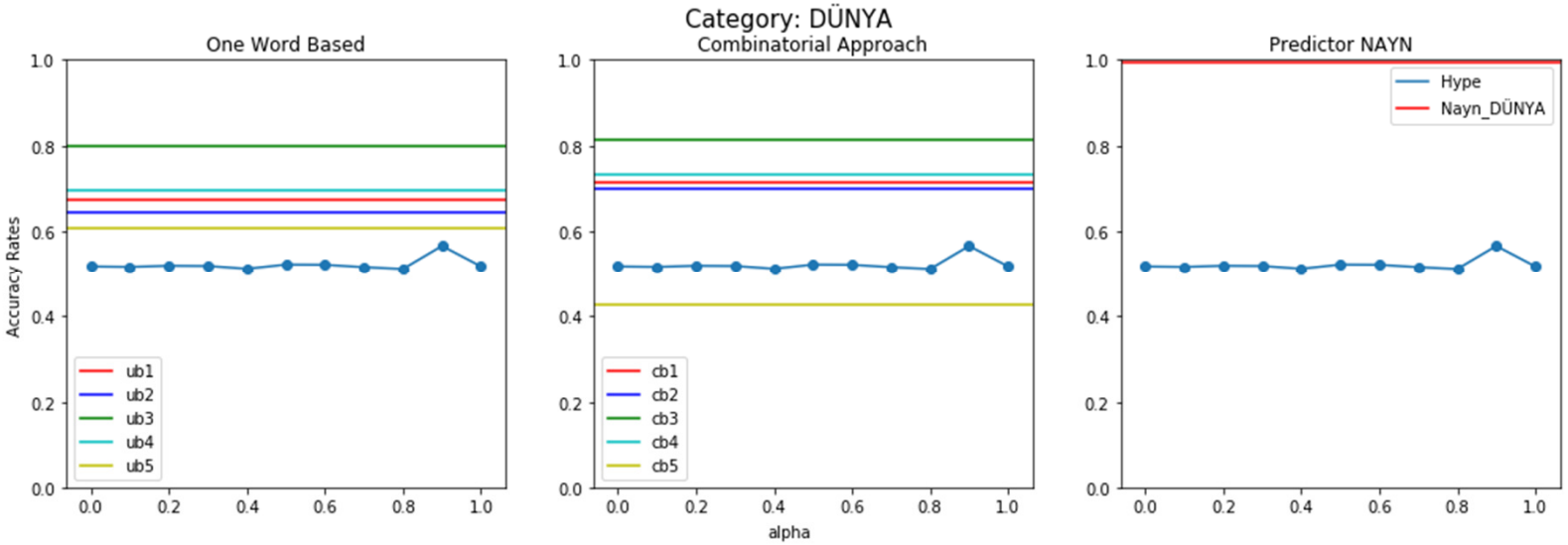
Result: Stable against samples

Note: Naturally Samples insersect wit each other at %80 level.

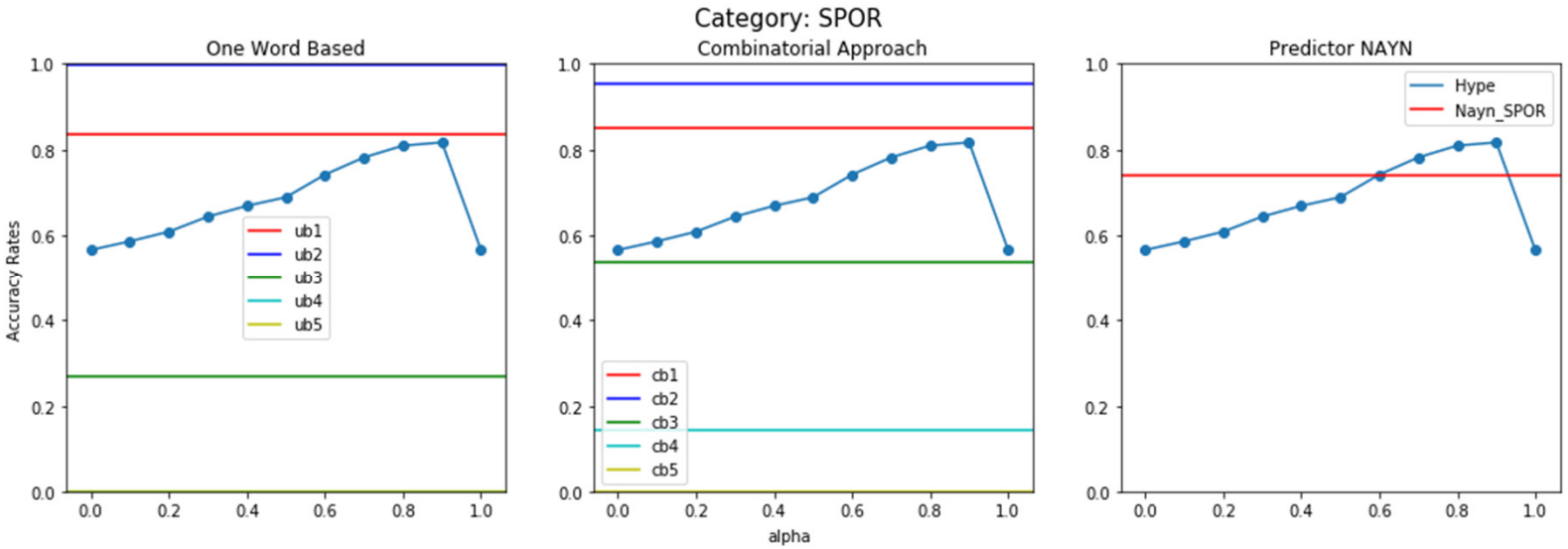
B) Let’s compare all methods for each category



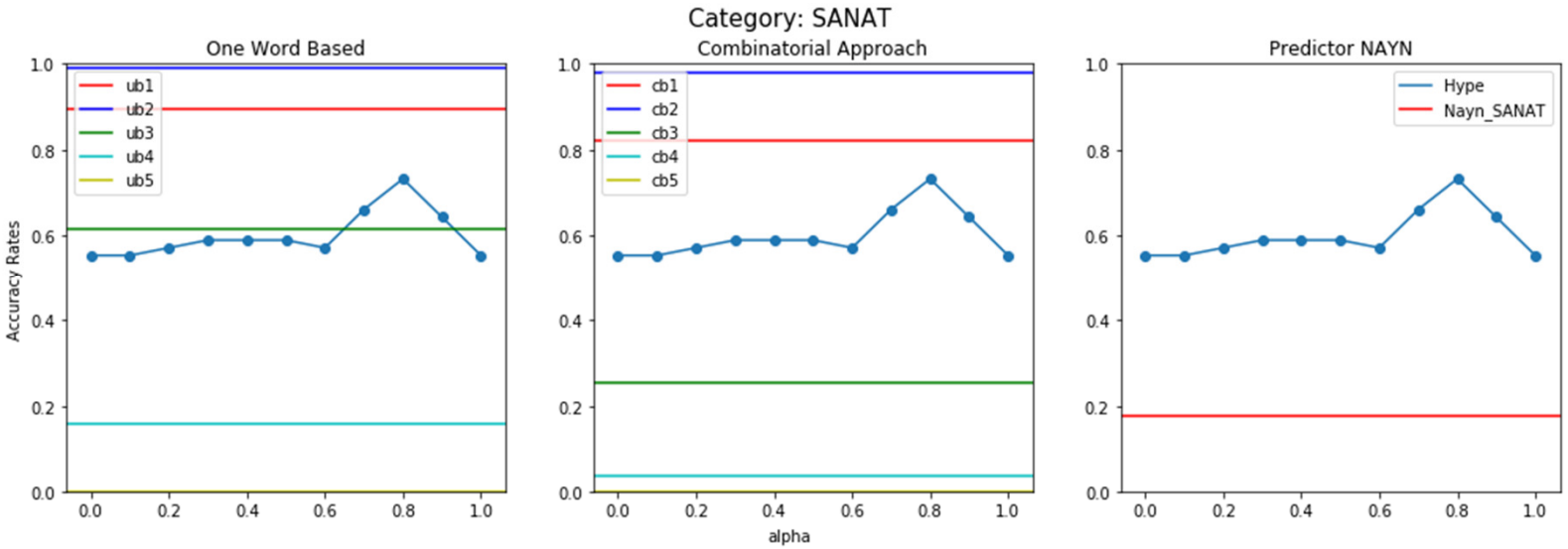
B) Let’s compare all methods for each category



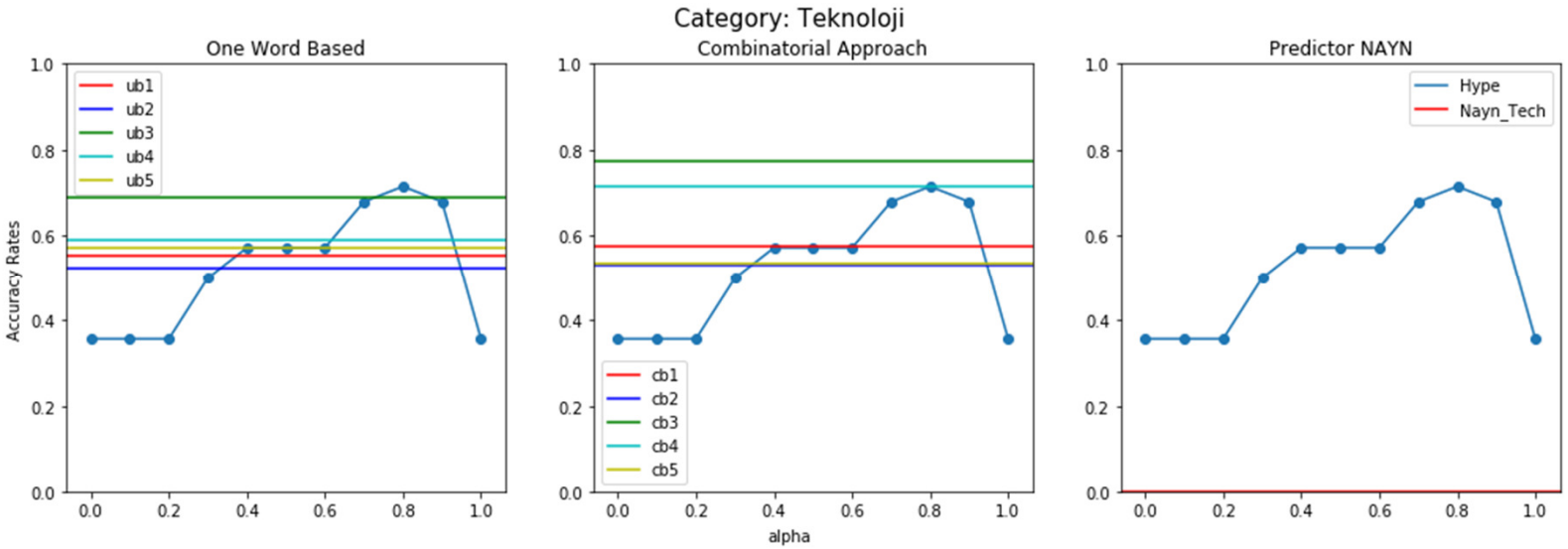
B) Let’s compare all methods for each category



B) Let’s compare all methods for each category



B) Let’s compare all methods for each category



B) Let’s compare all methods for each category

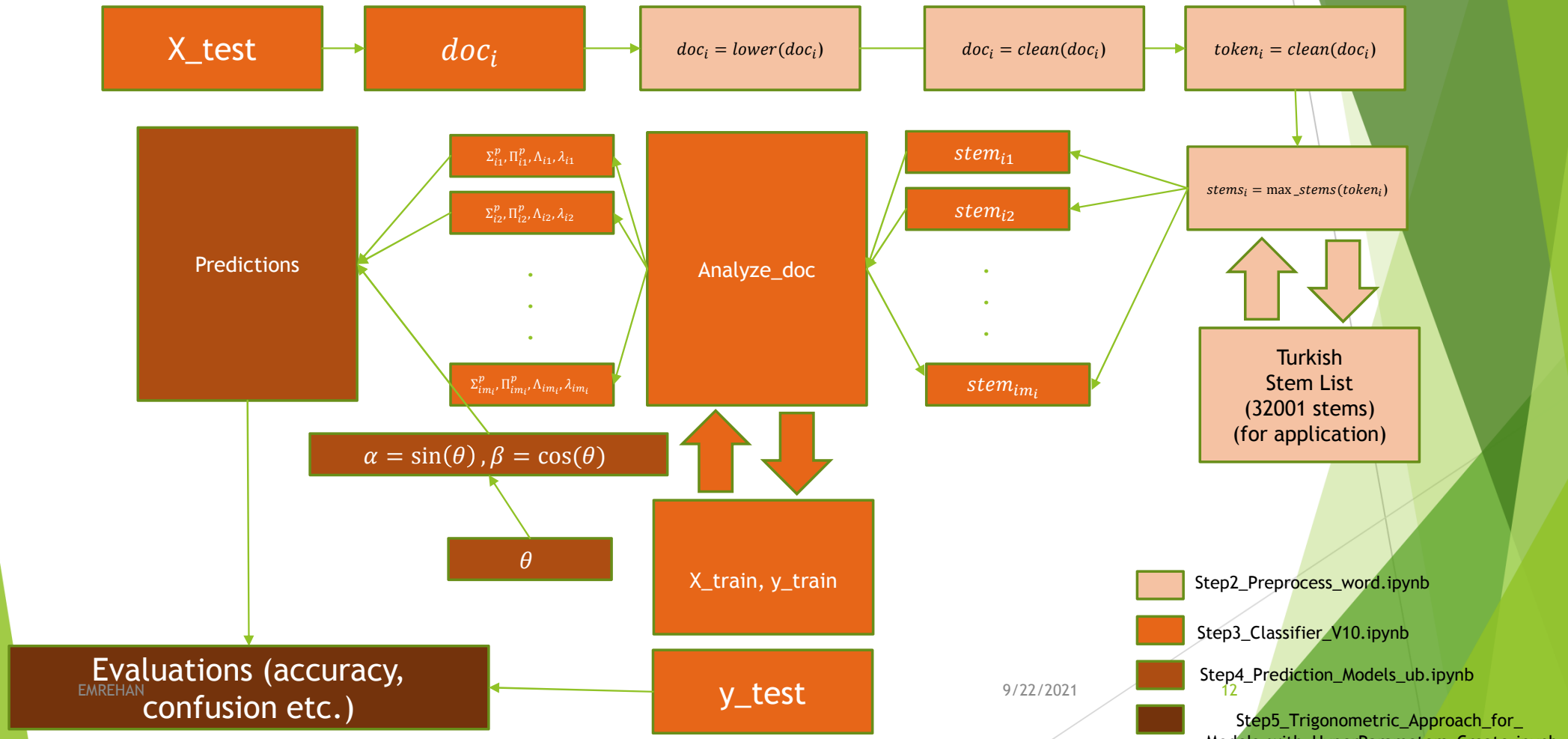
- Conclusion:
- Whereas NAYN predictor is successful at predicting documents labelled with «DÜNYA», other prediction models (one stems and combinatorial approach) is more successful at predicting documents labelled with other categories.

C) Can a Trigonometric Approach contribute to hyperparametic model ?

- In this section we use trigonometric approach to hyperparameters such that

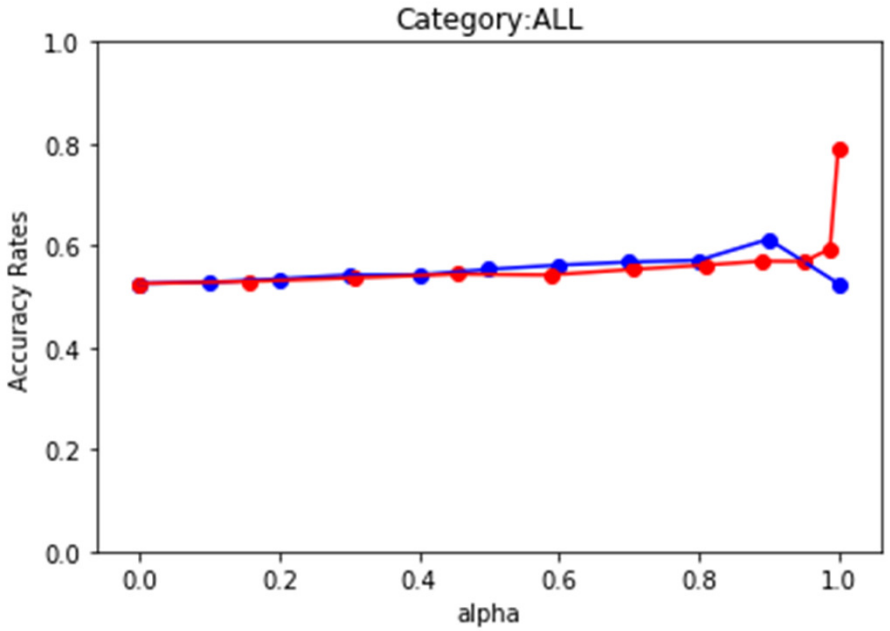
$$\alpha = \sin(\theta)\ and\ \beta = \cos(\theta)\ where\ \theta \in \left[0, \frac{\pi}{2}\right]$$

General Scheme for Application of Trigonometric Approach to Prediction Models with hyperparameter

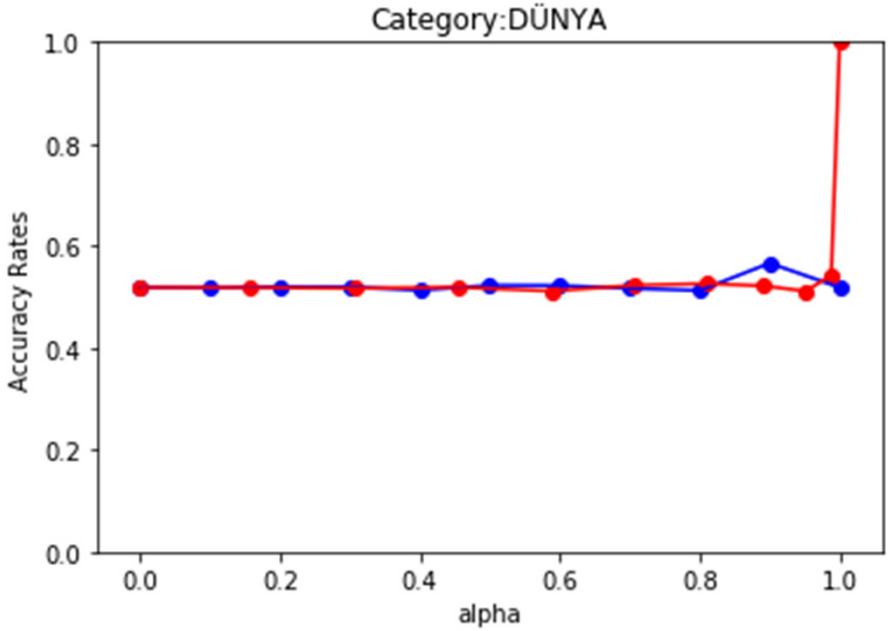


C) Can a Trigonometric Approach contribute to hyperparametic model ?

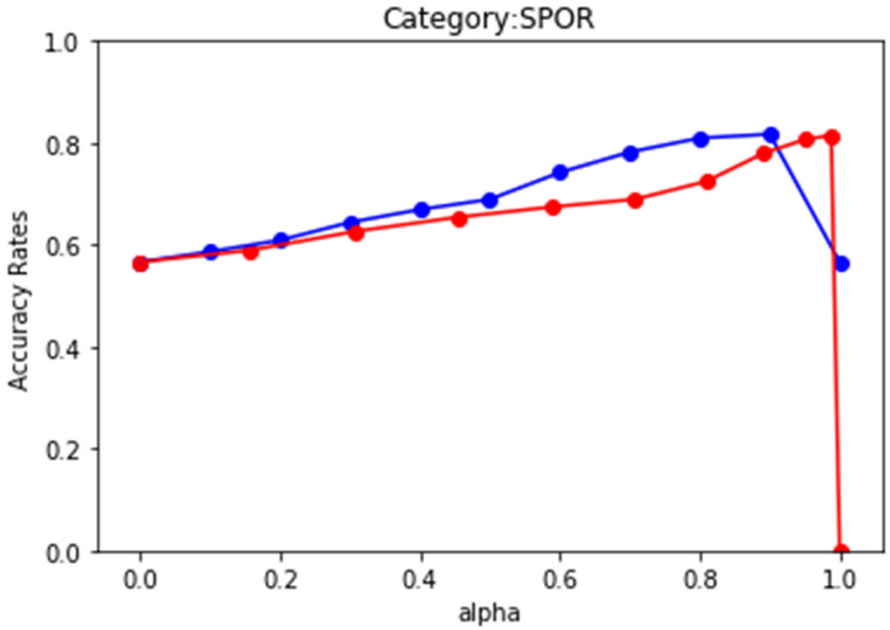
- Results



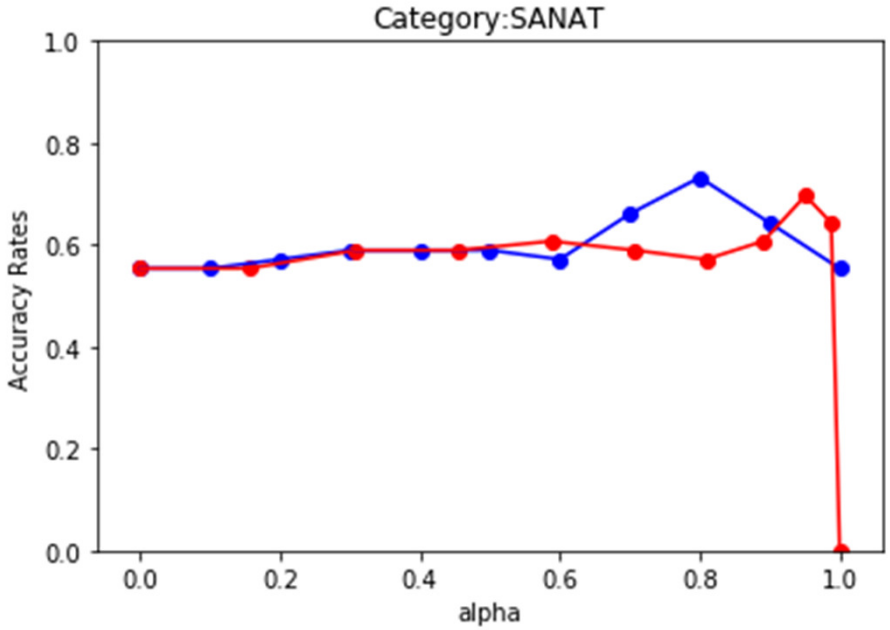
C) Can a Trigonometric Approach contribute to hyperparametic model ?



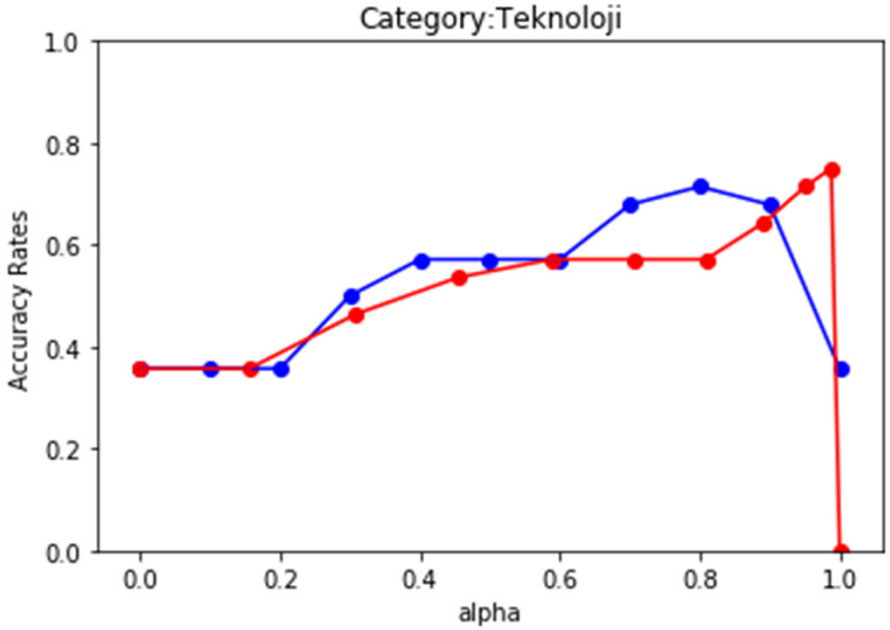
C) Can a Trigonometric Approach contribute to hyperparametic model ?



C) Can a Trigonometric Approach contribute to hyperparametic model ?



C) Can a Trigonometric Approach contribute to hyperparametic model ?



C) Can a Trigonometric Approach contribute to hyperparametic model ?

- Conclusion:
- 1) When $\theta = \frac{\pi}{2} \rightarrow \alpha = 0,\ \beta = 1$, model produces only «DÜNYA» label, then specificity fails. In other saying, predictions of other categories in test set become «DÜNYA»
- 2) Generally, performance of model based trigonometric functions falls behind that of model based uniform distribution in almost all categories. Only at prediction of documents labelled «Teknoloji», performance of trigonometric approach is a notch better than uniform approach.

THE END