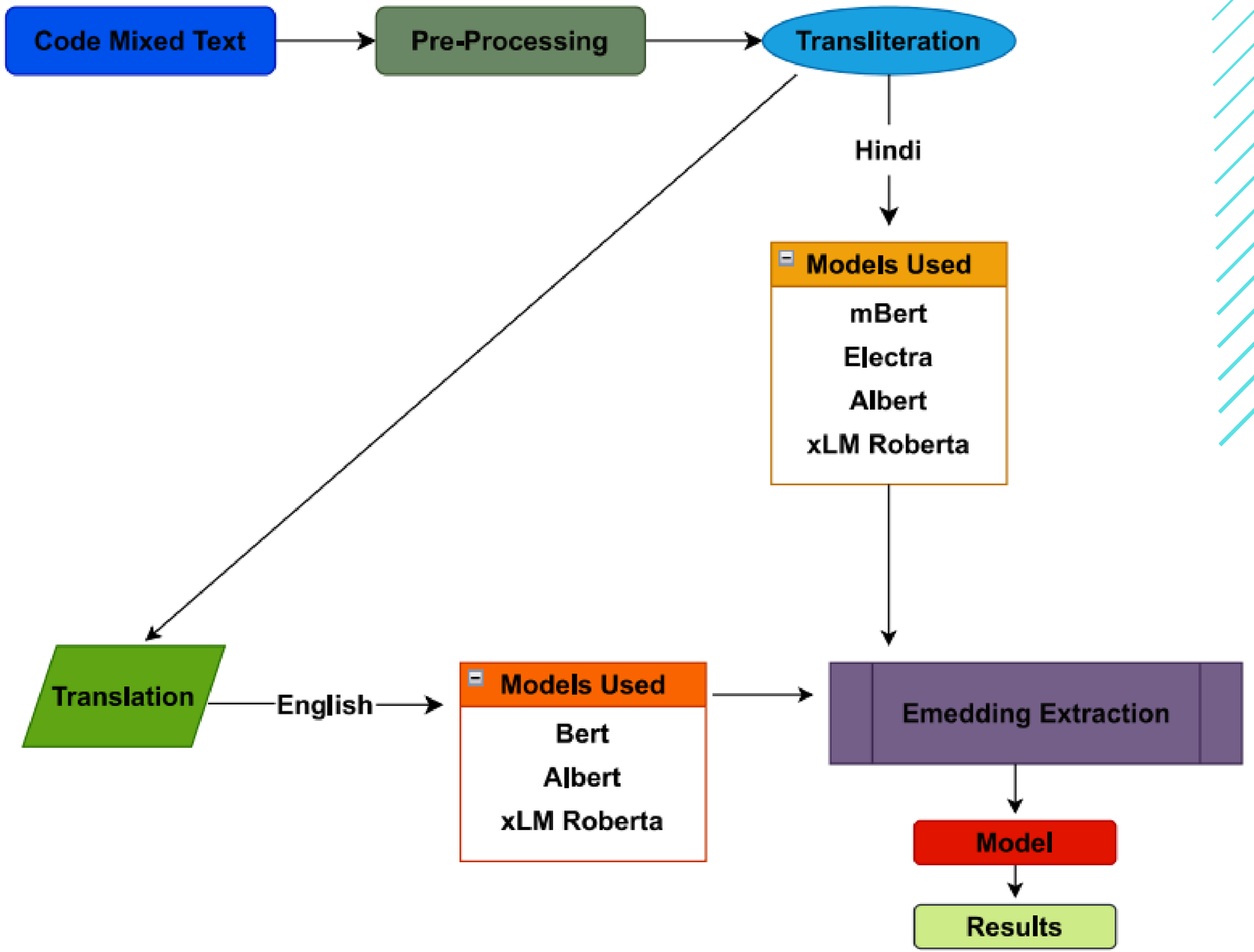
Indian Institute of Information Technology
Dharwad

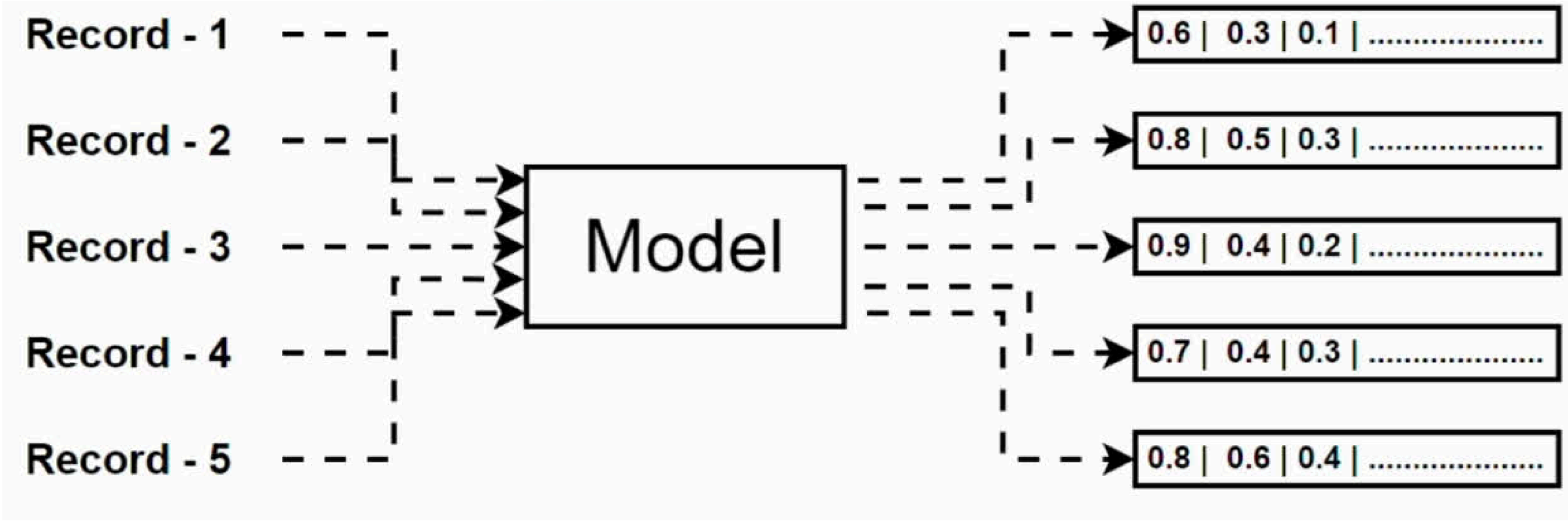
MACHINE LEARNING

FAKE HATE: INVESTIGATING THE ROLE OF FAKE NARRATIVES IN SPREADING HATEFUL REACTIONS

FLOW CHART OF THE MODEL

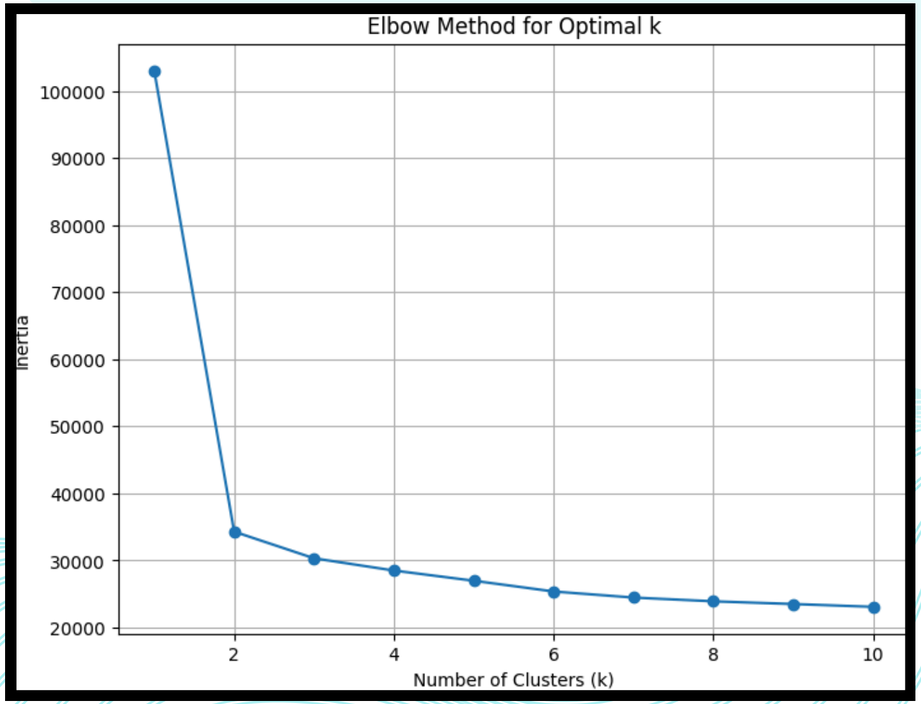


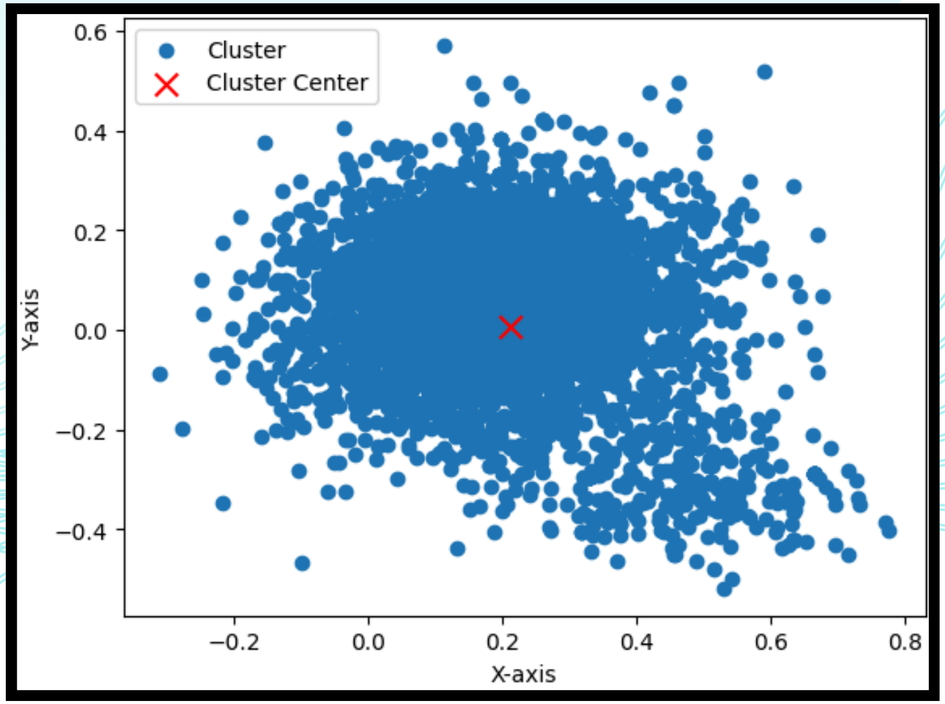
MODEL FLOW


Record - 1
Record - 2
Record - 3
Record - 4
Record - 5
Model
0.6 | 0.3 | 0.1 | .....
0.8 | 0.5 | 0.3 | .....
0.9 | 0.4 | 0.2 | .....
0.7 | 0.4 | 0.3 | .....
0.8 | 0.6 | 0.4 | .....

CLUSTER FORMATION AND CALCULATING THE DISTANCES

- We utilized the HASOC-2021, Constraint_Hindi datasets to create clusters for hate and fake speech analysis respectively.
- Creation of clusters followed by measurement of cluster centers.
- Implementation of a normal distribution for outlier detection; data points beyond the range of (μ - 3σ to μ + 3σ) are identified as outliers and subsequently removed.



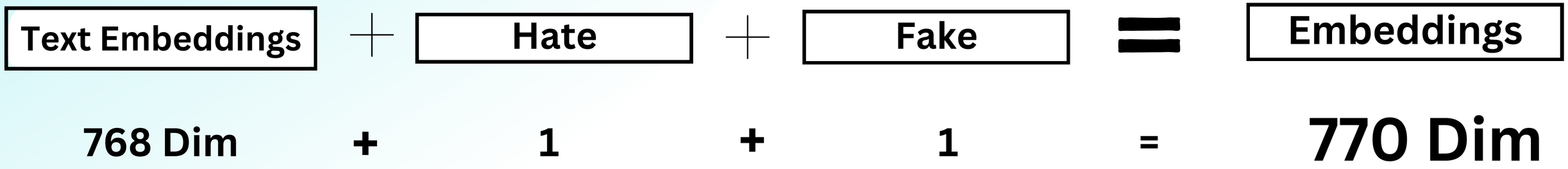


DISTANCE CALCULATION FOR OUR DATA

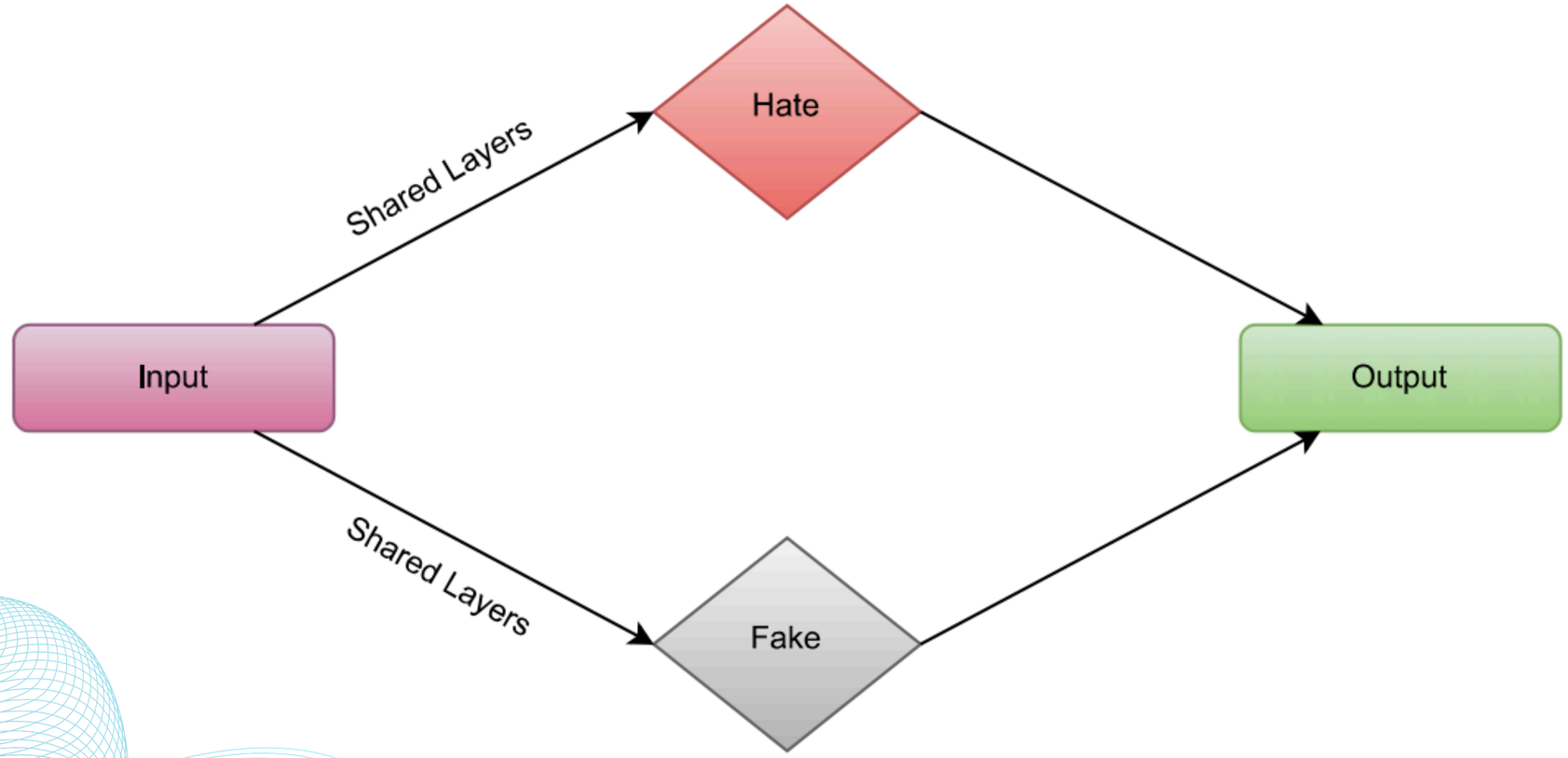
- Following the creation of clusters, we processed the data using embeddings, computing the distance of each record to the respective cluster center, and subsequently saved this information in a file.
- This procedure was applied to both translated and transliterated text datasets.

Pre_Processed_English_text	Fake	Hate	Distance_Eng_Fake	Distance_Eng_Hate
Just like coronavirus spread in China, students were brought in.	0	0	13.33701992	4.63313961
From its talk it appears that it is not the coronavirus but a scam that the Congress has done in collaboration with China	1	1	13.68127251	4.012666225

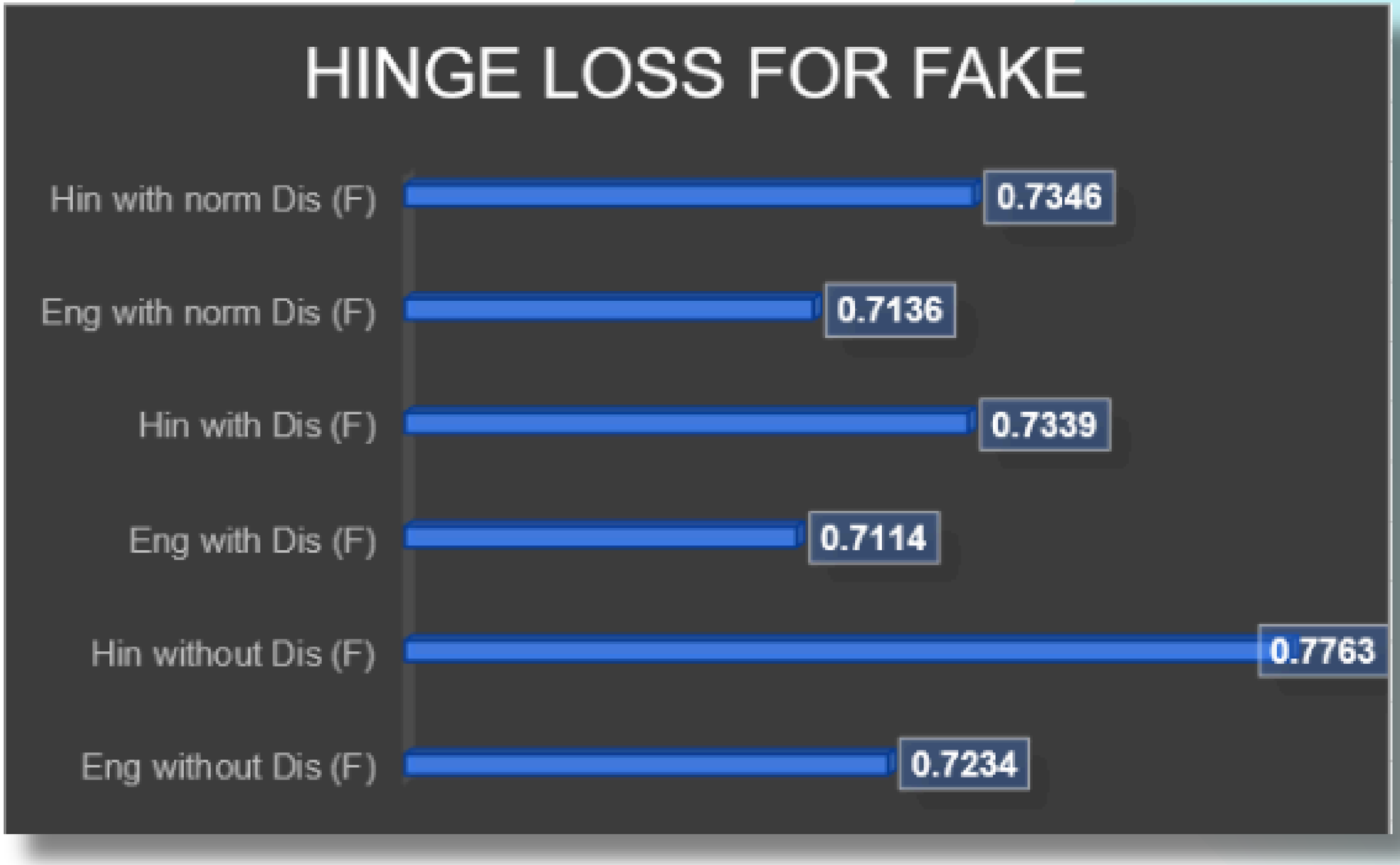
DIMENSIONS



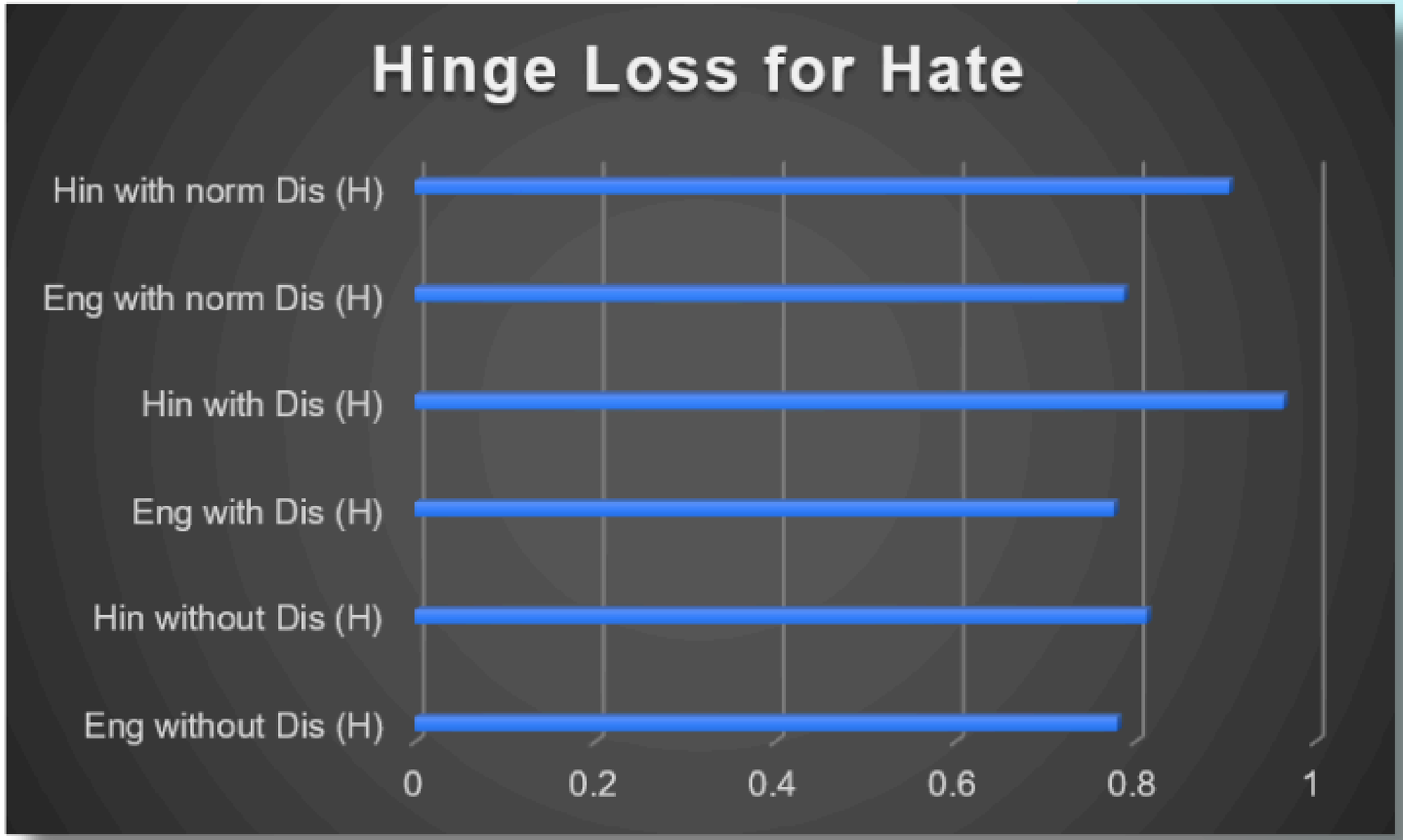
MODEL FLOW OF SHARED LAYERS



RESULTS


HINGE LOSS FOR FAKE
Hin with norm Dis (F)
0.7346
Eng with norm Dis (F)
0.7136
Hin with Dis (F)
0.7339
Eng with Dis (F)
0.7114
Hin without Dis (F)
0.7763
Eng without Dis (F)
0.7234

RESULTS


Hinge Loss for Hate
Hin with norm Dis (H)
Eng with norm Dis (H)
Hin with Dis (H)
Eng with Dis (H)
Hin without Dis (H)
Eng without Dis (H)
0
0.2
0.4
0.6
0.8
1