Uptrain.ai Hackathon

Submitted By-
Taresh Rajput
Keshav Tarun

Dataset - We hand-crafted our own training dataset and tried to include a diverse array of adjectives that can be considered in describing nike shoes.

Signal - In order to identify the edge cases, we used a combination of two signals -

"**nike_text_present**" : This signal determines if the samples contain the word "nike" or not. It ensures that the training sample is in the context of nike.

"**Is_negative_sentiment**" : This signal identifies the sentiment of the training sample. We used the SentimentIntensityAnalyzer from the nltk library. If the polarity score of a training sample comes below 0.20, it is identified as negative sentiment. Additionally, we checked for certain negative adjectives in the training sample. If the condition is met, then too that particular sample is identified as negative sentiment.

Using the uptrain framework, we provided a combination of the "nike_text_present" and negation of "Is_negative_sentiment" signal (negation as we want our training dataset to discard negative adjectives in context of nike shoes) to the "signal_present" argument in the cnfg used in the uptrain framework module.

We also spent a considerable time **tuning various hyperparameters** used in the model.

We didn't know about the **perplexity score** which is used in this model as a metric so we spent a considerable amount of time researching about this metric and came up with many insights.

We are grateful for the opportunity provided by the hackathon organizer to work on such an intriguing and challenging task. It not only allowed us to learn something new and specific to the industry, but also gave us the chance to delve deeper into the perplexity score metric. Our diligent research efforts have resulted in a wealth of valuable insights. Thank you for this enriching experience.