Indice:

Introduzione

1. Analisi del dataset
   1. Descrizione generale dei dati
      1. Descrizione generale feature (Quali sono, dimensione, numero di duplicati e nulli).
      2. Numero di elementi singoli da cui fare inferenze probabilistiche, spiegazione della sarcastic proportion e info rate
   2. Analisi del target (distribzione, bilanciamento (citando il paper))
   3. Analisi delle feature context
      1. Descrizione delle distribuzioni delle feature del context, affrontate singolarmente
      2. Descrizione del sarcastic proportion singolarmente
      3. Aggregazione e conclusioni sulle feature singole
   4. Analisi del testo
      1. Analisi delle distribuzioni dei singoli tipi di testo
      2. Descrizione del sarcastic proportion di ogni tipo di testo (ricorda approfondimento sulla punteggiatura)
      3. Aggregazione e conclusioni sul tipo di testo da adottare
   5. Analisi delle lunghezze del testo e del parent
   6. Preparazione dei dati
2. Modello
   1. Descrizione generale del modello adottato (spiegando un minimo cosa sono gli encoder, e perché sono stati scelti rispetto alle alternative (LSTM bidirezionali)
   2. Descrizione specifica del modello usato (spiegando specificatamente i vari rami a cosa servono, e in particolare il bert)
   3. Training:
      1. Divisione dei dati
      2. Funzionamento del training
      3. Problemi affrontati (spiegando perché non è stato possibile fare hypertuning)
3. Risultati
   1. Risultati con commento del training
   2. Risultati con commento del validation
   3. Risultati con commento del testing

Conclusioni

Appendice A: Elaborazione, modello e altri dettagli tecnici

* Descrizione organizzazione del file di progetto
* Descrizione sul sistema di training

Appendice B: Applicazione Dash

* Descrizione breve dell’organizzazione della pagina di dash
  + (note su problemi affrontati)
* Descrizione breve di ogni pagina