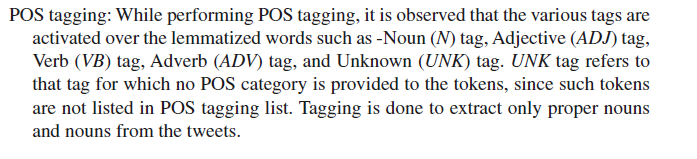
* **Medical persona classification in social media (2017):**

Dataset:

* + gathered using Twitter API.
  + 50 blogs and 30 tweets per query (50 queries).
  + A total of 1581 blogs and 1025 tweets were annotated.

Feature extraction:

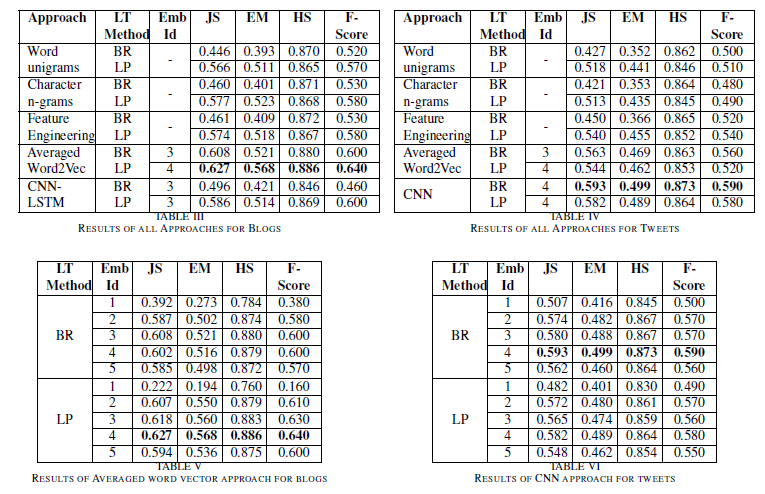
* + Label Powerset (LP) and Binary Relevance (BR) ->multi-label text classification.
  + TF-IDF
  + POS



Model selection:

SVM

CNN-LSTM



* **Persona classification of celebrity Twitter users (2020):**

Dataset:

* + gathered using Twitter API (Tweepy).

A table of numbers and a number of tweets

Description automatically generated A table with numbers and text

Description automatically generated

Feature extraction:

* + Finding topic words using LDA
  + Finding Hypernyms of topic words

Model selection:

* + Naïve Bayes -> best result
  + Decision Tree (Gini Index)
  + SVM
* **Persona traits identification based on Myers-Briggs Type Indicator (MBTI)-a text classification approach (2018)**

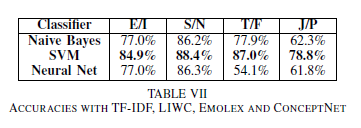
Dataset: tweets tagged with one of the 16 MBTI types (8660 rows).

Feature extraction:

* + Lemmatization, stemming
  + TF-IDF
  + LIWC (93 features)
  + emoSenticNet (10 emotions)
  + ConceptNet (300 features)
  + Dimensionality reduction using SVD

Model selection:

* + Naïve Bayes
  + NN
  + SVM



A table with numbers and text

Description automatically generated