

A Survey on the Recent Techniques for Sentiment Analysis on Twitter

Beril Besbinar, Dimitrios Sarigiannis, Panayiotis Smeros
(name.surname@epfl.ch)
École Polytechnique Fédérale de Lausanne (EPFL)

1. Problem

- Identify the **sentiment** of a group of people concerning a specific **topic** (based on hashtags and keywords)
- The **input** is a set of tweets and the **output** is their (positive or negative) sentiment

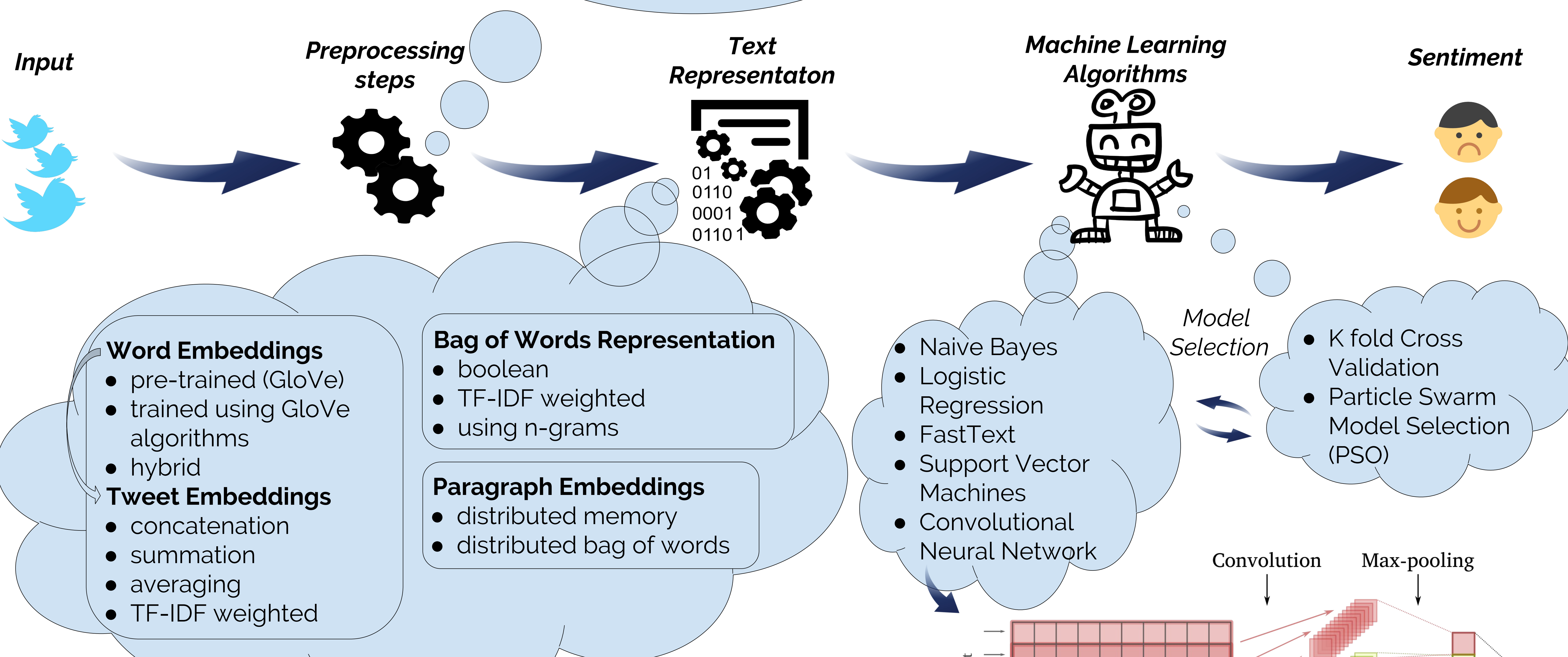
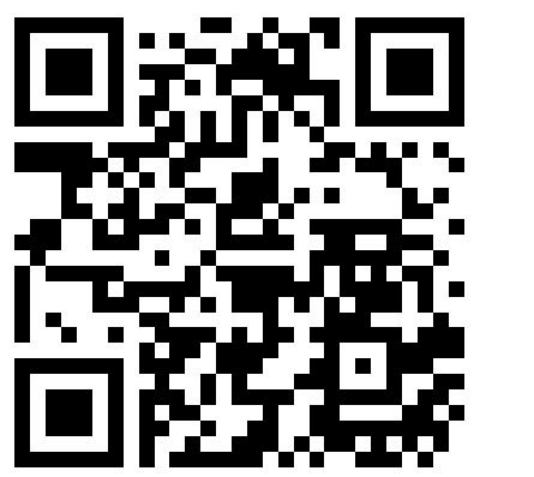
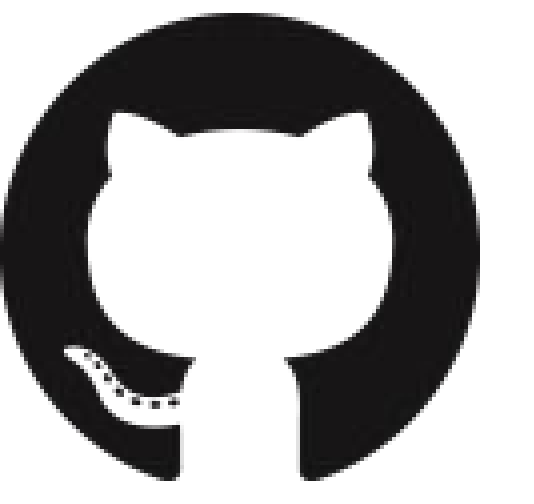
2. Use Cases

- **Prediction** of elections result
- **Public opinion** on global issues (e.g., refugees crisis, climate change)
- **Promotion** of products

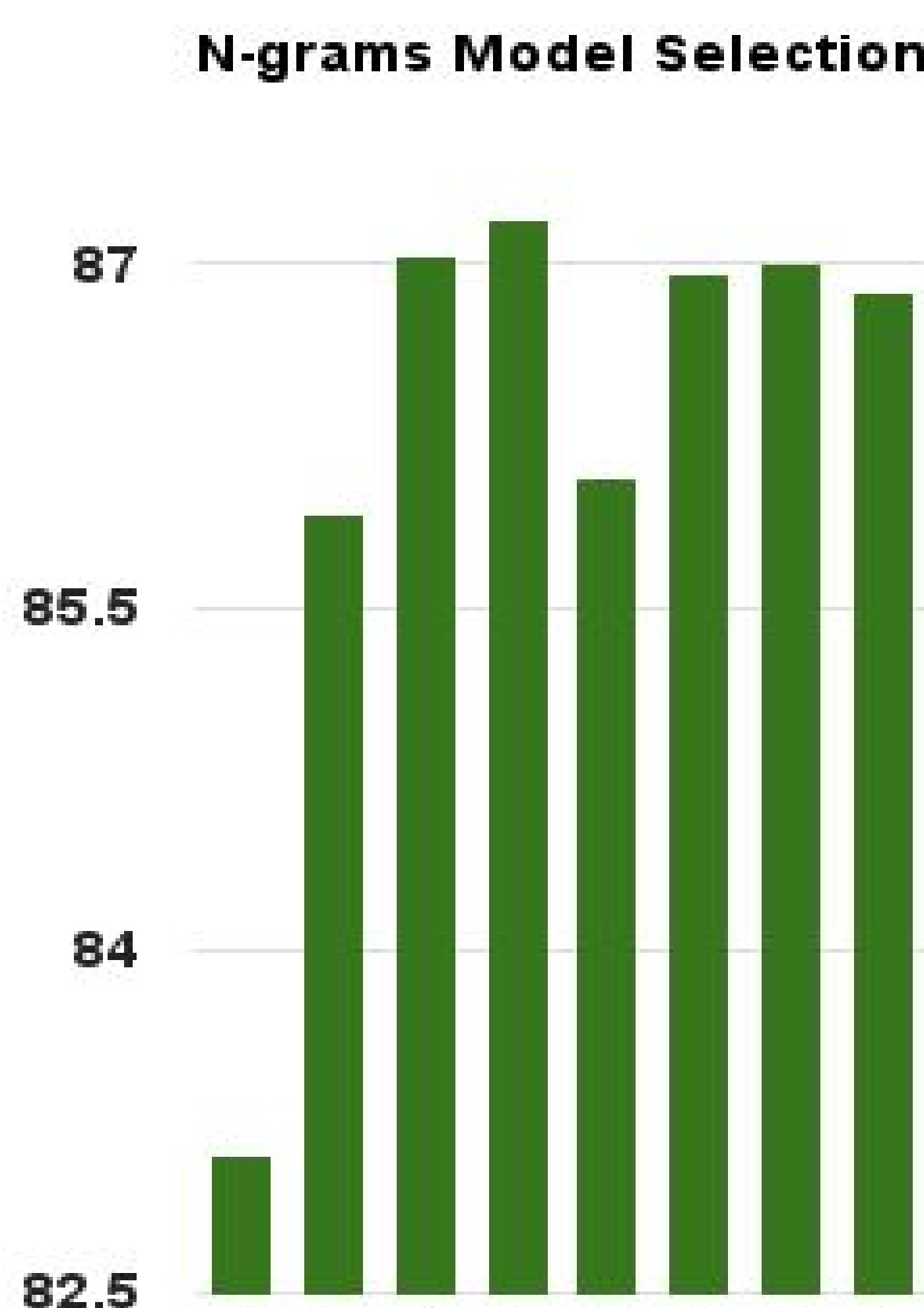
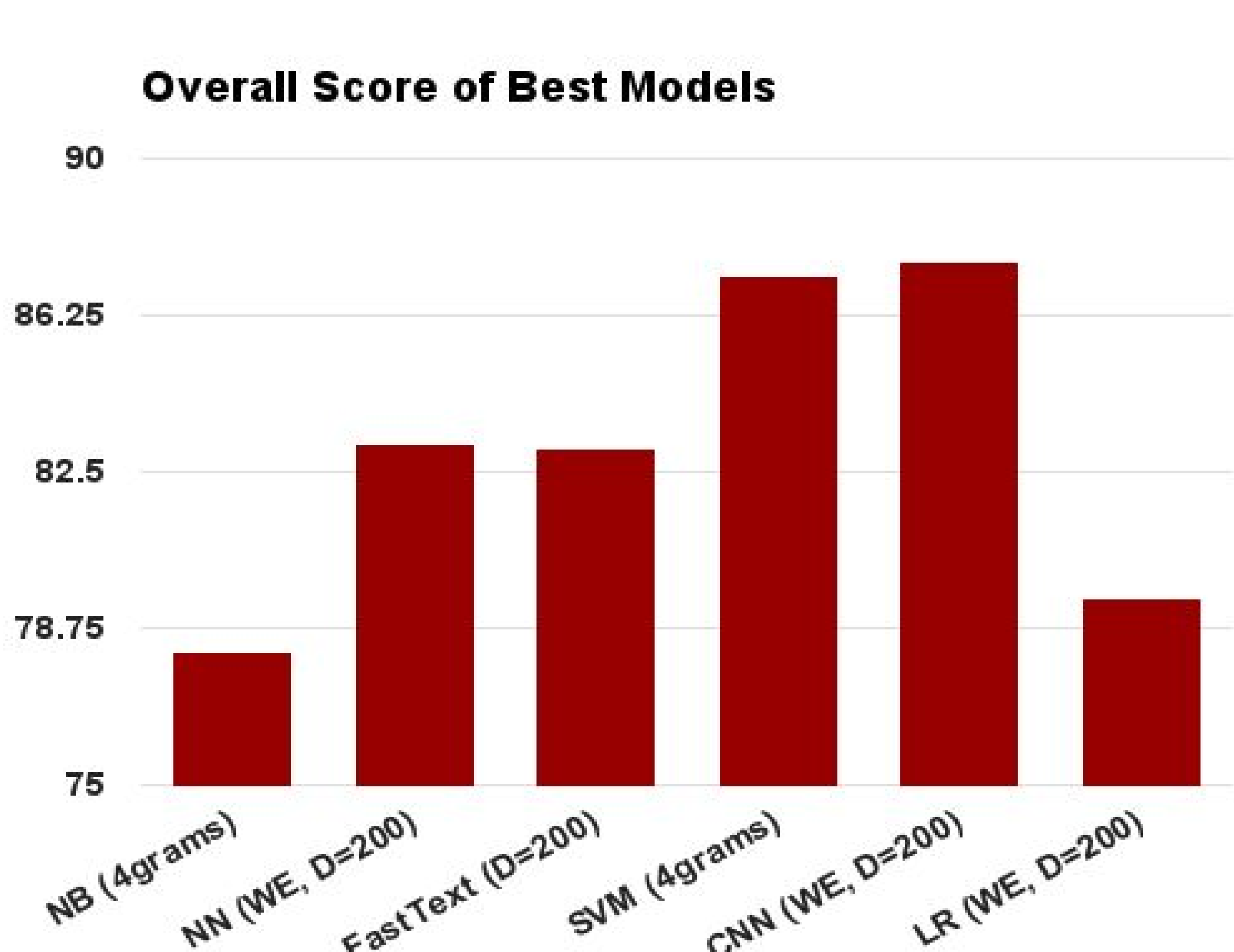
3. Approach

- **Contractions Expansion** (e.g., don't → do not)
- **Emojis Transformation** (e.g., :D → <smile face>, :(→ <sad face>)
- **Emphasize Repeated Punctuation** (e.g., yes !!!! → yes ! <repeated>)
- **Emphasize Repeated Last Characters** (e.g., I am hapyyyyy → I am happy <repeated>)
- **Filter Numerical Expressions** (e.g., 35 → <number>)
- **Split Hashtags** (e.g., #lovemyjob → love my job) FREQUENTLY USED ENGLISH WORDS DICTIONARY
- **Emphasize Sentiment Words** (e.g. I feel anxiety → I feel anxiety <negative>) SENTIMENT LEXICON
- **Part-Of-Speech Tagging** (e.g., guy → guy:Noun)
- **Lemmatization and Stemming** (e.g., followed → follow)
- **Stop-words Filtering** (e.g., the cat → cat)

Find Us



4. Experiments



Algorithm	without Preprocessing	with Preprocessing	Gain (from preprocessing)
TF-IDF (4grams)	85.28	87.20	1.92
FastText (D=200)	80.36	83.04	2.68