# What do you like in dating apps

## Text Mining and Sentiment Analysis Project

Alessia Cecere

## 1 Introduction

In the last years, dating apps have become more and more popular, and have caught the attention of researchers studying the interplay between new media **technologies** and **society**.

Studies have analyzed the **motives** driving people to use these apps, finding them ranging from casual sex to simply killing time [1], as well as involving romantic pursuits and other kinds of affiliation and information [2]. Interestingly, one overarching finding is that users are on the whole more motivated to find **love** on dating platforms, as opposed to **casual sex** [3]. Moreover, **gender** differences in app uses seem to be prominent, with men primarily pursuing hook-up sex, travelling and relationships, and women more prone to seek friendship and self-validation [4].

Few studies have tried to dive into these motivations and differences by automatically detecting them from text. In [5], topic modelling and dimensionality reduction are applied to dating apps reviews: the reasons for negative evaluations are found to be mainly concentrated in the charging mechanism, fake accounts, subscription, advertising and matching dynamics.

The aim of this project is to exploit the descriptive power of **aspect-based sentiment analysis** to try to automatically detect a polarity (and possibly an opinion) for these social impacting aspects.

## 2 Research question and methodology

An **opinion** (or sentiment) is a quadruple $(s, g, h, t)$, where $s$ represents the sentiment, $g$ the sentiment target, $h$ the holder (the one expressing the sentiment) and $t$ the time at which the opinion is expressed [6]. Sentiment analysis can be performed at the level of the document, sentence or aspect. Aspect-level sentiment analysis, addressed in this work, is the more fine-grained model, which extracts opinions expressed against different aspects/features of the entity [7].

More specifically, this research is conducted on the following aspects of dating apps.

- The **motives** driving people to use them, distinguishing between the search for a long-term relationship, casual sex, friendship and curiosity.

- **Gender** stereotypes and how those are propagated by apps use.
- The impact dating apps are having on **self esteem**, along with references to **appearence** and **intelligence** as factors that could make one perceived as attractive on these platforms.

This reduces to the two different tasks of **extracting** references to previous aspects and associating a **polarity** to them. Three different methods were conceived, implemented and evaluated for these purposes.

## 2.1 Rule/Embeddings based method

This first method exploits a **syntax** rule-based approach – combined with the use of **word embeddings** – for aspect-extraction.

At first, all reviews are preprocessed using spacy [8], in order to obtain part-of-speech tagging. A syntax tree is thus generated for all review sentences, and used to detect adjectives that are linked in the sentence structure, according to the following patterns.

- **ADJ-NOUN**, e.g. pairs like *nice friends*, where left-side adjectives are classified as children of the noun, but excluded if found after another noun.
- **NOUN-AUX-ADJ**, e.g. *friends are nice*, where the auxiliary verb lemma is *be*, the noun has a *subject* role and the adjective is found on the right of the noun.

In both cases, negations like *not nice friends* and *friends are not nice* are addressed by checking whether a subtoken with lemma *not* is found in adjective or verb children, in the first and second case respectively. In this event, the adjective is substituted by one of its antonyms selected using wordnet [9] (if any antonym is found).

The result after this procedure is a list of **nouns**, each with its own list of associated **adjectives**. To select the aspects of interest, a set of **queries** is defined, namely *casual sex*, *relationship romantic couple* (excluding *friendship*), *friendship friends networking* (excluding *love*), *curiosity exploration social* (excluding *space*), *women girls* (excluding *boy*), *man boy guy* (excluding *girls women*), *self esteem*, *attractiveness appearence beauty* and *witty intelligence*.

The GloVe [10] (specifically the *glove-twitter-200*) embeddings of these queries are taken and their cosine **similarity** with the embeddings of all the previously collected aspects is computed: only the aspects having a similarity up to a certain experimental threshold are kept. The **polarity** score of each associated adjective is computed by means of sentiwordnet [11]. More specifically, the final adjective score is set to the subraction between the average positive and negative polarity across all associated synsets. The polarity of relevant aspects is computed in terms of percentage: restricting the domain to adjectives having an absolute polarity score over 0.1, the percentage of positive and negative classified adjectives is associated to the aspect.

## 2.2 TF-IDF/PMI-based method

In this second approach, relevant aspects are treated as possible **categories** into which splitting the reviews. According to this intuition, the first step is setting up a **TF-IDF** retrieval on reviews, based on the queries *casual sex, relationship, friendship, curiosity, women, men, self esteem, attractiveness* and *intelligence*.

Once obtained the classification, two metrics are computed.

- A **Category-Word PMI**, describing – for each noun with a frequency over 10 – the relavance of the word for that category, defined as $\log \dfrac{P(w,c)}{P(w)P(c)}$, where
  - $P(w,c)$ is the probability of the word appearing in the reviews of that category, i.e. the number of times it appears in the category reviews devided by the number of times it appears in all categorized reviews.
  - $P(w)$ is the probability of the word, computed as the number of times it appears devided by the sum of the frequencies of all words.
  - $P(c)$ is the probability of the category, i.e. the sum of all word frequencies in that category devided by the sum of word frequencies in all categories.
- An **Adjective-Noun PMI**, defining the relevance of the adjective-noun pair for each category. This metric is defined as $\dfrac{P(n,a)}{P(n)P(a)}$, where
  - $P(n,a)$ is the probability of the adjective and the noun to appear together. This is approximated with the following procedure: the first 500 words by frequency are selected, then adjectives and nouns are classified using nltk part-of-speech tagging; all possible pairs of adjectives and nouns are generated, and their probability is computed as the number of reviews in which the adjective and the noun appear together, devided by the number of reviews.
  - $P(n)$ is the probability of the noun, i.e. its frequency devided by the sum of all words frequencies.
  - $P(a)$ is the probability of the adjective, similarly its frequency devided by the sum of all words frequencies.

A score is assigned to adjective-noun pairs by summing up the **adjective-noun PMI** and the **category-noun PMI**, after normalizing both scores in a range between 0 and 1 and excluding the pairs having score 0. Only the first 1000 pairs are taken, and the resulting polarity for the different categories is obtained by counting the percentage of adjectives having a sentiwordnet positive polarity and a negative one, again restricting the domain to adjectives having an absolute polarity higher than 0.1.

## 2.3 BM25/BERT-based method

This last approach starts from the same intuition as the previous one, and thus devides reviews in the same categories and using the same queries; it differs in the applied model, namely **BM25**, a popular ranking function in information retrieval systems [12]. In this case, no syntax distinction in adjectives and nouns is applied, but the sentiment score of each category is computed on the entire reviews text using *distilbert-base-uncased-finetuned-sst-2-english* [13], a checkpoint of **DistilBERT-base-uncased** [14] fine-tuned on SST-2. Positive and negative scores for aspects are thus computed as the percentage of reviews classified as positive and negative respectively, over a threshold of 0.70.

# 3 Experimental results

The three models were applied to the reviews contained in two different datasets.

- **Tinder Dating App - Google Play Store Review** [15], containing around 600.000 reviews of the app Tinder from the Google Play Store, written from 2013 to 2023. **Tinder** is a mobile dating app that enables users to connect with potential partners based on their location and mutual interest; users swipe right if they are interested or left to pass: if two users both swipe right, they are matched and can communicate within the app. Since its launch in 2012, Tinder has emerged as one of the most popular dating platforms globally.
- **Dating Apps Reviews 2017-2022 (all regions)** [16].
  This dataset contains around 620.000 Google Play Store reviews from 2017 to 2022 of three popular dating apps: Tinder, Bumble and Hinge. For the purposes of this project, Tinder ones (77%) were discarded in favour of the first dataset, while Bumble and Hinge reviews (102.384 and 52.994 respectively) were considered.
  **Bumble** was launched in 2014 and gained significant popularity as a female-friendly alternative to traditional dating apps: after a match, in fact, only women can initiate conversation with matched male users. It also offers various modes to cater to different relationship preferences: in addition to the dating mode, it also features Bumble BFF, which helps users find platonic friendships, and Bumble Bizz, focused on professional networking.
  On the other hand **Hinge** – which brands itself as *the dating app designed to be deleted* – aims to facilitate meaningful connections by focusing on users' interests and preferences. Launched in 2013, it utilizes a *like* or *comment* system, allowing users to express interest in a specific aspect of a person's profile, rather than simply swiping right or left based on pictures.

Table 1 shows a comparison between the positive polarity percentage scores of the Rule/Embeddings-based and TF-IDF/PMI-based models. Table 2, on the other hand, contains the positive scores obtained by the BM25/BERT-based model. Negative scores can be obtained as the complement, and were thus not reported.

Given the absence of a **ground thruth** on single aspects, and exploiting the presence of review **stars**, the first two models were evaluated in the following way: the pairs of noun/adjectives found with the respective methods were used as features; reviews were thus represented as one-hot-encoded vectors, where a feature was set to 1 if the corresponding *(adj, noun)* pair was present in the review, and to 0 otherwise. Review vectors obtained in this way were used for a **binary classification task**, to predict a positive ($\leq 2$ stars) or negative ($\geq 3$ stars) score for reviews belonging to the corresponding category (according to BM25). Table 3 shows the F1 score obtained in the classification task, applying logistic regression.

**Table 1** Comparison between positive polarity percentages for the three platforms, using first and second model.

| | Rule/Embeddings | | | TF-IDF/PMI | | |
|---|---|---|---|---|---|---|
| Aspect | Tinder | Bumble | Hinge | Tinder | Bumble | Hinge |
| Casual sex | 45.07% | 58.31% | 64.28%* | 52.16% | 39% | 45.93% |
| Relationship | 65.31% | 55.48% | 63.94% | 53.21% | 46.49% | 58.01% |
| Friendship | 83.94% | 67.36% | 74.40% | 46.04% | 45.21% | 54.87% |
| Curiosity | 42.10%* | 57.14%* | 53.84%* | 40.17% | 35.23% | 50% |
| Women | 67.39% | 69.24% | 76.33% | 48.95% | 52.97% | 59.56% |
| Men | 66.68% | 62.63% | 76.65% | 47.94% | 44.72% | 57.11% |
| Self esteem | 27.96% | 28%* | 33.33%* | 38.46% | 40.56% | 49.65% |
| Attractiveness | 60.52% | 70.37% | 56.52% | 31.77% | 44.80% | 51.66% |
| Intelligence | 34.71% | 45.21% | 63.07% | 30.52% | 37.24% | 53.57% |

Note: Values signed with * were computed with few data and therefore should not be considered significant.

**Table 2** Comparison between positive polarity percentages for the three platforms, using BM25/BERT model.

| | | BM25/BERT | |
|---|---|---|---|
| Aspect | Tinder | Bumble | Hinge |
| Casual sex | 39.53% | 32.51% | 33.01% |
| Relationship | 51.46% | 48.16 % | 57.54% |
| Friendship | 84.20% | 70.65% | 67.28% |
| Curiosity | 53.34% | 62.88% | 78.55% |
| Women | 46.49% | 69.24% | 41.18% |
| Men | 27.04% | 52.92% | 37.38% |
| Self esteem | 31.29% | 26% | 34.04%* |
| Attractiveness | 52.94%* | 52.94%* | 40%* |
| Intelligence | 16%* | 52.94%* | 100%* |

Note: Values signed with * were computed with few data and therefore should not be considered significant.

To evaluate the BM25/BERT model, given the presence of a score for each review, the Pearson correlation between the review and the number of stars was considered (and shown in Table 4).

# 4 Conclusions

One first thing we can notice, looking at Table 1 and Table 2, is that Rule/Embeddings scores tend to be more optimistic that TF-IDF/PMI ones, which in turn are more **optimistic** than BM25/BERT: it looks like the more we are detaching from the explicit assigment of adjectives to the aspects of interest, the more negative scores we are getting. This could happen for several reasons. First of all, TF-IDF/PMI model is considering the polarity of the adjectives whose PMI with the associated word is higher, but is not doing any check on whether the noun is actually referencing the aspect of interest: in the *causal sex* pairs, for example, we find values like *(access,*

**Table 3** Comparison between the F1 score obtained for models 1 and 2

| | Rule/Embeddings | | | TF-IDF/PMI | | |
|---|---|---|---|---|---|---|
| Aspect | Tinder | Bumble | Hinge | Tinder | Bumble | Hinge |
| Casual sex | 0.79 | 0.80 | 0.77* | 0.79 | 0.80 | 0.78 |
| Relationship | 0.77 | 0.78 | 0.77 | 0.79 | 0.80 | 0.77 |
| Friendship | 0.79 | 0.79 | 0.77 | 0.81 | 0.80 | 0.77 |
| Curiosity | 0.79* | 0.80* | 0.77* | 0.81 | 0.80 | 0.77 |
| Women | 0.78 | 0.78 | 0.77 | 0.80 | 0.80 | 0.77 |
| Men | 0.79 | 0.80 | 0.78 | 0.80 | 0.80 | 0.78 |
| Self esteem | 0.79* | 0.80* | 0.78* | 0.80 | 0.80 | 0.78 |
| Attractiveness | 0.73 | 0.80 | 0.78 | 0.82 | 0.80 | 0.78 |
| Intelligence | 0.77 | 0.79 | 0.76 | 0.82 | 0.80 | 0.76 |

Note: Values signed with * were computed with few data and therefore should not be considered significant.

**Table 4** Correlation between model 3 positive polarities and number of stars

| | | BM25/BERT | |
|---|---|---|---|
| Aspect | Tinder | Bumble | Hinge |
| Casual sex | 0.64 | 0.67 | 0.71 |
| Relationship | 0.51 | 0.72 | 0.66 |
| Friendship | 0.66 | 0.73 | 0.84 |
| Curiosity | 0.82 | 0.83 | 0.77 |
| Women | 0.65 | 0.66 | 0.70 |
| Men | 0.62 | 0.63 | 0.67 |
| Self esteem | 0.61 | 0.66 | 0.37* |
| Attractiveness | 0.78* | 0.55* | 0.72* |
| Intelligence | 0.28* | 0.55* | * |

Note: Values signed with * were computed with few data and therefore should not be considered significant.

*awful)*, which may be present in reviews regarding casual sex, but are not referecing the aspect of interest. Taking into account that previous literature tells us that users and mostly concerned with app functionalities and mechanisms, and that we are analyzing reviews coming from an app store, it could be that in these cases we are measuring the collateral experience of someone looking for casual sex, rather than his/her opinion on the aspect itself. This may be exacerbated by the BM25/BERT model, which is taking into account the sentiment of the whole review. Rule/Embeddings model, on the other hand, is considering specific simple syntactic patterns, which may be biased towards positive opinions, while users with a negative sentiment might articulate in a more complex way their experience in the hope to find ways to improve it. We should also have in mind that in this model we are addressing adjective negations and expanding the query using GloVe embeddings, while in the PMI/TF-IDF one we are not.

Despite the differences, some **common tendencies** can be found between models, especially when comparing the relative scores across platforms, which may be an indicator of how different technological environments are shaping social aspects. For example, if we look at **gender polarities**, we can see that in both Rule-based and

PMI-based approaches Hinge has the highest positive polarity score: when talking explicitly about the other/same gender, associated adjectives seem to be more favorable. This may mean that Hinge's attempt to encourage meaningful connections is having the effect of improving the relationship with the other; Bumble's feminist direction, on the other hand, may be reflected in the highest positive score gathered by the BM25/BERT model, which in turn may be able to capture better than others appreciations for the different maching mechanism. It is also interesting to notice that, in almost all models and platforms, women are assigned a more positive score than men (see Appendix A for further exploration on this).

When considering the **motives** driving people to use dating apps, another interesting trend can be found: across almost all platforms, the same model has higher scores on **relationship** than **casual sex**; this is consistent with the previous knowledge according to which users are looking for a relationship more than sex, when subscribing to these apps. A more positive sentiment towards casual sex was found on Tinder according to TF-IDF/PMI and BM25/BERT models, but the Rule/Embeddings model shows the opposite trend; same goes for relationships and Hinge.

The **friendship** aspect shows the highest positive scores for Rule/Embeddings and BM25/BERT, touching more than 80% when considering Tinder reviews. This may be due to the fact that people are not expecting to find friends on these platforms, and thus are more pleasantly surprised when it happens. Moreover, the term *friends* inherently carries a positive connotation, unlike *sex* and *relationship*. When referring to a friend, it is more likely that the sentiment will be positive.

*Curiosity*, *self esteem*, *attractiveness* and *intelligence* were the most challenging aspects to analyze, being more niche and difficult to find in text with explicit reference.

With regards to *intelligence* and *attractiveness*, BM25 selected less than 50 reviews for each platform: therefore, it should not be considered significant in evaluating the sentiment on those aspects. References to them were easier to find in the Rule/Embeddings approach, where query expansion captured related words like *beauty* and *smart*. In this setting, Hinge shows the highest polarity score towards intelligence (more than 60%), while Bumble has a slight prevalence towards negative sentiment (45% positive score) and Tinder is way more negative (34%). Attractiveness, on the other hand, always has positive scores (maybe partially because of the same bias as *friends*), with Bumble having the highest positive score, followed by Tinder and Hinge. Averall, these results seem to suggest again the idea that Hinge users are putting more emphasis on personality than appearence, when compared to Tinder ones.

Query expansion was not as useful for *curiosity*, which was better captured by the second an third model, but with contradictory results.

For *self esteem*, the only model considered significant was TF-IDF/PMI, showing overall mixed or negative sentiment scores, with best score obtained by Hinge, followed by Bumble and Tinder.

Despite the interesting findings, it is important to address the **limitations** of this work too. First of all, we must take into consideration that data come from the Google Play Store, and thus the reviews are more likely to be about the technology than the social aspects involving the app; a follow-up project could examine text coming from other sources, for example social network comments. Moreover, available datasets did not provide a ground thruth for evaluating the polarity associated to the aspects of interest: the evaluation reported in Tables 3 and 4 – although encouraging – is not complete; an annotated dataset would need to be created, to evaluate aspect as long as sentiment extraction. Additionally, the comparison between the sentiment extracted by models is interesting but not completely fair, since they are measuring different things: in the case of the Rule/Embedding and TF-IDF/PMI it is the percentage of positive and negative adjectives associated to aspects noun and adjective-noun pairs respectively, while in BM25/BERT the percentage is computed on the number of reviews.

From the methodological point of view, many refinements could be experimented, among which trying other lexicon polarity metrics – to reduce possible biases introduced by sentiwordnet –, including more syntactic patterns in the Rule/Embeddings model and applying its query expansion to the other models too.

## Appendix A What are users saying about men and women?

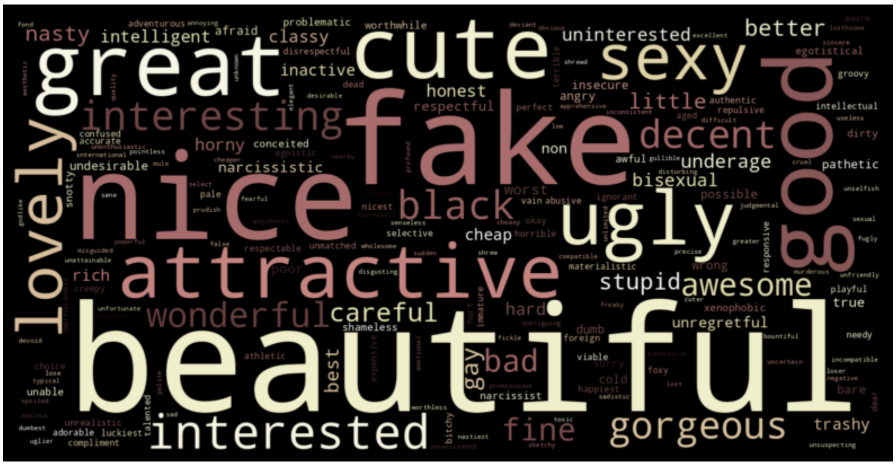


**Fig. A1** Wordcloud of adjectives referring to women

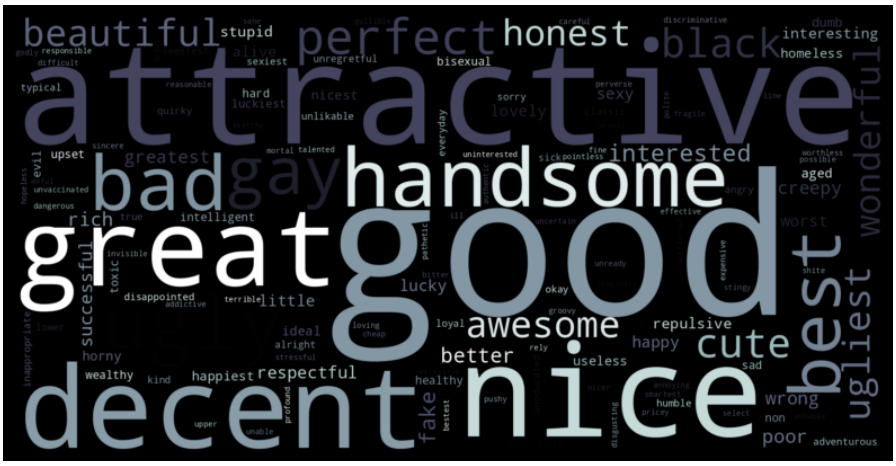


**Fig. A2** Wordcloud of adjectives referring to men

Figure A1 and Figure A2 show the wordclouds of adjectives referring to women and men respectively, as extracted by the Rule/Embeddings method on Tinder reviews. Both highlight the importance of beauty standards, reflected in adjectives like *attractive*, *beautiful* and *handsome*.

In comparison, women are more often referenced as *lovely*, *nice*, *uninterested* and *fake*; men as *perfect*, *respectful*, *decent* and *honest*.

## References


[1] Timmermans, E., De Caluwé, E.: Development and validation of the tinder motives scale (tms). Computers in Human Behavior **70**, 341–350 (2017) https://doi.org/10.1016/j.chb.2017.01.028

[2] Wu, S., Trottier, D.: Dating apps: a literature review. Annals of the International Communication Association **46**(2), 91–115 (2022) https://doi.org/10.1080/23808985.2022.2069046 https://doi.org/10.1080/23808985.2022.2069046

[3] Sumter, S.R., Vandenbosch, L., Ligtenberg, L.: Love me tinder: Untangling emerging adults' motivations for using the dating application tinder. Telematics and Informatics **34**(1), 67–78 (2017) https://doi.org/10.1016/j.tele.2016.04.009

[4] Ranzini, G., Lutz, C.: Love at first swipe? explaining tinder self-presentation and motives. Mobile Media & Communication **5**(1), 80–101 (2017) https://doi.org/10.1177/2050157916664559 https://doi.org/10.1177/2050157916664559

[5] Shen, Q., Han, S., Han, Y., Chen, X.: User review analysis of dating apps based on text mining. PloS one **18**, 0283896 (2023) https://doi.org/10.1371/journal.pone.0283896

[6] Schouten, K., Frasincar, F.: Survey on aspect-level sentiment analysis. IEEE Transactions on Knowledge and Data Engineering **28**(3), 813–830 (2016) https://doi.org/10.1109/TKDE.2015.2485209

[7] Rana, T., Cheah, Y.-N.: Aspect extraction in sentiment analysis: comparative analysis and survey. Artificial Intelligence Review **46** (2016) https://doi.org/10.1007/s10462-016-9472-z

[8] Honnibal, M., Montani, I.: spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear (2017)

[9] Miller, G.A.: WordNet: A lexical database for English. In: Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 8-11, 1994 (1994). https://aclanthology.org/H94-1111

[10] Pennington, J., Socher, R., Manning, C.: GloVe: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543. Association for Computational Linguistics, Doha, Qatar (2014). https://doi.org/10.3115/v1/D14-1162 . https://aclanthology.org/D14-1162

[11] Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (eds.)

LREC. European Language Resources Association, ??? (2010). http://nmis.isti.cnr.it/sebastiani/Publications/LREC10.pdf

[12] Robertson, S., Walker, S., Jones, S., Hancock-Beaulieu, M.M., Gatford, M.: Okapi at trec-3. In: Overview of the Third Text REtrieval Conference (TREC-3), pp. 109–126. Gaithersburg, MD: NIST, ??? (1995). https://www.microsoft.com/en-us/research/publication/okapi-at-trec-3/

[13] HuggingFace: DistilBERT base uncased finetuned SST-2. https://huggingface.co/distilbert-base-uncased-finetuned-sst-2-english

[14] Sanh, V., Debut, L., Chaumond, J., Wolf, T.: DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter (2020)

[15] Ganesh, S.K.: Tinder Dating App - Google Play Store Review. https://www.kaggle.com/datasets/shivkumarganesh/tinder-google-play-store-review

[16] Kriplani, S.: Dating Apps Reviews 2017-2022 (all regions). https://www.kaggle.com/datasets/sidharthkriplani/datingappreviews