**Data**

All data (i.e., posts) was scraped from incels.is and was pre-processed to anonymise users by removing IDs, usernames, and other identifiable features (comment number). The text was lowercased, and all special characters and emojis were removed.

**Labelling task**

Five annotators were given an annotation task, whereby each comment in a subset of the data was assigned one of the following labels by an annotator working independently:

1. Negative (INCEL TO OTHER): direct hate speech to an individual, group etc
2. Positive (INCEL TO INCEL): comments that are positive towards any other \*cel group
3. Neutral: no intended target audience

For 1650 examples, agreement (in terms of Cohen’s Kappa) between the annotators was calculated to be 0.8657. The data has not been made publicly available due to ethical issues.

Approximately, 15% of the comments were negative, 5% of the comments were positive, and 80% were neutral. This observation reveals that, despite the comments categorised as neutral containing offensive terms, these terms were devoid of direct targeting or any explicit intent. Instead, they appeared to be normalised and have been assimilated into the vernacular commonly used by the Incel community. These terms seem to function as an identifier as opposed to being employed with a specific purpose. Another observation is that there was minimal utilisation of positive language when communication occurred between different groups (both in-group and out-group).

**Modelling**

A model for automatically classifying posts according to the above three classes was trained. Specifically, a pre-trained transformer-based language model, RoBERTa, was fine-tuned on the labelled dataset. The class distribution is highly imbalanced, with different numbers of samples for each class. To address this issue, class weighting was applied during the training process. The dataset was split into an 80:20 ratio for training and evaluation, respectively. The training process was performed over 8 epochs, using a learning rate of 2e-5.

Evaluation of the trained model on the evaluation data set yielded a macro-averaged F1-score of **0.64.** The precision, recall and F1 score for each label can be seen in the table below.The model was unable to classify instances of positive labels correctly resulting in a 0% classification score. This is likely due to an insufficient number of data points with the positive label.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Precision | Recall | F1 |
| Positive | 0.00 | 0.00 | 0.00 |
| Negative | 0.93 | 0.97 | 0.95 |
| Neutral | 0.97 | 0.97 | 0.97 |

**Correlations between Incel-specific Terms**

Another conducted experiment involved quantifying the occurrences of terms from the Incel dictionary and subsequently classifying comments into respective categories based on these terms. The study compiled the frequencies of words belonging to specific categories. Notably, comments were susceptible to having multiple labels, contingent upon the presence of different terms within them. To investigate the relationship between these categories, the study calculated their correlations to identify which categories tended to co-occur most frequently. The findings revealed that the most prevalent category was Women, while the categories Men and Normie exhibited the highest co-occurrence rate among all analysed categories.

Terminology:

|  |  |  |
| --- | --- | --- |
| Category | Terms | Category Count |
| Women | 'stacy', 'stacie', 'foid', 'low inhib', 'sistercel', 'fesikh', 'gigastacy', 'femoid', 'roastie', 'becky', 'awalt', 'landwhale' | 3182 |
| Political | 'liberal soy', 'blackpill', 'redpill', 'bluepill', 'dogpill', 'jihadpill', 'jihadmaxxing', 'pinkpill', 'soyciety', 'statusmaxx' | 2243 |
| Men | 'high inhib', 'khhv', 'soy boy', 'betabuxx', 'chadlite', 'brad', 'mogged', 'beta orbiter', 'cucks', 'nice guy', 'normie' | 2167 |
| Normie | 'normie', 'normal' | 1697 |
| Origin/Race | 'arabcel', 'asian', 'indian', 'chaddam', 'chadpreet', 'chang', 'beancel', 'blackcel', 'currycel', 'ethnicel', 'francel', 'hapacel', 'italiancel', 'persicel', 'redcel', 'ricecel', 'turkcel', 'whitecel', 'jbb/just be black', 'seamaxx', 'seamaxxing', 'south-east-asia-maxx' | 980 |
| Fakecel | 'fakecel', 'fbicel' | 874 |
| Religion | 'christocel', 'muslimcel', 'muslim', 'hindu', 'jew', 'jews', 'jewish', 'islam' | 362 |
| Looks | 'lookism', 'looksmax', 'manlet', 'acnecel', 'baldcel', 'chincel', 'cutecel', 'dickcel', 'eyecel', 'fatcel', 'framecel', 'haircel', 'heightcel', 'nosecel', 'ramuscel', 'thincel', 'uglycel', 'wristcel', 'earcel' | 336 |
| Miscellaneous | 'cybercel', 'escortcel', 'incompetencel', 'gymcel', 'locationcel', 'countrysidecel', 'denialcel', 'nearcel', 'volcel', 'nymphocel', 'marcel', 'mismatchcel', 'oldcel', 'youngcel', 'shycel', 'standardcel', 'stoicel', 'voicecel', 'weebcel', 'protocel' | 258 |
| Disease and Mental | 'mentalcel', 'autistcel', 'spergcel', 'disabledcel', 'emcel', 'gynocel', 'manchildcel', 'medcel', 'pssdcel', 'stuttercel', 'aspie' | 186 |
| Job | 'neetcel', 'poorcel', 'stemcel', 'workcel', 'moneymaxx', 'neet' | 147 |
| Sex | 'mancel', 'femcel-lite', 'femcel', 'truefemcel', 'lesbocel', 'queercel', 'rainbowcel', 'transcel', 'intercel', 'kinkcel', 'fapfuel' | 47 |

Categories that occur together frequently:

|  |  |  |
| --- | --- | --- |
| Most frequent category combinations | | Correlation between the categories |
| Men | Normie | 0.741194 |
| Disease and Mental | Sex | 0.499505 |
| Men | Political | 0.114063 |
| Normie | Political | 0.110109 |
| Miscellaneous | Political | 0.081594 |
| Origin/Race | Religion | 0.080439 |
| Job | Men | 0.068126 |
| Looks | Women | 0.068126 |
| Men | Women | 0.061772 |
| Miscellaneous | Normie | 0.060645 |