Critical Information Extraction from Terms of Services Document

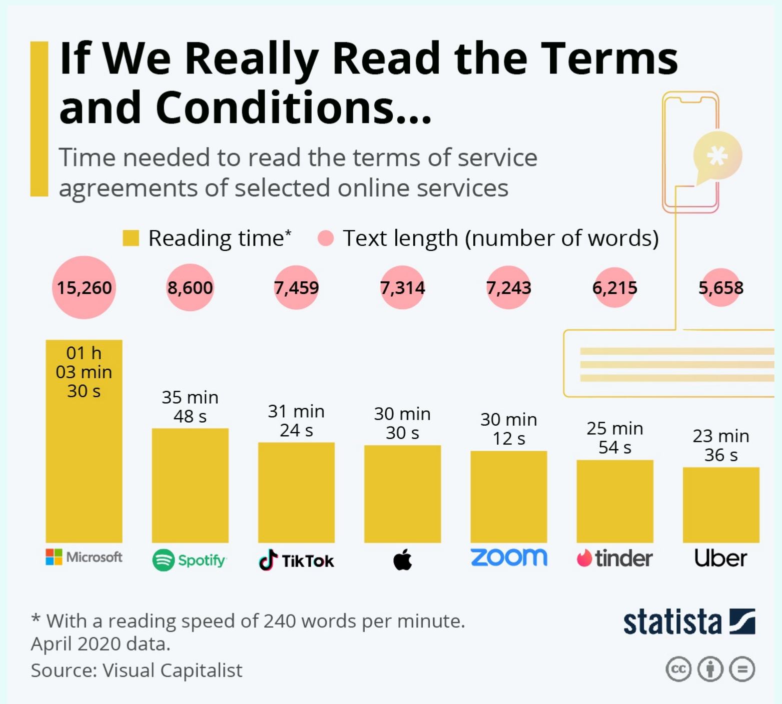
Group 18

Aditya Dave, Akanksha Nogaja, Lavina Agarwal, Shreya Prabhu, Yoshitha Akunuri

USC Viterbi School of Engineering

INTRODUCTION

- **Terms of Services (ToS)** are legal agreements between users and service providers.
- Since these documents are lengthy and use opaque jargon, users tend to sign the obligations which might expose them to unfair terms and practices.
- The solution proposes to extract critical information and present it to the users.
- We define **critical information** as obligations that users must comply with and clauses that are unfair when considering user interests.

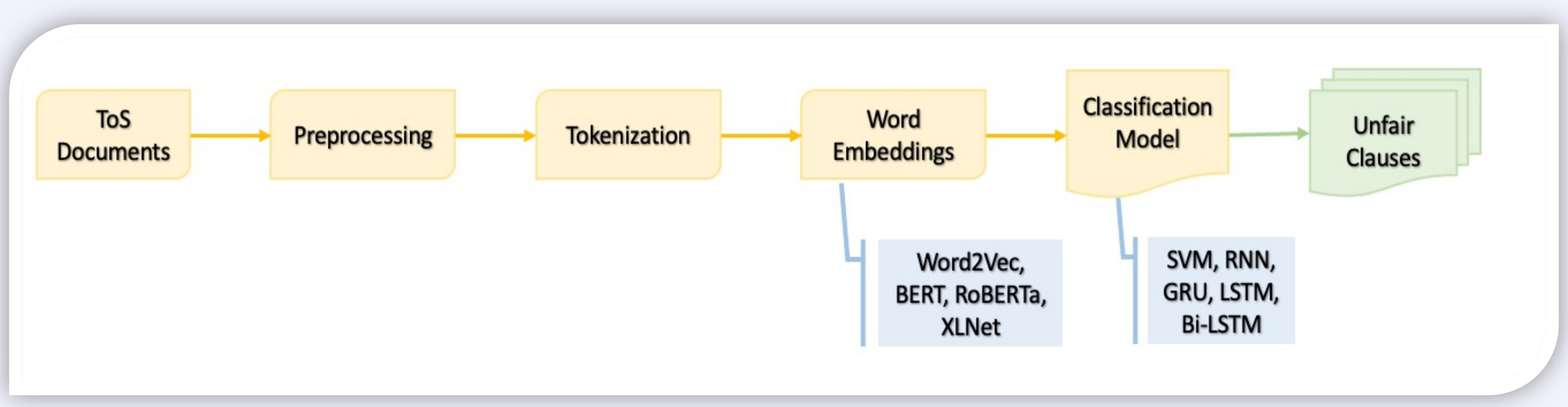


DATASET

- ToS dataset was created as a part of **Claudette** experimental study and is available open source.
- ToS clauses are categorized into clearly fair, potentially unfair and clearly unfair. We have merged these to 2 categories - **fair and unfair**.
- 9414 clauses out of which 1032 (11%) are unfair clauses.
- Dataset Size - Training: 7531(80%) | Testing: 1882 (20%)
- Preprocessing and cleaning of dataset included removing HTML tags, URLs, extra spaces, accented and non-alpha characters. Further contractions were fixed, and clauses were converted to lower case.

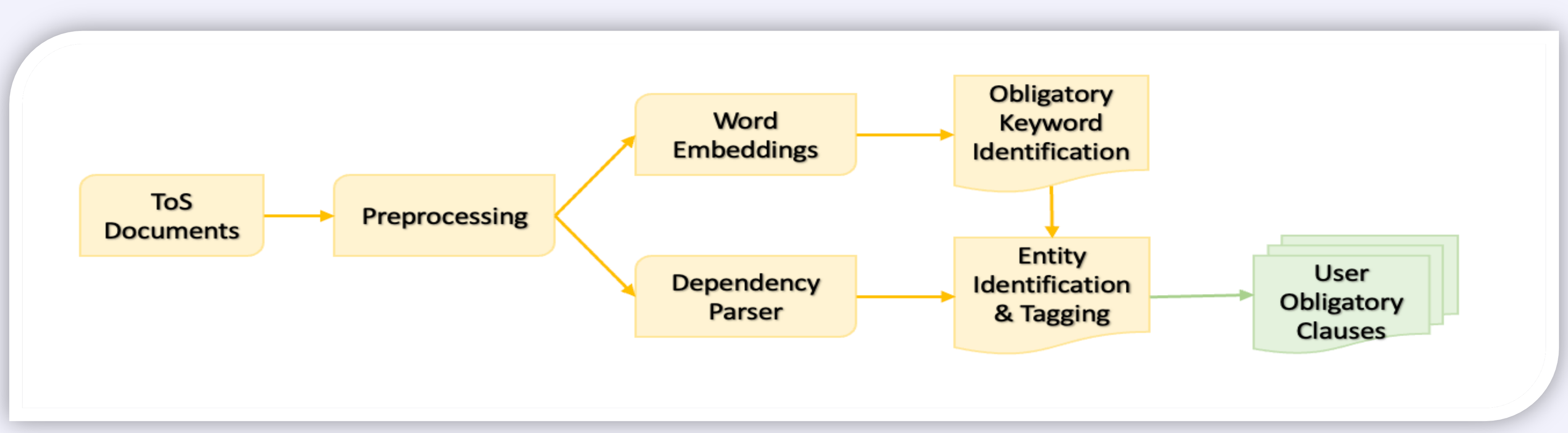
EXPERIMENT SETUP

Fairness Classification:



- We created word embeddings using Word2Vec, BERT, RoBERTa, and XLNet pre-trained models.
- Each of these embeddings were used to train RNN, GRU, LSTM and Bi-LSTM models.
- **Baseline**: We averaged the BERT embeddings and used it to train a SVM model and took it as our baseline model with a positive class F1 score of 0.35.
- **Best Model Selection**: The model with the lowest test loss was stored on each epoch as the representative weights for each model.

Obligatory Clauses Detection:



- For identifying obligations, we have generated custom word embedding model.
- This model is used along with Google word embeddings to identify the obligatory keywords list.
- We also used a dependency parser to identify the involved entities in each of the clauses.
- We mapped the identified entities to users and the organizations to output the user specific obligatory clauses.

EVALUATION METRICS

F1 Score (pos) : As the dataset is unevenly distributed and FP and FN are more crucial, we use F1 instead of accuracy. For fairness classification, identifying *unfair* clauses (i.e., positive class) is more important. So, we have considered F1 score of positive class to evaluate our models.

Accuracy : For obligations, we manually tagged the clauses which can be obligatory to the users and used that as a standard to measure accuracy of our models.

RESULTS

F1 Scores of **Unfair Classification Models** -

Models	Word2Vec	BERT	RoBERTa	XLNet
RNN	0.54	0.35	0.63	0.48
GRU	0.49	0.63	0.76	0.70
LSTM	0.50	0.39	0.65	0.64
Bi-LSTM	0.47	0.42	0.65	0.62

Accuracy for user-specific **Obligatory Clauses** detection - **72%.**

KEY FINDINGS

- While training different models, it was observed that recall was very high, and precision was relatively low. Some of the possible reasons are small dataset size, and high class-imbalance for unfair clauses.
- As observed from the results, RoBERTa embeddings outperform BERT embeddings. This may be because RoBERTa uses dynamic masking and hence it is more robust.

SAMPLE RESULT

demo video link

Fairness Classification:

9gag has discretion to terminate the account of any user upon 9gag 's own determination . → Classifier → Unfair

You are responsible for updating your user information → Classifier → Fair

Obligatory Clauses Detector:

is obligated to

You should visit this page regularly to review the current terms.

You are permitted to use your personal skype account at work for your own business communications .

FUTURE WORK

Data Augmentation – To mitigate class imbalance and increase dataset size, combination of NLP techniques like stop words removal, lemmatization, replacing words with synonyms can be done.

Summarization – If model marks all clauses as critical information, we are returning entire document to user which is not useful. Abstractive summarization can mitigate this issue.

Simplification – Vocabulary used in legal documents can be difficult to understand. Sentence simplification techniques can convert sentences into layman's terms.