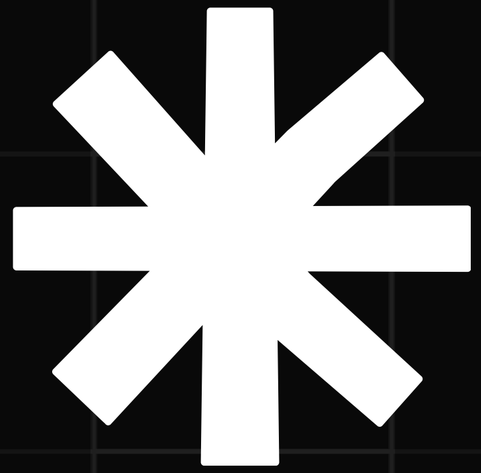
Presentation....



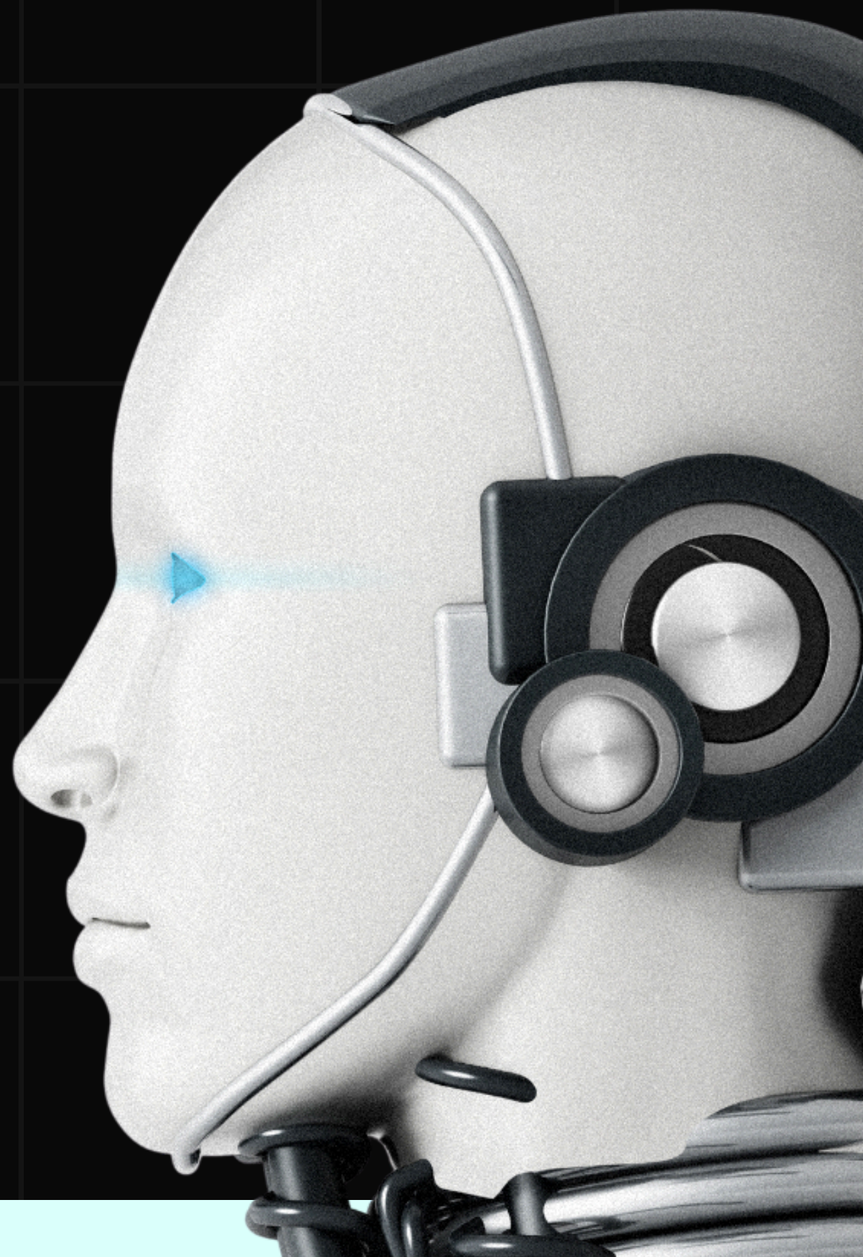
Natural Language Processing

Presented by **Joan Bayona, Luis Domene, Álvaro Sáenz-Torre and Alejandra Reinares**

Objective...

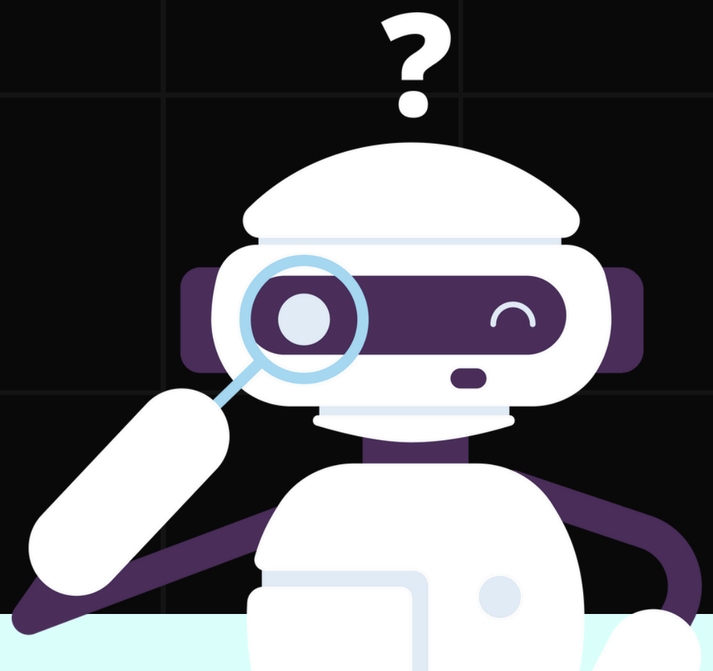
The objective of this project is to identify words expressing negation or uncertainty within a corpus of medical notes and determine their scope, using 3 different methods.

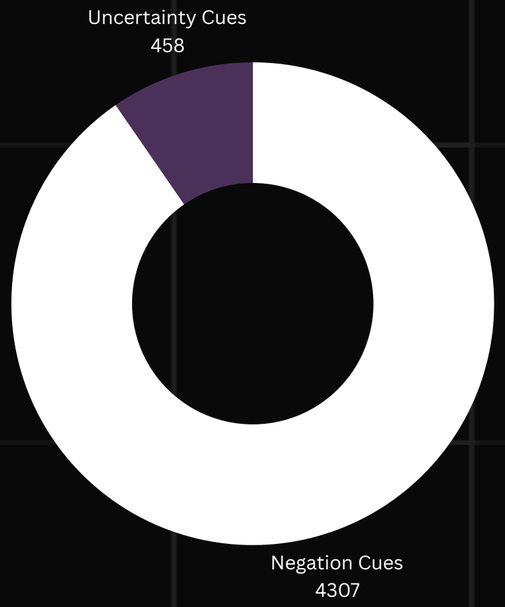
The methods used include Rule-based, Machine Learning and Deep Learning.

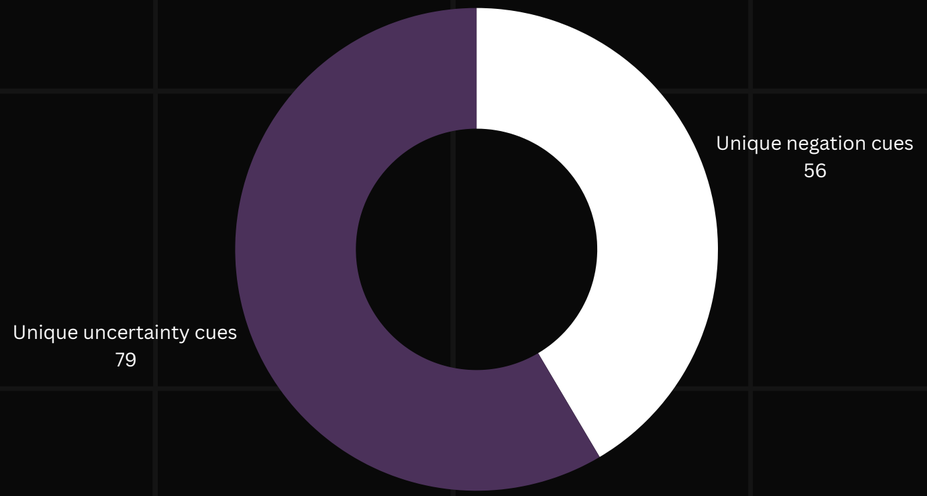


Data Exploration...

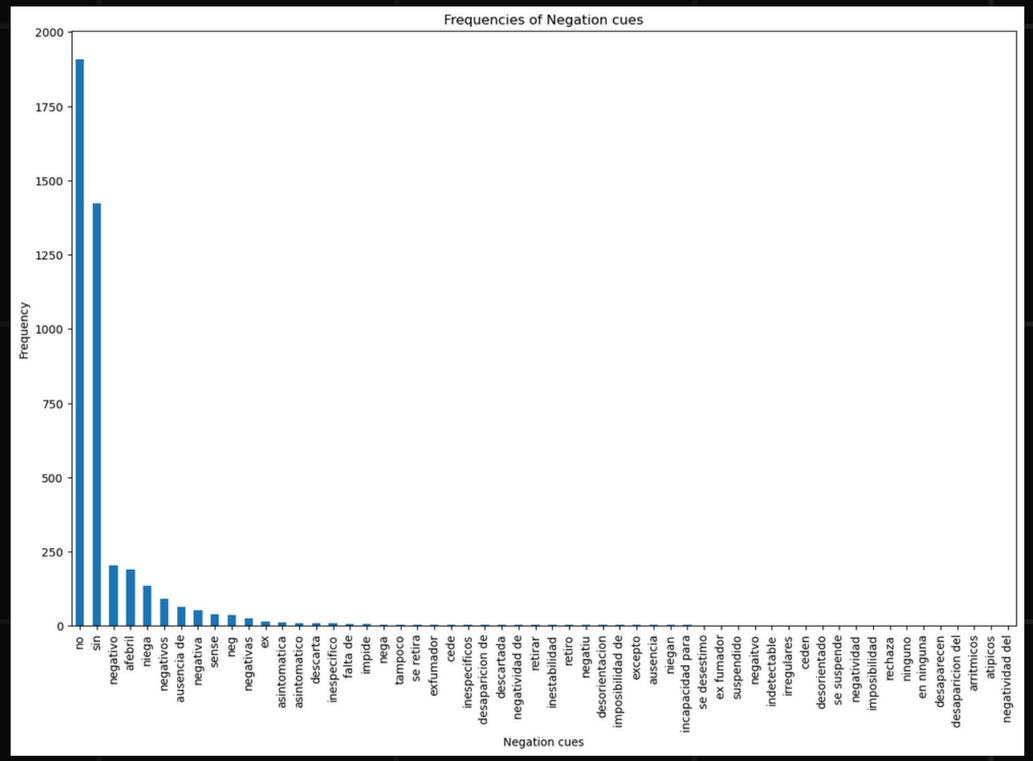
Necessary to know more about the structure and distribution of the data.

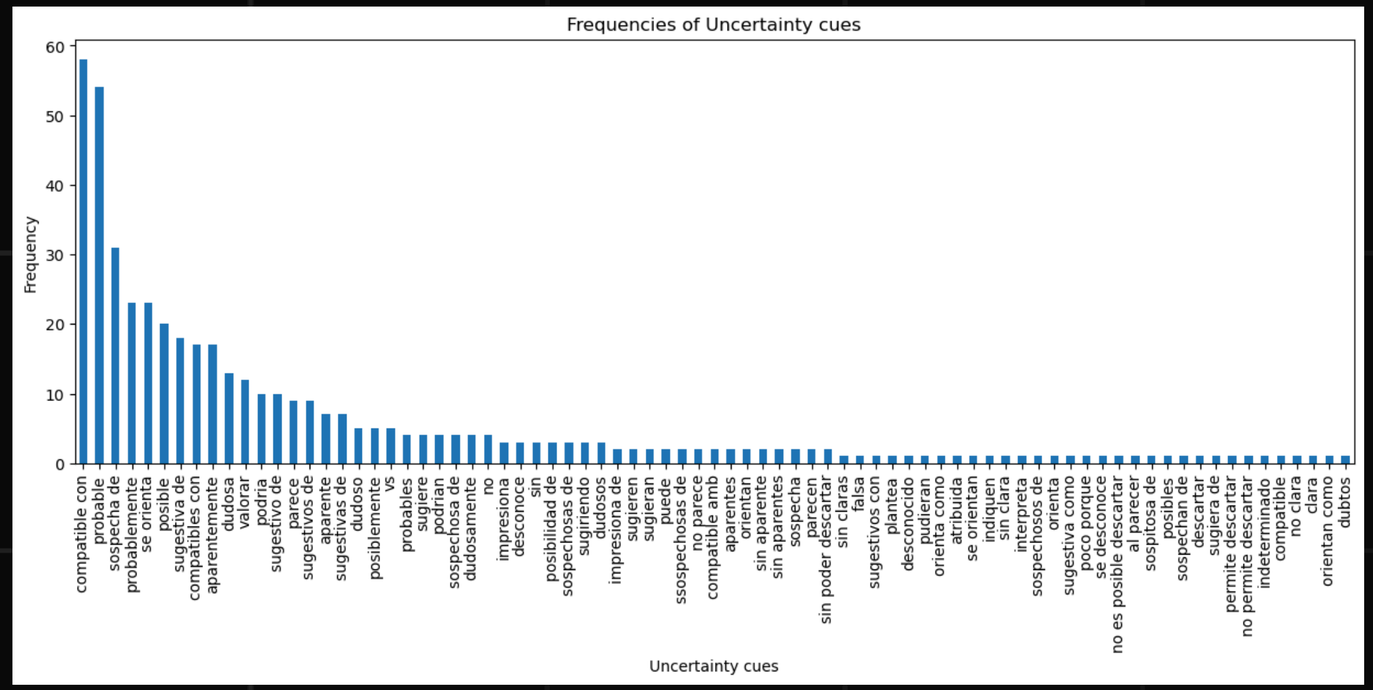




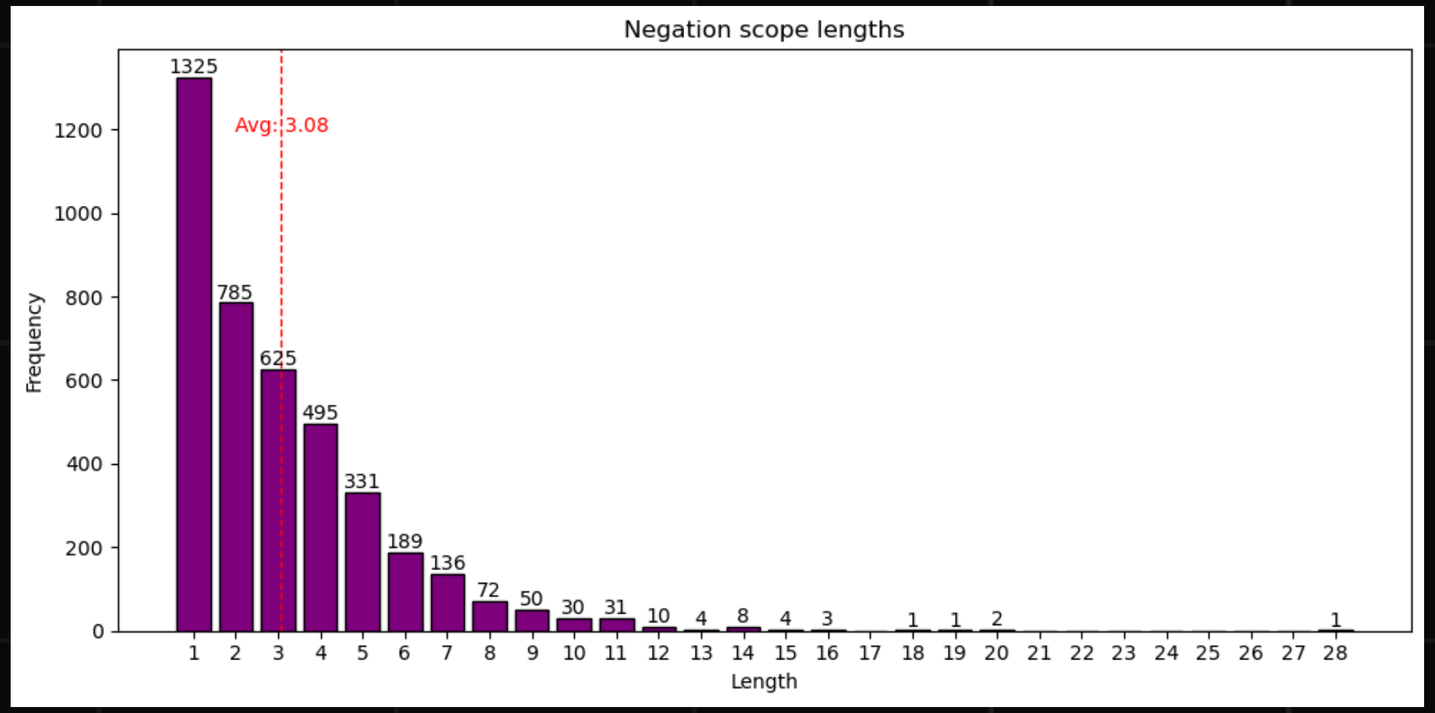


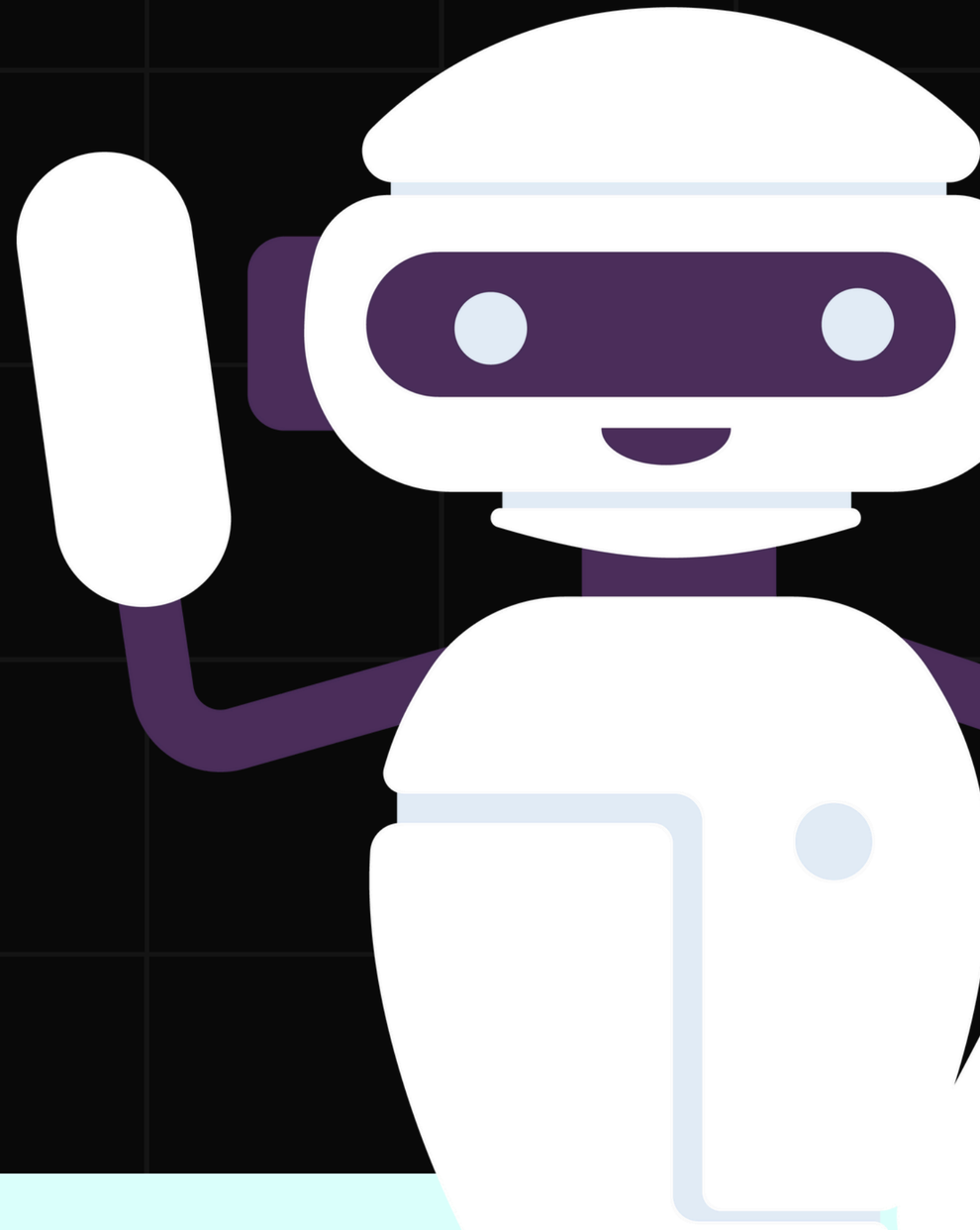
Data Exploration...



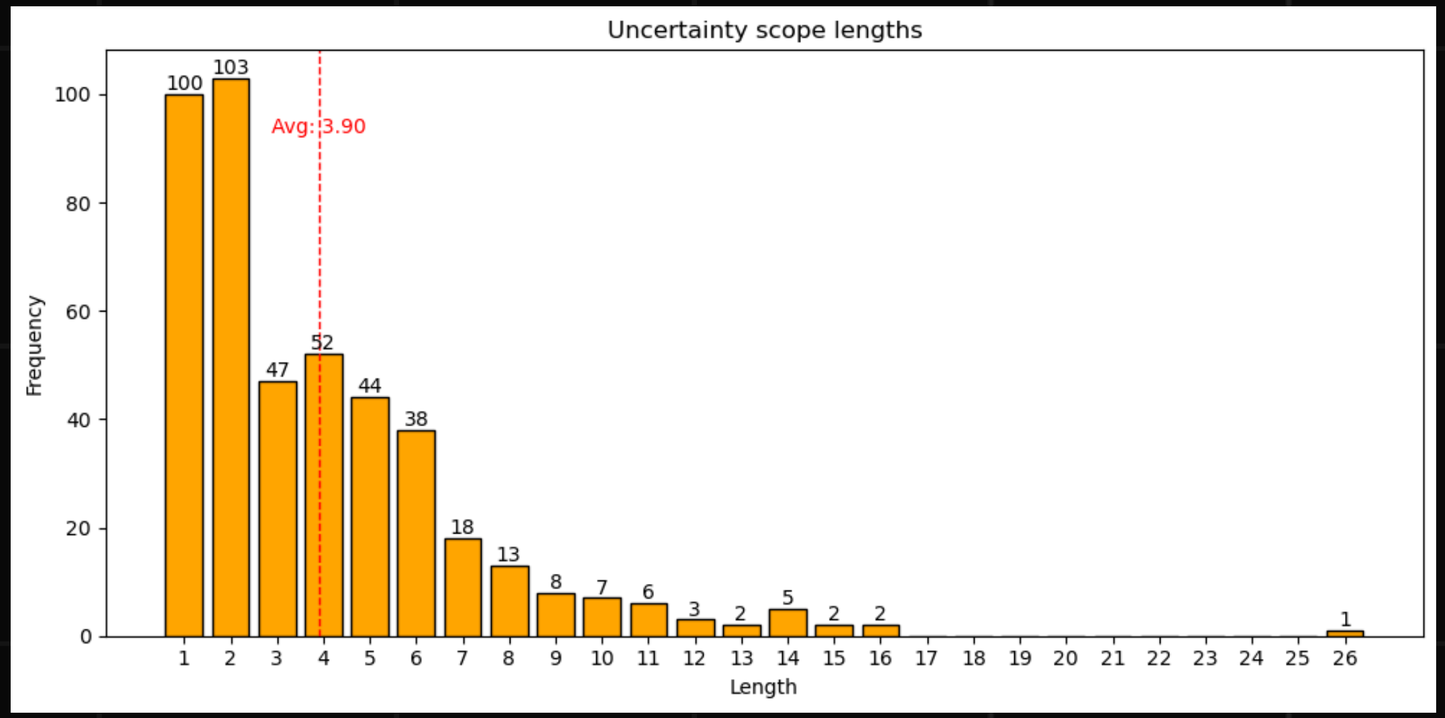


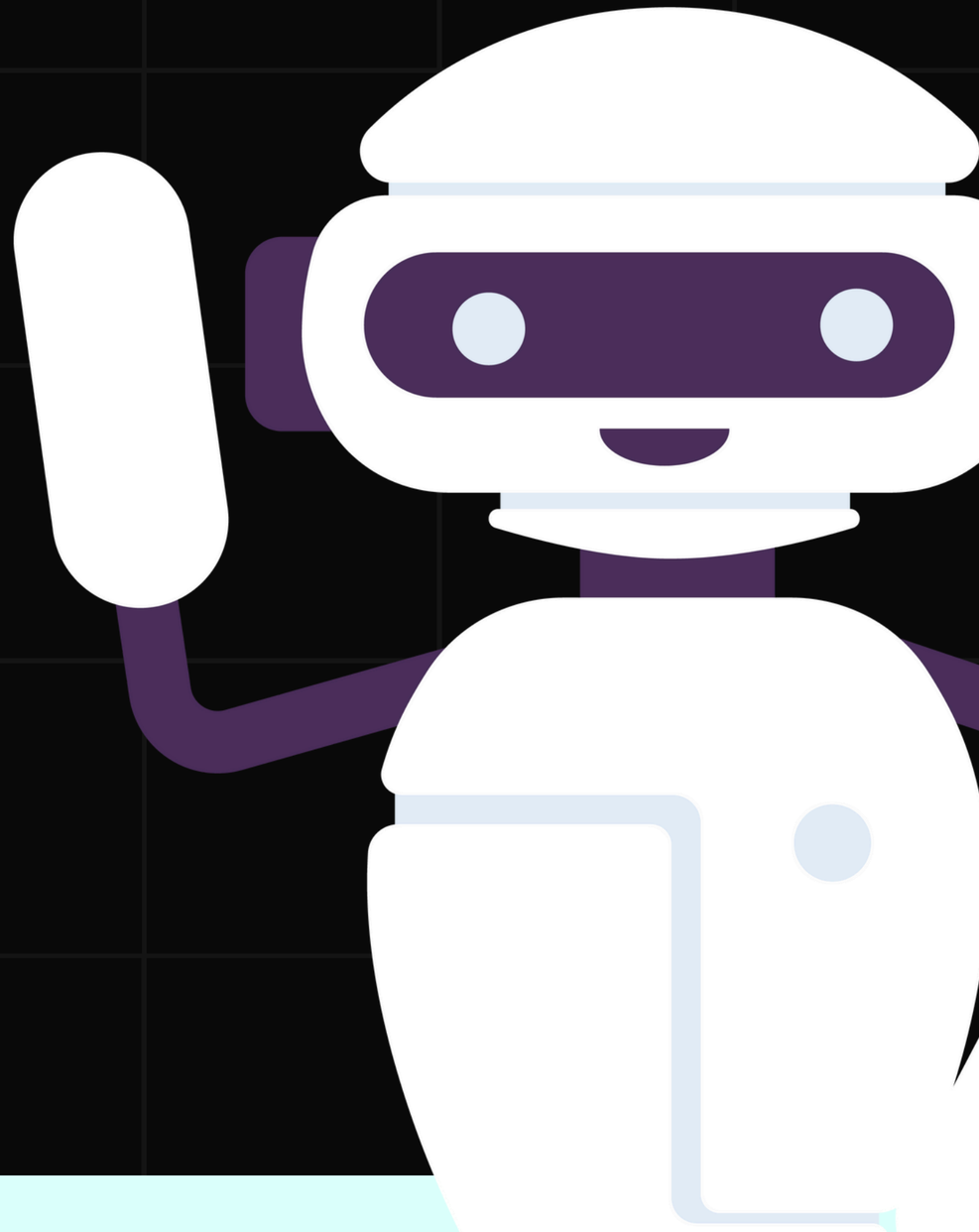
Data Exploration...





Data Exploration...





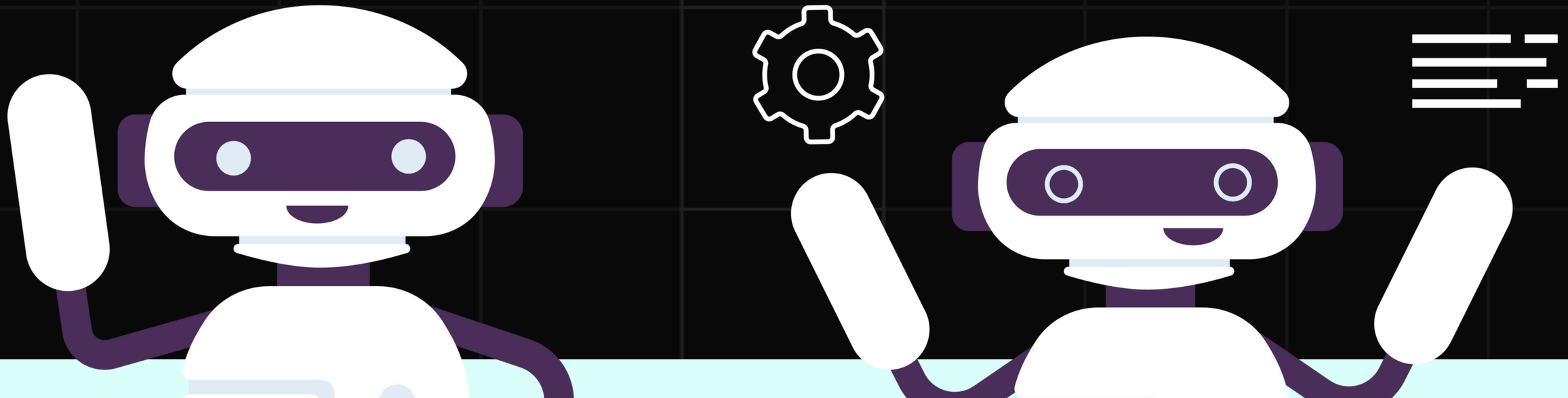
Preprocessing...

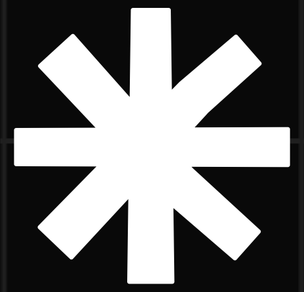
Patient Information

Such as name, gender or weight.

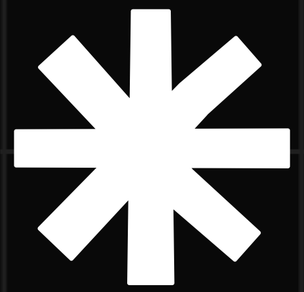
End of the document

Data about the hospital and the patient





Rule-based Model



Papers

A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries

Wendy W. Chapman,*,[1] Will Bridewell,† Paul Hanbury,* Gregory F. Cooper,*,† and Bruce G. Buchanan*,†

*Center for Biomedical Informatics and †Department of Computer Science, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

Received May 29, 2001; published online May 9, 2002

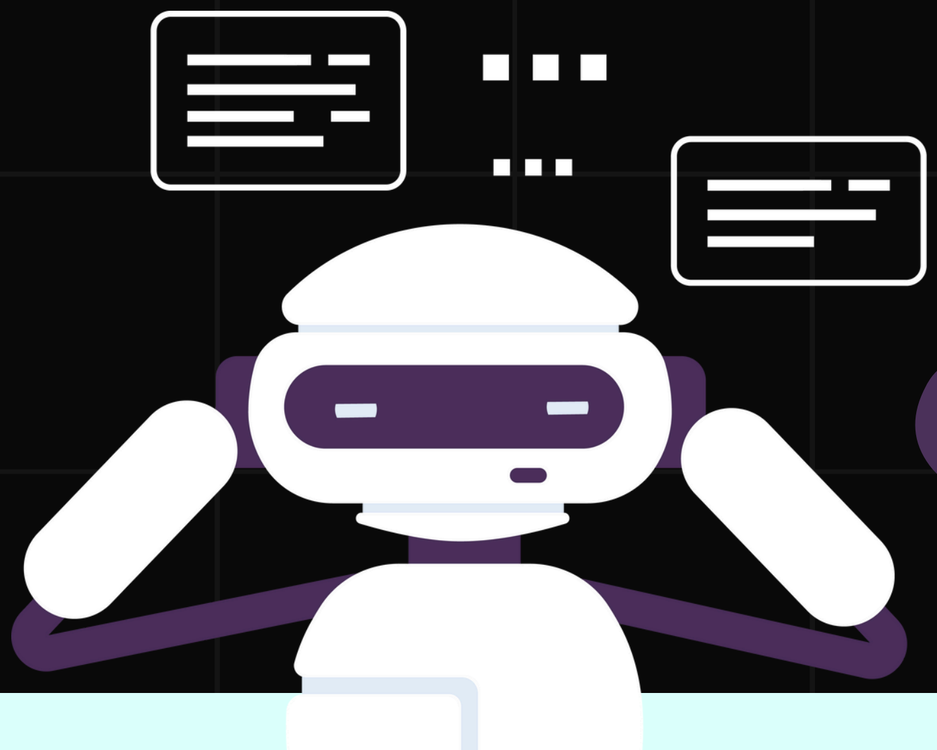
An Approach to Detect Negation on Medical Documents in Spanish

Roberto Costumero[1], Federico Lopez[2], Consuelo Gonzalo-Martín[1], Marta Millan[2], and Ernestina Menasalvas[1]

[1] Universidad Politécnica de Madrid - Centro de Tecnología Biomédica, Madrid, Spain

Simplicity

Similarity to the task



applied sciences

MDPI

Article

Integrating Speculation Detection and Deep Learning to Extract Lung Cancer Diagnosis from Clinical Notes

Oswaldo Solarte Pabón [1,2,*], Maria Torrente [3], Mariano Provencio [3], Alejandro Rodríguez-Gonzalez [1,2] and Ernestina Menasalvas [1,2]

Process

1 **Create 2 lists**

With the words from the corpus that express negation and uncertainty.

2 **Extraction**

We extract all sentences from all the documents.

3 **Rules**

Sentences are passed through all 5 rules

Rules

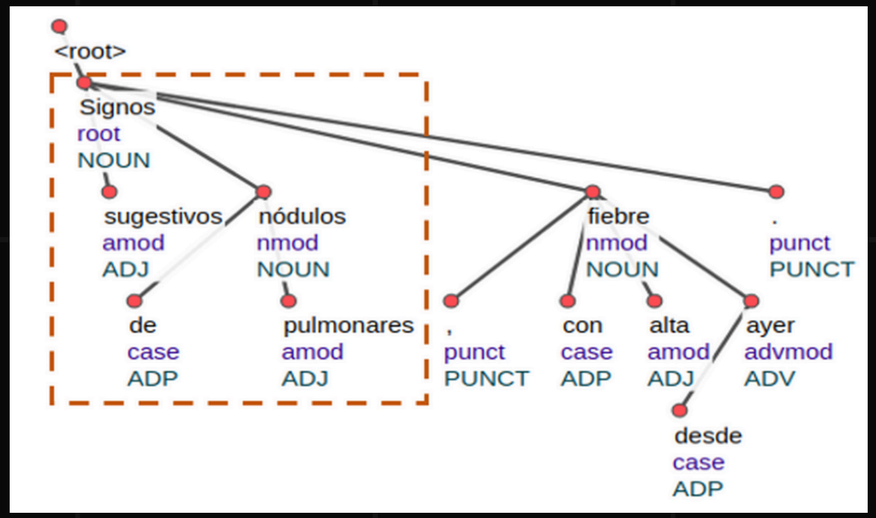
1 *Rule 1:* if the sentence contains a termination term, the scope is extracted using this term. A termination term is a word that indicates the end of the scope.

2 *Rule 2:* if a cue C_1 is detected in a sentence containing contiguous cues $C_1, C_2, C_3, \ldots, C_n$, the scope for each C_i will be given by the position of C_{i+1}.

3 *Rule 3:* if the sentence length corresponds to *"a short sentence heuristic"*, then the scope is given by the end of the sentence.

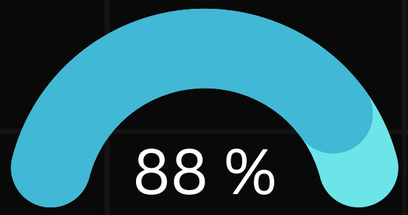
4 *Rule 4:* if the sentence contains a token POS tagged with a conjunction or verb category. In this case, the scope is determined by the position of this token.

5 *Rule 5:* if the sentence does not match the previous rules, the algorithm generates a sentence parse tree. In this case, the scope is given by the sub-tree that contains the uncertainty or negation cue, as is shown in the next sentence.

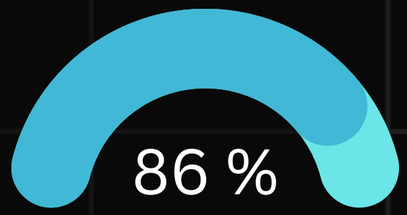


Results

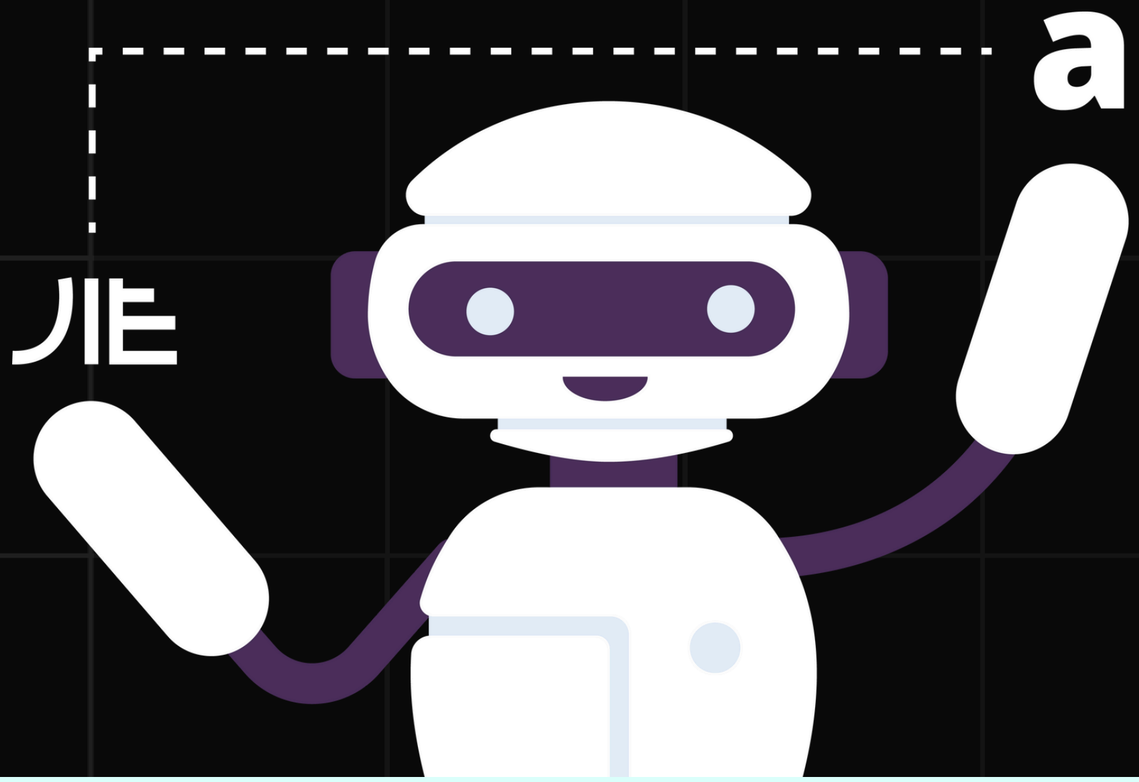
Accuracy, recall and F1-Score

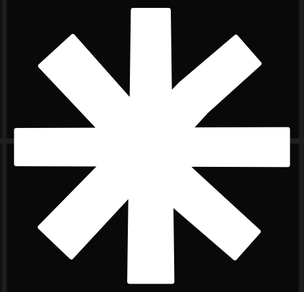


Train

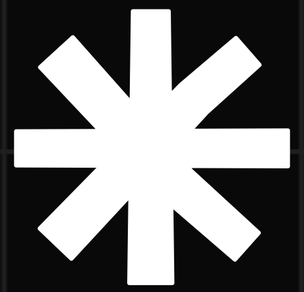


Test





Machine Learning Model



Papers

A metalearning approach to processing the scope of negation

Roser Morante, Walter Daelemans
CNTS - Language Technology Group
University of Antwerp
Prinsstraat 13, B-2000 Antwerpen, Belgium
{Roser.Morante,Walter.Daelemans}@ua.ac.be

An open-source tool for negation detection: a maximum-margin approach

Martine Enger **Erik Velldal** **Lilja Øvrelid**
University of Oslo, Department of Informatics
{marenger,erikve,liljao}@ifi.uio.no

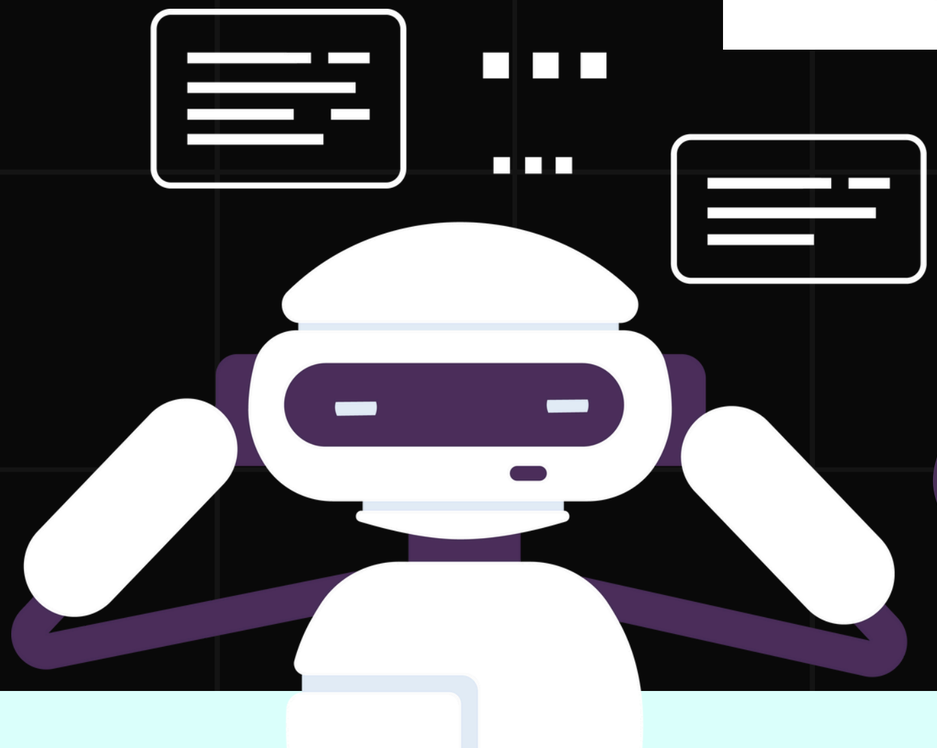
Detection of Negation Cues in Spanish: The CLiC-Neg System

Javier Beltrán[1,3] and Mónica González[2,3]

[1] javier.beltran@ub.edu
[2] monica.gonzalez.manzano@gmail.com
[3] Universitat de Barcelona, Spain

Details

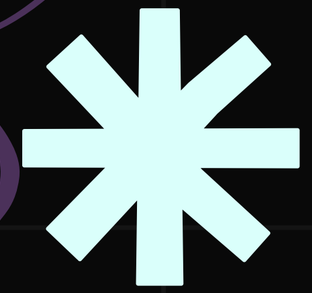
Specific



Negation Cues Detection Using CRF on Spanish Product Review Texts

Detección de Claves de Negación Usando CRF en El Texto de Revisión de Productos en Español

Henry Loharja[1], **Lluís Padró**[1], and **Jordi Turmo**[1]
[1]Universitat Politècnica de Catalunya
https://www.upc.edu/
{loharja, padro, turmo}@cs.upc.edu



Process

1 **Extract features from words**

Features based in the paper.

POS: the information of part of speech of the word.

INIT_CAP: word starts with capitalization.

ALPHANUM: word consists of alphanumeric characters.

HAS_NUM: word contains number.

SUFn: suffixes in the n character length ranged from two to four.

PREFn: prefixes in the n character length ranged from two to four.

2GRAMBEFORE: bigram of up to 6 word before the observed word.

2GRAMAFTER: bigram of up to 1 word after the observed word.

2 **Tagging**

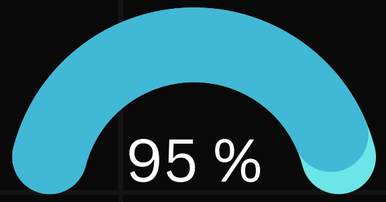
We decided to perform tagging using the same labels of the training set.

3 **CRF Model**

Created with the parameters that gave us the best results.

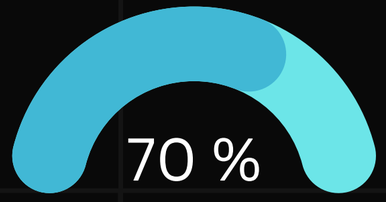
Accuracy

(counting every tag)



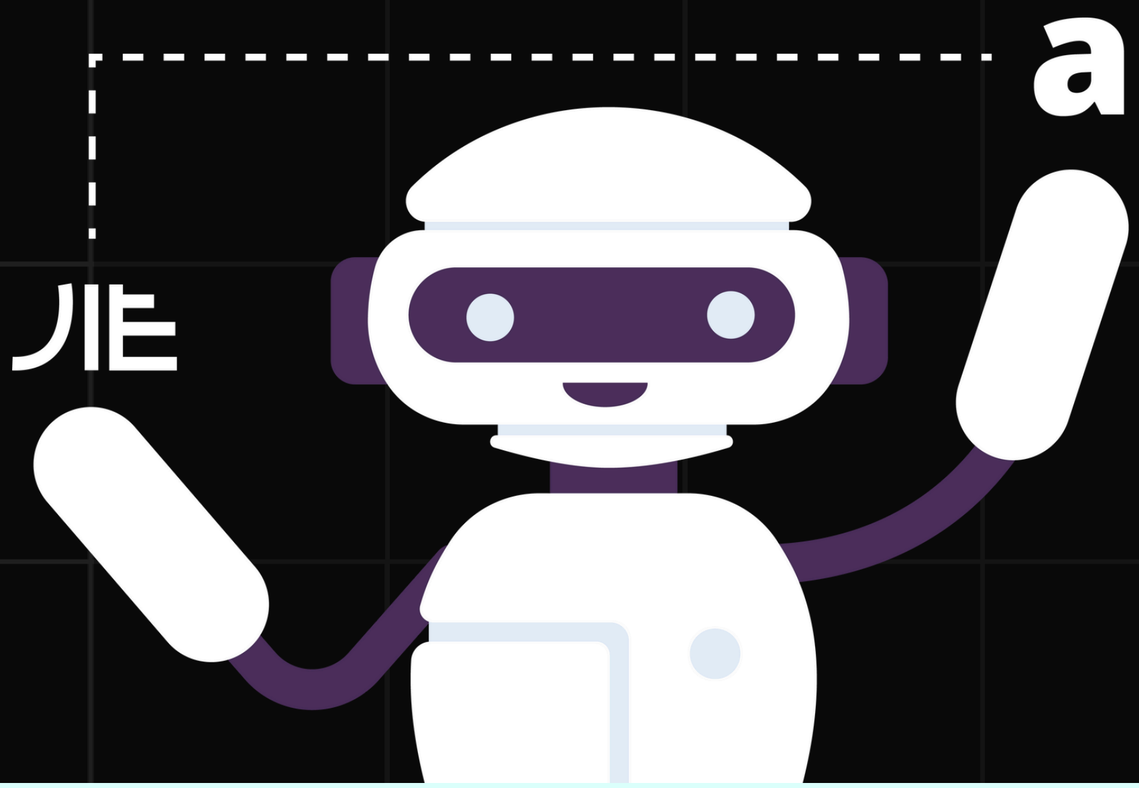
Train & Test

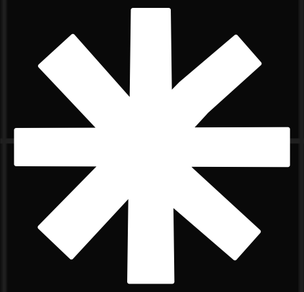
(not counting “O” tags)



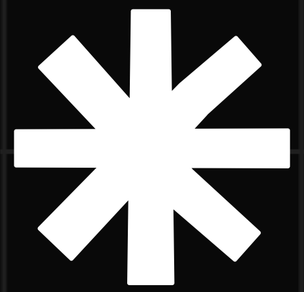
Train & Test

Results





Deep Learning Model



Papers

Deep Learning approach for Negation Cues Detection in Spanish

Aplicación Basada en Deep Learning para Identificación de Claves de Negación en Castellano

Hermenegildo Fabregat[1], **Juan Martinez-Romo**[1–2],**Lourdes Araujo**[1–2]
[1]Universidad Nacional de Educación a Distancia (UNED)
[2]IMIENS: Instituto Mixto de Investigación
{gildo.fabregat, lurdes, juaner}@lsi.uned

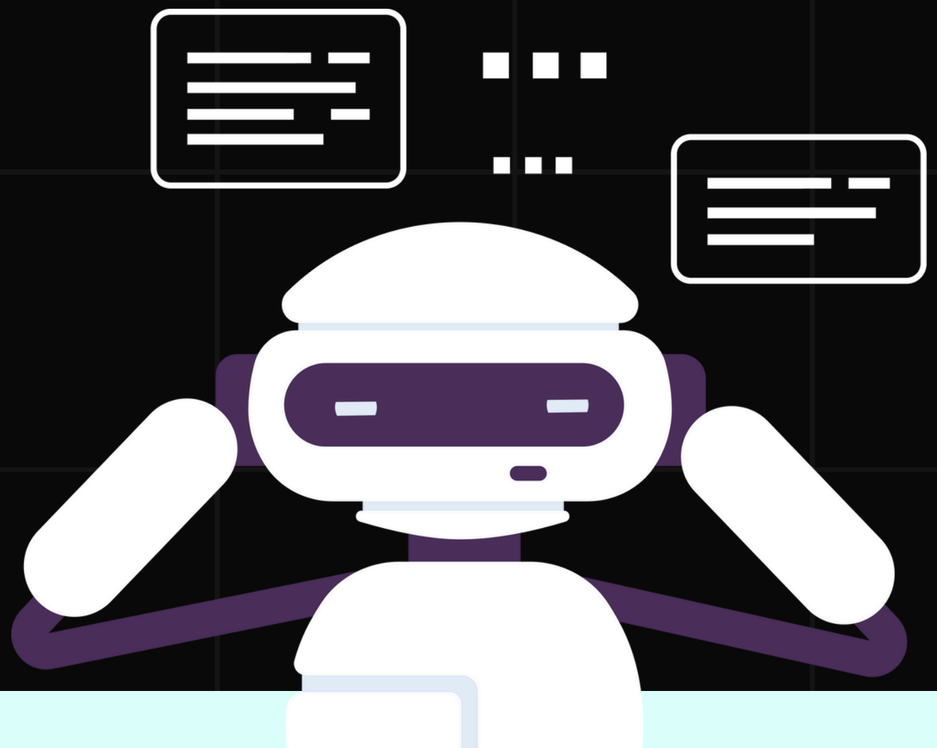
Insightful

Extending a Deep Learning Approach for Negation Cues Detection in Spanish

Hermenegildo Fabregat[1,3], **Andres Duque**[2,3,4], **Juan Martinez-Romo**[1,3,4], and **Lourdes Araujo**[1,3,4]

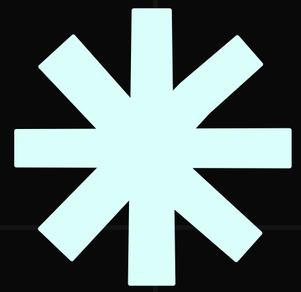
[1]NLP & IR Group, Dpto. Lenguajes y Sistemas Informáticos
[2]Departamento de Sistemas de Comunicación y Control
[3]Universidad Nacional de Educación a Distancia (UNED)
[4]Instituto Mixto de Investigación - Escuela Nacional de Sanidad (IMIENS)
gildo.fabregat@lsi.uned.es, aduque@scc.uned.es
juaner@lsi.uned.es, lurdes@lsi.uned.es

Better Results



Neural Networks For Negation Scope Detection

Federico Fancellu and **Adam Lopez** and **Bonnie Webber**
School of Informatics
University of Edinburgh
11 Crichton Street, Edinburgh
f.fancellu[at]sms.ed.ac.uk, {alopez,bonnie}[at]inf.ed.ac.uk



Process

1

Extract features from words

Sequence of characters, word, POS tag and case tagging

2

Apply padding

So all the sentences have the same length.

3

Change sentence representation

The padding doesn’t have features. The vectors that represent features are filled with zeros.

4

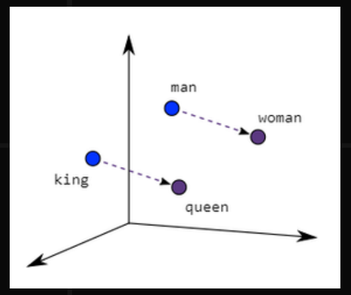
Training

We trained the model replicating the structure in the paper.

Features & Model

Sequence of characters → fastText

Word →



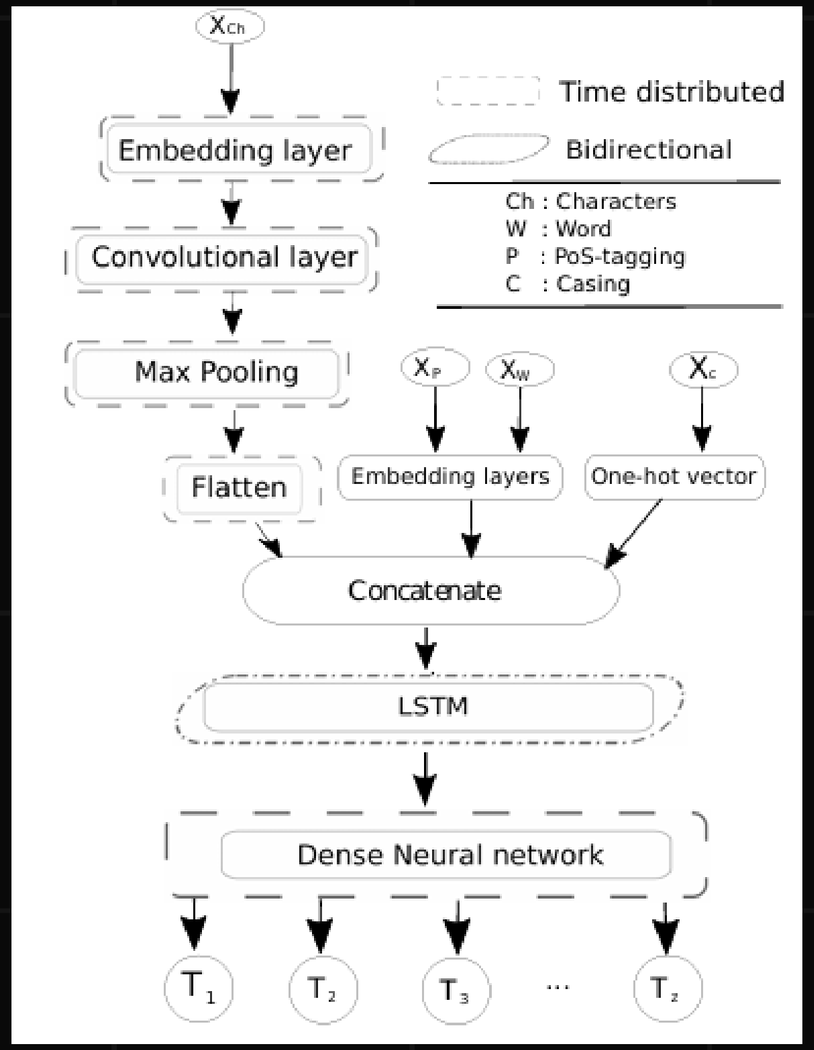
POS tag →

```
[1, 0, 0, 0, 0, 0, …, 0]
[0, 1, 0, 0, 0, 0, …, 0]
```

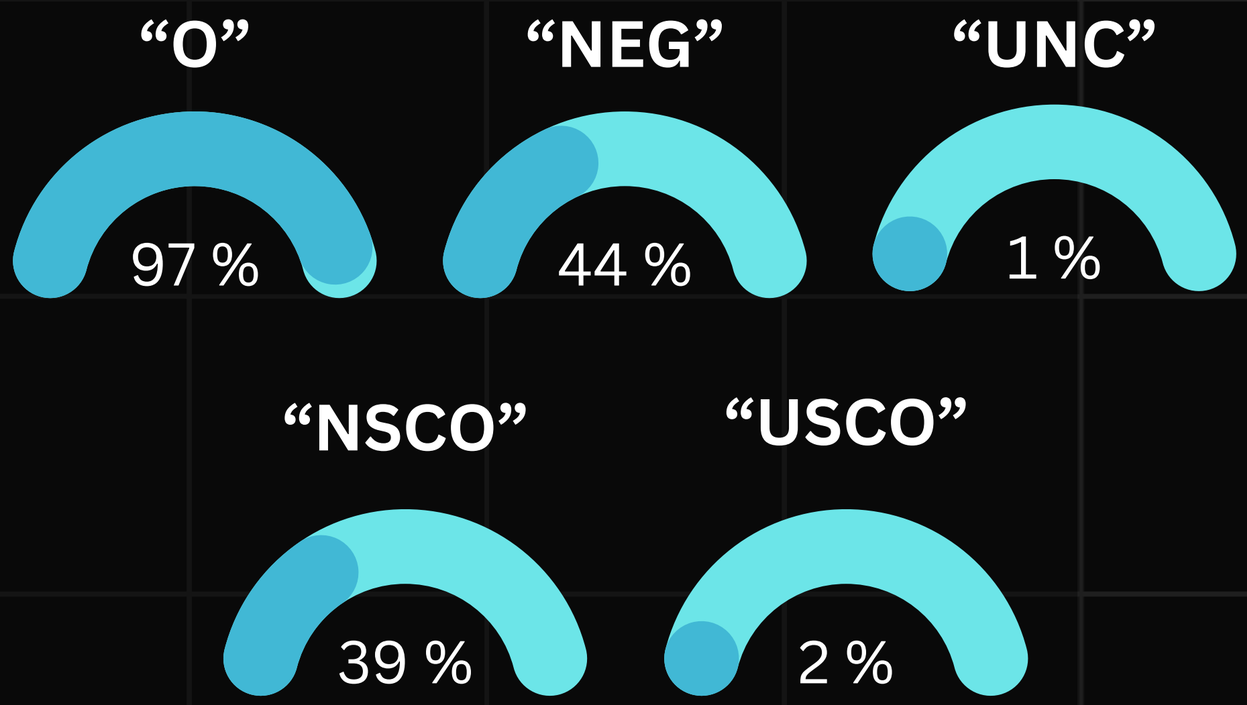
Case tagging →

```
[1, 0, 0, 0, 0, 0, …, 0]
[0, 1, 0, 0, 0, 0, …, 0]
```

Length of 8

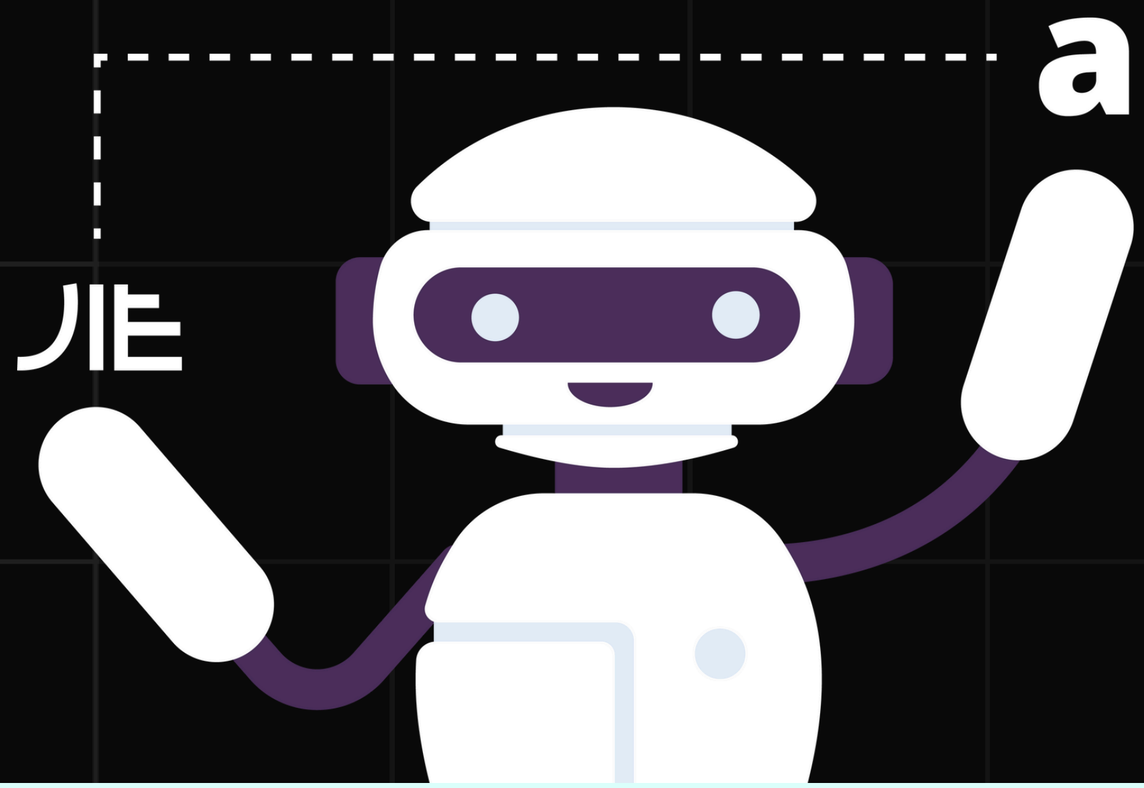


Accuracies by tag



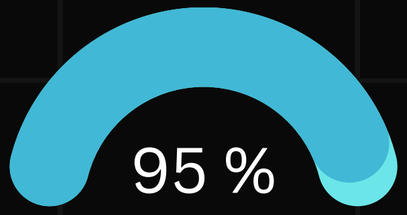
Test

Results



Our best model is...

Machine Learning Model!



Accuracy

→ **Test & Train**

1 CRFs are designed to work with sequential data, making them well-suited for tasks where the context of each word depends on its neighboring words.

2 CRFs allow for the integration of various features, POS tag, bigrams... Great for understanding the context in which these cues appear.

3 Hyperparameters can be chosen, that allowed us to choose the set of parameters that maximized accuracy.

4 CRFs excel at capturing sequential dependencies, which are critical for correctly identifying the start and end of scopes.

Thank you