ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
BUSINESS ANALYTICS Master of Science

MACHINE LEARNING PROJECT

Team members

- **Kasma Marianna**
- **Tsekouras Ilias**
- **Tsiolis Spyridon**

ARGUMENTATION MINING

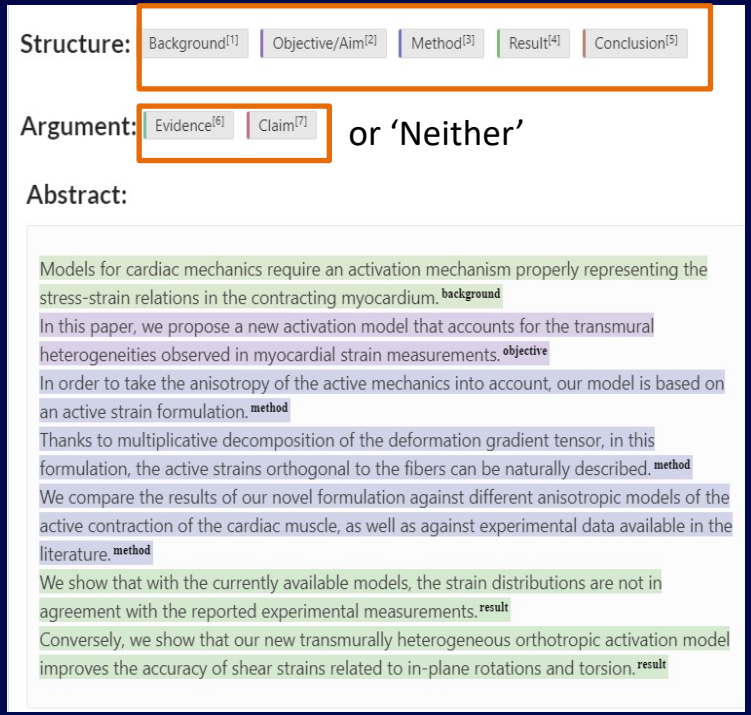
- Data Annotation
- Data pre-processing
- Classification
- Clustering
- Discussion/Comments

DATA ANNOTATION

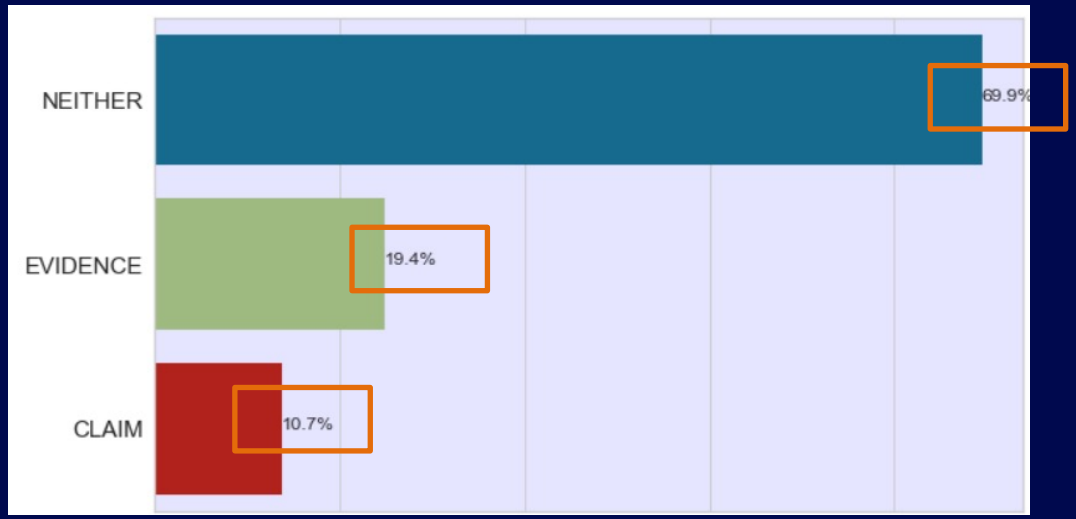
Data Source: Horizon 2020

- ✓ Gathered abstracts originated from different scientific fields, concerning environmental, biomedical and other scientific researches.

Data Labeling Tool: Label Studio

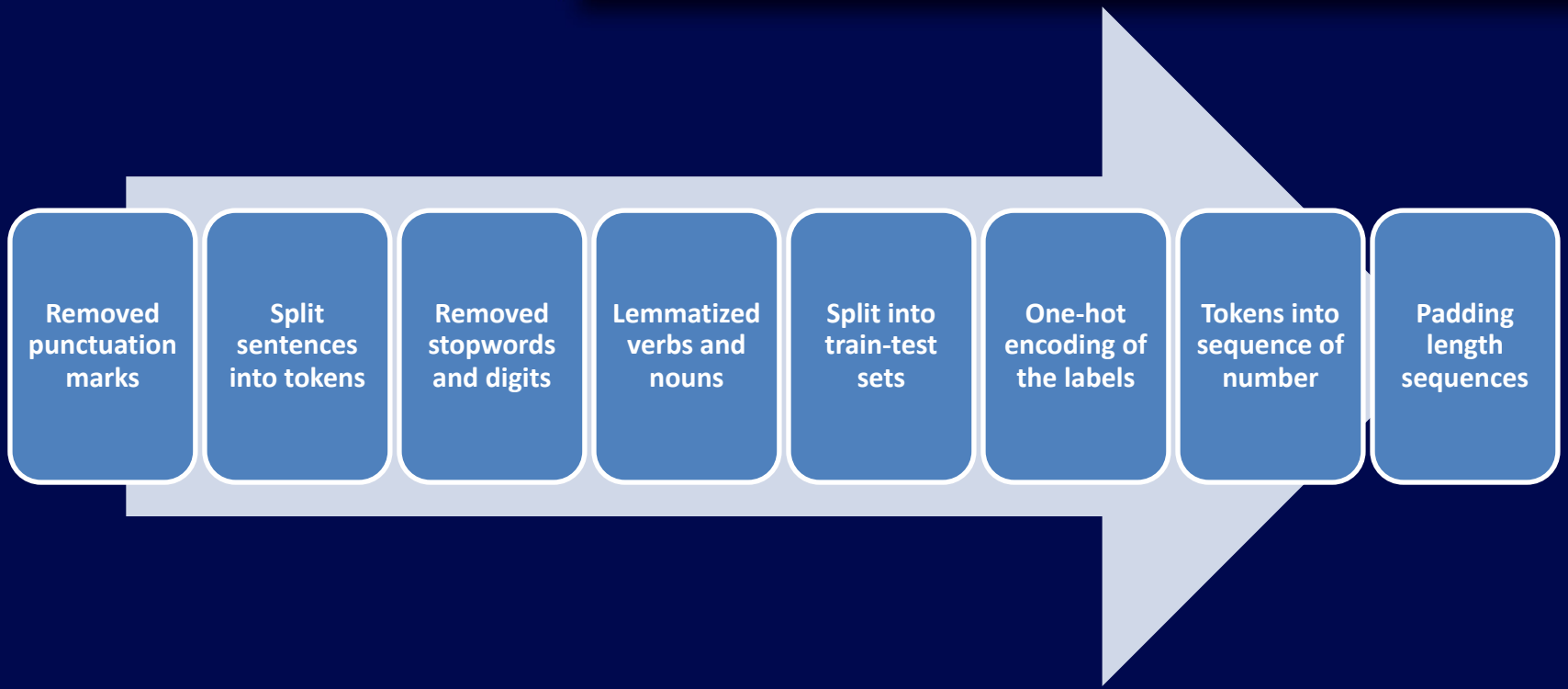


DATA OVERVIEW



- 2686 Abstracts
- 32004 Sentences
- 22375 (Neither)
- 6210 (Evidence)
- 3419 (Claim)

DATA PRE-PROCESSING



CLASSIFICATION

TWO CNN MODELS

1st : Custom Created Embeddings

2nd : Pre-Trained Embeddings (GloVe)

MODEL SUMMARY

- Sequential Model
- Embedding Layer
- Dropout Layer
- Convolutional Layer
- GlobalMaxPooling
- Dense Layer

CLASSIFICATION

1st Model

Confusion Matrix			
Label	Neither	Evidence	Claim
Neither	6094	491	128
Evidence	797	1018	48
Claim	594	182	250

Accuracy Score : 76.7%

2nd Model

Confusion Matrix			
Label	Neither	Evidence	Claim
Neither	6304	284	125
Evidence	994	820	49
Claim	666	98	262

Accuracy Score : 76.9%

CLASSIFICATION

BASELINE MODEL

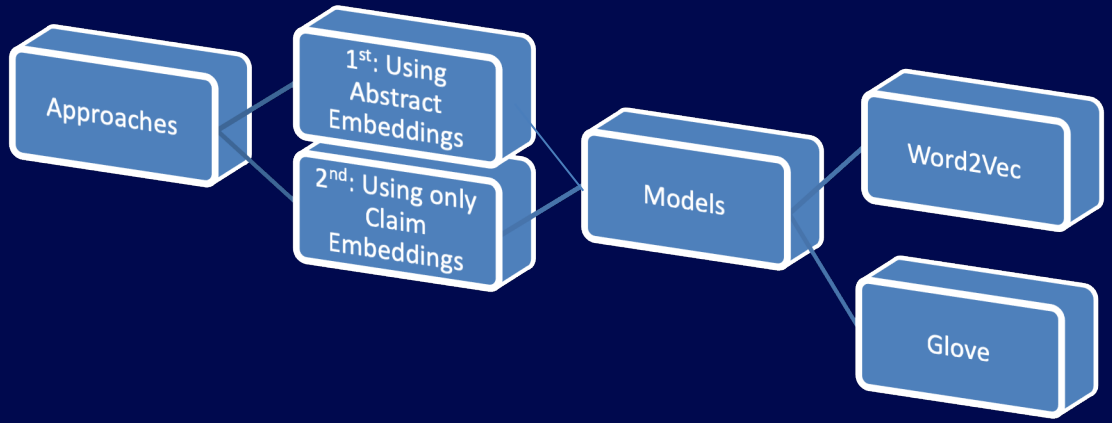
CREATION

- Word Lexicons
  - Claim (provide, confirm, suggest etc.)
  - Evidence (results, findings, etc.)
- Label all last sentences as claim

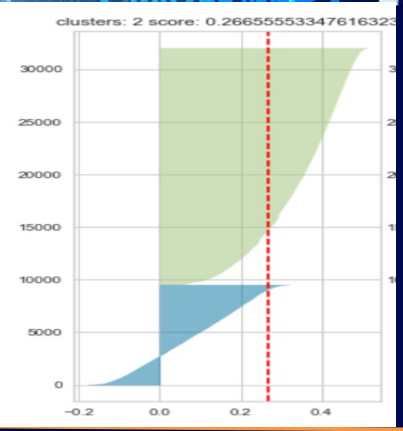
RESULTS

	Recall	Precision	F- Score
Claim	0.594	0.470	0.525
Evidence	0.084	0.333	0.135

CLUSTERING

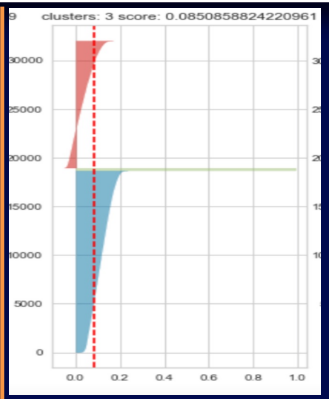

Approaches
1st: Using Abstract Embeddings
2nd: Using only Claim Embeddings
Models
Word2Vec
Glove

CLUSTERING



Abstract – Word2Vec

- K = 2
- Silhouette val = 0.26

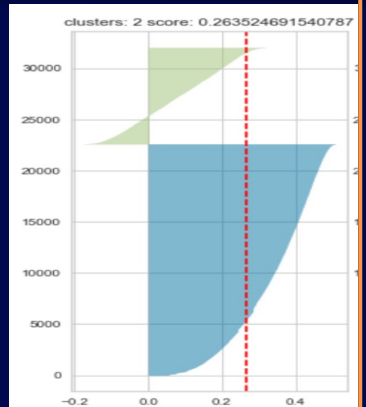


Abstract – GloVe

- K = 3
- Silhouette val = 0.08

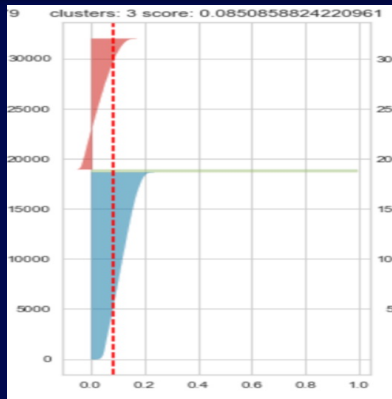
Claim – Word2Vec

- K = 2
- Silhouette val = 0.26



Claim – GloVe

- K = 3
- Silhouette val = 0.08



DISCUSSION/COMMENTS

- More attributes to be used in the future in our clustering
- Silhouette Values close to zero
- Compared to the dummy classifier we observed only 10% improvement in our models

Thank you for your attention!

Any Questions?