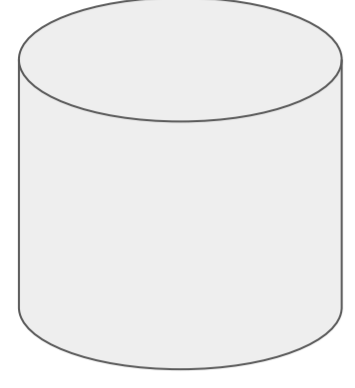


CrudeOilNews CORPUS



- English crude oil-related news annotated for **Event Extraction**
- First of its kind** for Commodity News
- Annotation aligned to **ACE/ERE** annotation schemas
- 21 Entity types** (Nominal and Named Entities)
- 18 Event types** (Geo-political, macro-economic, crude oil supply and demand events)
- 19 Argument roles**

Investing.com

DATASET CREATION #2: DATA AUGMENTATION

1. Trigger word replacement using candidates from FrameNet

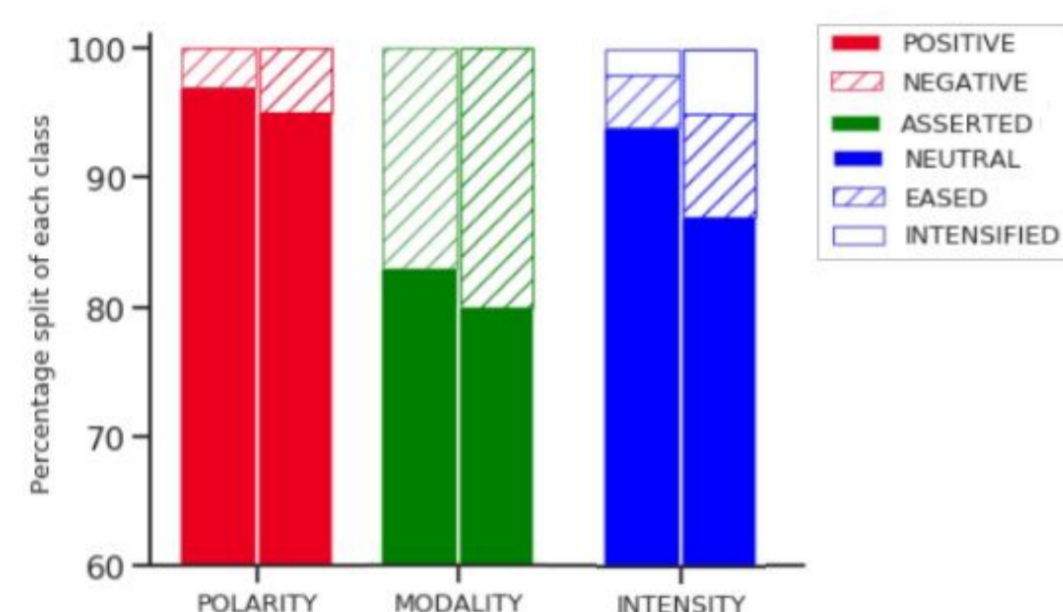
The benchmark for oil prices **advanced** 1.29% to \$74.71.
Candidates: **surged, rose, appreciated, climbed**

2. Event argument replacement - permutations of arguments within the dataset

.....after civil-unrest in **Libya**...
Candidates: **Iraq, Nigeria, Ukraine**

Outcome:

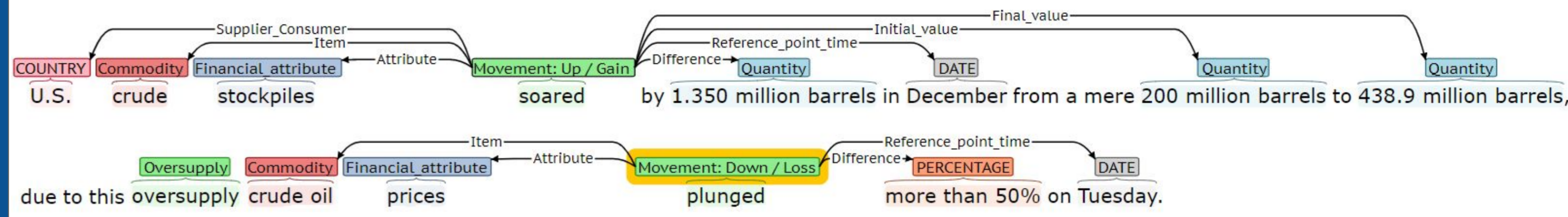
- Increase lexical richness** of the corpus while maintaining the semantic meaning.
- Minimize class imbalance** for event polarity, modality and intensity.



Statistics: - doc, 1,838 entities, 1,061 events, 1,693 arguments

DATASET CREATION #1: MANUAL ANNOTATION

- Layer 1:** Identify and annotate entity mentions
- Layer 2:** Annotate events by identifying event triggers
- Layer 3:** Identify and link arguments to their respective event triggers and identify argument roles each entity plays.



- Layer 4:** Annotate event properties: **Modality, Polarity** and **Intensity**:

- OPEC **cancelled** a **planned easing** of output **cuts** [NEGATIVE, OTHER, EASED]
- In order to end the global crisis, OPEC may **hesitate** to implement a **planned loosening** of output **curbs**. [NEGATIVE, OTHER, EASED]
- Oil prices rose \$110 a barrel on **rumours** of a **renewed strife**. [POSITIVE, OTHER, INTENSIFIED]

Inter-Annotator Agreement (IAA)

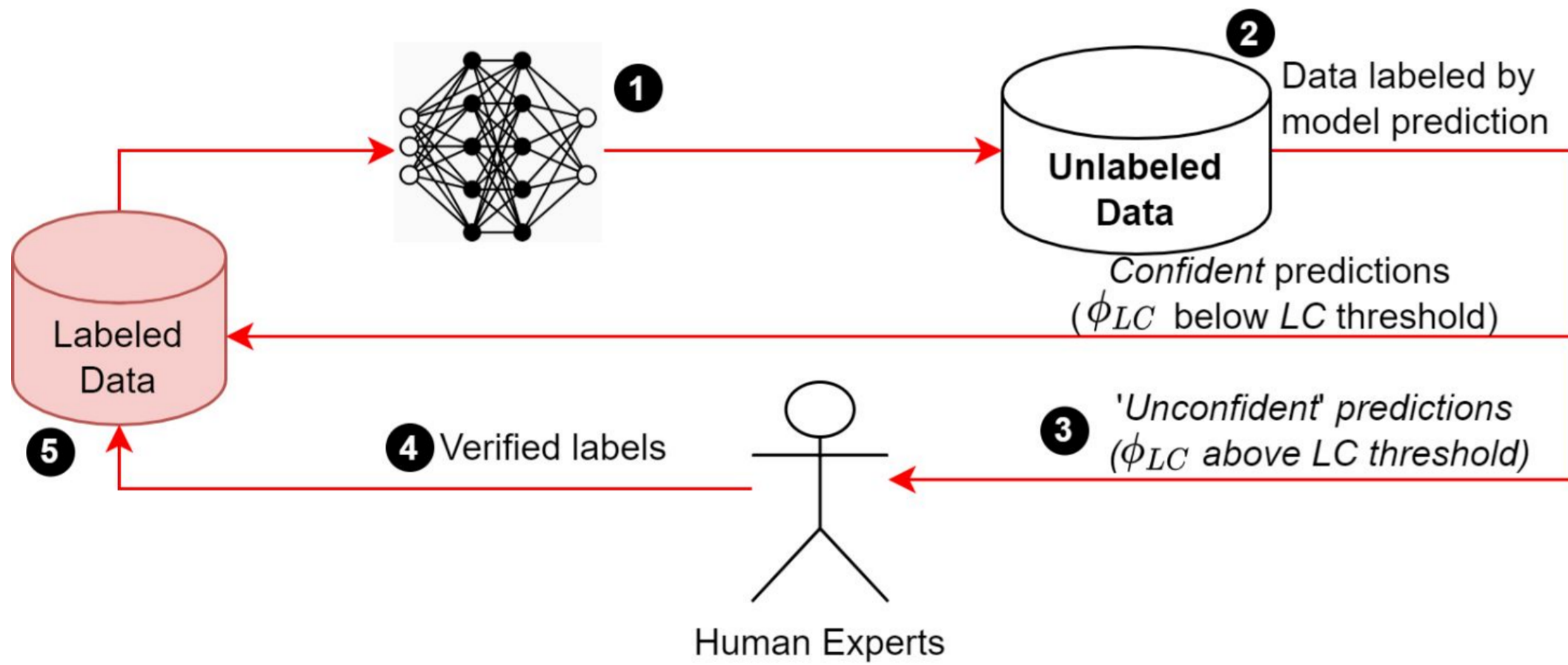
Task	Cohen's Kappa κ	F1 Score
Entity spans*	0.82	0.91
Trigger spans*	0.68	0.75
Entity Type	0.89	-
Event Type	0.79	-
Argument Role	0.78	-
Event Polarity	0.70	-
Event Modality	0.63	-
Event Intensity	0.59	-

IAA is calculated using Cohen's Kappa for all subtasks. For Entity and Trigger span, apart from Cohen's Kappa, they were evaluated on F1-Score as well (similar approach used in IAA for NER tasks).

Statistics: 175 doc, 9,090 entities, 3,520 events, 6,992 arguments

DATASET CREATION #3: HUMAN-IN-THE-LOOP ACTIVE LEARNING

Human-in-the-loop Active Learning



Event Extraction Sub-tasks:

- Entity Mention Detection
- Event Extraction
- Event Properties Classification

Human-in-the-loop Active Learning Cycle:

- train models with labeled data (from manually annotated data and augmented data)
- use the model to label new data via model prediction
- generate sample instances via uncertainty sampling
- validate these sample instance by human experts (relabeling if necessary)
- add checked instances to the pool of training data and re-train the models.

Steps 1 - 5 are repeated for each event extraction sub-task.

Uncertainty Sampling:

extract model inference where *Least Confidence Score*, $\phi_{LC} > LC \text{ threshold}$. ϕ_{LC} is calculated using this equation:

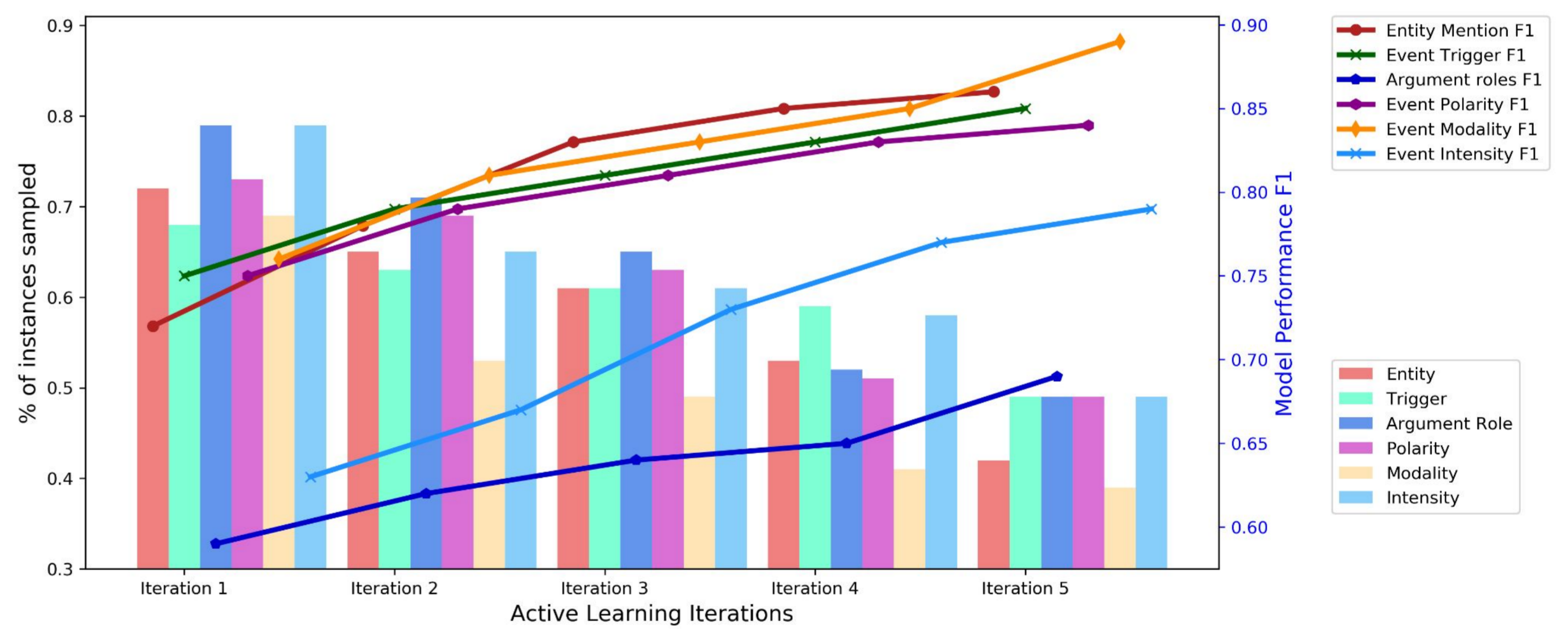
$$\phi_{LC}(x) = (1 - P(y^*|x)) \times \frac{n}{n-1}$$

where y^* = highest softmax score of an inference, n is number of classes

Statistics: 250 doc, 19,417 entities, 5,997 events, 13,582 arguments

- To find the best *LC threshold* individually for each sub-task:
 - Design the uncertainty sampling as a Binary Classification task
 - Repeat experiments with *LC threshold* from 0.0 - 1.0
 - Threshold* is chosen based on the best F1-score of *sampled* (uncertain data) and *non-sampled* data.

Results of five iterations of Active Learning:

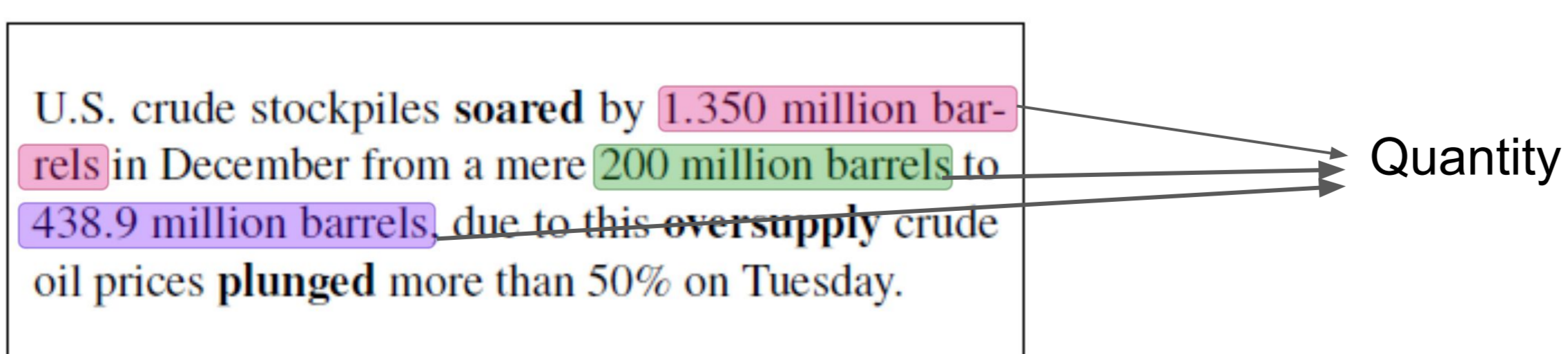


Result interpretation:

Initially, models are less confident, generating large 'uncertain' samples for each subtask. After each iteration, models' F1-score improve and become more confident with smaller 'uncertain' samples.

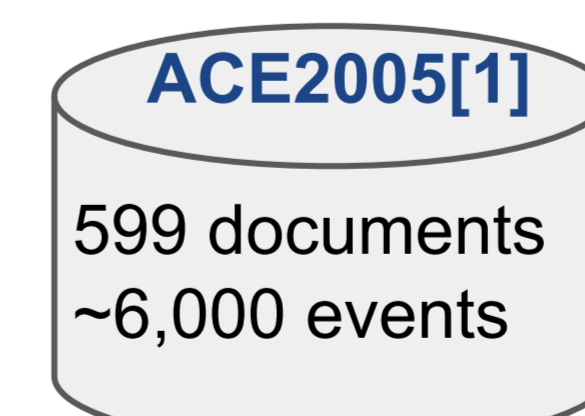
UNIQUE CHARACTERISTICS of CrudeOilNews:

- Number intensity**** - arguments are made up of numbers such as
 - Price - opening, closing price
 - Percentage - percentage of change
 - Dates
- Argument homogeneity** - arguments are made up of similar entity types
- Undifferentiated event types without its arguments** - hence event detection is insufficient.



** values and temporal expressions are considered as entities in this work.

COMPARISON WITH OTHER DATASET:



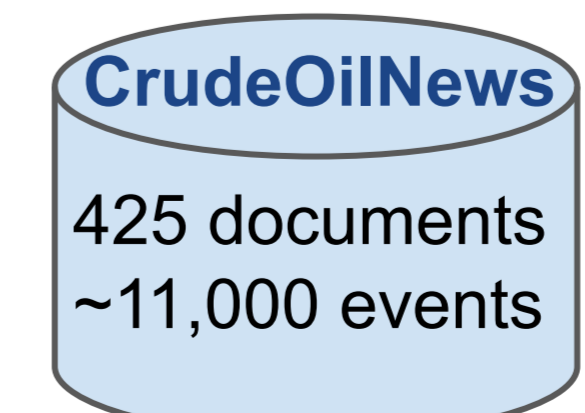
599 documents
~6,000 events

Domain: General
Source: newswire articles, broadcast news and etc
Event: General (33 types)



288 documents
6,194 events

Domain: Business / Economics
Source: Yahoo! Finance business news aggregator
Events: Company-related events (18 types)



425 documents
~11,000 events

Domain: Finance / Economics
Source: investing.com (business news aggregator)
Events: Macro-economics, Geo-political and supply & demand related. (18 types)

REFERENCES:

- <https://catalog.ldc.upenn.edu/LDC2006T06>
- Jacobs, G. and Hoste, V. (2021). Sentivent: enabling supervised information extraction of company-specific events in economic and financial news. Language Resources and Evaluation