



Team	Aspect category classification model			
	headline			
	accuracy	precision	recall	f1-score
CUKG_Tongji	0.2688	0.1512	0.1634	0.1399
IIT-Delhi	0.0537	0.0162	0.0201	0.0149

**Table 1: Aspect category classification model results for headlines.**

Systems are evaluated with regard to aspect classification, sentiment classification and aspect-sentiment attachment.

The Task 1 datasets include two types of discourse: *financial news headlines* and *financial microblogs*, with manually annotated target entities, sentiment scores and aspects. The financial news headlines dataset contains a total 529 annotated headlines samples (436 samples for the training set and 93 samples for the test set) while the *financial microblogs* contains a total 774 annotated posts samples (675 samples for the training set and 99 samples for the test set).

## 2.2 Task 2: Opinion-based QA over Financial Data

Given a corpus of structured and unstructured text documents from different financial data sources in English (microblogs, reports, news) build a Question Answering system that answers natural language questions. For this challenge, part of the questions are opinionated, targeting mined opinions and their respective entities, aspects, sentiment polarity and opinion holder.

The challenge takes both an Information Retrieval (IR) and a Question Answering (QA) perspective. Systems can rank relevant documents from the reference knowledge base with regard to a natural language question or generate their own answer. The relevant score information is implicit if you consider the question-doc matches information contained in the training FiQA\_question\_doc data source.

The Opinion QA test collection is built by crawling Stackexchange posts under the Investment topic in the period between 2009 and 2017. The final dataset contains a KB of 57.640 answer posts with 17.110 question-answer pairs for training and 531 question-answer pairs for testing.

## 3 EVALUATION MEASURES

In order to evaluate the sentiment scores models, regression model evaluation measures were used during the experiments, such as: **Mean Squared Error (MSE)**, **R Square ( $R^2$ )** and **Cosine**

To evaluate the financial aspect category models, classification model evaluation measures were used during the experiments: **Accuracy**, **Precision**, **Recall** and **F1-Score**

To evaluate the opinion question answering models, ranking evaluation measures were used during the experiments: **Normalized Discounted Cumulative Gain (nDCG)** and **Mean reciprocal rank (MRR)**

Team	Aspect category classification model			
	post			
	accuracy	precision	recall	f1-score
CUKG_Tongji	0.8484	0.5	0.4357	0.4619
NLP301	0.7575	0.3006	0.2678	0.2832
IIT-Delhi	0.2424	0.0274	0.0229	0.0250

**Table 2: Aspect category classification model results for posts.**

Team	Opinion question answering model	
	nDCG@10	MRR
eLabour	0.3052	0.1947
CUKG_Tongji	0.1682	0.0957

**Table 3: Opinion question answering results.**

Team	Sentiment score prediction model		
	headline		
	MSE	$R^2$	cosine
CUKG_Tongji	0.1345	0.4579	0.6768
IIT-Delhi	0.2039	0.1779	0.4401
Inf-UFG	0.2067	0.1665	0.4153

**Table 4: Sentiment score prediction results for headlines.**

Team	Sentiment score prediction model		
	post		
	MSE	$R^2$	cosine
Inf-UFG	0.0958	0.1642	0.5333
CUKG_Tongji	0.1040	0.0923	0.6063
IIT-Delhi	0.1049	0.0849	0.3422
NLP301	0.3058	-1.6669	-0.0685

**Table 5: Sentiment score prediction results for posts.**

## 4 EVALUATION RESULTS

Sentiment-based models were evaluated with regard to aspect category classification and sentiment score prediction. For question answering models, each team sent the output file containing the top 10 most relevant answers.

Tables 4 and 5 show the results for each sentiment score prediction models. Table 1 and 2 show the result for each aspect category models. For opinion question answering, the results were showed in Table 3.

## ACKNOWLEDGMENTS

This publication has emanated from research funded in part from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645425 SSIX.

