# MamaEdhaatSemEval-2017Task8: Stance Classiﬁcation with CNN and Rules

**Preprocessing**

* Stemming : snowballstemmer :python package
* Mentions and urls are deleted replaced with @ and http://
* POS tagging parsey mc-parseface (andor et al 2016) conll-U (nivre 2015)as data format for POS tagging
* Rule encoding feng et al 2012

**Feature extraction**

* Word to vec word embedding (skipgram model)

**Model building**

* Voting among several models
* CNN (kim 2014 architecture)– Automatic rule mining – handwritten rules-

**Results**

* 75% acc : grammatical structure had little impact

# NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter

**Preprocessing**

* NLTK Tweet Tokenizer tool used
* Removal of stop words, punctuation and abbreviations
* Stemming and lemmatization showed to worsen
* Removing urls showed not beneficial
* Using bigrams (no benefits)

**Feature extraction**

* Being a question :Question detection module : question mark and questioning keywords like wh- etc, utility classifier?.
* Denial term detection “not true ”
* Support words detection: “true”, “exactly”
* Hashtag existence
* url existence
* tweet/reply
* counting word sentiments , positive , negative balance
* tweet sentiments
* is user verified
* number of followers
* number of users past tweets
* number of users friends
* retweet ratio between source and reply
* photo existence
* days since user creation
* source tweet user is verified

Subtask A features

* User replied to is verified
* Cosine similarity with root rumorous tweet and the replied to tweet

Subtask B features

* Percentage of replying tweets classified as S D Q or C

**Feature selection**

* Chi squared feature selection
* Variance threshold feature selection
* Truncated svd dimensionality reduction

(non of them were used because making it worse in cross validation)

Additional features selelcted manually by seeing the result on training data

**Model building**

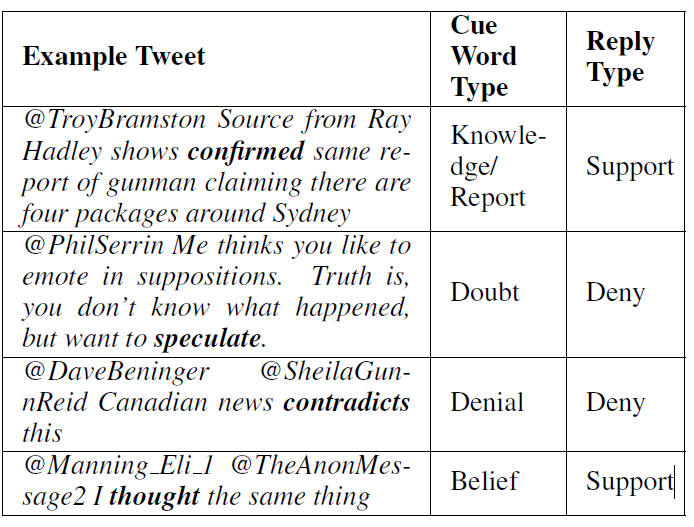
* Trying several scikit learn before final model
* for subtask A : linear SVC best, logistic regression

# UWaterloo at SemEval-2017 Task 8: Detecting Stance towards Rumours with Topic Independent Features

**Preprocessing**

* removal of quoted text (reply tweets at times quote the source tweet)
* discarding URLs, unicode characters, HTML tags
* stripping out the extra whitespaces and carriage returns

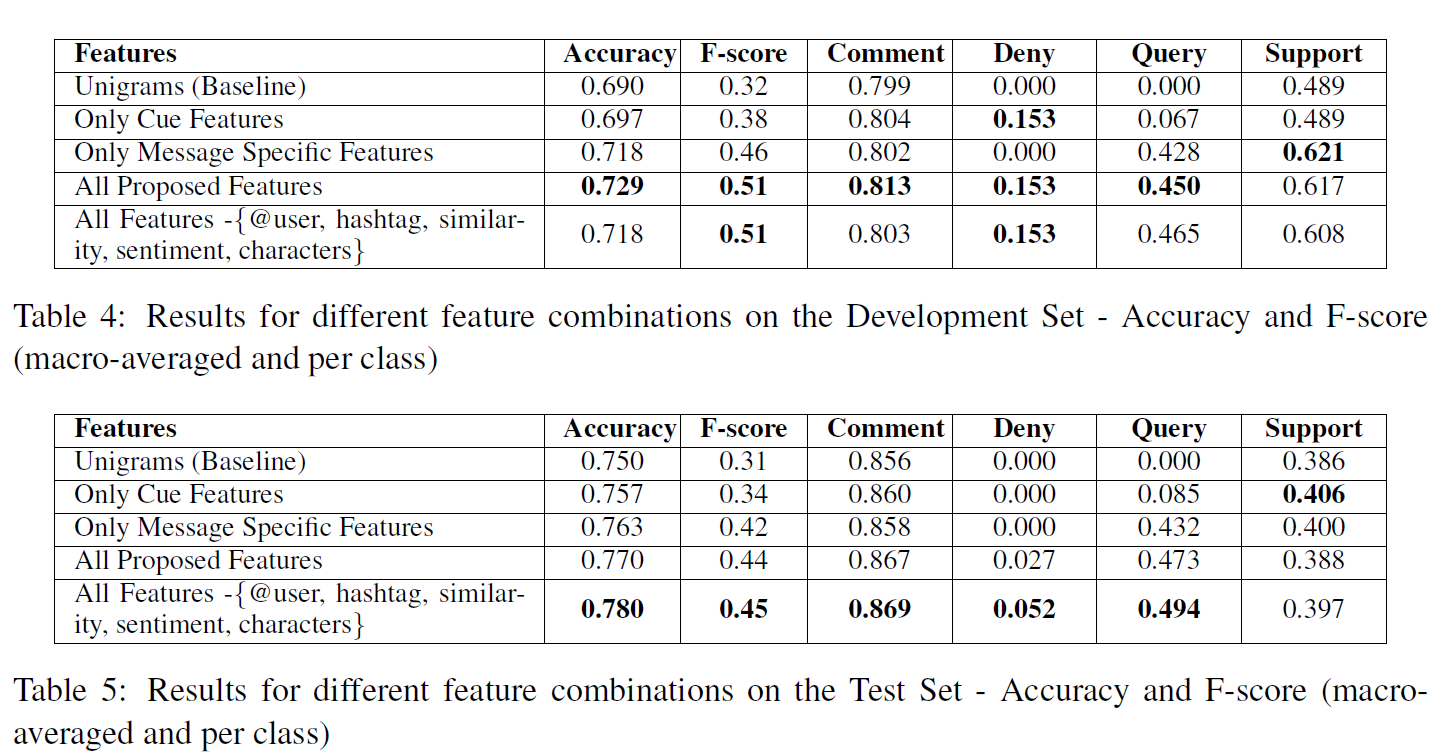
**Feature extraction**

* cue features :
* 
* certain other tweet specific features were also used as part of our model. Punctuation hastage etc.
* For calculating the sentiment polarity score, the lexicon based social media sentiment calculator, VADER, developed by Hutto and Gilbert (2014), was used

**Model Building**

supervised classification algorithm, specifically Gradient Boosting.

**Results**



# ECNU at SemEval-2017 Task 8: Rumour Evaluation Using Effective Features and Supervised Ensemble Models

Two step classifier by first classifying into comment(objective) and non-comment(subjective) which The 2-step is actually to classify a non-comment tweet that expresses positive (support), negative (deny) or doubtful (query)

**Preprocessing**

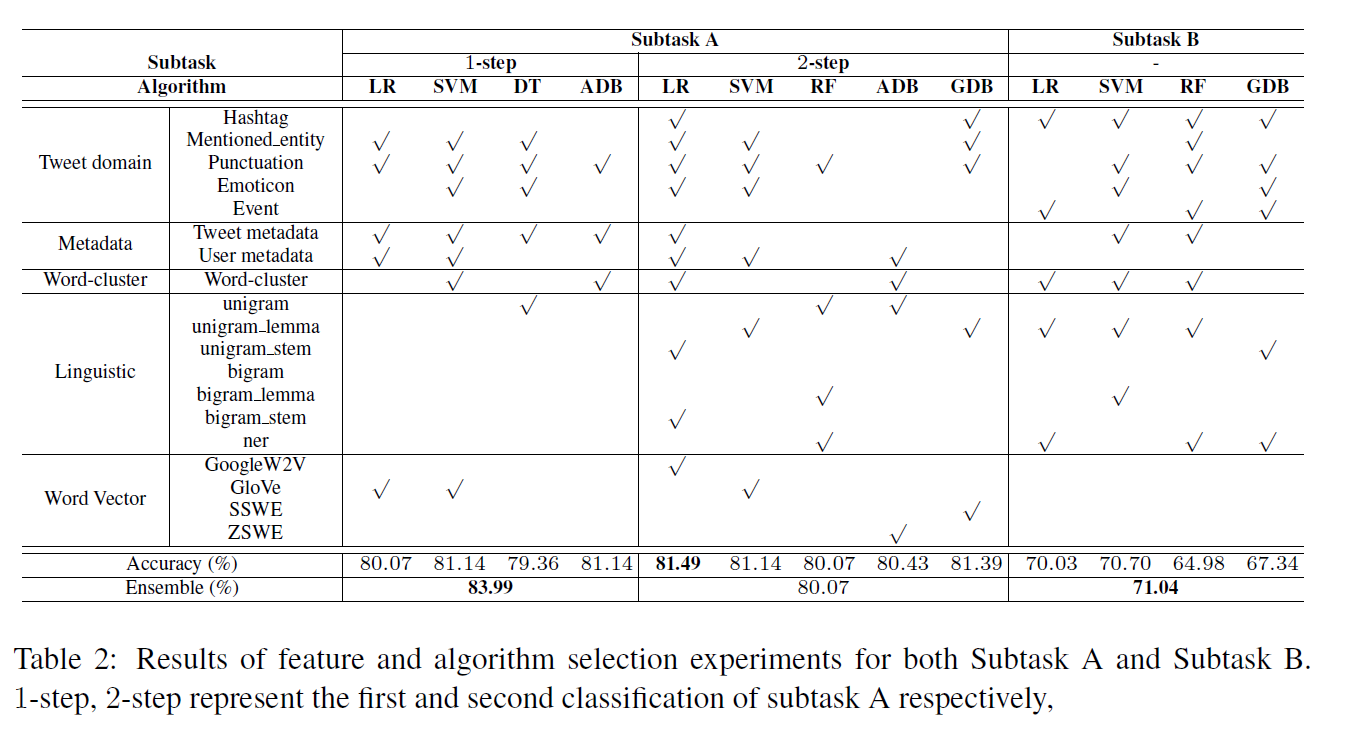
**Feature Extraction**

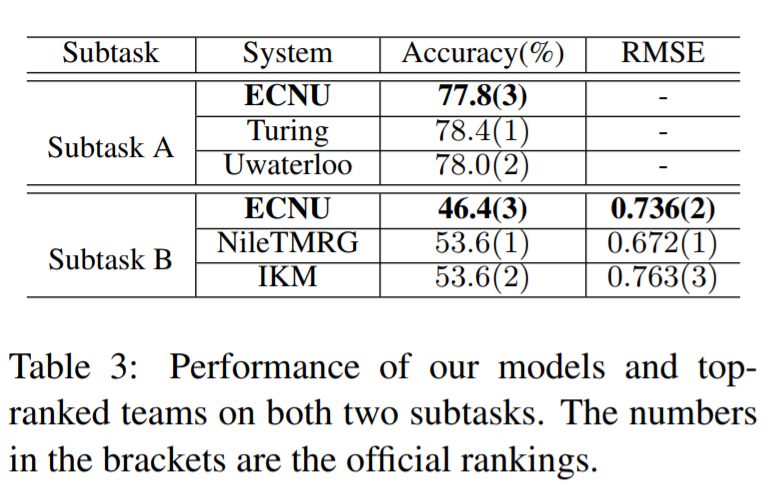
* Word N-grams:
* NER : The 12 types (i.e., DURATION, SET, NUMBER,LOCATION, PERSON, ORGANIZATION, PERCENT, MISC, ORDINAL, TIME, DATE, MONEY) named entities are labeled by Stanford CoreNLP tools.
* There are some particular elements in tweets, that can help to predict labels of tweets. For instance, hashtag and mentioned entity (e.g., “#semeval”, “@YouTube”) express the topic information of the tweets, and several special punctuation and emotions (e.g., “!”, “?”, and “:)”) reveal the sentiment information of users.
* Tweet domain features like hashtags and urls and mentions : punctuations – emoticon – events
* Metadata: tweet metadata , user metadata
* Word-cluster Feature: To further group similar words into a small set and to make better use of word semantic information

**Model building**

* Logistic Regression (LR), supplied in liblinear tools6, Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), AdaBoost (ADB), and Gradient Tree Boosting (GDB), implemented in scikit-learn7.

**Results**





# DFKI-DKT at SemEval-2017 Task 8: Rumour Detection and Classification using Cascading Heuristics

**Preprocessing**

* standard tokenisation scripts
* it was discovered that cleaning the tweets actually impacted the classification algorithm in a negative way.

**Feature-extraction**

* Message is a retweet, i. e., begins with RT
* Message is a reply (@usernames)
* Message contains exclamation marks
* Message is a question (question mark or whword:
* who/what/when/why/where/how)
* Message contains emoticons (smileys)
* At least 70% of the message is in uppercase
* Message contains negations (not, doesn’t)
* Message contains expletives or abuse

**Feature-selection**

**Model-Building**

* used heuristic rules as a post process (doesn’t seem like a great job because others have used smarter rules in a smarter way) Subtask A rules:
  + If a tweet begins with a wh-word (where,when, how, what, why, which) and/or ends with a question mark, then classify it as query
  + If a tweet contains a negation, then classify it as denial
  + If a tweet is a retweet, then classify it as support
  + If more than 70% of the text is all uppercase,
  + then classify it as comment

Subtask B rules

* Maximum Entropy classification (MaxEnt) (Malouf, 2002), also known as Multivariate
* Logistic Regression.
* Naive Bayes classification (Frank and Bouckaert, 2006) assumes independence of the features while counting.
* Winnow classification (Winnow2) (Littlestone, 1988) is similar to the perceptron model but uses a multiplicative weight update
* Ensemble of these models

# IITP at SemEval-2017 Task 8 : A Supervised Approach for Rumour Evaluation

# Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM

# IKM at SemEval-2017 Task 8: Convolutional Neural Networks for Stance Detection and Rumor Verification