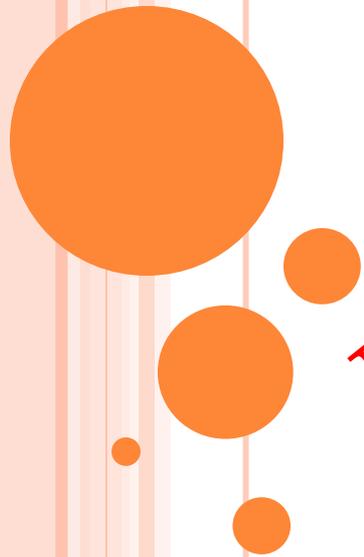


**TOWARDS REAL-TIME, COUNTRY-LEVEL LOCATION  
CLASSIFICATION OF WORLDWIDE TWEETS**

**MICANS INFOTECH**



# ABSTRACT

- The increase of interest in using social media as a source for research has motivated tackling the challenge of automatically geolocating tweets, given the lack of explicit location information in the majority of tweets.
- In contrast to much previous work that has focused on location classification of tweets restricted to a specific country, here we undertake the task in a broader context by classifying global tweets at the country level, which is so far unexplored in a real-time scenario.
- We analyse the extent to which a tweet's country of origin can be determined by making use of eight tweet-inherent features for classification.
- Furthermore, we use two datasets, collected a year apart from each other, to analyse the extent to which a model trained from historical tweets can still be leveraged for classification of new tweets.

# EXISTING SYSTEM

- Choosing an appropriate combination of both tweet content and metadata can actually lead to substantial improvements of between 20% and 50%.
- We observe that tweet content, the user's self-reported location and the user's real name, all of which are inherent in a tweet and available in a real-time scenario, are particularly useful to determine the country of origin.
- We also experiment on the applicability of a model trained on historical tweets to classify new tweets, finding that the choice of a particular combination of features whose utility does not fade over time can actually lead to comparable performance, avoiding the need to retrain
- However, the difficulty of achieving accurate classification increases



# DISADVANTAGE

- There is no big difference between the two approaches based on GeoNames when we look at micro-accuracy
- Note that while higher values are desired for micro-accuracy and macro-accuracy,
- lower values are optimal for MSE.

MICANS INFOTECH



# PROPOSED WORK

- What motivates the present study is the increasing interest in inferring the geographical location of either tweets or twitter users.
- The automated inference of tweet location has been studied for different purposes, ranging from data journalism [21], [34] to public health [15].
- Table 1 shows a summary of previous work reported in the scientific literature, outlining the features that each study used to classify tweets by location, the geographic scope of the study,
- The languages they dealt with, the classification granularity they tried to achieve and used for evaluation, and whether single tweets, aggregated multiple tweets and/or user history were used to train the classifier.



# ADVANTAGES

- Choosing an appropriate combination of both tweet content and metadata can actually lead to substantial improvements of between 20% and 50%
- Our methodology enables us to perform a thorough analysis of tweet geolocation, revealing insights into the best approaches for an accurate country-level location classifier for tweets.
- The micro-accuracy is computed for the test set as a whole.
- For macro-accuracy, we compute the accuracy for each specific country in the test set, which are then averaged to compute the overall macro accuracy.



# SYSTEM REQUIREMENT

## ○ **HARDWARE REQUIREMENT:**

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Sony.
- Ram : 512 Mb.

**MICANS INFOTECH**



## CONT..

### ○ **SOFTWARE REQUIREMENT:**

- Operating system : Windows XP.
- Coding Language : ASP. Net with C#
- Data Base : SQL Server 2005.

**MICANS INFOTECH**



# CONCLUSION

- To the best of our knowledge, this is the first study performing a comprehensive analysis of the usefulness of tweetinherent features to automatically infer the country of origin of tweets in a real-time scenario from a global stream oftweets written in any language.
- Most previous work focusedon classifying tweets coming from a single country and hence assumed that tweets from that country were already identified.
- Where previous work had considered tweets from all over the world, the set of features employed for the classification included features,
- such as a user's social network, that are not readily available within a tweet and so is not feasible in a scenario where tweets need to be classified in real-time as they are collected from the streamingAPI.

