# MACHINE LEARNING
# CSL - 603
# Project Proposal

SMS Spam Detection

By -
Manish Kumar
(2015csb1019)
&
Shikhar Jaiswal
(2015csb1116)

Instructor -
Dr. Narayanan C Krishnan

**Aim:** Our aim is to apply various Machine Learning algorithms to detect SMS Spam in mobile phone. We would explore the problem up to a deep extent by comparing the performance of those algorithms in different scenario and finally design an application based on the algorithm that filters the spam with high accuracy.

**Motivation:** In recent years, popularity of mobile phones has increased and SMS cost has reduced significantly. This resulted in growth in unsolicited commercial advertisements (spams) sent to mobile phones.
In 2012 more than 30% of SMS in Asia were spam.!!!
Limited availability of mobile phone spam-filtering softwares makes spam-detection an interesting problem to look into.

**Introduction:** Spam filtering in SMS and emails are quite different. There is a large amount of dataset available for emails but real datasets for SMS are very limited. Also SMS are relatively smaller, which makes extracting good amount of features difficult. Also language of SMS are very informal and full of abbreviations and idioms. These factors cause existing email filtering algorithms to underperform. We will use a database of real SMS from UCI Machine Learning repository. Apply different machine learning techniques, compare the results and introduce best algorithm for SMS spam filtering.

Probable list of classifiers:

- Basic Naive Bayes
- Multinomial Boolean NB
- Multivariate Bernoulli NB
- Boolean NB
- Multivariate Gauss NB
- Flexible Bayes
- Boosted NB
- Minimum Description Length

Linear Support Vector Machine
Logistic Regression
K-Nearest Neighbors
Random Forest  etc.

**Database:** We will use a database of 5574 real text messages from UCI Machine Learning repository gathered in 2012
Link : https://archive.ics.uci.edu/ml/datasets/sms+spam+collection

**References:**

[1] Almeida, T.A., Gomez Hidalgo, J.M., Yamakami, A. Contributions to the study of SMS Spam Filtering: New Collection and Results. Proceedings of the 2011 ACM Symposium on Document Engineering, Mountain View, CA, USA, 2011.
http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/doceng11.pdf

[2] **http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/**