Insult Detection in Hindi
Chetan Dalal(11218) Shivyansh Tandon(11690) Guide: Prof. Amitabha Mukherjee
Indian Institute of Technology – Kanpur, India

Problem Statement
- To detect Insults in one-liner Hindi comments in blogs and conversations.
- Given a tagged corpus of M sentences and unlabeled of N sentences, we use the tagged data to obtain good accuracy on the unseen to identify if the new comment is insult or not.
- Insults may be racial slurs, abuses, taunts, sarcasm, provocative words etc.

Eg. To determine if “प्यारा घर से जाना बुद्ध हरमी जीव है” is an insult or not.

Present Importance
- With the increase of technology and availability of ‘Hindi’ language in mobiles and computers, it has become necessary to also have systems capable of detecting insults.
- The exponential growth of social networking sites, blogs etc. have encouraged us to do build such a system, because a human cannot monitor all the data that is flooded on the internet.
- For Eg. Over 409 million people view more than 14.7 billion pages each month and users produce about 44.5 million new posts and 56.7 million new comments each month. (source: WordPress)
- Hence, a need arises to filter all this data.

Past Works
- No work has been done for Insult detection in Hindi.
- Part of Speech n-grams, skipgrams, pattern matching approaches have been used.

Challenges with Dataset
- No proper dataset available.
- Used Google Translate which was not very efficient. (~1000 comments)
- Manually copied comments from various Hindi blogs (~250 comments)

 Implementation

Google Translate is not that awesome!!!!

Word Translations
- Severity of words reduced
  - “Dickhead” → “रोगी”
  - Meaning lost
  - “Slut” → “फू हड़ी”
  - Some words not translated
  - “Shitbag” → “शिटबैग”
  - Word to Word translation

Sentence Translation
- Long sentences contain error
  - “Define the war on women please. You mean requiring me to pay for your birth control? That’s about it. You are dead on with the talking points.”
- Short sentences are better
  - “You’re an idiot.” → “गुस मुसूद हो”

Idiom Translation
- Some idioms are fine
  - “Every Cloud Has A Silver Lining” → “नूतनता बने कारण में सुधार की भी आई होती है”
  - “A Piece Of Cake” → “कोई आसान सा काम”
- Some are translated word to word and meaning is lost
  - “Icing On The Cake” → “नीचे गले राख”
  - “A Wolf In Sheep’s Clothing” → “एक मिसाल में शेष के सर”

Normalisation
- Removing random characters like ‘@’, ‘’ etc.
- Removing punctuations, numbers etc.
- Removing words which come only once.
- Removing the words which come very frequently.

Feature Extraction
- Sentences converted into words vector.
- Bag of Words model
- N-gram, Skip Grams
- Tfidf score
- Special cases of Second-Person Narrative
- Taking the special case of negation.
  - For Eg: “गुना जिनके की अपनी ही भूमिका बन सकता”

Feature Selection
- Since the number of features are large, Chi-Square test is used to find the co-relation of features with the label.
- Features with maximum value of Chi-Square statistics are selected.

Classification
- Logistic Regression and SVM models are combined to evaluate the insult.

Results
- No past results as such to compare with.
- Accuracy without applying second person feature and negation feature: 85.96
- Accuracy with negation feature: 87.96
- Accuracy with second-person feature: 86.96
- Accuracy with both second-person and negation feature: 87.71

Suggestions for Future
- To have an even larger dataset.
- To have a bigger corpus to improve accuracy.
- To identify sarcasm.
- To identify insults that have been tampered. For Eg: “केक कुंज”

References
- Sentiment Analysis For Hindi Language, MS Thesis IIIT 2013.
- Piyush Arora.Vestibulum turpis quam, tristique vel dapibus et, scelerisque luctus erim.
- For starter code:
- For dataset:
- Extra: