REVA UNIVERSITY

# Identifying Kannada and English Code Switch Text

Ramesh Chundi

Vishwanath R. Hulipalled

J.B Simha

# Agenda

- Introduction
- Reasons for Code-Switch
- Applications of Code-Switch Text
- Literature review
- Data Pipeline
- Modelling
- Findings
- Conclusion

# Introduction

- **NLP =** Linguistics + ML + AI

- **Plain Text or Monolingual Text**

[EX1] ಗಿರಿಶ ಕಾರ್ನಾಡ್ ಅವರು ಆತ್ಮಕ್ಕೆ ಶಾಂತಿ ಸಿಗಲಿ ಎಂದು ಹಾರೈಸುತ್ತೇನೆ

***English Translation:*** *Let Girish Karnad's soul rest in peace.*

[EX2] I am proud of you, I'm fan of you from this minute.

- **Code-Switching Text or Bilingual Text**

[EX3] You are great sir zÉÃªÀgÀ£ÀÄß £Á£ÀÄ £ÉÆÃr®è but ¤ÃªÀÅ ªÀiÁqÀÄªÀ PÉ® ¸ÀzÀ°è žÉÃªÀgÀ£ÀÄß £ÉÆÃrzÉ lots of love to you sir.

***English Translation:*** you are great sir I didn't see God but I saw God in your work lots of love to you sir.

[EX4] Deshakkagi tyaga madiddare jeevana

***English Translation:*** Sacrifice the life for country.

# Reasons for Code-Switch

There are several reasons to switch codes in a single conversation:

- **Quoting someone:** People have to switch codes while quoting another person.
- **Abuse/Negative Sentiment:** Language is switched to either abuse or express a negative sentiment.
- **Sarcasm:** A simple opinion about a particular topic is expressed in a language and a switch to another to express a sarcastic opinion about the same.
- **Reinforcement:** Code-Switch is used for reinforcing a sentiment/opinion by a related one.
- **Translation:** A fact or opinion expressed is translated to the other language.

# Applications of Code-Switch Text

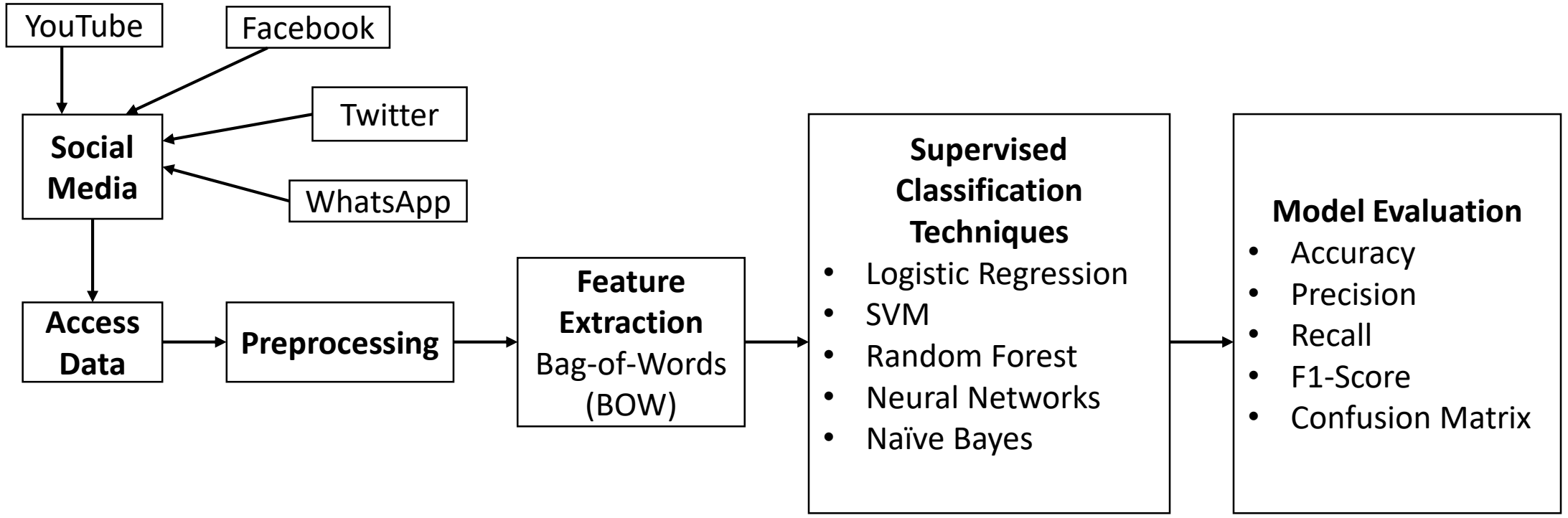
REVA UNIVERSITY

- To improve call centres efficiency.
- Better understanding about government policies implementation.
- More effective Q & A systems.
- Better understanding of product reviews.
- Effective sentiment analysis about public events, etc.

# Literature review

- Research on Code-Switch has been following from 1970s.
- Communication Accommodation Theory, Markedness model.
- Conversational Analysis Model, Code-Switch Archives.
- predicting code-switched points in Spanish – English.
- Identifying Code-Switched Tokens.
- Adding code-switched support to language models.
- Developing POS tagging for code-switching text.
- Performed analysis in English-Hindi posts from facebook and found that at least 4.2% of the data is code-switched.
- Language identification was done on code-mixed English-Kannada social media text.

# Data Pipeline



# Modelling

- Our data set has 2266 text comments.
- In that 1263 code-switch (CS) text comments and 1003 plain text (PT) comments.
- 10-fold cross validation is used for splitting the data set into 70-30 (support 680).

| CLASSIFICATION TECHNIQUE | CS / PT | PRECISION | RECALL | SUPPORT |
|---|---|---|---|---|
| LOGISTIC REGRESSION | CS | 0.82 | 0.90 | 381 |
| | PT | 0.85 | 0.74 | 299 |
| S V M | CS | 0.82 | 0.74 | 381 |
| | PT | 0.71 | 0.79 | 299 |
| RANDOM FOREST | CS | 0.88 | 0.79 | 427 |
| | PT | 0.70 | 0.82 | 253 |
| NEURAL NETWORK | CS | 0.80 | 0.88 | 381 |
| | PT | 0.82 | 0.72 | 299 |
| NAÏVE BAYES | CS | 0.83 | 0.88 | 381 |
| | PT | 0.85 | 0.74 | 299 |

# Findings

- F1-Score – This is a harmonic mean of precision and recall metric. Closer the numerator and denominator, better will be the model.

| CLASSIFICATION TECHNIQUE | CS / PT | F1-SCORE |
|---|---|---|
| LOGISTIC REGRESSION | CS | 0.86 |
| | PT | 0.79 |
| NAÏVE BAYES | CS | 0.85 |
| | PT | 0.79 |
| S V M | CS | 0.78 |
| | PT | 0.75 |
| RANDOM FOREST | CS | 0.83 |
| | PT | 0.75 |
| NEURAL NETWORK | CS | 0.84 |
| | PT | 0.77 |

CS: Code-Switch

PT: Plain Text

# **Findings**(continued..)

REVA UNIVERSITY

- CLASSIFICATION ACCURACY

| CLASSIFICATION TECHNIQUE | CLASSIFICATION ACCURACY |
|---|---|
| LOGISTIC REGRESSION | 83 |
| NAÏVE BAYES | 83 |
| NEURAL NETWORK | 81 |
| RANDOM FOREST | 80 |
| SVM | 76 |

# Conclusion

- Identified code-switch text and plain text.
- The Code-Switch text have emotion words and these emotion words are important for sentiment analysis.
- This work can be extended for detecting emotions in English-Kannada code-switch text.
- It can be extended further to sentiment analysis in English-Kannada code-switch text.

REVA UNIVERSITY

# Thank You