MB899 ADVANCED TEXT MINING

Unit 1

Basics of Text Mining - Different types of text data - Plotting commercial value of types of text data - insights from text data -Unigrams, Bigrams and Trigrams - converting structured to unstructured data -Removing stop words - stemming - spell correction

Unit 2

Text classification -Supervised - Unsupervised - Semi Supervised Algorithms - regex - normalizing text - Measuring accuracy of classification - Precision, recall and F score- Scikit Learn in Python- Case study with Semi Supervised Algorithm

Unit 3

Sentiment Analysis - Lexicon based - POS tagging based approaches - Supervised Approaches- Wordnet - Sentinets - Semantic Orientation based approaches - Combining supervised and unsupervised approaches for Sentiment Analysis - Case Study with Sentiment Analysis - Detecting Sarcasms

Unit 4

Unsupervised Approaches - K-means Clustering - Discovering topics using LDA- LDA in Python- Affinity Propagation - Word2Vec - Doc2vec - webcrawling using urlib2 and beautiful soup

Unit 5

Natural Language basics - Similarity approaches - Semantic Similarities -NLP with python NLTK - POS Tagging - Chunking - Parsing Trees - NER- Standford Parsers - Similarity using synsets in Wordnet - NLP using SpaCY - Text Summarization

References:

Text Books

Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications- Gary Miner, John Elder, Andrew Fast, Thomas Hill, Robert Nisbet, Dursun Delen

Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data

Online Articles:

- http://rinuboney.github.io/2016/01/19/ladder-network.html
  https://en.wikipedia.org/wiki/Semi-supervised_learning
- http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/
- https://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners
- http://www.dummies.com/programming/natural-language-processing-machine-learning/
- https://monkeylearn.com/blog/definitive-guide-natural-language-processing/
- http://stp.lingfil.uu.se/~bea/publ/megyesi-BrillsPoSTagger.pdf
- http://www.inf.ed.ac.uk/teaching/courses/icl/nltk/tagging.pdf
- https://blog.algorithmia.com/introduction-natural-language-processing-nlp/

N. Murali

Department of Management Studies
National Institute of Technology
Tiruchirappalli - 620 015

Senate