

RSM 317: Text Mining and Natural Language Processing

Introduction to a diverse collection **text mining** techniques and **natural language processing** using **machine learning**.

Identification and **quantification** of various structures in the text data to **answer business problems** and **provide managerial insights**.

Model validation and effective communication of model-based results will be stressed.

The course will employ a “**white-box**” methodology, which emphasizes an **understanding** of the **algorithmic and statistical model** structures and how they apply to text analysis.

Following leading industry standards, this **hands-on** course will use **Python** to apply a number of different algorithms to real-world big text data.

Topics:

Natural Language and Text Processing

Regular Expressions

Spellchecking

Tokenization and Normalization

Bag of Words, Document-Term Matrix, TF-IDF

Classification Algorithms with Neural Networks

Sentiment Analysis

Similarity Analysis, Clustering

Topic Modelling (LSI, LDA)

Embeddings: CBOW and Skip grams