**Outline**

**Introduction + Business Values + Exec Summary:*(March 27(Documentation))***

Business Case

Hate Speech Data on social media and company website

Business Value and impact

**Data Introduction + data cleansing + Preprocessing(Optional) + data augmentation: Atsu, Hafriza**

***(March 14(Code), March 20(Documentation))***

Data Description

Preprocessing

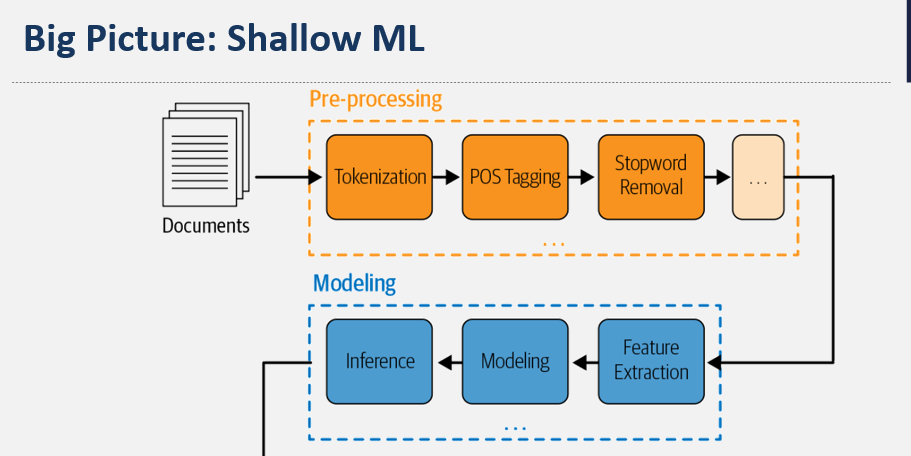
Data Summary

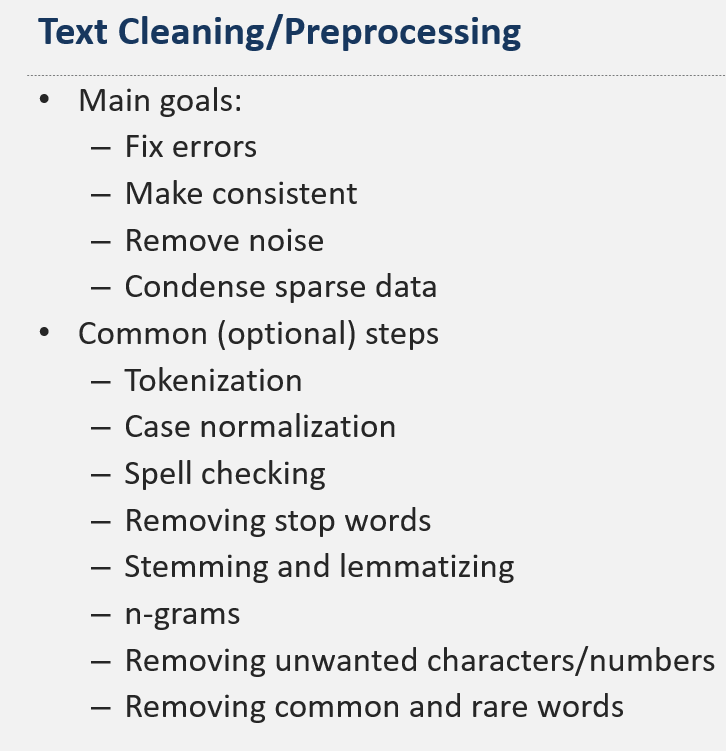
Text Preprocessing technique

* Tokenization
* Case Normalization
* Spell Checking + Spelling Normalization
* Stop word removal + remove non-ascii character
* Pruning Rare and Common Words
* Removing Patterns
* Stemming and Lemmatization
* Etc.

**Handling Class Imbalance**

*Visualization: Word cloud of the dataset, distribution of text length, etc.*





<https://github.com/stepthom/NLP_course/blob/main/preprocessing/slides_preprocessing.ipynb>

**Albert Model Design: Hafriza, Jason**

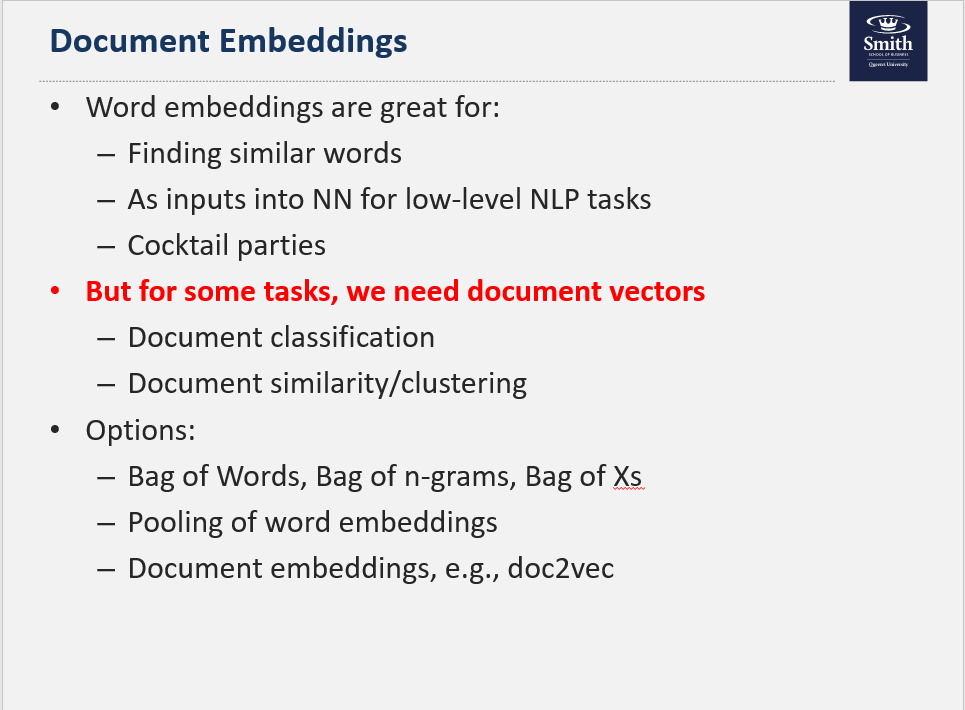
***(March 19(Code), March 24(Documentation))***

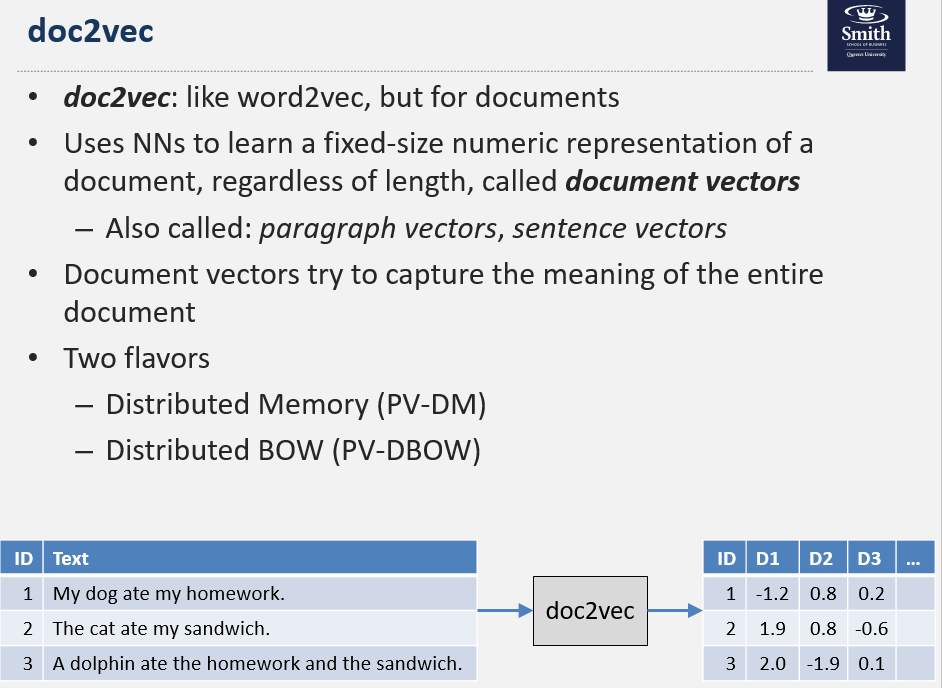
General Design

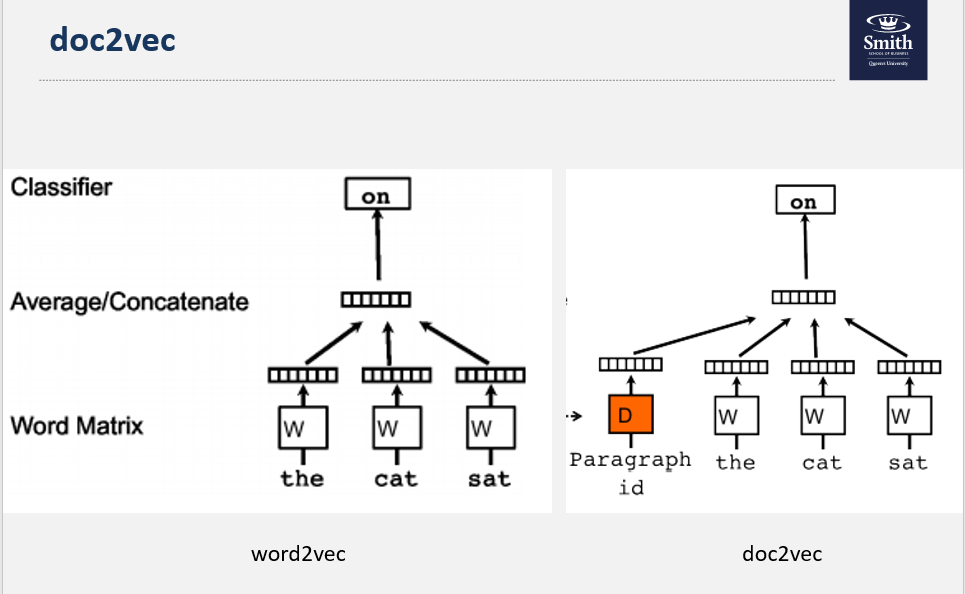
Framework Selection (Python NLTK, Keras, Pytorch, etc.)

**Word Embedding introduction: (Word2vec, Glove, etc.)**

**Document Embedding**







**Vectorization**

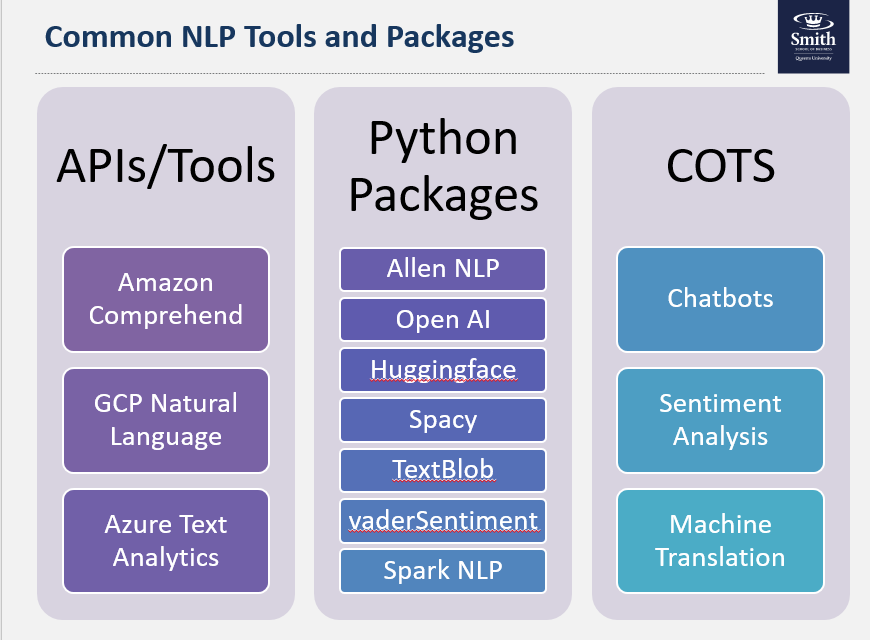
**Model (LSTM, CNN, RNN, DNN, GRU) or ensemble with Rule-based**

Modeling Architecture

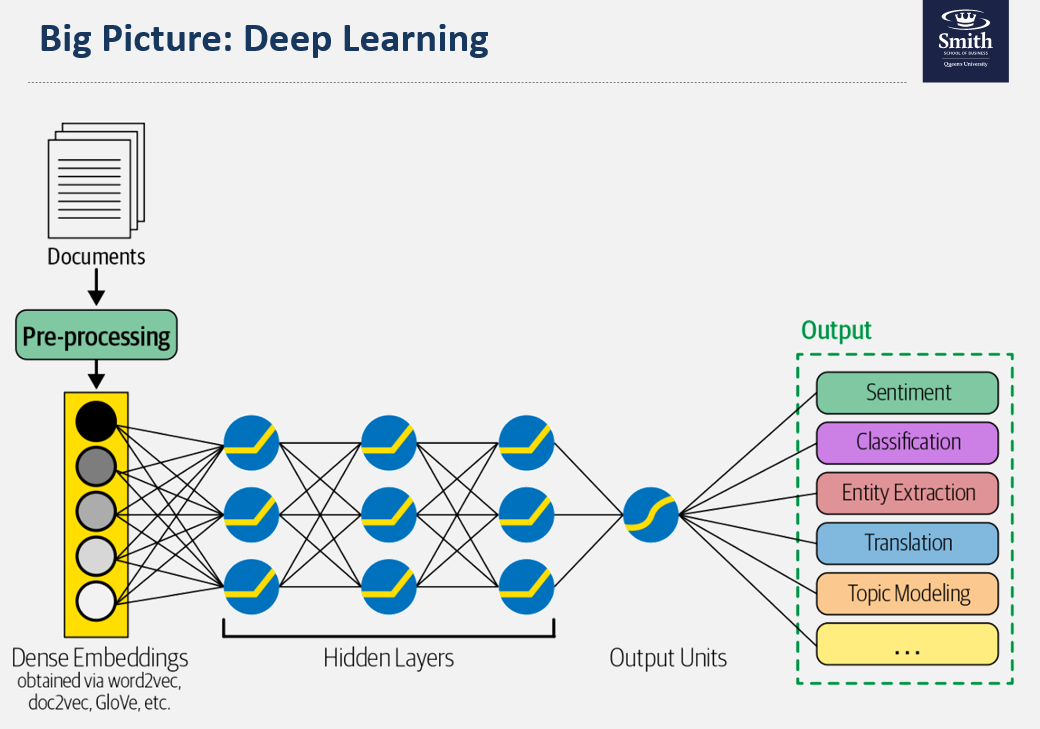
Performance Metric Selection (F1, Accuracy, recall, precision?)

Hyper-parameter selection and other parameters and explanation

*Visualization: Albert Model Architecture*



* *Python packages*
  + *NLTK*
  + *Spacy*
  + *Gensim*
  + *Textblob*



**Transfer Learning Jason, Ajay**

***(March 19(Code), March 24(Documentation))***

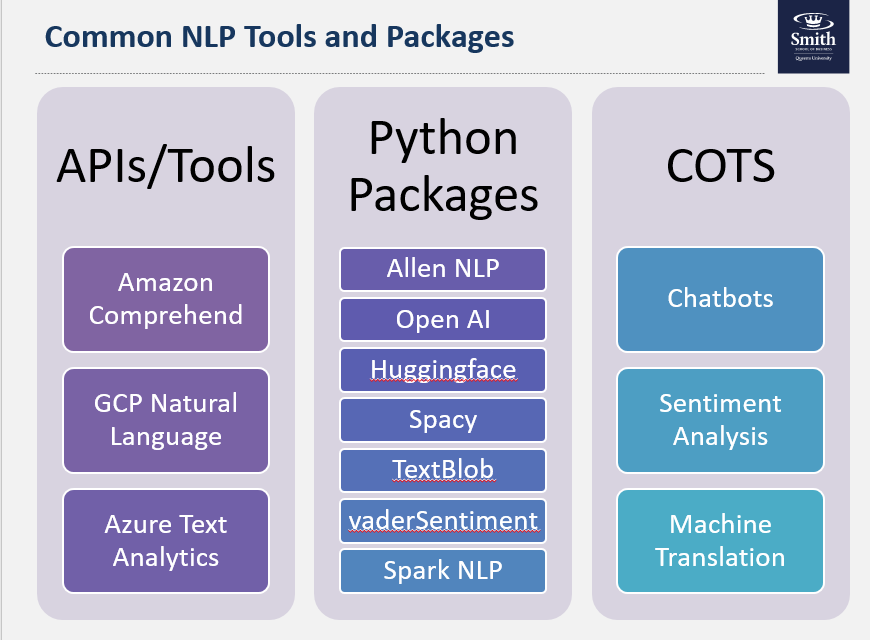
Pre-trained neural networks

**? word embeddings (Word2vec, Glove, etc.)**

**Pretrained: BERT, ElMo, GPT-2, GPT-3**

What is their architecture?

*Visualization: Transfer learning model Architecture*



**Experimentation and Model Comparison: Ajay**

***(March 19(Code), March 24(Documentation))***

Hyper-parameter and Miscellaneous Tuning

Addressing Biases and Overfitting

Albert model Training and Validation

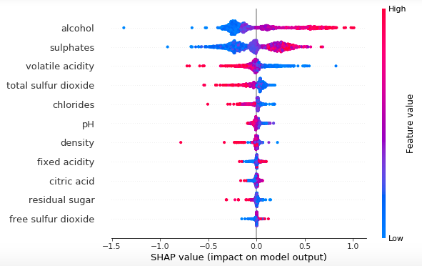
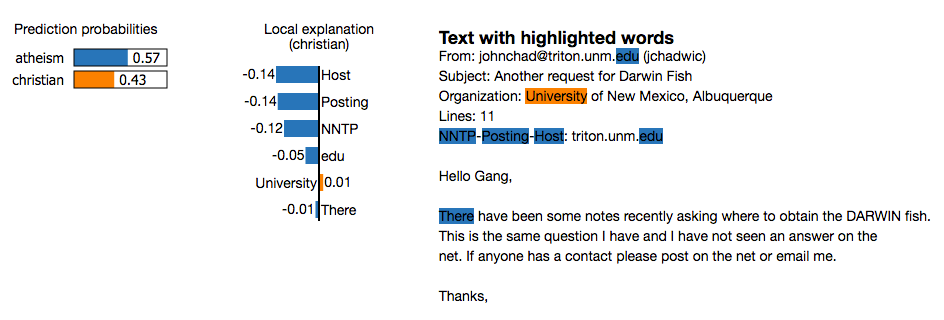
Transfer Learning Training and Validation

Performance Comparison

Model Selection for Deployment

*Visualization: Training epochs with accuracy, summary of model performance*

*Ideas: Lime? SHAP Values?*



**Implementation with Github: Ben**

***(March 27(Documentation))***

GitHub Repository [Bonus]

High Level Description of Code

Required Packages

Implementation Instructions [ Github Readme]

*Visualization: GitHub Readme Snapshot*

**Recommendation and Model Production**

***(March 27(Documentation))***

Model’s Effectiveness for Business Use Case basing on metric

Business Recommendations

Extension to Production

Challenges before Operationalization, and Solutions

Approaches to Operationalization and Scaling

**Future consideration and challenges**

***(March 27(Documentation))***

Future Considerations

Dataset Challenges

Processing Power

Other limitation

Privacy issue?

*Challenges:*

*Managing environments: Backwards compatibility, Libraries, modules, dependencies, Load balancing*

*Serving Models: A/B testing, Storing, updating, Versioning, lineage*

*Monitoring: Input drift, Output/performance drift*

*Scheduling: Retraining, Batch jobs*

**References and Appendix *(March 31(Documentation))***

Links

Appendix