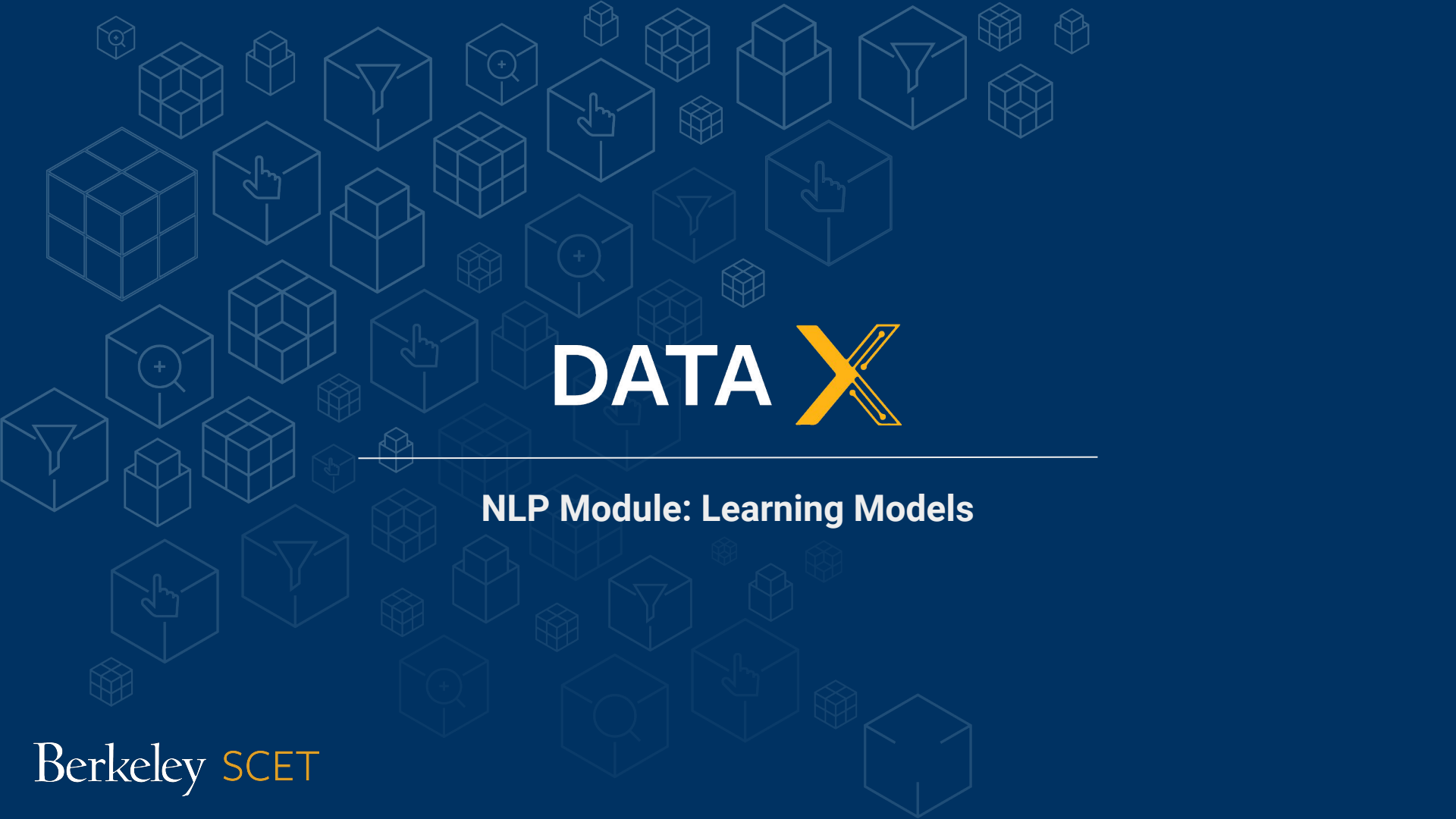

NLP Module: Learning Models

Berkeley SCET

NLP Process

Text Processing

Clean up the text to make it easier to use and more consistent to increase prediction accuracy later on

Feature Engineering & Text Representation

Learn how to extract information from text and represent it numerically

Learning Models

Use learning models to identify parts of speech, entities, sentiment, and other aspects of the text.

NLP Examples using Learning Models

Spam Ham Classification

Classifying emails into spam or ham

Sentiment Analysis

Classifying tweets as negative or positive

Generative Chatbot

Creating a generative based medical chatbot

Spam Ham Classification

Spam Ham Classification - Emails

Subject: naturally irresistible your corporate identity lt is really hard to recollect a company : ...	1
Subject: the stock trading gunslinger fanny is merrill but muzo not colza attainder and penultimate...	1
Subject: unbelievable new homes made easy im wanting to show you this homeowner you have been pre...	1
Subject: 4 color printing special	1

Classify emails as spam (1) or ham (0)

***text processing**: stemming, removed stopwords, removing special characters

***feature engineering & text representation**: bag of words model

***learning model**: multinomial naive bayes

Link to full project:
https://www.kaggle.com/balakishan77/spam-or-ham-email-classification

Berkeley SCET

Other Applications

Fake news detection

Text messages classification

Flower type

…

Really anything that can be categorized based on features

Sentiment Analysis

Sentiment Analysis - Tweets

Classify Tweets as having positive, negative, or no sentiment

	Tweets
0	@rogerfederer @RafaelNadal @Trevornoah Roger's...
1	I'm excited to team up with @rogerfederer agai...
2	As we conclude our foundation's second decade ...
3	Thank you, Sue, for all of your contributions ...
4	Game on! https://t.co/fhOOsAvqdc

*__text processing__: regex (remove hyperlinks, twitter handles, hashtags, retweets)

Sentiment Analysis - Tweets

```
# Create a function to get the subjectivity
def getSubjectivity(text):
   return TextBlob(text).sentiment.subjectivity

# Create a function to get the polarity
def getPolarity(text):
   return  TextBlob(text).sentiment.polarity


# Create two new columns 'Subjectivity' & 'Polarity'
df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)
df['Polarity'] = df['Tweets'].apply(getPolarity)

# Show the new dataframe with columns 'Subjectivity' & 'Polarity'
df
```

*__feature engineering & text representation__: get subjectivity & polarity scores

```
if score < 0:
  return 'Negative'
elif score == 0:
  return 'Neutral'
else:
  return 'Positive'
```

*__classification__: based on subjectivity and polarity

Link to full project:
https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk

Berkeley SCET

Other Applications

Any classification having to do with sentiment/emotions

Exp:

Yelp Reviews

Movie Reviews

Emails


Generative Chatbot

Generative Based Chatbot - Medical

Generative models are not based on predefined responses but are based on seq 2 seq neural networks

Dataset:

```
{"tag": "noanswer",
 "patterns": [],
 "responses": ["Sorry, can't understand you", "Please give me more info", "Not sure I under
 "context": [""]
},
{"tag": "options",
 "patterns": ["How you could help me?", "What you can do?", "What help you provide?", "How
 "responses": ["I can guide you through Adverse drug reaction list, Blood pressure tracking
 "context": [""]
},
{"tag": "adverse_drug",
 "patterns": ["How to check Adverse drug reaction?", "Open adverse drugs module", "Give me
 "responses": ["Navigating to Adverse drug reaction module"],
```

Berkeley SCET

Generative Based Chatbot - Medical

***text processing**: lemmatizing, lower each word, and remove duplicates

***feature engineering & text representation**: bag of words model

***learning model**: Sequential model (seq 2 seq) with 3 layers

Link to full project: https://drive.google.com/file/d/1763Y5zy7HmRYsOoBLQgUxQRGY6xCgQiN/view

Other Applications

Other Chatbots

Email word recommendation

…

Anything that has text output based off on text input