



UNIVERSITY
of VIRGINIA

EdPolicyWorks

Sentiment Analysis and Topic Modeling Validation Training

Brian Heseung Kim

University of Virginia

August 26, 2021

Session Outline

- Setting the Stage
 - What is Natural Language Processing?
 - How is it being used in this study?
 - What is validation, and why is it necessary?
- Definitions, Instructions, and Examples
 - Sentiment
 - Topic Modeling
- Phrases and Terms
- Next Steps and Timeline
- Questions and Wrap-up

Setting the Stage

What is Natural Language Processing?

The process by which naturally-produced
language is processed into a form more
interpretable by computers

What is Natural Language Processing?

Text Documents

District with Secretary Johnson from the Department of Homeland Security. All these delegations have spent hours keeping in close contact with one another and with officials in Arkansas. in particular. regarding the tornado that ripped through Vilonia. Mayflower. El Paso. and Paron.

Taliban in their own country the way they have always fought and the way the Taliban fights. and they defeated them. Then we did an unfortunate thing. We helped them with a constitution that centralized the government. In a very regional federalist area. tribal area. we should have helped

contraception. a pharmacists denial can be an unsurmountable obstacle to access within the limited timeframe. Under the ABC Act. if a requested product is not in stock. but the pharmacy stocks other forms of contraception. the pharmacy must help the woman obtain the

What is Natural Language Processing?

Text Documents

District with Secretary Johnson from the Department of Homeland Security. All these delegations have spent hours keeping in close contact with one another and with officials in Arkansas. in particular. regarding the tornado that ripped through Vilonia. Mayflower. El Paso. and Paron.

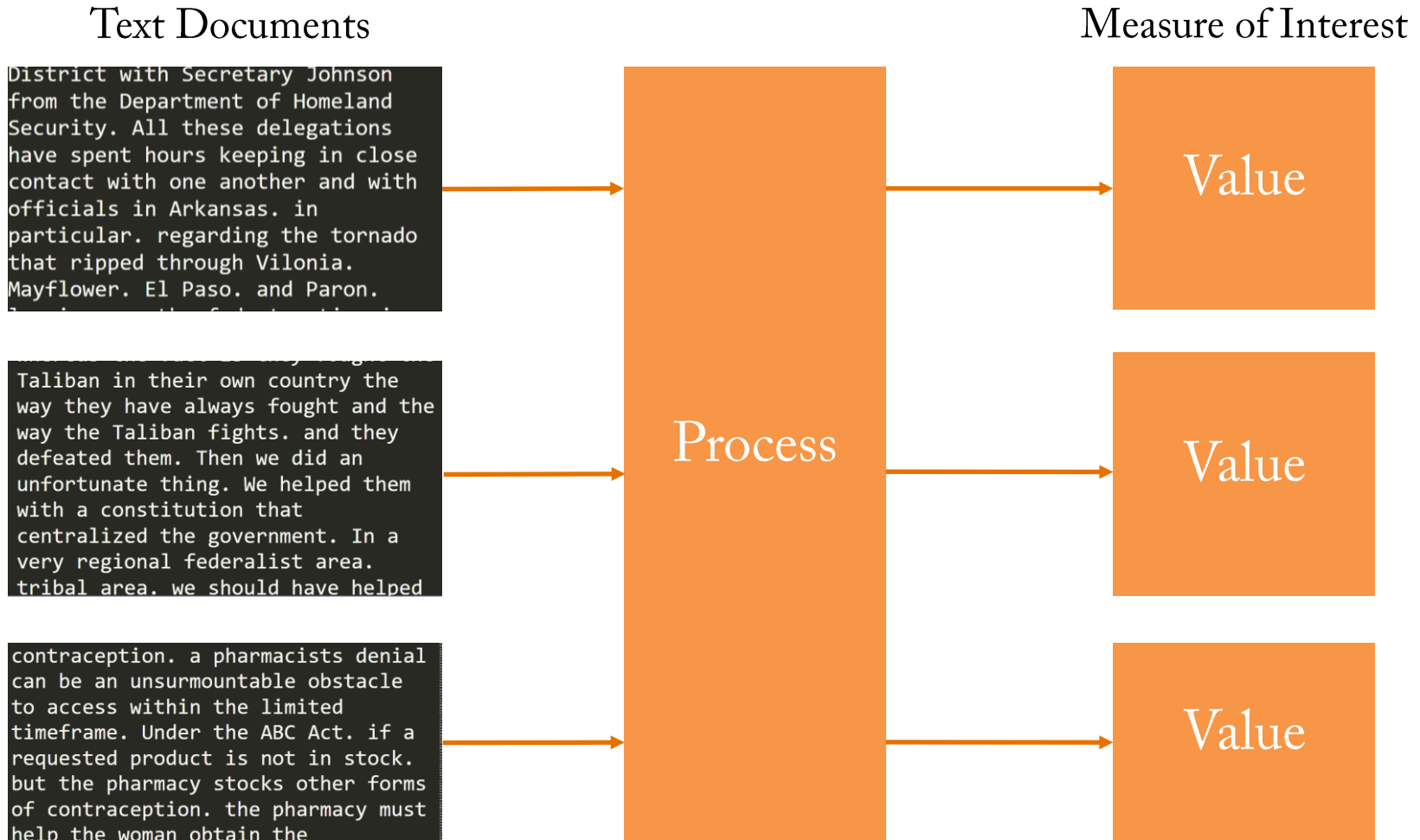
Taliban in their own country the way they have always fought and the way the Taliban fights. and they defeated them. Then we did an unfortunate thing. We helped them with a constitution that centralized the government. In a very regional federalist area. tribal area. we should have helped

contraception. a pharmacists denial can be an unsurmountable obstacle to access within the limited timeframe. Under the ABC Act. if a requested product is not in stock. but the pharmacy stocks other forms of contraception. the pharmacy must help the woman obtain the

```
graph LR; A[Text Document 1] --> D[Process]; B[Text Document 2] --> D; C[Text Document 3] --> D;
```

Process

What is Natural Language Processing?



What is Natural Language Processing?

Senate Speeches

District with Secretary Johnson from the Department of Homeland Security. All these delegations have spent hours keeping in close contact with one another and with officials in Arkansas. in particular. regarding the tornado that ripped through Vilonia. Mayflower. El Paso. and Paron.

Taliban in their own country the way they have always fought and the way the Taliban fights. and they defeated them. Then we did an unfortunate thing. We helped them with a constitution that centralized the government. In a very regional federalist area. tribal area. we should have helped

contraception. a pharmacists denial can be an unsurmountable obstacle to access within the limited timeframe. Under the ABC Act. if a requested product is not in stock. but the pharmacy stocks other forms of contraception. the pharmacy must help the woman obtain the

Word
Counter

Measure of Interest

634

201

517

What is Natural Language Processing?

Senate Speeches

District with Secretary Johnson from the Department of Homeland Security. All these delegations have spent hours keeping in close contact with one another and with officials in Arkansas. in particular. regarding the tornado that ripped through Vilonia. Mayflower. El Paso. and Paron.

Taliban in their own country the way they have always fought and the way the Taliban fights. and they defeated them. Then we did an unfortunate thing. We helped them with a constitution that centralized the government. In a very regional federalist area. tribal area. we should have helped

contraception. a pharmacists denial can be an unsurmountable obstacle to access within the limited timeframe. Under the ABC Act. if a requested product is not in stock. but the pharmacy stocks other forms of contraception. the pharmacy must help the woman obtain the

Fancy
Positivity
Checker

(Sentiment
Analysis)

Measure of Interest

Very
Negative

Very
Positive

Slightly
Negative

How is it Being Used in This Study?

- Researchers ran a texting-based college advising program to support students approaching graduation
- Advisors regularly reached out to students with pre-set updates and reminders via text message, and students were able to directly ask for support, more information, etc.
- We want to know more about the ways advisors and students interacted throughout the intervention
 - Scheduled messages from advisors
 - Ad-hoc messages from students
 - Ad-hoc responses from advisors
- We fed the texts of students and their advisors through a natural language processing tool to measure:
 - How positive/negative/neutral was each text message?
 - What did students and their advisors talk about?

What is “Validation?” Why is it Necessary?

Question: How do we know the tool actually did what we thought it did?

What is “Validation?” Why is it Necessary?

Question: How do we know the tool actually did what we thought it did?

Answer: Compare it to what humans would think!

Definitions and Instructions

Sentiment Analysis

- To measure the positivity, negative, or neutrality of a given text message, we use a technique called “Sentiment Analysis”
- In this framework, the tool “reads” the text and tries to determine how positive, negative, or neutral it is based on grammar, words used, context, etc.
- It ultimately tries to measure the *perceived* positivity of emotions *and* ideas in the text. This includes:

Sentiment Analysis

- To measure the positivity, negative, or neutrality of a given text message, we use a technique called “Sentiment Analysis”
- In this framework, the tool “reads” the text and tries to determine how positive, negative, or neutral it is based on grammar, words used, context, etc.
- It ultimately tries to measure the *perceived* positivity of emotions *and* ideas in the text. This includes:

Concept	Negative Example	Neutral Example	Positive Example
Stated Emotions	“I feel awful.”	“I am here.”	“I am fantastic!”
Intentions	“I hope you die.”	“I’ll see you there.”	“I wish you well.”
Topical Content	“I am struggling with finances right now.”	“There are several banks in town.”	“I can’t wait to celebrate with you!”

Sentiment Analysis

- By the end of the process, every individual text message is assigned a score:
 - Very Negative (-2)
 - Negative (-1)
 - Neutral/Factual (0)
 - Positive (1)
 - Very Positive (2)

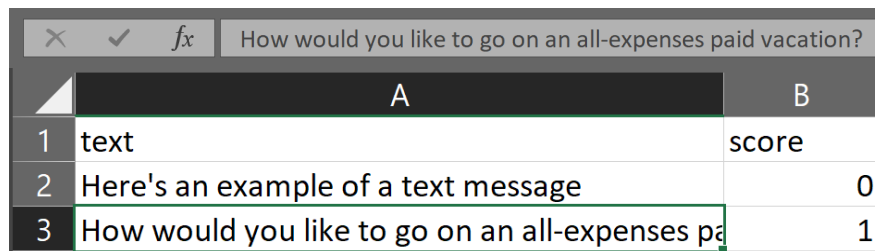
Text Message (sic)	Sentiment Score
Terrible I have to find a class or two to sign up for. I'm so behind.	Very Negative (-2)
Bad news. according to UTSA student unable to achieve a "C" or better shall be required to change his major to a filed outside of College of Engineering. Sorry!	Very Negative (-2)
I was trying to drop a class and it doesnt allow me	Negative (-1)
Hi, quick correction from my earlier txt: Spring 2019 course registrations starts *Friday*, 11/9. Also, our records show that you have an advising hold on your account. Holds must be resolved to register for the classes you need to graduate. Can I help?	Negative (-1)
I am in school Tuesdays and Thursdays	Neutral (0)
You won't need to declare a minor because those are 18 credits each. You only need two courses, 6 credits.	Neutral (0)
Hey. Congrats on finishing this term - you're that much closer to your goal! I won't send any more texts going forward, but you can still reach me at [EMAIL] if you need help :-)	Positive (1)
Each advisor has various availabilities, however, all advisors are eager to help you reach your academic goals, as well as graduation.	Positive (1)
Good Morning! Keep up the good work! December is closer than you think! Have a Great Day.	Very Positive (2)
It was very helpful thank you so much!	Very Positive (2)

Sentiment Analysis: Your Task

- For each of the texts in the spreadsheet, rate your perception of its *overall* sentiment from -2 to 2. Some tips:
 - Think of each text entirely on its own, unrelated to the others
 - Think of very negative/very positive on an absolute scale, not relative
 - Differences in opinion are already *built into the process* – choose what you think is best!
 - It may be helpful to imagine: if you received this text, how would it *generally* make you feel in response? What headspace would it put you in?
 - If you think there are multiple sentiments present (e.g., negative *and* positive, or negative *and* neutral), record the more dominant sentiment rather than some “average”

Sentiment Analysis: Your Task

- For each of the texts in the spreadsheet, rate your perception of its *overall* sentiment from -2 to 2. Some tips:
 - Think of each text entirely on its own, unrelated to the others
 - Think of very negative/very positive on an absolute scale, not relative
 - Differences in opinion are already *built into the process* – choose what you think is best!
 - It may be helpful to imagine: if you received this text, how would it *generally* make you feel in response? What headspace would it put you in?
 - If you think there are multiple sentiments present (e.g., negative *and* positive, or negative *and* neutral), record the more dominant sentiment rather than some “average”



	A	B
1	text	score
2	Here's an example of a text message	0
3	How would you like to go on an all-expenses paid vacation?	1

Sentiment Analysis Practice Examples

Sentiment Analysis Examples

1. Hi. FAFSA is open for the 2019-2020 school year. If you plan to take classes next year, getting your FAFSA in early can maximize your aid \$\$\$. Have you started?

Very Negative (-2)

Negative (-1)

Neutral/Factual (0)

Positive (1)

Very Positive (2)

Sentiment Analysis Examples

2. Hi! Congrats on finishing this term - you're making great progress toward graduation! Want to work together to make sure your Spring 2019 courses count toward your degree?

Very Negative (-2)

Negative (-1)

Neutral/Factual (0)

Positive (1)

Very Positive (2)

Sentiment Analysis Examples

3. Hi! I was thinking of transferring to Lehman because what I want to major in isn't actually offered at KCC. Would you know if I'd have to come to the campus to begin the process? I think it's too late to apply for fall.

Very Negative (-2)

Negative (-1)

Neutral/Factual (0)

Positive (1)

Very Positive (2)

Sentiment Analysis Examples

4. Great thanks.

Very Negative (-2)

Negative (-1)

Neutral/Factual (0)

Positive (1)

Very Positive (2)

Topic Modeling

- To measure the content of a given text message *conversation*, we use a technique called “Topic Modeling”
- In this framework, the tool “reads” a set of related texts and tries to determine what topics of discussion are present based on the words being used
- It ultimately tries to measure how *prevalent* each topic is in each given text conversation.

Topic Modeling

- To measure the content of a given text message *conversation*, we use a technique called “Topic Modeling”
- In this framework, the tool “reads” a set of related texts and tries to determine what topics of discussion are present based on the words being used
- It ultimately tries to measure how *prevalent* each topic is in each given text conversation.

Text Conversation	Financial Aid	Course Registration	Meeting Planning
Hey, I was wondering how to register for my classes this fall? I want to enroll in ECON515. Sure thing, go to tokenurl to see available courses. Are you available on Thursday for an appointment to talk more?	0%	70%	30%

Topic Modeling

- To measure the content of a given text message *conversation*, we use a technique called “Topic Modeling”
- In this framework, the tool “reads” a set of related texts and tries to determine what topics of discussion are present based on the words being used
- It ultimately tries to measure how *prevalent* each topic is in each given text conversation.

Text Conversation	Financial Aid	Course Registration	Meeting Planning
I need help! I want to register for my classes, but the website says I haven't paid my tuition fees and there's a hold. I thought I filled out my FAFSA last week to get my student loans. I don't have enough money to pay right now.	80%	20%	0%

Topic Modeling: Your Task

- For each of the text conversations in the spreadsheet, select the *most prevalent* and *second-most prevalent* topics present in the conversation from the provided list of topics. Some tips:
 - You will select your chosen topics from a drop-down menu
 - Think of each conversation entirely on its own, unrelated to the others
 - Differences in opinion are already *built into the process* – choose what you think is best!
 - Who says what is not important; think of the conversation as a whole “unit”
 - If the conversation doesn’t seem to be about anything in particular, choose “General Conversation/Other” for both Most and Second-Most prevalent topics. If you think there isn’t really a Second-Most prevalent, select General Conversation/Other for only that one.

Topic Modeling: Topic Definitions

- **Academic Planning**: Anything related to broad plans related to academics. Graduation, degree/major requirements, transfer/admissions requirements, applying to other programs/schools, etc. Big picture!
- **Academic Resources**: Anything related to asking for general academic help and/or info on academic support services and resources. Tutoring centers, questions about finals, how courses are going (including grades), counseling services, campus services, etc.
- **Advising Logistics**: Anything related to setting up advising meetings or communications with advisors. Email logistics, walk-in/office hours, planning an advising appointment, how to call the advising center, out-of-office responders, etc.
- **Course Selection/Enrollment**: Anything related to course registration and accompanying logistics. Course prerequisites, enrollment deadlines (especially spring/fall/summer/winter), withdrawing from courses, instructor permissions, etc.
- **Finances**: Anything related to loans, financial aid, applying/filing for aid, paying bills, account holds, monetary issues, etc.
- **General Conversation/Other**: Pretty much everything else that doesn't fall into the above. Pleasantries, how-are-you's, enjoy-your-break's, don't-hesitate-to-get-in-touch's, I'm-going-on-vacation..., etc. etc.

Topic Modeling Practice Examples

Topic Modeling Examples

1. Yes of course! Are you still pursuing Mathematics as your major? Yes i am Okay, so I advise for Health and Bioscience majors, and Math falls under STEM, but I can try to assist you, but we need to get you switched a STEM advisor. What university do you want to transfer to? Okay, yeah sure. I was thinking UTSA I emailed you at your aces email; I sent you the transfer guide for Math from here at SAC to UTSA, and listed the courses still needed.

Academic Planning

Academic Resources

Advising Logistics

Course Selection/Enrollment

Finances

General Conversation/Other

Topic Modeling Examples

2. Actually yes I was curious of possibly taking a course for the summer That's great! Let me know if you'd like to make an appointment to discuss this further. Yes I'd like to make an appointment to further discuss summer courses. Thank you very much. Great. So i'm on campus Monday - Wednesday. You can stop by my office in the Political Science Dept tomorrow between 1pm-5pm for walk-in advisement. or you can log onto AdvisorTrac and reserve yourself a 30min advisement appointment with me. Great. I can do a walk in tomorrow during community hour to further talk more about summer classes. Thank you very much. Sounds good! See you tomorrow afternoon! Thank you. Likewise

Academic Planning

Academic Resources

Advising Logistics

Course Selection/Enrollment

Finances

General Conversation/Other

Topic Modeling Examples

3. Hi, and no, not at the moment. Thank you for asking though. Thank you for taking the time to respond to this text. If you need assistance later on, please do not hesitate to reach out. Enjoy the rest of your semester! Np! And ok, will do and same to you!

Academic Planning

Academic Resources

Advising Logistics

Course Selection/Enrollment

Finances

General Conversation/Other

Topic Modeling Examples

4. Yes I will. Thank you. I need six classes to complete my degree. I will put them in my shopping cart, that way I will be ready when it's time to register. I need to take an ID class, I choose PSY 3405 ID. Do you know if I would have to go to the PSY coordinator in order for me to register for this ID class. Good I have an issue. My Welfare Policy Professor made the mistake and put a WN for my class grade. I always attended his class. I emailed him about the his mistake and he told me he rectified the grade with the Registra but they told me that the issue is not resolved. They said my Professor didn't hand in the paper work. My professor also said he did but Cuny is slow with paper work. I am currently in the class and has handed in all required work. Can you please help me with this issue before it is too late for me. Thank you.

Academic Planning

Academic Resources

Advising Logistics

Course Selection/Enrollment

Finances

General Conversation/Other

Phrases and Terms

Term	Meaning
token____	A placeholder word we used to replace specific mentions of names, phone numbers, institution names, etc.
FAFSA	The form students must fill out to apply for federal financial aid
SIS	Student Information System – things like Blackboard, Canvas, etc. to manage online student financial accounts
Bursar	Office of most institutions that handle student finances/payments
Apply to graduate	The process by which students need to submit their intention to graduate ahead of time (usually a semester or two in advance)
Drop-in hours	Unstructured time where students can seek support from their advising center or advisor
B304, 5306, etc. etc.	Likely room numbers or course numbers. Use your best judgment!
Apply to transfer	The process by which students apply to transfer to another institution, usually to a four-year institution (bachelor's) after completing their degree at a two-year institution (associate's)

Next Steps and Timeline

- BK to send coding datasets by August 30
 - Approximately 50 text messages for sentiment analysis
 - Approximately 50 conversations for topic modeling analysis
- Complete individual codes by September 10th
- Week of September 13th (if possible)
 - Regroup to discuss results of the validation process and figure out next steps
 - i.e., whether we need to do another round
- BK to send full coding datasets by September 20th
 - Approximately 150 more text messages for sentiment analysis
 - Approximately 150 more conversations for topic modeling analysis
- Complete individual codes by October 1st

Any Questions?

- Setting the Stage
 - What is Natural Language Processing?
 - How is it being used in this study?
 - What is validation, and why is it necessary?
- Definitions, Instructions, and Examples
 - Sentiment
 - Topic Modeling
- Phrases and Terms
- Next Steps and Timeline
- Questions and Wrap-up



UNIVERSITY
of VIRGINIA

EdPolicyWorks

Thank you!

Brian Heseung Kim



@brhkim



brian.kim
@virginia.edu



brhkim.com



@brhkim

For the most up-to-date information from EdPolicyWorks,
please contact us at:
EdPolicyWorks@virgina.edu