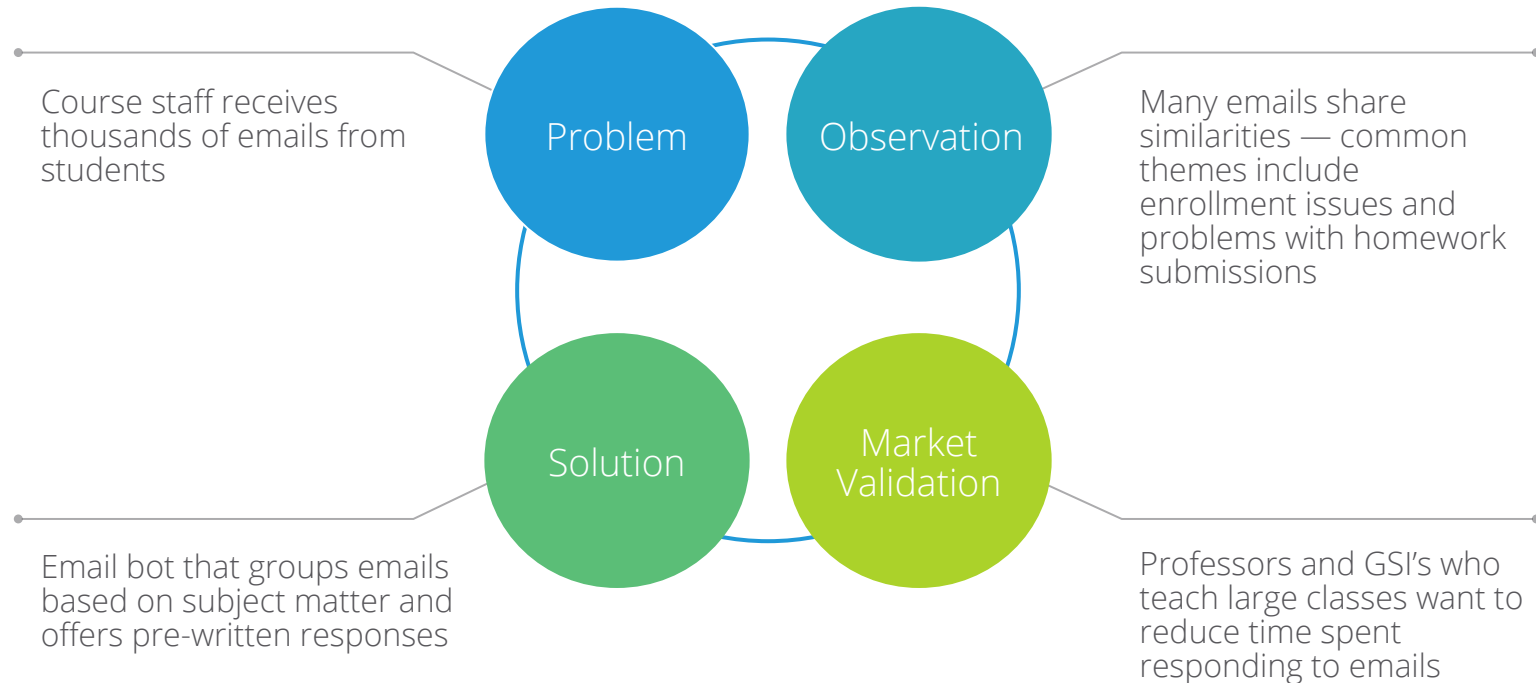




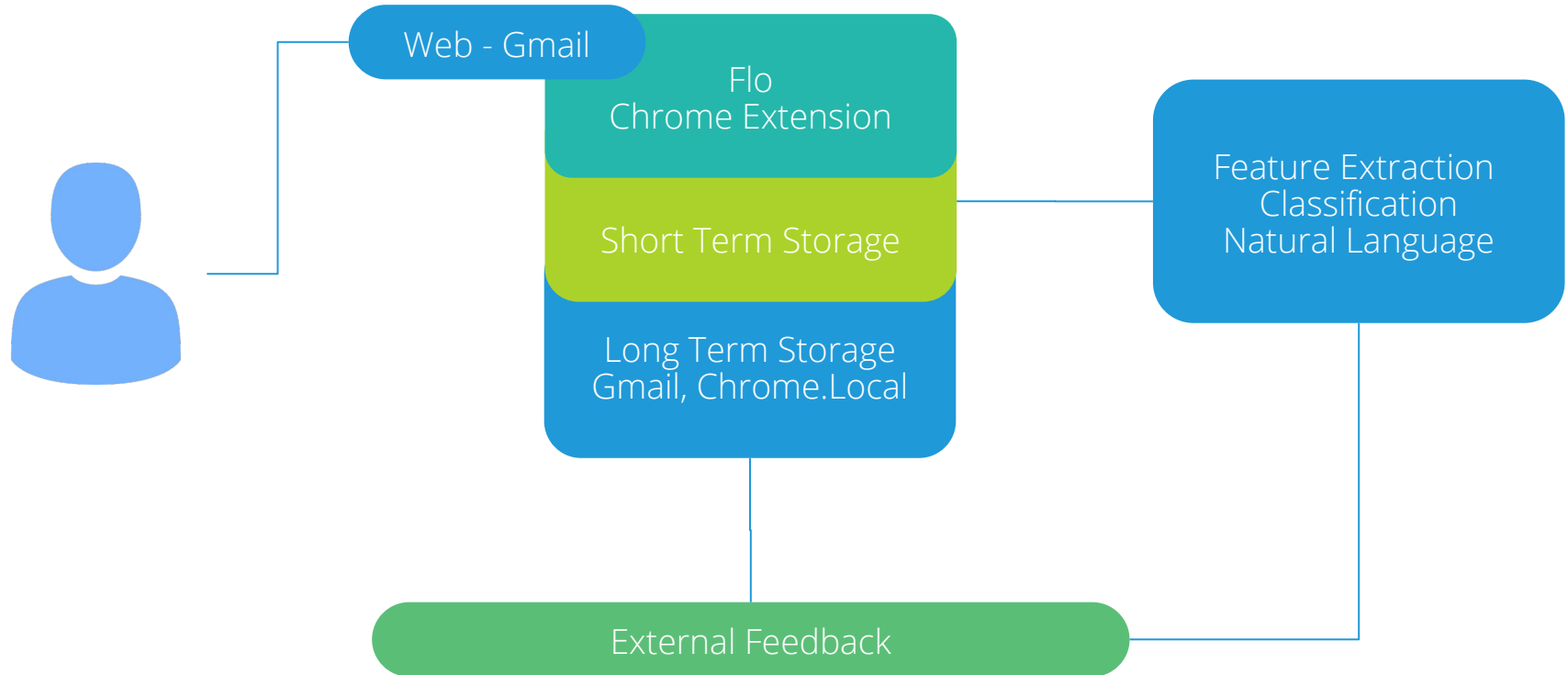
An Email Management System

Ndeye Fatou Diop, Keiko Kamei, Rohan Lageweg,
Ting Chih Lin, Joyce Siu Ying Lo, Kristian Rolland

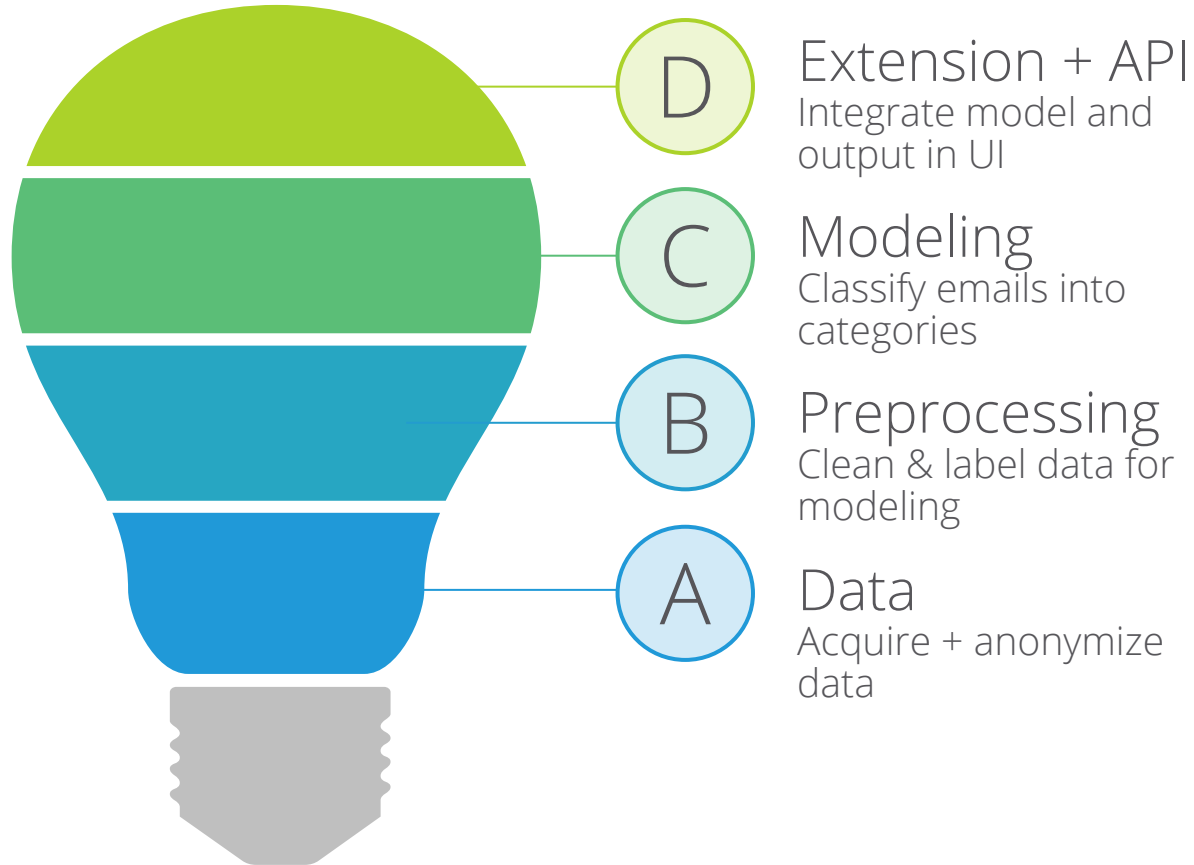
Flo: An Email Management System



Architecture of Solution

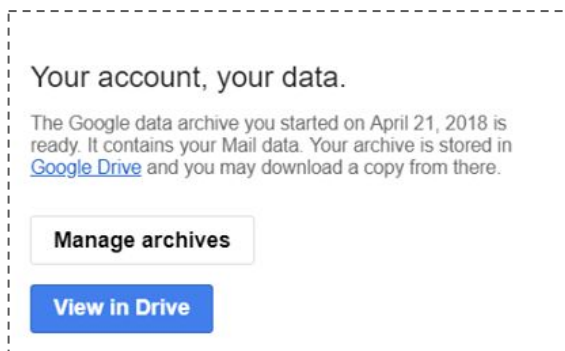


Project Outline



Data Acquisition

- With permission from course staff, downloaded MBOX file containing roughly 3 months worth of course emails

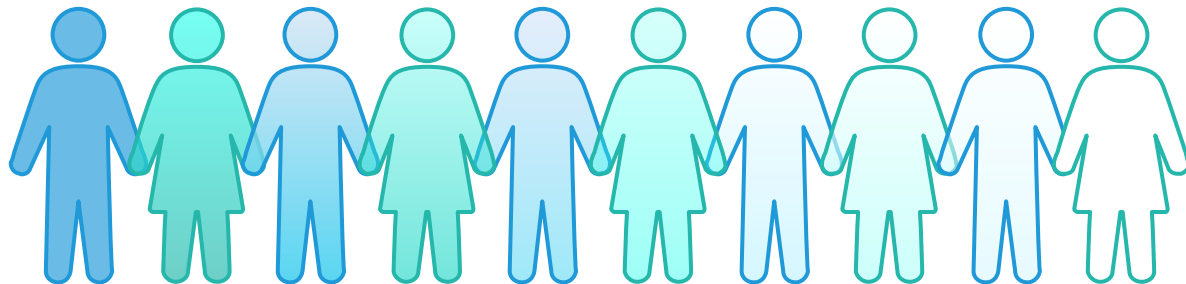
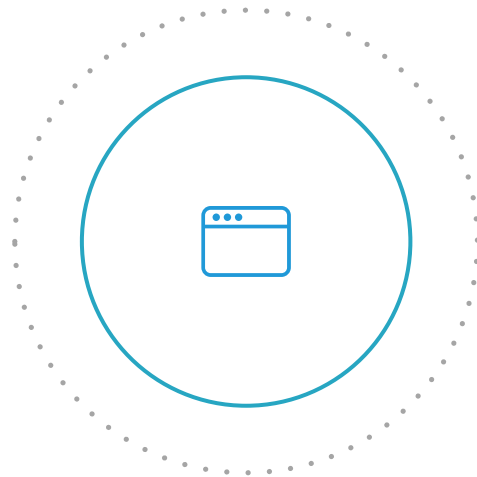


- Dataset:
 - Approximately 350 unique conversations
 - Equates to 1520 total course emails (including replies and forwards)
- Reached out to instructors & teaching staff of other courses but could not obtain data due to student privacy concerns



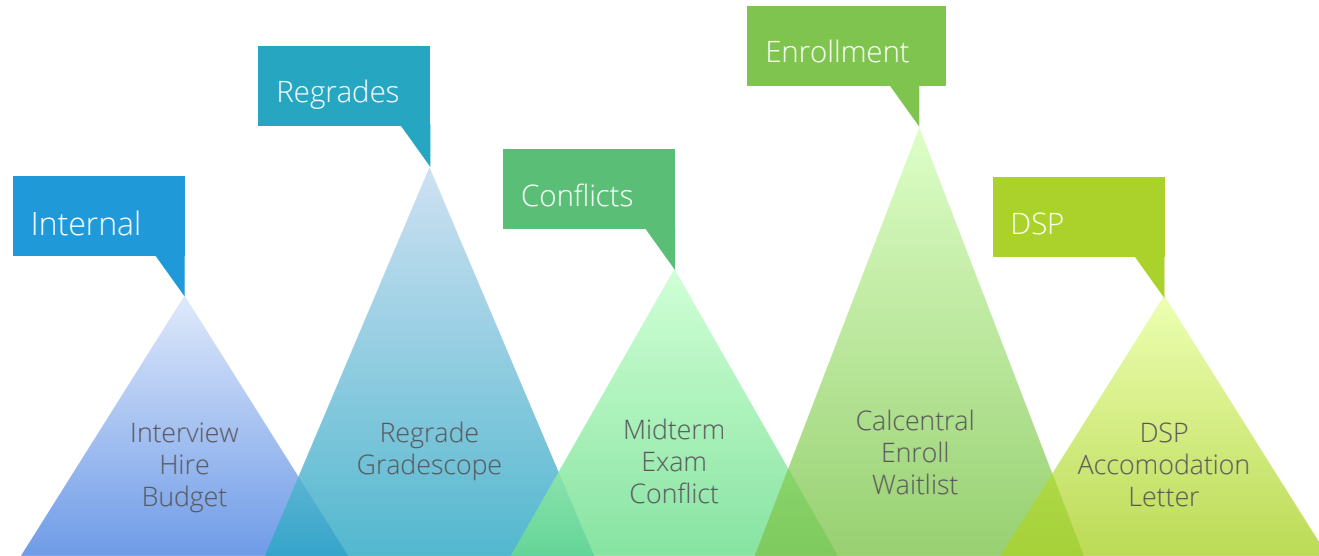
Cleaning & Anonymizing

- Removed any instance of course-identifying words/email ID's
- Filtered out indicators of forwarded messages and unwanted thread attachments via RegEx



Labeling

1. Misl.
2. Conflicts
3. Attendance
4. Assignments
5. Enrollments
6. Internal
7. DSP
8. Regrades



Our Models

Cos Classifier with LDA topics

LDA topics
↓
Cosine Similarity

- Validation accuracy: 38.6%
- Pros: No training needed
- Cons: Depends heavily on topic vectors and too much variability

Random Forest

LDA topics
↓
Random Forest Classifier

- Validation accuracy: 61.4%
- Pros: Easy to train
- Cons: Depends heavily on topic vectors

Cos Classifier with Word2vec

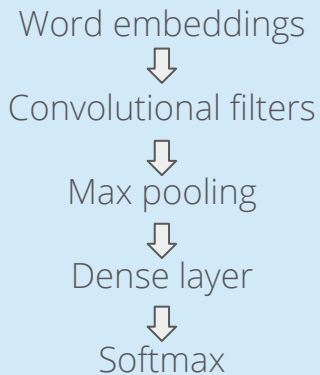
Wor2Vec vectors
↓
Cosine Similarity

- Validation accuracy: 59.8%
- Pros: No training needed
- Cons: Depends heavily on topic vectors

Baseline accuracy : 37.9%

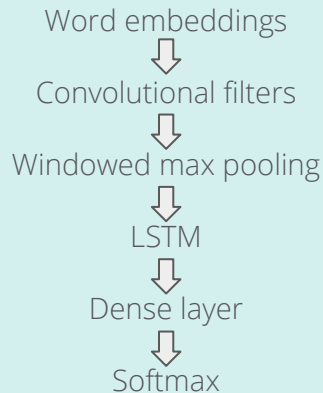
Our Models

Convolutional Neural Network



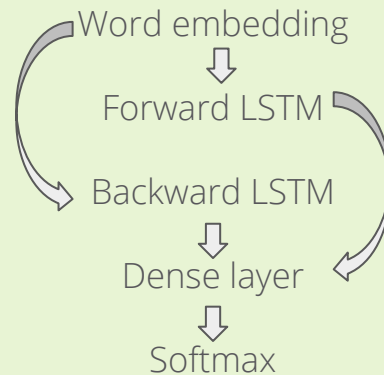
- Validation accuracy: 48.8%
- Pros: Easy to train
- Cons: Can't capture long-term dependencies

C-LSTM



- Validation accuracy: 51.9%
- Pros: CNN can learn short-term and LSTM can capture long-term dependencies
- Cons: Operates in a single direction

Bidirectional LSTM



- Validation accuracy: 74.1%
- Pros: Can effectively use past and future information
- Cons: Harder to train

Baseline accuracy : 37.9%

Chrome Extension

The screenshot shows a Chrome browser window with the address bar displaying `https://mail.google.com/mail/u/0/#inbox?compose=1631f40c556f0609`. The page is titled "bConnected powered by Google". The left sidebar shows the "Mail" menu with options like "Compose", "Inbox (1)", "Starred", "Important", "Sent Mail", "Drafts (1)", and "Categories". The main content area displays a list of emails, each with a star icon, a subject line, and a snippet of the body text. A "New Message" popup is visible on the right side of the email list, showing the "Recipients" field, the "Subject" field, and the body text "Hello, This is an automated email generated by Flo. Thank you for using our services. -Flo". The bottom of the browser window shows the Chrome toolbar with the address bar, search bar, and various icons.

Chrome File Edit View History Bookmarks People Window Help

Extensions x Inbox (1) - tc.lin@berkeley.edu x Ting Chih

Secure | <https://mail.google.com/mail/u/0/#inbox?compose=1631f40c556f0609>

bConnected powered by Google

Mail

COMPOSE

Inbox (1)

Starred

Important

Sent Mail

Drafts (1)

Categories

Auer Lab

TING CHIH

No Hangouts contacts
[Find someone](#)

CS 70 on Piazza [Miscellaneous] [Instr Note] Summer / Fall AI interest form - Instructor James Hulett posted a new Note. Your instructor selected tc 9:43 pm

Launchpad Team [DSP] [Launchpad] Spring 18 Showcase Tomorrow - Hello Ting Chih, Come check out the Launchpad Spring 2018 Project Show 5:52 pm

Parker from Interview Ca. [Regrades] Interview Cake Weekly Problem #189: Implement 5:52 pm

Sreenivas .. Joaquin (17) [Homework] [Conflicts] Intro to Microscopy solutions pl

billingnotice@berkeley.e. [Enrollment] UC Berkeley Student Account Statement is A

Tanya, me (3) [Miscellaneous] [Enrollment] share your CV/resume with

UC Berkeley Career Cente. [Conflicts] Bank of America Merrill Lynch Summer Interns

Jamie from MLH [Attendance] Learn, Build, and Share with MLH Localhost

Parker from MLH [Conflicts] Coding Interview Hacks day 8: How to get t

Music Now (Spring 2018) [Conflicts] Review Session tomorrow Wed May 2nd 2:30-4

Aubrey, me (3) [Enrollment] [Internal] Green grant final points - Forward

Mark from Kaggle [Homework] TrackML Competition Launch: Can ML Assist

noreply@kivuto.com [Miscellaneous] Your software subscription is expiring - De

Berkeley Lab Training [DSP] Your training either has expired or will expire soon

New Message

Recipients

Subject

Hello,

This is an automated email generated by Flo. Thank you for using our services.

-Flo

Sans Serif

Send



Server-side Implementation

~~aws~~



~~Google Cloud Platform~~



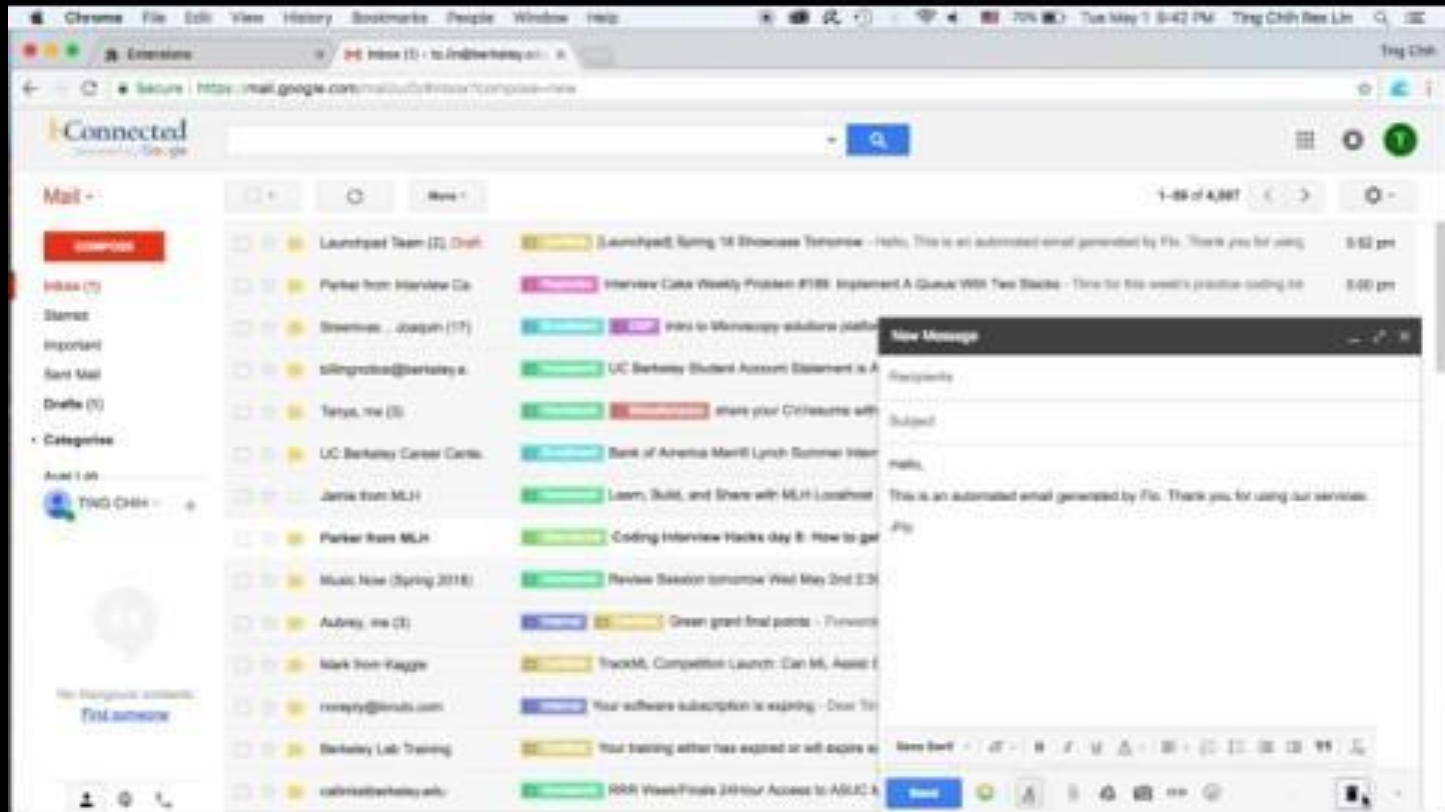
Flask

web development,
one drop at a time



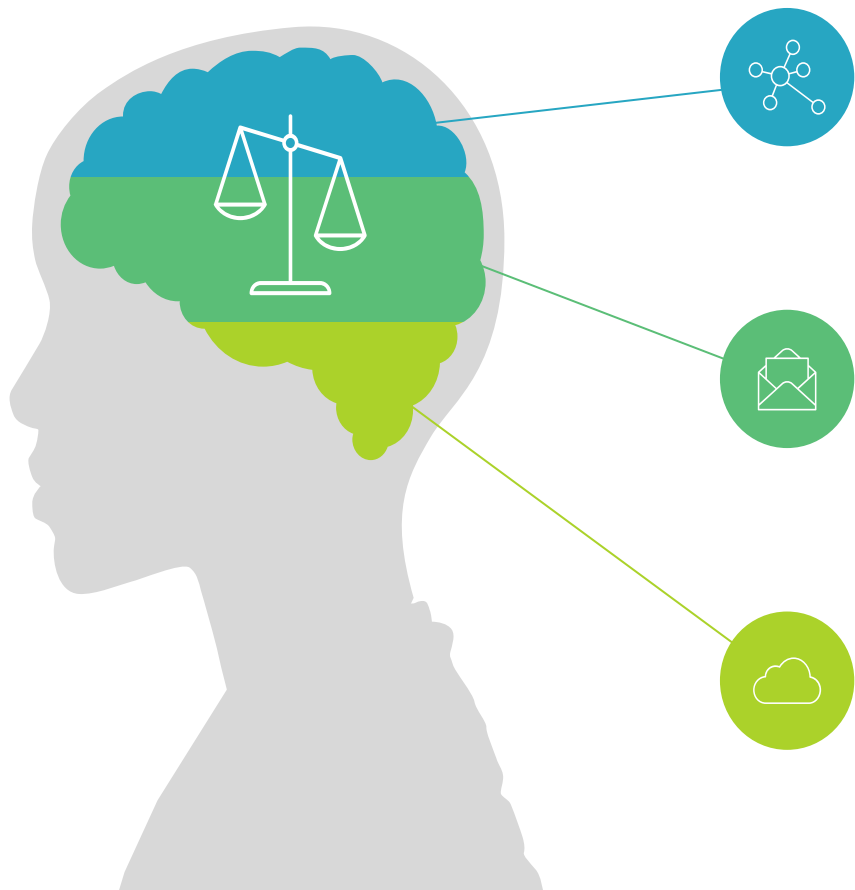
HEROKU

Demo



https://drive.google.com/open?id=1_FE-fZObZVVT0mqfBnZzw8zZ0jtDIP-E

Learning Path Summary



Model

- Corpus
 - Google News Vectors
- Word2Vec / Gensim
- Features of Model
 - Time correlation
 - Pre-assign weights

Email Data

- Understand & work with MBOX file type
- Extract relevant data
- Clean and anonymize
- Labeling

Chrome Extension

Looking Forward: Project Release For Summer 2018

1. Data

- More data
- Further exploration for feature extraction
- Labeling variations

2. Models

- Improve categorization model
- Explore different models

3. Extension

- Launch model
- Feedback feature
- Advance Chrome Extension capabilities

4. Response

- Draft responses based on category

Team Flo



Ting-Chih (Rex) Lin
Fourth year MCB.



Keiko Kamei
Hi! I'm a third year studying Applied Math & Data Science and I will be a Team Lead for the Data Modules Program next semester -- Go Bears!



Rohan Lageweg
Hi! I'm a third year studying EECS/MSE and I'm fairly interested in data science. I'm currently a TA for an undisclosed course and Go Bears!



Joyce Lo
Hi, I'm a senior studying statistics and I'm passionate about data science. I'm currently a GSI for DS100 and I'll be working at Facebook after graduation!



Ndeye Fatou Diop
I am a Master of Engineering in IEOR and did my undergrad in France in Applied Math & Computer Science. I plan to work as a Software Engineer after graduation.



Kristian Rolland
Hi, I'm a junior studying Economics at Cal. My other interests include data science, coding, and music producing!

Notable References

- GloVe: Global Vectors for Word Representation
Stanford Word Vector Paper Published In 2014

Authors: Jeffrey Pennington, Richard Socher, Christopher D. Manning

- Smart Reply: Automated Response Suggestion For Email
A Google Paper Published in June 2016

Authors: Anjuli KAnnan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs Marina Gaenea

- Multi-Task Sequence To Sequence Learning
A Google Brain Paper Published From March 2016

Authors: Minh-Thang Luong, Quoc V. Le, Ilya Sutskever, Oriol Vinyals, Lukasz Kaiser