EMAIL MANAGEMENT & NLP

PRESENTED BY:

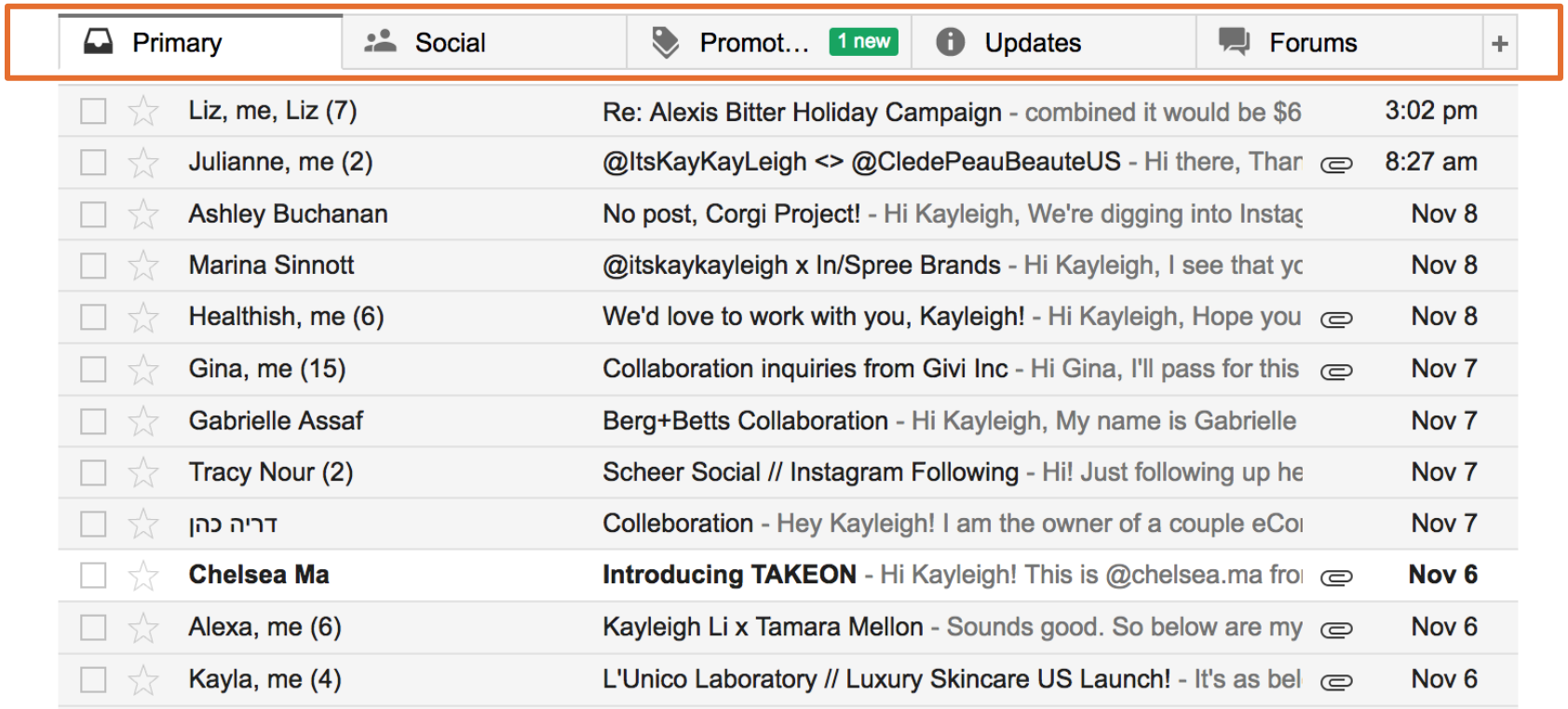
KAYLEIGH LI

PROBLEM

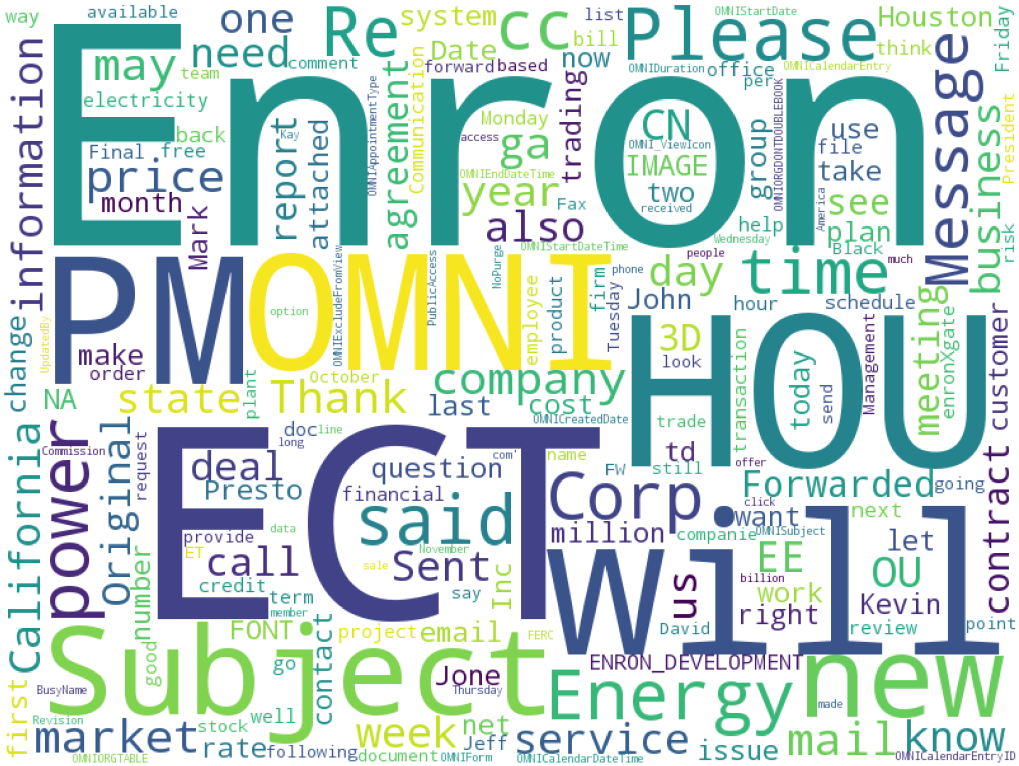
- Time wasted on other people’s priorities
- Less time for real work
- Inefficient way of handling urgent tasks

OBJECTIVE

- **Work Smart and Not Hard**
- **Personalized email management system**
  - 2-5 buckets based on email content



DATA



- ✓ **0.5 million emails**
- ✓ **150 users**
  - ✓ Mostly senior management

TEXT PREPROCESSING

10,000

TEXT PREPROCESSING

- Punctuation ✖

  >?<!{}-)@*&%$#

- Numbers ✖

  1234567890

- Stop words ✖

  the, a, about, an, at …

- Lower case ✔

  Million → million

- Lemmatizing / Stemming ✔

  buying → buy

- Tokenization ✔

  “let me know” → “let”, “me”, “know”, “let me”, “me know”

TEXT PREPROCESSING

10000

843299

- Vectorization ✔

sell, call, fun
1 1 0

TEXT PREPROCESSING

10000

843299

5000

TOPIC MODELING

LDA	(Latent Dirichlet Allocation)
LSA	(Latent Semantic Analysis)
NMF	(Non-Negative Matrix Factorization)

TEXT PREPROCESSING

10000

843299

Non-Negative
Matrix Factorization

4

5000

TOPIC MODELING

Topic: ' Internal_Corp ' **Internal / Corporation Related**

enron, enron enron, corp, corp enron, forwarded, enronxgate, enron enronx gate, america, enron_development, north, north america, enron north, man n, jeff, communication, houston, mark, america corp, enron communication, employee

Topic: ' Meeting_Call ' **Meetings / Calls / Deals**

please, thanks, message, original, original message, know, mail, attache d, call, question, need, agreement, email, deal, information, copy, conta ct, file, please know, review

Topic: ' IT_Updates ' **IT-Related**

schedule, hour, hourahead, final, date, start, start date, hourahead hou r, date hourahead, variance, detected, variance detected, file, final sch edule, message, california, scheduling, portland, scheduling final, file portland

Topic: ' Industry_Market_Business ' **Industry / Market / Business**

would, time, energy, power, http, market, company, price, like, year, ser vice, said, week, also, need, business, know, state, california, informat ion

TOPIC MODELING

31%

Topic: ' Meeting_Call '

Meetings / Calls / Deals

please, thanks, message, original, original message, know, mail, attached, call, question, need, agreement, email, deal, information, copy, contact, file, please know, review

- ✓ **Save time**
- ✓ **More efficient**

APPLICATION

- A new dataset:
  - More personal
  - Less work-related

Calls / Meetings / Appointments

```
Topic  0
call, later, free, sorry, call later, sorry call, mobile, claim, prize, p
lease, urgent, phone, free call, text, stop, please call, reply, meeting,
customer, contact
```

Friends & Family Chatting

```
Topic  1
good, come, time, like, know, home, want, love, take, going, back, still,
dear, morning, well, hope, send, dont, care, give
```

thank you!

APPENDIX

BUCKET 2: IT-RELATED

```
**********
start date: 3601; hourahead hour: 5;  <codesitestart date: 3601; hourahea
d hour: 5;  no ancillary schedules awarded.  variances detected.variances
detected in generation schedule.    log messages:parsing file  o:\portlan
d\westdesk\california scheduling\iso final schedules\2001030605.txt gener
ation schedule $$$ variance found in table tblgen_schedule.     details:
 hour: 5  preferred:   1.00  final:   0.97  trans_type: final  sc_id: ect
stca  mkt_type: 2  trans_date: 3601  unit_id: badlnd_7_pl1x2
**********
schedule crawler: hourahead failurestart date: 11602; hourahead hour: 18;
  hourahead schedule download failed. manual intervention required.    lo
**********
start date: 42901; hourahead hour: 15;  <codesitestart date: 42901; houra
head hour: 15;  no ancillary schedules awarded.  no variances detected.
**********
schedule crawler: hourahead failurestart date: 122001; hourahead hour: 8;
  hourahead schedule download failed. manual intervention required.
**********
schedule crawler: hourahead failurestart date: 11902; hourahead hour: 18;
  hourahead schedule download failed. manual intervention required.    lo
g messages:parsing file  o:\portland\westdesk\california scheduling\iso f
```