Alpha Presentation
Leveraging SPAM to Make Bold Societal Predictions
The Capstone Experience
Team Proofpoint
Matthew Xu
Ben Kandel
Connor Southwell
Scott Newhard
Cameron Smith
Department of Computer Science and Engineering
Michigan State University
Fall 2020
Project Overview

• Reason for project
  ▪ Proofpoint stores terabytes of SPAM data which is costly
  ▪ Improvements on current SPAM engines

• Goal of our project
  ▪ Show that SPAM has meaning
    o Signals
    o Trends

• Why did you do it this way?
  ▪ Proof of concept that SPAM does contain signals and those signals can be meaningful
  ▪ Ultimately, this is only one of the measurements we are exploring
    o Data influx
    o Effectiveness of lures
    o Popular topics being used
System Architecture

Back end

flair
Sentiment Analysis & Machine Learning

Email Parser

proofpoint Input

Front end

React

{REST}

Users

Apache Web Server

Django

Database

Django

SQLite
Landing Page

Leveraging SPAM to Make Bold Societal Predictions

Proofpoint, headquartered in Sunnyvale, California, is a cybersecurity company that provides solutions to top research universities, banks and over half of the Fortune 100 corporations. Proofpoint protects sensitive data across every domain including email, the web, the cloud, social media and mobile messaging.

Proofpoint shields their clients from millions of spam emails per day. By analyzing terabytes of email data, Proofpoint can predict future cyber security attacks and prevent them before they happen, increasing the security and reliability of their system and the web.

Our System
SQLite
Apache

Select a topic

ELECTION 2020
System Health

The Capstone Experience

Team Proofpoint Alpha Presentation
System Health 2

### System Health and Status

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Emails Processed</td>
<td>8933</td>
</tr>
<tr>
<td>Last Data Upload</td>
<td>2 days ago</td>
</tr>
<tr>
<td>Runtime</td>
<td>0.45 ms</td>
</tr>
<tr>
<td># of Positive Sentiment Emails</td>
<td>6006</td>
</tr>
<tr>
<td># of Neutral Sentiment Emails</td>
<td>0</td>
</tr>
<tr>
<td># of Negative Sentiment Emails</td>
<td>2927</td>
</tr>
</tbody>
</table>

---

The Capstone Experience

Team Proofpoint Alpha Presentation
Election Data Visualization
What’s left to do?

• Test our code on Proofpoint's data
• Train our model to analyze sentiment regarding two other selected topics
• Continue to improve accuracy of our model
• Develop more front-end features
• Reduce runtime
Questions?