Visualizing Facebook Data

Noah Duggan Erickson
Dept. of Computer Science
Western Washington University
Bellingham, WA, USA
██████████████

Peter Hafner
Dept. of Computer Science
Western Washington University
Bellingham, WA, USA
██████████████

Carter Jacobs
Dept. of Computer Science
Western Washington University
Bellingham, WA, USA
██████████████

Trevor Le
Dept. of Computer Science
Western Washington University
Bellingham, WA, USA
██████████████

Dustin O'Hara
Dept. of Computer Science
Western Washington University
Bellingham, WA, USA
oharad@wwu.edu

Abstract—**This is the abstract for our report. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque.**

Index Terms—**Facebook, Data Visualization, Social Media, Big Data**

I. INTRODUCTION

The goal for our project was to create a consumer-grade application for Facebook Users to easily interpret the data that Facebook has collected on them. We did this by creating an application that takes the user's data, in json format, and creates visualizations to give the user a greater understanding and in depth insight of their data. From our client, Dustin O'hara, we were instructed to implement a story-like format when creating our visuals so that users who view their own data apply their own interpretation. To establish the final scope of the project, we finalized on the idea of having our visualizations implemented onto a website where users can view our findings and the comparisons of us as Facebook users. The website will be implemented by Carter Jacobs and if the users are interested they are about to download a package which allows them to implement their own Facebook Profile data and view their own visuals.

Our project was only made possible from the General data Protection Regulation (GDPR) forcing companies to allow the users to view, edit and delete the data collected on them. The motivation behind this project was powered by the want of unraveling the contents of Facebook's data collection. We wanted users of all ages and skill levels to be able to make sense of the data.

The collection of our hard work will provide opportunities for education of one's digital footprint and spread awareness on the mysteries on how Facebook or companies similar to Facebook collect data on their users.

II. RELATED WORK

With the ubiquity of social media platforms in the modern age [1], it is no surprise that these monolithic services collect/hold unfathomable quantities of data on their users. This data can then be used by Meta and their advertising partners to infer a user's interests, preferences, and behaviors. However, these inferences can frequently be made erroneously or in ways that do not align with the user's true self or values.

A prime example of where these discrepancies can manifest lies in the algorithmic profiling for ads interests, which [2] found to at least "somewhat accurate[ly]" reflect the user in only 47.6% of survey participants. On the other hand, textual data such as posts, comments, and messages can be used in a variety of ways, including in depression detection, where [3] shows an F1 score (see Eqn. 1) of 0.89 on a Twitter-based dataset.

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{1}$$

Where divides occur, they manifest from a variety of sources such as the difficulty of retrospective geolocation of IPv4 addresses [4], and the false persona we present

while online [5]. This "garbage in, garbage out" phenomenon highlights the inherent challenges in deriving accurate insights from social media data, as the input data may be inherently flawed or misleading.

Despite these shortcomings, much of the literature approaches user-centric Facebook data analysis with the intent to surprise the user about the sheer quantity of data, without considering the quality thereof [6], [7]. In this project, we aim to take a different approach - one that centers on digital self-reflection by encouraging users to interact with their data in a way that motivates and self-generates individual discovery and insight.

III. METHODOLOGY

We focused on making sure that all the visualizations that we created were going to have some meaning behind them. We looked at all the files to find files that we found the most interesting. Our final selection of files includes all the messaging files, topics, (SPECIFIC FILE NAMES HERE). We also wanted to provide the user with an overall interactive view of the size of their data, as each individual user's data has differing sizes from a few megabytes to several gigabytes.

We created six main methods, one for each type of visualization we decided to use. There is Facebook activity, off Facebook activity, filesize sunburst, location via IP addresses, notification charting, and interests. Each creating unique and in-depth visualizations on their respective files. There are two groups of methods that we have, there are three word based methods and three statistical methods. The word based methods are on and off facebook activity and interests and the statistical methods are filesize sunburst, location via IP addresses, and notification charting.

Going more in depth, the word based methods are focused on the words that are contained inside the files, rather than specific numeric values in the files. Facebook activity uses the users messaging data to create sentiment plots along with word clouds of the users most messaged words over time. Off Facebook activity uses what other sites report to Facebook that the user has visited and creates a visual showing the top 5 most visited sites and all sites the user has visited over 50 times. Interests takes the data of what ads the user is interested in and creates a collage of pictures of those interests to show the user the variety that Facebook thinks they are interested in.

The statistical methods use the numerics of the data to give the user more insight. Filesize sunburst uses all the users data files and creates a sunburst showing the size of each folder in comparison to all the other folders. That includes an interactable site that allows the user to click on folders to enhance the composition of that folder. Location via IP addresses takes all the IP addresses that Facebook collects from the user and creates a map showing all the specific locations and how many times the user has been tracked at that location. Notification charting uses the past 30 days(the amount of days that Facebook holds onto this type of data per user) of notifications that a user has received and plots them in a similar manner to the github contributions graph.

All of the methods use a variety of Python libraries, ranging from Pandas to beautifulsoup4 to matplotlib. (IM NOT SURE WHAT ELSE TO ADD HERE, IF ANYTHING)

IV. RESULTS & ANALYSIS

This is Peter's Results section. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

An example of how to include a figure is shown in Fig. 1. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.
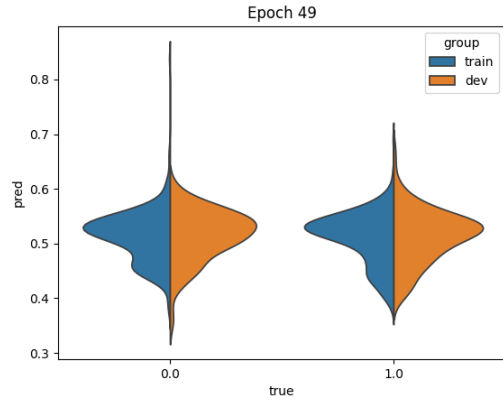
Fig. 1. ABIDE CNN Model predictions at epoch 49 from [8].

Even more figures are shown in Fig. 2. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

An example of how to include a table is shown in Table I. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.
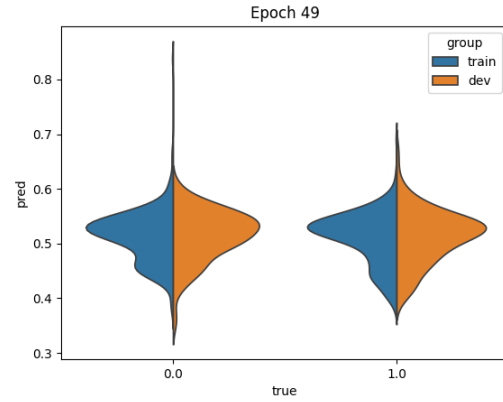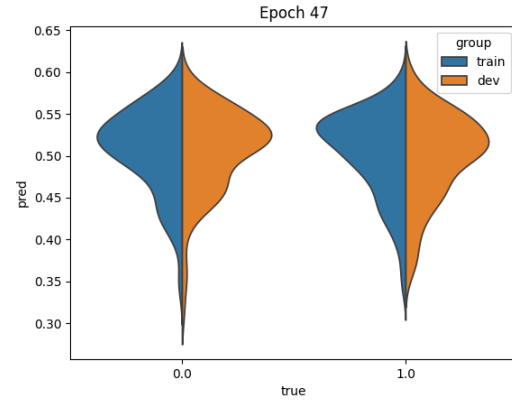

Fig. 2. ABIDE CNN Model predictions at multiple epochs from [8].

TABLE I
POINTLESS TABLE.

C1	C2	C3
1	2	3
4	5	6
7	8	9

NOTE: Positioning of the figs/tables known-bad - Will be refined when actual content is available.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pul-

vinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu,

tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in,

ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi.

V. Conclusion

This is Trevor's Conclusion section. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Acknowledgments

References

[1] Meta, *Meta reports fourth quarter and full year 2023 results; indicates quarterly dividend*, Press Release, 2024.

[2] M. Büchi, E. Fosch-Villaronga, C. Lutz, A. Tamò-Larrieux, and S. Velidi, "Making sense of algorithmic profiling: User perceptions on facebook," *Information, Communication & Society*, vol. 26, no. 4, pp. 809–825, 2023. DOI: 10.1080/1369118X.2021.1989011.

[3] J. d. J. Titla-Tlatelpa, R. M. Ortega-Mendoza, M. Montes-y-Gómez, and L. Villaseñor-Pineda, "A profile-based sentiment-aware approach for depression detection in social media," *EPJ Data Science*, vol. 10, no. 54, 2021.

[4] D. Komosny, "Retrospective ip address geolocation for geography-aware internet services," *Sensors*, vol. 21, no. 15, 2021, ISSN: 1424-8220. DOI: 10.3390/s21154975. [Online]. Available: https://www.mdpi.com/1424-8220/21/15/4975.

[5] O. Gil-Or, Y. Levi-Belz, and O. Turel, "The "facebook-self": Characteristics and psychological predictors of false self-presentation on facebook," *Frontiers in Psychology*, vol. 6, 2015, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2015.00099. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2015.00099.

[6] A. M. V. Venkata Sai and Y. Li, "A survey on privacy issues in mobile social networks," *IEEE Access*, vol. 8, pp. 130 906–130 921, 2020. DOI: 10.1109/ACCESS.2020.3009691.

[7] M. A. Kabir, J. Han, J. Yu, and A. Colman, "User-centric social context information management: An ontology-based approach and platform," *Personal and Ubiquitous Computing*, vol. 18, pp. 1061–1083, 2014.

[8] N. Duggan Erickson, *On the dangers of ai autism diagnosis*, 2024.