Project Proposal

Group 7

Harshitha Peddi, Manikya Deepika Eadara, Taravat Ashabi

Data Analytical Engineering, George Mason University

AIT 664: Represent, Process & Visualize Applied Information Technology

Prof. Ebrima N Ceesay

Behind the Screen: The Damaging Effects of Algorithmic Bias on social media.

Overview:

This study aims to investigate the connection between social media usage and the impacts it has on peoples' personal lives. Personal handheld devices and the widespread use of social media platforms have caused people to spend more time on their smartphones, leading to concerns about the impact of social media on emotional, mental, and physical well-being. This project aims to investigate the relationship between social media usage and its consequences on individuals' private lives using Kaggle's Social Media Influence dataset. Our research will provide insights into the implications of social media usage on individuals' personal lives and highlight the need for strategies to mitigate the potential negative impacts on mental and physical health.

Literature review: [2]

Our daily lives have become increasingly reliant on social media platforms, which give us a digital forum for connection, communication, and knowledge sharing. The development of social media technologies has changed how individuals communicate with one another in both their personal and professional lives. Users can now collaborate and communicate online thanks to social media platforms like weblogs, blogs, wikis, broadcasts, photographs, and videos. In contrast to typical one-way communication in conventional media, social media has made two-way contact between users possible.

Javed conducted a study to investigate the impact of social media on employee performance, productivity, job satisfaction, and organizational commitment. The study found that social media has a positive effect on employee performance and productivity, and it also facilitates knowledge-sharing and innovation within organizations. However, excessive use of social media can lead to potential risks to mental health and well-being, such as anxiety, depression, and other mental health problems, as well as negative impacts on physical health, such as poor sleep quality and increased sedentary behavior. Therefore, it is crucial to encourage healthy social media use and develop strategies to mitigate the potential negative impacts of social media on mental and physical health. [5]

Research question:

According to research, how much time people spend on social media relates to their personal life. We will specifically investigate the following issues:

1. How long do people typically spend on social media in a day?
2. How is social media impacting user's mood?
3. Which community members use social media platforms more frequently?
4. Are people prepared to give up social media in return for faster professional advancement?
5. Do individuals blame social media for compromising their own health?
6. Which profession has the highest likelihood of having users on social media?

Data source:

The dataset, which comprises information on social media engagement and usage across numerous platforms, such as Instagram, Facebook, YouTube and Twitter, is available from Kaggle [3]

The data also contains particulars on the users' location, gender, and age.

Objectives:

1. To ascertain the typical amount of time people spend using their and social media based off their age and gender.
2. To research and establish the link between personal life satisfaction and social media use.
3. To find any patterns or trends in the data that can assist us in comprehending how social media affects our daily lives.

Methods:

1. Data Cleaning and Preprocessing: [7]

Data cleaning and preprocessing are essential steps in preparing a dataset for analysis. In the context of the Social Media Influence dataset from Kaggle, data cleaning involves identifying and correcting inaccurate, damaged, duplicate, or incomplete data points. Conversely, preprocessing consists in transforming the data into a more usable format, such as converting categorical variables into numerical ones or scaling variables to have similar ranges.

Several techniques can be used to clean and preprocess the Social Media Influence dataset with the help of python libraries. The preprocessing of data achieves the following:

1. Fixing structural errors.
2. Filtering unwanted outliners.
3. Handling missing data.
4. Transforming dataset into quality data.

Quality data:

Various measures must be taken to ensure quality data in the Social Media Influence dataset. Firstly, the validity of the data needs to be checked to ensure that it measures what it intends to measure and conforms to the dataset's purpose. Accuracy should be verified to ensure that the data is error-free and precise. Completeness of the data is crucial to ensure that all values are present in the results. Consistency is also vital to ensure that the data is uniform and free of contradictions and that all the data points follow the same format. Finally, uniformity should be maintained throughout the dataset to ensure consistency across all its attributes. Ensuring these qualities are met, the Social Media Influence dataset can be considered quality data, effectively used for various analyses and applications.

Importance of data cleaning:

1. Removal of inaccuracies when several data sources are involved.
2. Clients are happier and employees are less annoyed when there are fewer mistakes.
3. Monitoring mistakes and improving reporting make it easier to resolve incorrect or damaged data for future applications by allowing users to identify where issues are coming from.
4. Using data cleaning solutions will result in more effective corporate procedures and quicker decision-making.

2. Data Analysis: [6]

Various technical tools and techniques can be utilized to perform data analysis on the dataset available at Kaggle [3]. These may include statistical software such as Python and tableau which can be used to manipulate and analyze the data and create visualizations such as charts, graphs, and tables. Additionally, machine learning algorithms such as regression analysis can be applied to identify patterns and relationships within the data. Other techniques, such as data mining and text analytics, can also extract valuable insights from the data.

Data analysis process:

Gathering all the information, processing it, studying the data, and using it to uncover patterns and other insights are all parts of the data analysis process, often known as the data analysis phases. The steps involved in data analysis are as follows:

1. **Data collection**: The first step is to collect the data from various sources, such as databases, web services, or files. In the case of the social media influence dataset, the data may be obtained from sources such as Twitter or Facebook APIs.
2. **Data preprocessing**: This step involves cleaning, transforming, and structuring the data to make it suitable for analysis. This may include removing duplicates, handling missing values, and converting data types.
3. **Exploratory data analysis**: In this step, we perform a statistical and visual analysis to understand the relationships and patterns in the data. This helps to identify any outliers or anomalies and gain insights into the data.
4. **Data modeling**: This step involves creating mathematical models or algorithms to make predictions or gain insights from the data. Regression analysis is a commonly used technique in data modeling, which consists in finding the relationship between a dependent variable and one or more independent variables. [1]
5. **Data visualization**: Visualization helps to present the insights and results from the data analysis in an easily understandable format. Tableau is a powerful tool for creating interactive and visually appealing dashboards and reports.
6. **Communicating results**: Finally, the insights and results obtained from the data analysis are displayed to stakeholders clearly and concisely.

Data analysis methods: [4]

Although there are multiple data analysis methods available, they all fall into one of two primary types : qualitative analysis and quantitative analysis. We can use these data analysis approaches individually or in combination to get business insights from various data kinds.

1. **Quantitative data analysis**: In the context of the social media influence dataset on Kaggle, quantitative data analysis can be performed using technologies such as Python and regression analysis.

2. **Qualitative data analysis**: On the other hand, qualitative data analysis can be conducted using tools such as Tableau, which allows for the visualization of non-numerical information.

Timeline:

The project is anticipated to be finished in 10 to 12 weeks, with the following timeline:

- Weeks 1-2: Data preprocessing and project setup
- Weeks 3–4: Descriptive analysis and data visualization
- Weeks 5–6: Interpretation of Statistical Analysis and Findings
- Weeks 7-8: Report writing and Recommendation Developments
- Weeks 9–10: Creation and Presentation of the Final Report

Expected Results:

The proposed research aims to investigate the impact of social media on individuals' private lives and to provide insights into developing healthy usage patterns. The research findings can be used to create interventions to increase awareness of the potential adverse effects of excessive social media use and to promote positive behavior change. These recommendations can guide individuals, groups, and policymakers in creating policies and guidelines that support healthy social media use and foster overall well-being and productivity. The research may employ various data analysis techniques, such as qualitative and quantitative methods, and utilize technology tools and software to analyze and interpret data.

Conclusion:

The study's results contribute significantly to understanding the relationship between social media usage and its impact on individuals' personal lives. The insights gained from this research could inform the development of effective strategies for promoting healthy social media usage patterns and creating interventions to raise awareness of the potential negative consequences of excessive social media use. The recommendations generated by this analysis could guide individuals, groups, and policymakers in implementing measures to manage social media usage to enhance productivity and well-being while minimizing any adverse effects on personal life.

References:

[1] Courses AIU. (n.d.). *Correlation and Regression*. Retrieved from Courses AIU: https://courses.aiu.edu/STATISTICS/10/SESSION%2010%20CORRELATION.pdf

[2] Frontiers. (n.d.). *The Impact of Social Media Usage on Work Efficiency: The Perspectives of Media Synchronicity and Gratifications*. Retrieved from Frontiers: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.693183/full

[3] Kaggle. (n.d.). *Social Media Influence*. Retrieved from Kaggle: https://www.kaggle.com/datasets/apoorva1225/social-media-influence

[4] Maryville University. (n.d.). *Top 4 Data Analysis Techniques That Create Business Value*. Retrieved from Maryville University: https://online.maryville.edu/blog/data-analysis-techniques/#:~:text=The%20two%20primary%20methods%20for,insights%20from%20different%20data%20types

[5] Science Direct. (n.d.). *Does time spent using social media impact mental health?: An eight year longitudinal study*. Retrieved from Science Direct: https://www.sciencedirect.com/science/article/abs/pii/S0747563219303723?via%3Dihub

[6] SimpliLearn. (n.d.). *What is Data Analysis? Methods, Process and Types Explained*. Retrieved from SimpliLearn: https://www.simplilearn.com/data-analysis-methods-process-types-article

[7] Tableau. (n.d.). *Guide To Data Cleaning: Definition, Benefits, Components, And How To Clean Your Data*. Retrieved from Tableau: https://www.tableau.com/learn/articles/what-is-data-cleaning