

Studying the Impact Of D.P.SL Model on Online Identity Management

Pratik Emmanuel¹ and Olufemi Isiaq²

¹Creative Computing Institute, Postgraduate Research, University of the Arts London, UK

²Creative Computing Institute, Computer Data Science and AI, University of the Arts London, UK

p.emmanuel1220231@arts.ac.uk

f.isiaq@arts.ac.uk

Abstract: This paper investigates online identities through social media language use, with a focus on classifying online identity within textual conversations. It sheds light on how Demographic (D) groups and Personality (P) impact the use of Social-Media Language (SL) for identity representation online. This study was conducted in 2023 when there were 4.76 billion social media users worldwide, making it essential to study how social media language is used in textual conversations to convey online identity. The study defines social media language as consisting of emoticons/emojis, abbreviations, and mixed language within textual conversations, which have become essential for expressing feelings and emotions during conversations. The D.P.SL based survey conducted for this work aimed to understand how demographic groups and personality are related to social media language. Based on the total number of social media user worldwide, 400 responses (required based on Cochran and Yamane's formulae sample-size calculation) and the survey was distributed across various verified online survey exchange platforms. However, 406 responses were recorded with young people in age groups of 18-24 and 25-34 using social media language more as it has become a part of their social media habits. The study also found that emoticons/emojis and slang abbreviations with letter reduplication were quite common, making conversations lively and funny. Additionally, individuals whose primary language is not English use their native language but type in English for quick communication. Subsequent study is to be conducted using online mock group conversations between participating respondents to further understand correlations, causation, and concurrency on how 'online identity' is managed during online communications via social media language, its context of use, and polarity sentiment.

Keywords: Demographic groups, Personality, Social media language, Online identity management, Mixed languages, Textual conversations

1. Introduction

In recent times, the Internet has now become an indispensable part of our lives, with individuals and businesses relying on it for effective functioning. Social media, which is an offshoot of the Internet, has allowed people to connect with others, be it individuals, organizations, or businesses. Social media users leave behind a trail of their online activity, which can be studied by social media researchers. As of 2023, there were 4.76 billion social media users worldwide, making it crucial to study Social Media Language (SL), which enables people to express their feelings and emotions in textual conversations (Kemp, 2023).

Social media users actively manage their online identity by presenting themselves in a way that influences the opposite person with whom they are conversing. Warburton and Hatzipanagos (2015) define online identity as an individual's social identity that creates a self-impression on others based on the conversation topic. Social media users tend to display their positive self to gain attention from friends and non-friends during textual conversations. Studies show that a person's self-impression varies based on the settings of a particular online platform (McCabe et al., 2005; Gibbs et al., 2006; Heino et al., 2006).

To gain a better understanding of how social media users utilise language to present their online identity, this work is investigating how social media language is used strategically and performatively within textual conversations. This outcome can be used to develop computer-based tools for identifying social media language and understanding online identity. Furthermore, studying a person's online identity formation within text conversation can help researchers understand how the results can be used to improve human-computer interaction, human language translation and text to speech amongst others (Pietro, 2020).

A critical question that arises is whether there is a correlation between demographic groups and personalities with social media language, which can be used to represent identity online. To establish this correlation, we have defined a Null Hypothesis (H0), which states that *demographic groups and personalities do not impact the use of social media language for identity representation online*. An Alternate Hypothesis (H1) states that *demographic groups and personalities do have an impact on the use of social media language for identity representation online*.

2. Method

Before designing the online survey for this work, a literature analysis to investigate related works of other authors was conducted. The academic contribution of this work is D.P.SL (Demographic groups, Personality, Social-Media Language) model to help understand how demographic groups and personalities impact the use of social media language. The survey is cross-sectional meaning it was easy to administer and cost-effective. Also, a JISC Survey tools which is considered safe with regards to data protection was adopted (JISC, 2023).

To begin data collection, the survey was published on survey exchange groups on social media platform Facebook alongside survey exchange websites with multiple social media accounts for verification of their real-world existence and checking their review ratings for further verification (SurveyCircle, 2023; Survey Swap, 2023). Survey exchange is where researchers gain participants for their surveys outside of their friend group, the survey exchange groups on Facebook work by completing surveys of other researchers and sharing a screenshot of the completed survey as proof of completion, and then asking them to complete the needed survey by sharing the survey link with them to which they will respond back with proof of completion. Survey exchange groups on Facebook was the main source of gaining more participants as everyday surveys were exchanged throughout the day and night and the cycle repeated each day for two months. The survey link was also further distributed online by sharing among various groups. Informed consent was also presented to the participants at the beginning of the online survey so they could first read about the research project, then data safety assurance and informed consent before deciding if they wanted to take part in the survey. The contents of the informed consent were derived from the university's informed consent document (Solent, 2023)

2.1 Online Survey Sample Size Calculations (SSC)

Social media users worldwide in 2023 at the time of data collection were 4.76 billion according to a report from 'DataReportal' (Kemp, 2023; Statista, 2023). Industry standard confidence level at 95% was adopted (Hazra, 2017). The margin of error was between 4 – 8% meaning that the range of the population's response may deviate from sample is between 4 – 8%. By adopting these formulas, the Cochran's 1963 formula showed that 384 responses were required for the online survey (Asenahabi and Ikoha, 2023) (Equation 1), whilst Yamane's 1967 formula showed that 399 responses were required (Divakar and Nanjundeswaraswamy, 2021) (Equation 2). Hence, in-total 406 responses were obtained for the online survey during data collection.

2.1.1 Online survey SSC: Cochran's Formula

$$n = \frac{z^2 pq}{e^2}$$

Equation 1: Cochran's Formula.

$z = 1.96$ for a confidence level of 95%, (e) margin of error 5% which is 0.05 and (p) population proportion is 0.5 which is 50% (left at default if not known).

$$z = 1.96, p = 0.5, e = 0.05$$

$$n = 1.96^2 * 0.5 * (1 - 0.5) / 0.05^2$$

$$n = 384.16$$

2.1.2 Online survey SSC: Yamane's Formula

$$n = \frac{N}{K + N(e)^2}$$

Equation 2: Yamane's Formula.

N = Population of study, K = Constant (1), e = Degree of error expected and n = Sample size.

$N = 4760000000$ (Social media users worldwide).

K = 1

e = 0.05 (5% margin of error (default)).

n = 4760000000 / (1 + 4760000000 * 0.05^2)

n = 399.99

2.2 Online Survey Data Analysis

Statistical tests for hypotheses testing (Section 1.2) were conducted for the survey data analysis. Since, the phase one variables were categorical in-nature, a non-parametric ‘Wilcoxon Signed-rank test (WSR)’, ‘Chi-Square Test of Independence (CSTI)’ and ‘Correlation Coefficient’ test was applied (Table 1).

Table 1: Predictor variables & Target variable

	Variables	Analysis	Tools
Demographic groups. (Predictor)	Language, Age-Groups, Education qualification Vs. Social media Language.	Correlation Coefficient, CSTI & WSR test.	Jupyter Notebook. Python programming language.
Personality. (Predictor)	Personality (OCEAN) Vs. Social media Language.	Correlation Coefficient, CSTI & WSR test.	Jupyter Notebook. Python programming language.
Social media Language. (Target)	All Social media languages (emoticons/emoji, abbreviations and mixed-languages).	Correlation Coefficient, CSTI & WSR test.	Jupyter Notebook. Python programming language.

3. Results

3.1 Correlation Coefficient

The interpretation of the correlation values was carried out as per Table 2 below (Bhandari, 2023).

Table 2: Correlation Coefficient interpretation table

Correlation ranges	Interpretation	Correlation type
-.7 to -1	Very strong	Negative
-.5 to -.7	Strong	Negative
-.3 to -.5	Moderate	Negative
0 to -.3	Weak	Negative
0	None	Zero
0 to .3	Weak	Positive
.3 to .5	Moderate	Positive
.5 to .7	Strong	Positive
.7 to 1	Very strong	Positive

3.1.1 Demographic groups and social media languages

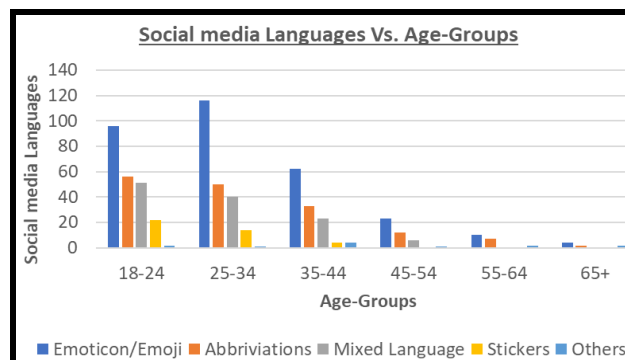


Figure 1: Number of users using social media language as per 'Age-Groups'

The preliminary observation of ‘Age-Groups’ and ‘Social media Language’ suggested that as ‘Age-Groups’ increase so the use of social media language decreases, but the use of ‘Emoticons/Emoji’, ‘Abbreviations’ and ‘Mixed Language’ remains prevalent (Figure 1). Therefore, to understand the significance between variables represented by the P-values, ‘Demographic Groups’ was broken down, hence Table 3 below shows correlation values and the significance represented by their P-value of each age-groups and ‘Social media Language’. The interpretation of the P-value was that if it would be less than $P < 0.05$, then there is a significant relationship (Zach, 2021). The Table 3 below suggests that within ‘Demographic Groups’, ‘Age-Groups’ has a significant relation with ‘Social media Languages’ observed by their P-values. The Table 3 also shows a strong correlation between all six age-groups and ‘Social media Languages’.

Table 3: Each of the ‘Age Groups’ versus the ‘Social media Language’

Age Groups	Social media Language				
	Emoticons/Emoji	Abbreviations	Mixed Languages	Stickers	Gifs
18-24 (32.8%)	0.069 (Weak, Positive) P-value: 0.223	0.841 (Very Strong, Positive) P-value: 0.492	0.917 (Very Strong, Positive) P-value: 0.023	0.443 (Moderate, Positive) P-value: 0.021	0.232 (Weak, Positive) P-value: 0.018
25-34 (32.5%)	0.068 (Weak, Positive) P-value: 0.234	0.835 (Very Strong, Positive) P-value: 0.013	0.923 (Very Strong, Positive) P-value: 0.024	0.448 (Moderate, Positive) P-value: 0.074	0.234 (Weak, Positive) P-value: 0.025
35-44 (20.7%)	0.048 (Weak, Positive) P-value: 0.398	0.642 (Strong, Positive) P-value: 0.054	0.768 (Very Strong, Positive) P-value: 0.032	0.632 (Strong, Positive) P-value: 0.237	0.330 (Moderate, Positive) P-value: 0.024
45-54 (7.6%)	0.027 (Weak, Positive) P-value: 0.635	0.324 (Moderate, Positive) P-value: 0.046	0.420 (Moderate, Positive) P-value: 0.076	0.866 (Very Strong, Positive) P-value: 0.019	0.602 (Strong, Positive) P-value: 0.029
55-64 (4.4%)	0.020 (Weak, Positive) P-value: 0.724	0.241 (Weak, Positive) P-value: 0.012	0.313 (Moderate, Positive) P-value: 0.034	0.645 (Strong, Positive) P-value: 0.027	0.816 (Very Strong, Positive) P-value: 0.013
65+ (2%)	0.012 (Weak, Positive) P-value: 0.821	0.158 (Weak, Positive) P-value: 0.041	0.205 (Weak, Positive) P-value: 0.073	0.423 (Moderate, Positive) P-value: 0.058	0.811 (Very Strong, Positive) P-value: 0.026

3.1.2 Personality and social media languages

Table 4: Correlation Coefficient between ‘Personality’ and ‘Social media Language’

Personality	Social media Language				
	Emoticons/Emoji	Abbreviations	Mixed Languages	Stickers	Gifs
Openness	-0.018 (Weak, negative) P-value: 0.723	0.021 (Weak, positive) P-value: 0.671	0.009 (Weak, positive) P-value: 0.848	0.026 (Weak, positive) P-value: 0.593	0.033 (Weak, positive) P-value: 0.503
Conscientiousness	-0.032 (Weak, negative)	0.039 (Weak, positive)	-0.035 (Weak, negative)	0.006 (Weak, positive)	-0.006 (Weak, negative)

Personality	Social media Language				
	Emoticons/Emoji	Abbreviations	Mixed Languages	Stickers	Gifs
	P-value: 0.514	P-value: 0.437	P-value: 0.491	P-value: 0.912	P-value: 0.934
Extroversion	-0.037 (Weak, negative) P-value: 0.465	-0.032 (Weak, negative) P-value: 0.526	0.028 (Weak, positive) P-value: 0.572	-0.039 (Weak, negative) P-value: 0.424	-0.039 (Weak, negative) P-value: 0.431
Agreeableness	-0.008 (Weak, negative) P-value: 0.862	0.028 (Weak, positive) P-value: 0.577	0.015 (Weak, positive) P-value: 0.762	0.029 (Weak, positive) P-value: 0.554	0.059 (Weak, positive) P-value: 0.238
Neurotic	-0.054 (Weak, negative) P-value: 0.28	-0.019 (Weak, negative) P-value: 0.71	-0.063 (Weak, negative) P-value: 0.20	-0.040 (Weak, negative) P-value: 0.42	-0.091 (Weak, negative) P-value: 0.07

From Table 4, it can be observed that the correlation coefficient values between ‘Personality’ and ‘Social media Language’ suggests that there is a weak relationship between the two variables.

3.2 Wilcoxon Signed-rank (WSR) test for Demographic Groups and Social Media Language

According to the Table 5 below, there is a significant relation between ‘Age-Groups’ and ‘Stickers’, and ‘Educational Qualification’ and all of the ‘Social media Languages’ as their P-values are less than $P < 0.05$. A significant relation was also observed between ‘Language’, ‘Emoticons/Emoji’, ‘Mixed Languages’ and ‘Stickers’. Overall, there is a significant relation between ‘Demographic Groups’ and ‘Social media Language’ supported by the WSR test.

Table 5: WSR test between ‘Demographic Groups’ and ‘Social media Language’

Demographic Groups	Social media Language				
	Emoticons/Emoji	Abbreviations	Mixed Languages	Stickers	Gifs
Age-Groups	Statistic: 8789.0 P-value: 0.09	Statistic: 4816.0 P-value: 0.08	Statistic: 3003.0 P-value: 0.06	Statistic: 1824.0 P-value: 0.01	Statistic: 486.5 P-value: 0.06
Educational Qualification	Statistic: 0.0 P-value: 0.01	Statistic: 0.0 P-value: 0.04	Statistic: 0.0 P-value: 0.02	Statistic: 0.0 P-value: 0.04	Statistic: 0.0 P-value: 0.01
Language	Statistic: 4958.5 P-value: 0.01	Statistic: 10450.0 P-value: 0.58	Statistic: 5264.0 P-value: 0.04	Statistic: 2002.0 P-value: 0.02	Statistic: 841.5 P-value: 0.07

3.2.1 WSR test on personality vs. social media language

According to Table 6 below, it was observed that there is a significant relation between ‘Personality’, and ‘Social media Language’ supported by the WSR test.

Table 6: WSR test between ‘Personality’ and ‘Social media Language’

Personality	Social media Language				
	Emoticons/Emoji	Abbreviations	Mixed Languages	Stickers	Gifs
Openness	Statistic: 0.0 P-value: 0.07	Statistic: 0.0 P-value: 0.08	Statistic: 0.0 P-value: 0.03	Statistic: 0.0 P-value: 0.01	Statistic: 0.0 P-value: 0.01
Conscientiousness	Statistic: 0.0 P-value: 0.02	Statistic: 0.0 P-value: 0.06	Statistic: 0.0 P-value: 0.02	Statistic: 0.0 P-value: 0.01	Statistic: 0.0 P-value: 0.06

Personality	Social media Language				
	Emoticons/Emoji	Abbreviations	Mixed Languages	Stickers	Gifs
Extroversion	Statistic: 0.0 P-value: 0.03	Statistic: 0.0 P-value: 0.01	Statistic: 0.0 P-value: 0.06	Statistic: 0.0 P-value: 0.02	Statistic: 0.0 P-value: 0.02
Agreeableness	Statistic: 0.0 P-value: 0.01	Statistic: 0.0 P-value: 0.06	Statistic: 0.0 P-value: 0.07	Statistic: 0.0 P-value: 0.02	Statistic: 0.0 P-value: 0.01
Neurotic	Statistic: 0.0 P-value: 0.01	Statistic: 0.0 P-value: 0.03	Statistic: 0.0 P-value: 0.02	Statistic: 0.0 P-value: 0.01	Statistic: 0.0 P-value: 0.01

3.3 Chi-square Test of Independence

3.3.1 Demographic groups vs. social media language

The Chi-square Test of Independence is a non-parametric test to determine if the categorical variables are associated (Jain, 2020). In this phase one the categorical variables are ‘Demographic Groups’, ‘Personality’ and ‘Social media Language’ (Table 1). The Chi-square Test of Independence was performed between ‘Demographic groups’ and ‘Social media Language’ to assess the statistical associations between them (Table 7). This test was also performed between ‘Personality’ and ‘Social media Language’ (Table 8).

Table 7: The Chi-square Test of Independence between ‘Demographic groups’ and ‘Social media Language’

Demographic Groups	Social media Language				
	Emoticons/Emoji	Abbreviations	Mixed Languages	Stickers	Gifs
Age-Groups	Chi-square value: 5.31 P-value: 0.37 Degrees of Freedom: 5	Chi-square value: 14.68 P-value: 0.01 Degrees of Freedom: 5	Chi-square value: 2.04 P-value: 0.84 Degrees of Freedom: 5	Chi-square value: 4.82 P-value: 0.04 Degrees of Freedom: 5	Chi-square value: 0.93 P-value: 0.96 Degrees of Freedom: 5
Educational Qualification	Chi-square value: 4.23 P-value: 0.37 Degrees of Freedom: 4	Chi-square value: 1.49 P-value: 0.82 Degrees of Freedom: 4	Chi-square value: 10.69 P-value: 0.03 Degrees of Freedom: 4	Chi-square value: 0.93 P-value: 0.92 Degrees of Freedom: 4	Chi-square value: 3.04 P-value: 0.55 Degrees of Freedom: 4
Language	Chi-square value: 0.21 P-value: 0.01 Degrees of Freedom: 1	Chi-square value: 0.38 P-value: 0.53 Degrees of Freedom: 1	Chi-square value: 1.42 P-value: 0.23 Degrees of Freedom: 1	Chi-square value: 3.18 P-value: 0.07 Degrees of Freedom: 1	Chi-square value: 1.87 P-value: 0.17 Degrees of Freedom: 1

In the interpretation of the Chi-square Test of Independence the P-value was observed in which if the P-value would be less than $P < 0.05$, then there is a statistical association between the variables (Brian, 2023). From the Table 7 above, the Chi-square Test of Independence shows that there is an association between ‘Age-Groups’, ‘Abbreviations’ $X^2(5, N = 406) = 14.68, p = .01$ and ‘Stickers’ $X^2(5, N = 406) = 4.82, p = .04$. The test also shows that there is an association between ‘Educational Qualification’ and ‘Mixed Languages’, $X^2(4, N = 406) = 10.69, p = .03$. There is also an association observed between ‘Language’ and ‘Emoticons/Emoji’, $X^2(1, N = 406) = .21, p = .01$. However, the rest of the results showed that there was no association between ‘Demographic groups’ and ‘Social media Language’.

3.3.2 Personality vs. social media language

Table 8: The Chi-square Test of Independence between ‘Personality’ and ‘Social media Languages’

	Social media languages				
Personality	Emoticons/Emoji	Abbreviations	Mixed Languages	Stickers	Gifs
Openness	Chi-square value: 6.55 P-value: 0.16 Degrees of Freedom: 4	Chi-square value: 6.47 P-value: 0.16 Degrees of Freedom: 4	Chi-square value: 2.88 P-value: 0.57 Degrees of Freedom: 4	Chi-square value: 4.89 P-value: 0.29 Degrees of Freedom: 4	Chi-square value: 2.75 P-value: 0.59 Degrees of Freedom: 4
Conscientiousness	Chi-square value: 3.61 P-value: 0.46 Degrees of Freedom: 4	Chi-square value: 1.08 P-value: 0.89 Degrees of Freedom: 4	Chi-square value: 3.32 P-value: 0.50 Degrees of Freedom: 4	Chi-square value: 1.76 P-value: 0.78 Degrees of Freedom: 4	Chi-square value: 4.08 P-value: 0.39 Degrees of Freedom: 4
Extroversion	Chi-square value: 5.13 P-value: 0.27 Degrees of Freedom: 4	Chi-square value: 2.95 P-value: 0.56 Degrees of Freedom: 4	Chi-square value: 1.72 P-value: 0.78 Degrees of Freedom: 4	Chi-square value: 3.16 P-value: 0.53 Degrees of Freedom: 4	Chi-square value: 10.43 P-value: 0.03 Degrees of Freedom: 4
Agreeableness	Chi-square value: 0.64 P-value: 0.95 Degrees of Freedom: 4	Chi-square value: 4.14 P-value: 0.39 Degrees of Freedom: 4	Chi-square value: 3.24 P-value: 0.51 Degrees of Freedom: 4	Chi-square value: 1.56 P-value: 0.81 Degrees of Freedom: 4	Chi-square value: 3.92 P-value: 0.41 Degrees of Freedom: 4
Neuroticism	Chi-square value: 6.65 P-value: 0.15 Degrees of Freedom: 4	Chi-square value: 1.29 P-value: 0.86 Degrees of Freedom: 4	Chi-square value: 3.10 P-value: 0.54 Degrees of Freedom: 4	Chi-square value: 3.50 P-value: 0.47 Degrees of Freedom: 4	Chi-square value: 12.64 P-value: 0.01 Degrees of Freedom: 4

From the Table 8 above, the Chi-square Test of Independence shows that there is an association between ‘Extroversion’ and ‘Gifs’, $X^2(4, N = 406) = 10.43, p = .03$. The test also shows that there is an association between ‘Neuroticism’ and ‘Gifs’, $X^2(5, N = 406) = 12.64, p = .01$. However, the rest of the results show that there is no association between ‘Personality’ and ‘Social media Language’. Overall, the Chi-square Test of Independence revealed that besides few variables, there is no association between ‘Demographic Groups’ and ‘Social media Language’, and there is no association between ‘Personality’ and ‘Social media Language’.

4. Discussion

The two hypotheses for this paper relate to understanding the impacts of demographic groups and personality on social media language, it was necessary to prove one of the hypotheses correct by supporting evidence. Hence, the data analysis supports the alternate hypothesis which states that demographic groups and personality does impact the use of social media language, this answers the question. The study also demonstrates a strong positive correlation between demographic groups and personality with social media language.

Therefore, in line with the alternate hypothesis the correlation coefficient values of ‘Demography groups’ and ‘Social media Language’ showed that there is a high external validity in terms of the overall population validity, hence the findings can be generalised in real-life settings of the social media user population. However, due to the correlation between the input and output variables such as observed between: ‘Age Groups’ and ‘Social media Language’, ‘Educational Qualification’ and ‘Language’, it can be deduced that there is also a high internal validity. A high internal validity means a causal relationship between ‘Demography groups’ and ‘Social

media Language' further meaning that demography influences the user of social media language. This finding seems to agree with the findings of Veenstra et al (2017) and Abbasova (2019) in which it was found that 60% of their participants preferred English for communication on social media because it was the quickest way to type responses as compared to their own native language. However, a study by Mubarak (2016) showed that people do communicate in their native language but occasionally use English words alongside their native words, or sometimes they type native words in English. Furthermore, the use of abbreviations was also common among the social media users because it enabled them to express their feelings and emotions in text messaging. The similarity between this study and Roni et al (2019) is that both have identified, mixed-languages and abbreviations as common trend among social media users. However, the results also suggests that as age-groups increase so does the use of social media language decreases, but the use of emoticons/emoji, abbreviations and mixed-language remains prevalent, furthermore the data also reveals that the use of mixed-language only remains prevalent up to the age-group 45-54, whereas emoticons/emoji and abbreviations continue to prevail up to the age-group 65+. This finding also agrees with the study of Christa et al (2020) in which 80% of their data collected show that people use abbreviate words like 'But' to 'Bt', 'You' to 'U', 'Text' to 'Txt' etc. for quick and better communication and for quick response.

The analysis of 'Personality' and 'Social media Language' data in this study suggests a high external validity, hence this finding can be generalised to real life settings of social media user population. However, due to some correlation between the input and output variables such as between 'Openness', 'Agreeableness' and 'Social media Language', it can be deduced that there is also a high internal validity. A high internal validity means that there is causal relationship between 'Personality' and 'Social media Language' further meaning that personality does influence social media language. This finding seems to agree with a study from Teresa et al (2009) and Thomas et al (2020), these authors have studied and shown some link between demography, personality and social media, but the focus was only on extraversion when it came to personality. Whilst it could be that extraversion does impact the use of social media as even agreed by Andreassen et al (2012) and Wilson et al (2010), however the results share a new insight between personality and social media language that it is not only extraversion/introversion, but also openness, conscientiousness, agreeableness and neurotic traits which also impacts social media language (Samuel et al, 2003).

5. Conclusion

The study aimed to investigate 'Demographic groups' and 'Personality' impact on social media language which will lead to further research on the development of a computer-sophisticated tool to help identify online identity. The findings of the online survey obtained through statistical analyses and tests provide enough evidence to support the alternate hypothesis which states that 'Demographic groups and Personality impacts the use of social media language which can be used to represent 'identity' online.'. This conclusion was achieved by breaking down demographic groups and personality into independent variables and comparing each of their correlation values to understand their cause-and-effect relation to social media language. These findings will be used further in the research project by focusing on the use of social media language for online identity management in textual conversations. The outcome of the further research will be compared to the phase one data to obtain demographic groups and personality data for the respective participants in the mock conversations, this can also be used to formulate a theory and communication model headed in the path of social media communication using textual conversations and online identity.

References

- Abbasova, M. (2019). Language of Social Media: An investigation of the changes that social media has imposed on language use. *9th International Conference on Education, Language and Literature*. Tbilisi, Georgia, pp. 309-313.
- Asenahabi, B. and Ikoha, P. (2023). Scientific Research Sample Size Determination. *The International Journal of Science and Technoledge*, vol. 11, pp. 08-12.
- Andreassen, S., Torsheim, T., Brunborg, G. and Pallesen, S. (2012). Development of a Facebook Addiction Scale. *Psychological Reports*, vol. 110, pp. 501-517
- Brian, B. (2023, March 05). P-Value: What It Is, How to Calculate It, and Why It Matters [Online]. Available from: <https://www.investopedia.com/terms/p/p-value.asp#:~:text=Key%20Takeaways,significance%20of%20the%20observed%20difference>.
- Christa, I. and Mercy, O. (2020). Language use in social media and natural language. *International Journal of Humanities and Social Science*, vol. 10, pp. 1-09.
- Divakar, S. and Nanjundeswaraswamy, T. (2021). Determination of Sample Size and Sampling Methods in Applied Research. *Proceedings on Engineering Sciences*, vol. 03, pp. 25-32.

- Gibbs, J., Heino, R. and Ellison, N. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, vol. 11, pp. 415-441.
- Hazra, A. (2017). Using the confidence interval confidently. *Journal of Thoracic Disease*, vol. 9, pp. 4125–4130.
- Heino, D., Ellison, B. and Gibbs, J. (2006). Self-presentation in online personals: The role of anticipated future interaction, self-disclosure, and perceived success in Internet dating. *Communication Research*, vol. 33, pp. 152-177.
- JISC, (2023, October 10). Online Surveys (formerly BOS) [Online]. Available from: <https://www.onlinesurveys.ac.uk/>
- Jain, R. (2020, May 31). Correlation between Categorical Variables [Online]. Available from: <https://medium.com/@ritesh.110587/correlation-between-categorical-variables-63f6bd9bf2f7>
- Kemp, S. (2023, January 26). Digital 2023: Global Overview Report [Online]. Available from: <https://datareportal.com/reports/digital-2023-global-overview-report#:~:text=There%20are%20now%204.76%20billion,growth%20of%20just%203%20percen>
- Mccabe, D., Yurchisin, J. and Watchravesringkan. (2005). An exploration of identity re-creation in the context of Internet dating. *Social behaviour and Personality*, vol. 33, pp. 735-750.
- Mubarak, A. (2016). Learning English as a second language through social media: Saudi Arabian Tertiary context. *International Journal of Linguistics*, vol. 8, pp. 112-119.
- Pietro, M. (2020, June 09). Text Analysis & feature engineering with NLP [Online]. Available from: <https://medium.com/towards-data-science/text-analysis-feature-engineering-with-nlp-502d6ea9225d>
- Roni, R. and Perdhani, W. (2019). Generational differences related to linguistic and discourse features of WhatsApp users of texting. *First International Conference on Advances in Education and language*, Malang, Indonesia, pp.1-09.
- Samuel, G., Rentfrow, P., Swan, W. (2003). A very brief measure of the Big-Five personality domains. *Methods in Ecology and Evolution*, vol. 27, pp. 504- 528.
- Solent, (2023, April 12). Solent Informed Consent [Online]. Available from: <https://www.solent.ac.uk/research-innovation-enterprise/documents/informed-consent-form.pdf>
- Statista, (2023, March 21). Number of internet and social media users worldwide [Online]. Available from: <https://www.statista.com/statistics/617136/digital-population-worldwide/>
- SurveyCircle, (2023, April 14). Find Survey Participants with Survey Circle [Online]. Available from: <https://www.surveycircle.com/en/>
- SurveySwap, (2023, April 14). Find survey respondents now [Online]. Available from: <https://surveyswap.io/>
- Teresa, S., Amber, H., and Homero, Z. (2009). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behaviour*, vol. 26, pp. 247-253.
- Thomas, B., Hinds, J. and Joinson, A. (2016). How extraversion is related to social media use. *Personality and Individual differences*, vol. 164, pp. 01-11.
- Veenstra, A. and Hossain, M. (2017). The uses and gratifications of language in social media among the Indian subcontinental diaspora in the US. *Journal of Intercultural communication research*, vol. 46, pp. 478-496.
- Walther, J. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, vol. 23, pp. 2538-2557.
- Warburton, S. and Hatzipanagos, S. (2015). *Digital Identity and Social-Media*. 1st ed. Pennsylvania: Information Science Reference.
- Wilson, K., Fornasier, S and White, K. (2010). Psychological predictors of young adults' use of social networking sites. *Cyberpsychology, behaviour and social networking*, vol. 13, pp. 173-177.