

# The Devil is in The Details: Measuring Sensory Processing Sensitivity Using Natural Language Processing

Lingyao (Ivy) Yuan  
Iowa State University  
[lyuan@iastate.edu](mailto:lyuan@iastate.edu)

Wenli Zhang  
Iowa State University  
[wzhang@iastate.edu](mailto:wzhang@iastate.edu)

Kevin P. Scheibe  
Iowa State University  
[kscheib@iastate.edu](mailto:kscheib@iastate.edu)

## Abstract

*Personality traits play a strong role in our perceptions, attitudes, and decision-making behaviors in our daily lives, including our choices of words and writing patterns. While prior Information Systems (IS) research on personality typically used the Big Five personality traits as a theoretical framework, we look into measuring a comparatively new inherent personality trait, sensory processing sensitivity, using natural language processing. We collect data on twenty general essay questions from along with self-reported sensory processing sensitivity survey questions from 241 participants. We categorize participants based on survey questions with multiple methods and derive different features from the textual data. Our results show almost perfect agreement among the different methods categorizing a highly sensitive person versus a non-highly sensitive person. The initial analysis demonstrates that certain features can be of great potential in measuring sensory processing sensitivity in written text.*

## 1. Introduction

Undoubtedly, personality influences an individual's behavior by capturing the holistic picture of the "whole, integrated, coherent unique individual" [13]. Genetic variables set the main tone of one's personality. However, personality does have a certain elasticity, meaning it can be changed by the environment, parenting, and social factors, especially at a young age [25]. Different traits make up personalities. Among different trait taxonomy theories, the Big Five personality traits model has been widely adopted in research areas such as marketing, management, sociology, and psychology to study human behavior and decision making. Organizations have commonly used this model as a way to understand employees and customers better.

The Big Five model identifies five broad personality dimensions: agreeableness, conscientiousness, extroversion, neuroticism, and

openness [33, 38]. Since these five dimensions influence human behaviors, we can "peek" into a person's personality by observing one's behavior. Instead of using the traditional self-reported surveys to measure personality, prior research has found that personality decides how we write. Therefore, much research has been done to infer personality from word choice and writing habits [10, 11, 28, 31, 37]. With the help of machine learning and deep learning methods, particularly natural language processing techniques, we can derive a personality profile from written text even in real-time. For instance, prior research has developed personality-based deep learning models to detect online hate speech [28]. Researchers also used machine learning to derive personality traits and observe the users' acceptance of word of mouth [3].

The Big Five personality model has been well established and widely adopted. The five dimensions describe an individual's personality at a very high cognitive level. This study is interested in a more fundamental personality trait, Sensory Processing Sensitivity (SPS). SPS is an innate personality trait associated with individuals' sensitivity to environmental and social stimuli [7, 8]. SPS leads people to be more prone to receive and respond to stimuli, especially social stimuli. We believe understanding inherent reactions and acceptance of sensory stimuli have different implications than the Big Five. From a theoretical perspective, SPS helps us understand human personalities that are not easily influenced by external factors. From a practical perspective, in the information age, sensory stimuli are delivered at high speed, in large amounts, and at a low cost, especially in the virtual context. Our lives are easily filled with sensory stimuli in a wide range of formats. However, people react to stimuli differently. Some may find the stimuli overwhelming, while others can handle them perfectly at ease.

On the one hand, people with high levels of SPS may generate different behaviors than those without. For example, sensory-sensitive people would behave differently to extreme comments or content on social media. On the other hand, people with high levels of SPS may evoke others through their words, drawing, or other creative works due to their strong sense of the

word. Those creative expressions can influence others' behaviors at the individual level or even the social level. Thus, people with high levels of SPS becoming social influencers.

Based on the level of SPS, researchers have categorized people into different groups, Highly Sensitive People (HSP) and non-Highly Sensitive People (non-HSP). Capturing such an innate personality trait has great potential for both research and practice. Thus, we seek to answer the question of whether people with different levels of SPS exhibit different writing patterns. To answer this question, we explore whether we can differentiate SPS people through their writing patterns. Specifically, this study aims to answer the following research questions:

**RQ:** *Can we categorize people into HSP and non-HSP based on writing patterns?*

We collected textual data from essay questions and conducted an initial analysis using natural language processing techniques to answer this question. We also used existing survey items to measure individual SPS levels. We categorized participants into two groups, HSP and non-HSP, using clustering and  $n$ -th percentile techniques. At the current stage, we extracted distinguishable features from textual data for categorizing SPS and non-HSP groups.

We contribute to the limited research on this topic with the hope of attracting greater interest from fellow researchers. From a practical perspective, the current design of the survey measuring SPS (27 items) is not convenient or possible to be administrated by companies. We plan to continue this research to establish an efficient model to measure SPS from written text. Our study will provide a practical way for companies to understand their employees and customers better.

## 2. Theoretical Background

### 2.1. Sensory Processing Sensitivity

We adopt the original definition of Sensory Processing Sensitivity (SPS). It is “a basic temperament/personality trait categorized by sensitivity to both internal and external stimuli, including social and emotional cues” [2, p. 1]. We separate SPS into three parts: sensory, processing, and sensitivity to better explain the concept. Sensory processing indicates the process human brains undertake in handling stimuli received by sensory organs. Sensory stimuli can be received through the skin (touch), ears (hearing), eyes (vision), mouth (taste), and nose (smell). Processing captures the

activity where the social cues received from sensory organs are reorganized, converted, and transformed into information or environmental/social cues. Sensitivity, the most important part of the definition of SPS, describes the levels of arousal and responsiveness, both positive and negative, in a person's brain upon receiving the sensory information and social cues. People with high levels of SPS may find certain smells bothersome, while others with lower levels of SPS may barely notice. People with high levels of SPS may notice micro facial expressions from strangers and thus, feel overwhelmed with social cues upon entering a new environment. In short, SPS captures differences in the way the brain processes sensory and social cues.

Prior research has categorized people with a high level of SPS into HSP. However, people not categorized as HSP are not labeled as low sensitive people, but simply as non-HSP [7]. SPS has been found to associate with other types of personality traits closely. For instance, HSP are predominantly introverted. HSP is also associated with neuroticism. Both are dimensions of Big Five personality traits. Despite the correlation with other personality traits, especially Big Five, prior research has established that SPS is a temperament trait of its own [5, 7].

High SPS means more sensory information in more vibrant colors flood the brain. Common details that usually go unnoticed may “jump out” to HSP. With more sensory information and social cues entering the brain simultaneously, more time and cognitive resources must be allocated to process all the stimuli. For HSP at a young age, this can be overstimulating, overwhelming, and even stressful. Researchers have called for parents to be more attentive and aware of their highly sensitive children coping with the environment [6]. On the other hand, highly sensitive children appear to be more attentive to details, have a great depth of processing, are more cognizant of subtleties in the environment, and are capable of making better long-term judgments [5, 8, 7, 6].

HSP may be highly sensitive to environmental cues and even hyper- or hypo-responsiveness to stimuli. This does not mean that HSP are cognitively atypical or suffer from neurological disorders. HSP do not bear structural differences in their brain from others. By integrating multiple fMRI studies, Acevedo, et al. [2] found that HSP are neurologically different from clinical disorders of Autism Spectrum Disorder (ASD), Schizophrenia, and Post-Traumatic Stress Disorder.

Similar to other innate and hereditary [7, 29] personality traits, such as the Big Five, SPS lead to differences in behavior, judgment, and perception.

Even though psychologists and neurologists have studied SPS and HSP, research on this topic is scarce in information systems. Yuan and Barlow (2021) examined the impact of SPS bidding decisions towards humanized products in online auctions. They found that SPS has a positive effect on perceived anthropomorphism, which strongly influences willingness to pay. In other words, people with high levels of SPS are more likely to anthropomorphize non-human objects and treat them like other humans [43]. In their study, SPS was measured using a self-report survey and was treated as a continuous measurement.

## 2.2. Natural Language Processing

Natural languages have evolved in humans through use without conscious premeditation [16]. Natural Language Processing (NLP) is a research area that explores how natural language can be understood and manipulated by machines to perform desired tasks [14]. The foundations of NLP lie in linguistics, psychology, computer and information sciences, and artificial intelligence [14].

NLP researchers program algorithms to process and analyze large amounts of natural language data that would otherwise be inaccessible. Moreover, NLP can help resolve ambiguity in natural language and structure highly unstructured data for many downstream applications, which includes several fields of study, such as speech recognition [22, 42], machine translation [9, 40], natural language understanding [39, 18], and natural language generation [20]. In the last decade, pattern recognition and deep learning-based approaches became mainstream for their state-of-the-art performance in many NLP tasks [17, 34].

With the increasing development of NLP, more and more researchers attempt to use NLP techniques to understand people's emotional states and personalities [19, 21, 24, 41]. The underlying theory is that the ways people use words and language can reflect basic psychological processes, including clues to their thoughts, feelings, perceptions, and personalities [4].

A growing number of studies show that we can use reliable language patterns for psychological states and personality predictions. For example, [15] discovered that when facing a collective upheaval (e.g., September 11) or a local disaster, people increase their use of first person plural pronouns and decrease the use of first person singular. [32] show that function words (processed differently in the brain compared with nouns and regular verbs) can mirror people's psychological states. When depressed or

emotionally vulnerable, individuals exhibit increases in pronouns (especially first person singular), drops in articles, and increases in their use of present tense auxiliary verbs. Another example is IBM Watson Personality Insights [1]. The service uses linguistic analytics to infer individuals' intrinsic personality characteristics, including Big Five traits [1], needs, and values, from digital communications such as email, text messages, tweets, and forum posts.

## 3. Data Collection

### 3.1. Participants

Two hundred and forty-one undergraduate and graduate students from a large mid-western public university in the US participated in this study. Out of the 241 participants, 155 were male (64%). The average age of all participants was 21.9 years. On average, it took subjects 38 minutes to finish the data collection. All participants received extra credit for their participation.

### 3.2. Measurements

An online survey was delivered via Qualtrics. In the survey, participants answered demographic questions, sensory processing sensitivity measurements, and 20 essay questions randomly selected from a total essay question bank of 42 essay questions.

To measure *Sensory Processing Sensitivity (SPS)*, we adopted the 27-item highly sensitive person scale [8], the standard measure of SPS in adults.

All essay questions covered general topics, such as weather, food, childhood memory, etc. The essay questions were designed to collect textual input from participants.

<i>Per essay question</i>	<i>number of sentences</i>	<i>number of words</i>	<i>number of characters</i>
<b>Mean</b>	261.60	3670.23	15324.79
<b>Standard Error</b>	5.97	104.92	420.63
<b>Median</b>	256	3643	15223
<b>Mode</b>	268	3045	~
<b>Standard Deviation</b>	39.16	688.03	2758.29
<b>Sample Variance</b>	1533.62	473392.61	7608199.64
<b>Kurtosis</b>	0.31	0.0008	-0.12
<b>Skewness</b>	0.5762	0.0457	0.0009

<b>Range</b>	177	3277	13066
<b>Minimum</b>	194	2127	9113
<b>Maximum</b>	371	5404	22179
<b>Sum</b>	11249	157820	658966

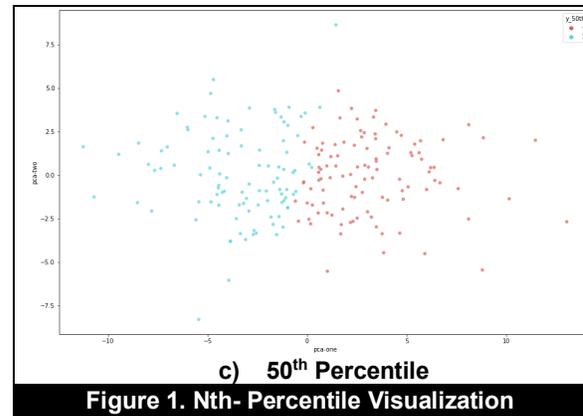
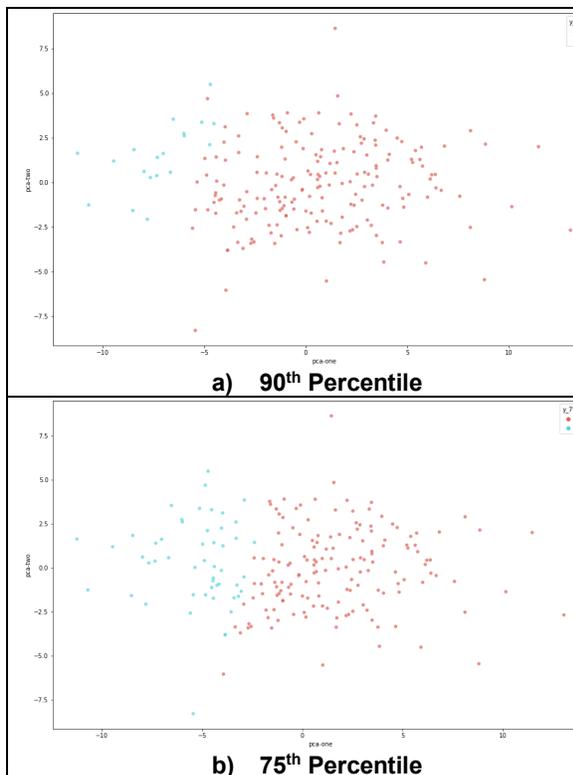
## 4. Analysis and Results

### 4.1. Clustering vs. $n$ -th Percentile

Following [8], we took multiple explorative approaches to categorize participants into the HSP and non-HSP groups.

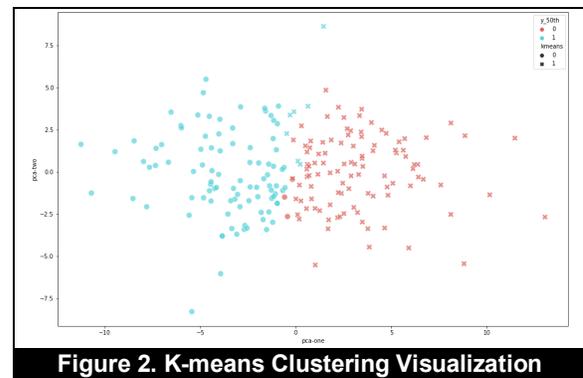
**4.1.1. 90th, 75th, and 50th percentiles.** A percentile is a score below which a given percentage of SPS scores in its frequency distribution falls. For example, the 50th percentile is the score below which 50% of the scores in the distribution are found. Prior research has consistently found that about 20-30% of the population are HSP [8, 7].

In conjunction with principal component analysis (PCA), we visualized the high dimensional survey data of the 27 SPS measurements. In Figure 1, we can easily observe that the two groups (HSP and non-HSP) are clearly divided into 50th percentile, 75th percentile, and 90th percentile.



**4.1.2. K-means Clustering.** The K-means algorithm clusters data by separating data records in 2 groups (i.e., HSP vs. non-HSP) of equal variance, minimizing the criterion (within-cluster sum-of-squares). Formally, the k-means algorithm divides a set of  $n$  samples into  $K$  disjoint clusters  $C$ , each described by the mean  $\mathbf{u}_j$  (aka cluster centroids) of the samples in the cluster. The K-means algorithm aims to choose centroids that minimize the within-cluster sum-of-squares criterion:

$$\sum_{i=0}^n \min_{\mathbf{u}_j \in C} (\|x_i - \mathbf{u}_j\|^2)$$



**4.1.3 Inter-rater reliability measurement.** Inter-rater reliability is the extent to which two grouping methods agree. It addresses the issue of consistency of the implementation of the grouping system. We use Cohen's kappa coefficient to measure inter-rater reliability for the  $n$ -th percentiles and k-means clustering methods.

Note: we have 241 respondents, and 36 of them are non-native English speakers. We removed the non-native English speakers for better robustness.

Table 2 shows that the number of observed agreements is 197, accounting for 94.71% of all observations. The number of agreements expected by

chance is 104.1 (50.06% of the observations). Cohen’s kappa is 0.894 (SE=0.031), indicating an almost perfect agreement (between 0.81 and 1).

		k-means		Total
		HSP	non-HSP	
50 <sup>th</sup> Percentile	HSP	103	3	106
	non-HSP	8	94	102
Total		111	97	205

## 4.2 Text Preprocessing and Feature Extraction

In NLP, text preprocessing can reduce lexical noise and maintain a reasonable feature size for the textual data. The text preprocessing steps adopted in this study include:

(a) Tokenization: we apply English-specific tokenization (note, we only include English native speakers in this study). Abbreviations (e.g., U.S.A.), words with optional internal hyphens (e.g., no-questions-asked), numbers, currency, and percentages (e.g., \$12.40), emoticons (e.g., :) are separated as single tokens.

(b) Lowercase conversion: Tokens are transformed to lowercase characters, allowing a reduction in the number of tokens while retaining useful information, e.g., “Good” will be identified as “good.”

(c) Stemming: Words with the same stem are families of derivationally related words with similar meanings. Stemming reduces inflectional forms and derivational forms of a word to a common base form. In this work, tokens are stemmed by applying Porter stemming algorithm<sup>1</sup>. It is a process for removing the more common morphological and inflectional endings from words in English, e.g., “run” and “running” are recognized as one token.

(d) Negation: Consider two different text sentences like “I like potato chips” and “I don’t like potato chips.” To make machine learning algorithms differentiate between these two sentences, we need to analyze negation. We prefix all words between the negation word (“not,” “didn’t,” etc.) and the first punctuation sign following the negation word with the tag (NOT\_), e.g., “I don’t like potato chips” will be identified as “I don’t NOT\_like NOT\_potato NOT\_chips.”

Feature extraction is used to transform textual data into numerical vectors for classification using

Machine Learning techniques. Following the state-of-the-art techniques, we explore the use of the following feature groups and their combinations:

(a) Token-based features: Including (i) the number of punctuation signs, (ii) the ratio of characters in upper cases, (iii) the number of words with all characters in upper case, and (vi) the number of elongated words (e.g., “happppy”). Words with more characters in upper case, repeated punctuation signs, and elongated words are signs of strong emotions [26].

(b) Word n-grams ( $n \in [1,4]$ ) and character n-gram ( $n \in [3,5]$ ) features: Word n-grams are contiguous sequences of words from a given sequence of texts. Character n-grams are contiguous sequences of characters, which can help detect linguistic noise; especially, when  $n=1$ , the features are unigrams. A unigram is essentially an individual token in a text. When  $n=2$ , the features are bigrams. Bigrams are contiguous sequences of two items from a given sequence of texts. Bigrams can help with phrases, i.e., a group of words carrying a special meaning, e.g., “New York.” The feature space of bigrams tends to be sparse as compared to unigram features.

(c) Part-of-Speech (POS) analysis as the lexical feature including (i) the number of each POS type and (ii) POS trigram. Relevant and irrelevant texts tend to have very different lexical features. Figure 3 shows an example of a lexical feature with POS tags.

<b>Text:</b> I got an <b>asthma</b> attack.		
<b>Part-of-Speech tag:</b>		
<i>Tokens</i>	<i>Part-of-speech</i>	<i>Tags</i>
I	List item marker	LS
got	Verb, past tense	VBD
an	Determiner	DT
asthma	Noun, singular or mass	NN
attack	Noun, singular or mass	NN
<b>Feature extracted:</b>		
an_DT \ asthma_NN \ attack_NN		

**Figure 3. Lexical Feature (Part-of-Speech Tag)**

(d) Linguistic Inquiry and Word Count (LIWC) features [36]: LIWC is a text analysis process that counts words in psychologically meaningful categories. Empirical results using LIWC demonstrate its ability to detect meaning in various experimental settings, including showing attentional focus, emotionality, social relationships, thinking styles, and individual differences.

(e) Word and document embedding features: Embedding representation resulted in dense feature representations for text. We explore word and document-level embeddings for this study. (i) word

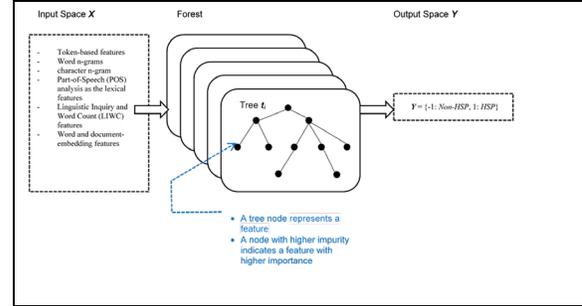
<sup>1</sup> <https://tartarus.org/martin/PorterStemmer/>

embeddings [30]: We train the embeddings using the entire training corpus, and we tune the number of dimensions using cross-validation. We experiment with both CBOW (continuous bag-of-words) and skip-gram. Three different methods are used to get the vector representations for documents using these word embeddings: summed vector of all the word embeddings, averaged vector for all the word embeddings, and a combination of average and standard deviation. (ii) In addition to word embeddings, vector representations were also generated for an entire document (i.e., document embedding) with distributed memory (DM) and distributed bag-of-words (DBOW) architectures [27].

### 4.3. Feature Importance Assessment

This study uses Random Forest to measure feature importance based on the decrease in average feature impurity. Using Random Forest for feature importance measuring has received attention in many domains [35]. The advantages of using forest rather than multiple linear regression methods are that (1) it allows nonlinearities and interactions among the data without explicitly identifying them [23]; and (2) it requires very little feature engineering and parameter tuning [12].

The Gini importance is adopted for this study [12]. It is operationalized as follows: a Random Forest is a classifier consisting of a collection of tree-structured classifiers  $\{h(x, \theta_k), k = 1, 2, 3, \dots\}$  where the  $\{\theta_k\}$  are independent and identically distributed random vectors and each tree  $h(x)$  casts a unit vote for the most popular class for input  $x$  [12]. Every node in  $h(x)$  is a condition on a single feature. In a two-way classification problem (i.e., HSP or non-HSP), every node in the tree  $h(x)$  splits the input data  $x$  into two so that data records with similar response values are in the same set. Gini importance measures how often a randomly chosen data record from the input  $x$  is incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. In a two-way classification problem, the Gini importance can be calculated as  $I_G(p) = 1 - \sum_{i=1}^{\{-1,1\}} p(i|t)^2$ , where  $p_i$  is the fraction of data records annotated with class  $i$  in the data set. Gini impurity ( $I_G$ ) reaches zero (the minimum) when all the data records in a tree node fall into a single class. Thus, when training a tree  $h(x)$ , we can calculate how much each feature decreases the weighted impurity in a tree. For a forest  $\{h(x, \theta_k), k = 1, 2, 3, \dots\}$ , the impurity decrease from each feature can be averaged, and the features are ranked according to Gini importance.



**Figure 4. Random Forest for HSP-related Feature Importance Assessment**

In this study, we have already identified the HSP (see section 4.1), and the Random Forest classification target is already known (i.e., HSP or non-HSP). The focus of this study is to measure feature importance based on Gini importance. The figure above (Figure 4) illustrates the use of Random Forest in this study.

**Table 3. Feature Group Importance**

Feature Set	Group	Feature Importance (Sum of Gini Index)
word_ngram (n=1)	FG 2	0.0980
char_ngram (n=5)	FG 2	0.0870
pos_ngram (n=3)	FG 3	0.0490
word_ngram (n=2)	FG 2	0.0484
skip_gram (sum)	FG 5	0.0440
LIWC (percept)	FG 4	0.0418
LIWC (drives)	FG 4	0.0369
Word_ngram (n=3, inside word boundaries)	FG 2	0.0265
LIWC (function words)	FG 4	0.0258
char_ngram (n=4, inside word boundaries)	FG 2	0.0158
LIWC (tone)	FG 4	0.0148
char_ngram (n=4)	FG 2	0.0139
LIWC (punctuation marks)	FG 4	0.0119
skip_gram (average)	FG 2	0.0118
# of words with all characters in upper case	FG 1	0.0004
<b>SUM</b>		<b>0.5260</b>

*FG = Feature Group*

*FG 1 = Token-based Feature*

*FG 2 = Word n-grams (n ∈ [1, 4]) and character n-gram (n ∈ [3, 5])*

*FG 3 = POS as lexical features*

*FG 4 = LIWC feature*

*FG 5 = Word and document embedding feature*

Table 3 summarizes the importance of different feature groups. The results show that, although the selected features span all the five feature groups, the vast majority are taken from the word and character n-grams (FG 2) features. Lexical features (FG 3) also

contribute a lot to distinguishing HSP and non-HSP. The LIWC (FG 4) features reveal subjects' thoughts, feelings, personality, and motivations. They show the potential to identify HSP and non-HSP people effectively. Also, the word embedding features (FG 5) add value to the feature set. The identified feature groups have an abundance pattern that can be analyzed and used to categorize HSP and non-HSP people using written data, which is easier, faster, and inexpensive to collect than survey data.

## 5. Discussion

Our results on categorizing participants into HSP and non-HSP yield almost perfect agreement among different approaches. For the NLP feature extraction, our results are still preliminary but show great promise. Out of the top 200 features from the feature importance selection process, some were particularly interesting and yield potential for further investigation. Those features can be categorized as follows:

- Features indicating the sense of self and the need to withdraw from others or the environment: “ppron” (LIWC feature: personal pronouns); space, I, me, “authentic,” “differ,” “difference,” and “special”
- Features emphasizing rational/analytical thinking: “analytic,” “number,” “function,” and “prep” (prepositions prep such as to, with, and above)
- Features emphasize responses: “affect,” “swear,” “very,” and “awesome”
- Features indicating strong will: “certain,” “absolutely,” and “focus future”
- Features indicating social tie or connection: “family,” “social,” and “football”
- Other features with potential implications: “bitcoin,” and “money”

Our sample size is comparatively small for studies using natural language processing. Therefore, we are collecting more data on various platforms (e.g., survey, chatbot, and social media).

### 5.1. Limitations and Future Research

As with all research, this study has limitations that should be considered. First, as aforementioned, our results and interpretation are limited by the current sample size. Through our current analysis, our goal is to see whether research on this topic has potential. Even though our results are promising, more data is needed. Overfitting is a potential issue. We conducted cross-validation, which provides robust metrics to measure the performance to control the over-fitting

issues. For future research, we plan to have several more rounds of data collection to get our sample size of at least 500 more participants and 10,000 documents for analysis.

Second, we used a survey method to collect our data. Text data were collected using essay questions. Individuals may have minor differences in their choice of words or written words answering survey questions versus communicating in conversations. We are interested in comparing data collected from essay composition with those collected from conversations. For future research, we plan to use a conversational chatbot to collect more data. We believe this data collection method can provide more realistic conversation scenarios. Besides, we believe answering essay questions is not as engaging and interesting as chatting with chatbots.

Third, our participants are undergraduate and graduate students. We believe SPS is an inherent characteristic applicable to everybody. Therefore, students as our sample should not bias our results, and the findings of this sample would not suffer from generalizability issues. For future research, however, we plan to collect data from a more general population in addition to students.

### 5.2. Implications

We believe understanding the written patterns of SPS has great potential to benefit both academia and practice. The concept of Sensory Processing Sensitivity was not proposed until 1997, and it has not been studied in the field of IS. As a unique inherent personal characteristic, it acknowledges how we receive, process, and respond to environmental cues differently. In the IS discipline, especially on topics involving using technology to create artificial cues in the virtual and digital environment, understanding SPS has a great potential to explain different behaviors and perceptions. In the age of artificial intelligence, our findings can be utilized by conversational AI in identifying the user personality through text-based conversations. A better understanding of the users, including their SPS, will generate more opportunities for companies to deliver improved customized and targeted content.

On the other hand, HSP may have undiscovered value to companies and society. Are there contain jobs more suitable for HSP? Are artists more likely to be HSP? Will HSP have a stronger impact as social influencers? The answers to those questions have great potential for practitioners and are worth pursuing. This study is just the beginning, and we hope it will attract more research on this topic.

The research on SPS also encourages discussions from both academia and practitioners on the ethical use of technology. For HSP, a virtual environment may pose greater risks. Prominent sensory stimuli may invoke severe negative responses from HSP, and such negative effects could be long-lasting. The first step towards protecting HSP from being overwhelmed or harmed by overstimulating virtual environments is identifying them. Companies also need to be mindful of collecting textual data from consumers, protecting consumers' privacy, and using consumer textual data ethically.

To HSP customers, our research suggests that they should be mindful of their sensitivity towards the stimuli. Companies have always used stimuli, especially visual stimuli, to attract consumers' attention and nudge their behavior in subtle ways. However, those subtle cues may have a different impact on HSP. Understanding oneself is a good start to mitigate against those types of manipulation.

## 6. Conclusion

Research on Sensory Processing Sensitivity, especially in the discipline of information systems, is still young. In this study, we collected text data and SPS survey measurements. We used two methods to categorize participants into HSP and non-HSP groups based on the survey measurements with an almost perfect agreement. We also analyzed the text data to extract important features from the text. Even though our results are preliminary, there is great promise for future research. More data collection is necessary, and we hope our study will attract more attention in the IS field.

## 7. References

- [1] IBM Cloud. (2020). "The science behind the service". Retrieved from <https://cloud.ibm.com/docs/services/personality-insights?topic=personality-insights-science>
- [2] B. Acevedo, E. Aron, S. Pospos and D. Jessen, "The functional highly sensitive brain: a review of the brain circuits underlying sensory processing sensitivity and seemingly related disorders", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373 (2018), pp. 20170161.
- [3] P. Adamopoulos, A. Ghose and V. Todri, "The impact of user personality traits on word of mouth: Text-mining social media platforms", *Information Systems Research*, 29 (2018), pp. 612-640.
- [4] S. Argamon, S. Dhawle, M. Koppel and J. W. Pennebaker, Lexical predictors of personality type, *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America*, 2005, pp. 1-16.
- [5] A. Aron, S. Ketay, T. Hedden, E. N. Aron, H. Rose Markus and J. D. Gabrieli, "Temperament trait of sensory processing sensitivity moderates cultural differences in neural response", *Social cognitive and affective neuroscience*, 5 (2010), pp. 219-226.
- [6] E. Aron, *The highly sensitive child: Helping our children thrive when the world overwhelms them*, Harmony, 2002.
- [7] E. Aron, *The highly sensitive person*, Kensington Publishing Corp., 2013.
- [8] E. N. Aron and A. Aron, "Sensory-processing sensitivity and its relation to introversion and emotionality", *Journal of personality and social psychology*, 73 (1997), pp. 345.
- [9] D. Bahdanau, K. Cho and Y. Bengio, "Neural machine translation by jointly learning to align and translate", *arXiv preprint arXiv:1409.0473* (2014).
- [10] C. P. Barlett and C. A. Anderson, "Direct and indirect relations between the Big 5 personality traits and aggressive and violent behavior", *Personality and individual differences*, 52 (2012), pp. 870-875.
- [11] M. R. Barrick and M. K. Mount, "The big five personality dimensions and job performance: a meta-analysis", *Personnel psychology*, 44 (1991), pp. 1-26.
- [12] L. Breiman, "Random forests", *Machine learning*, 45 (2001), pp. 5-32.
- [13] D. Cervone and L. A. Pervin, *Personality: Theory and research*, John Wiley & Sons, 2015.
- [14] G. G. Chowdhury, "Natural language processing", *Annual review of information science and technology*, 37 (2003), pp. 51-89.
- [15] M. A. Cohn, M. R. Mehl and J. W. Pennebaker, "Linguistic markers of psychological change surrounding September 11, 2001", *Psychological science*, 15 (2004), pp. 687-693.
- [16] D. Davidson and G. Harman, *Semantics of natural language*, Springer Science & Business Media, 2012.
- [17] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", *arXiv preprint arXiv:1810.04805* (2018).
- [18] L. Dong, N. Yang, W. Wang, F. Wei, X. Liu, Y. Wang, J. Gao, M. Zhou and H.-W. Hon, "Unified language model pre-training for natural

- language understanding and generation”, arXiv preprint arXiv:1905.03197 (2019).
- [19] L. A. Fast and D. C. Funder, “Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior”, *Journal of personality and social psychology*, 94 (2008), pp. 334.
- [20] A. Gatt and E. Kraemer, “Survey of the state of the art in natural language generation: Core tasks, applications and evaluation”, *Journal of Artificial Intelligence Research*, 61 (2018), pp. 65-170.
- [21] A. Gill, S. Nowson and J. Oberlander, What are they blogging about? Personality, topic and motivation in blogs, *Proceedings of the International AAAI Conference on Web and Social Media*, 2009.
- [22] A. Graves, A.-r. Mohamed and G. Hinton, Speech recognition with deep recurrent neural networks, 2013 IEEE international conference on acoustics, speech and signal processing, Ieee, 2013, pp. 6645-6649.
- [23] U. Grömping, “Variable importance assessment in regression: linear regression versus random forest”, *The American Statistician*, 63 (2009), pp. 308-319.
- [24] J. B. Hirsh and J. B. Peterson, “Personality and language use in self-narratives”, *Journal of research in personality*, 43 (2009), pp. 524-527.
- [25] O. P. John, R. W. Robins and L. A. Pervin, *Handbook of personality: Theory and research*, Guilford Press, 2010.
- [26] K. S. Jones and P. Willet, *Readings in information retrieval*. San Francisco, 1997, Morgan Kaufmann.
- [27] Q. Le and T. Mikolov, Distributed representations of sentences and documents, *International conference on machine learning*, PMLR, 2014, pp. 1188-1196.
- [28] K. Lee and S. Ram, “PERSONA: Personality-Based Deep Learning for Detecting Hate Speech”, (2020).
- [29] C. L. Licht, E. L. Mortensen and G. M. Knudsen, “Association between sensory processing sensitivity and the 5-HTTLPR Short/Short genotype”, *Biol. Psychiatry*, 69 (2011), pp. 152S-153S.
- [30] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, “Distributed representations of words and phrases and their compositionality”, arXiv preprint arXiv:1310.4546 (2013).
- [31] J. W. Pennebaker, M. E. Francis and R. J. Booth, “Linguistic inquiry and word count: LIWC 2001”, Mahway: Lawrence Erlbaum Associates, 71 (2001), pp. 2001.
- [32] J. W. Pennebaker and T. C. Lay, “Language use and personality during crises: Analyses of Mayor Rudolph Giuliani’s press conferences”, *Journal of Research in Personality*, 36 (2002), pp. 271-282.
- [33] R. A. Power and M. Pluess, “Heritability estimates of the Big Five personality traits based on common genetic variants”, *Translational psychiatry*, 5 (2015), pp. e604-e604.
- [34] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever, “Language models are unsupervised multitask learners”, *OpenAI blog*, 1 (2019), pp. 9.
- [35] C. Strobl, A.-L. Boulesteix, A. Zeileis and T. Hothorn, “Bias in random forest variable importance measures: Illustrations, sources and a solution”, *BMC bioinformatics*, 8 (2007), pp. 1-21.
- [36] C. Tang and L. Guo, “Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication”, *Marketing Letters*, 26 (2015), pp. 67-80.
- [37] Y. R. Tausczik and J. W. Pennebaker, “The psychological meaning of words: LIWC and computerized text analysis methods”, *Journal of language and social psychology*, 29 (2010), pp. 24-54.
- [38] E. C. Tupes and R. E. Christal, “Recurrent personality factors based on trait ratings”, *Journal of personality*, 60 (1992), pp. 225-251.
- [39] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy and S. R. Bowman, “GLUE: A multi-task benchmark and analysis platform for natural language understanding”, arXiv preprint arXiv:1804.07461 (2018).
- [40] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao and K. Macherey, “Google’s neural machine translation system: Bridging the gap between human and machine translation”, arXiv preprint arXiv:1609.08144 (2016).
- [41] T. Yarkoni, “Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers”, *Journal of research in personality*, 44 (2010), pp. 363-373.
- [42] D. Yu and L. Deng, *Automatic Speech Recognition*, Springer, 2016.
- [43] L. Yuan and J. Barlow, “Sensitive to the Digital Touch? Exploring Sensory Processing Sensitivity and Its Impact on Anthropomorphized Products in E-Commerce”, *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021, pp. 4104.