



Avi Verma
Dr. Kaustubh Supekar

Mental Health Outcome Prediction among Youth

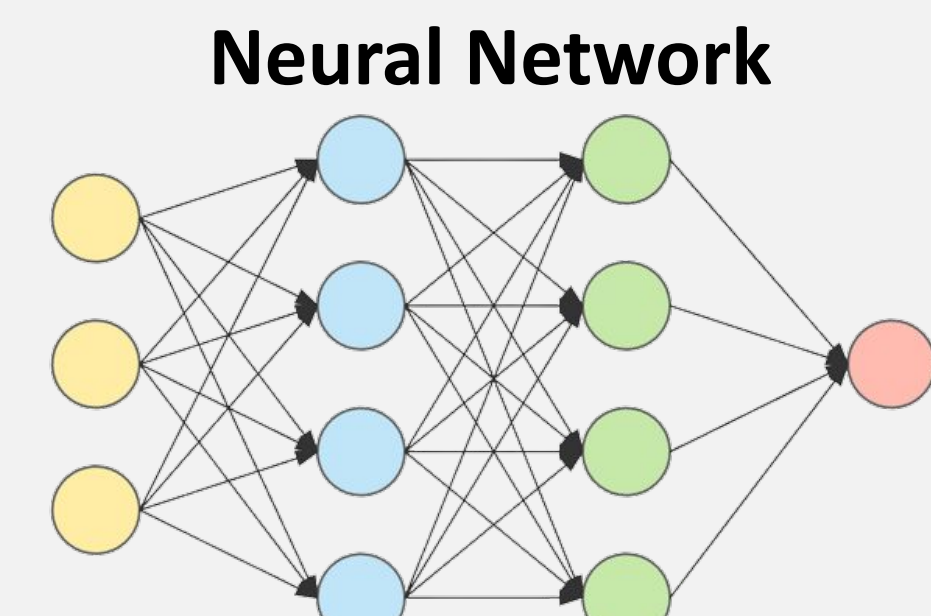
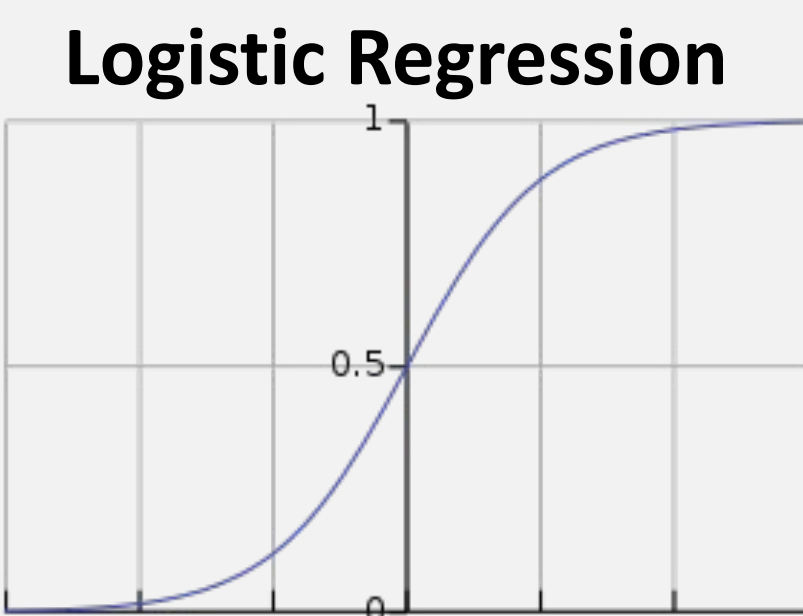
Comparing the Accuracy of Logistic Regression and Neural Network Models

Palo Alto High School
Stanford University



INTRODUCTION

- In 2019, **more than a third of high school students persistently felt sad or hopeless** (CDC, 2021)
- Traditional intervention is psychotherapy consultation, which tends to be held back by financial barriers and stigma (Rowan et al., 2014)
- Even **psychologists are only able to accurately diagnose 46% of depressive and 50% of anxiety disorder cases** (Al-Huthail, 2008)
- We can utilize **novel advances in statistics and machine learning** to help psychologists improve their diagnoses and proactively identify cases in urgent need of intervention
- These methods have been used successfully to predict other types of diagnoses in healthcare settings (Davenport & Kalakota, 2019)
- For this study, we compared the cross-test accuracies of two models in predicting anxiety and depression outcomes:



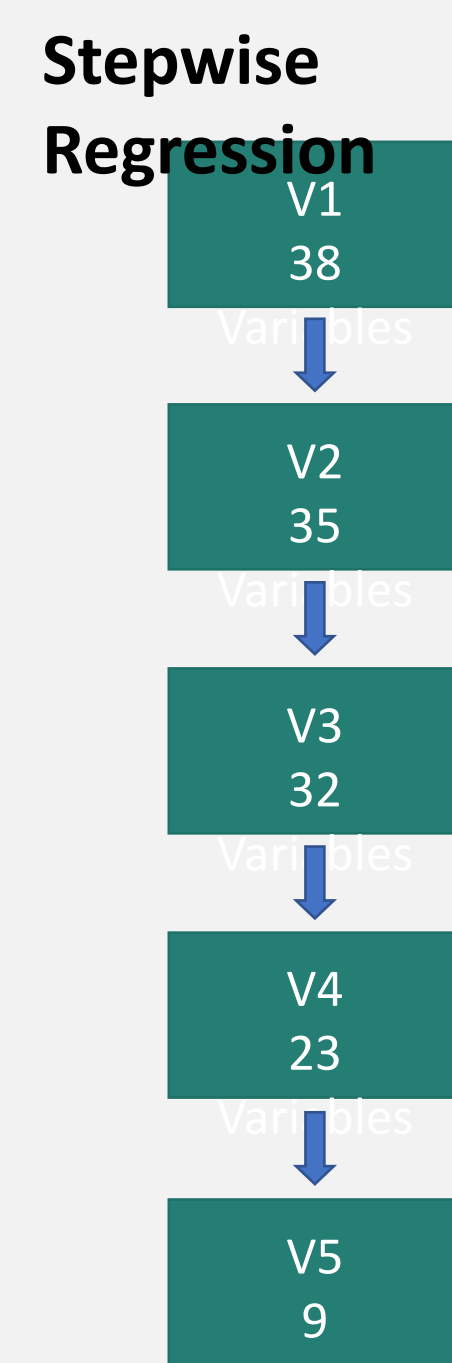
METHODOLOGY

1: Data Collection and Setup

- This study was conducted on aggregated, anonymized survey responses of 6.9 million patients in state hospitals in 2020 (Substance Abuse and Mental Health Services Administration's Mental Health Client-Level Data (MH-CLD))
- The data was cleaned and summarized in the Jupyter Notebooks environment using the pandas library (see Table 1 & 2)
- In the cleaned data for young adults, aged 15 to 24, 26.5% had anxiety and 35.7% had depression

2: Model Creation and Evaluation

- The logistic regression model was set up in Jupyter Notebooks and neural network in Google Colab with an "Adam" optimizer and "ReLU" activator
- Logistic regression was trained on a randomized three k-fold subset of the dataset, whereas the neural network was trained on 10% of the data and evaluated against the remaining 90%
- For logistic regression and neural network models, the first training variable set, V1, included 38 of 40 available variables
- **As we aimed to determine the model with the highest cross-test accuracy**, we stepwise removed variables with the least predictive correlation in sets V2 to V5
- All sets (V1 to V5) included the minimum of the following variables, as suggested by previous literature:
 - Race, gender, demographics, overall living situation, and stressor events (Afifi, 2007; Businelle et al., 2013; CDC, 2021; Koneru et al., 2007)



DATA AND RESULTS

Table 1. Summary of Raw Data for Most Correlated Variables

	Age	Education	Race	Gender	Mental Health Diagnosis 1	Mental Health Diagnosis 2	Mental Health Diagnosis 3	Marital Status	Employment
Count	$6.9 \cdot 10^6$	$6.9 \cdot 10^6$	$6.9 \cdot 10^6$	$6.9 \cdot 10^6$	$6.9 \cdot 10^6$	$6.9 \cdot 10^6$	$6.9 \cdot 10^6$	$6.9 \cdot 10^6$	$6.9 \cdot 10^6$
Min	-9	-9	-9	-9	-9	-9	-9	-9	-9
25%	3	-9	3	1	1	-9	-9	-9	-9
50%	7	-9	5	2	6	-9	-9	1	-9
75%	10	4	5	2	7	2	-9	1	4
Max	14	5	6	2	13	13	13	4	5

Table 2. Summary of Clean Data for Most Correlated Variables

	Age	Education	Race	Gender	Mental Health Diagnosis 1	Mental Health Diagnosis 2	Mental Health Diagnosis 3	Marital Status	Employment
Count	$3.8 \cdot 10^5$	$3.8 \cdot 10^5$	$3.8 \cdot 10^5$	$3.8 \cdot 10^5$	$3.8 \cdot 10^5$	$3.8 \cdot 10^5$	$3.8 \cdot 10^5$	$3.8 \cdot 10^5$	$3.8 \cdot 10^5$
Min	3	1	1	1	0	0	0	1	0
25%	3	3	5	1	2	0	0	1	0
50%	4	3	5	2	6	0	0	1	4
75%	5	4	5	2	7	2	0	1	5
Max	5	5	6	2	13	13	13	4	5

Figure 1. Logistic Regression v.s. Neural Network Accuracy for Anxiety

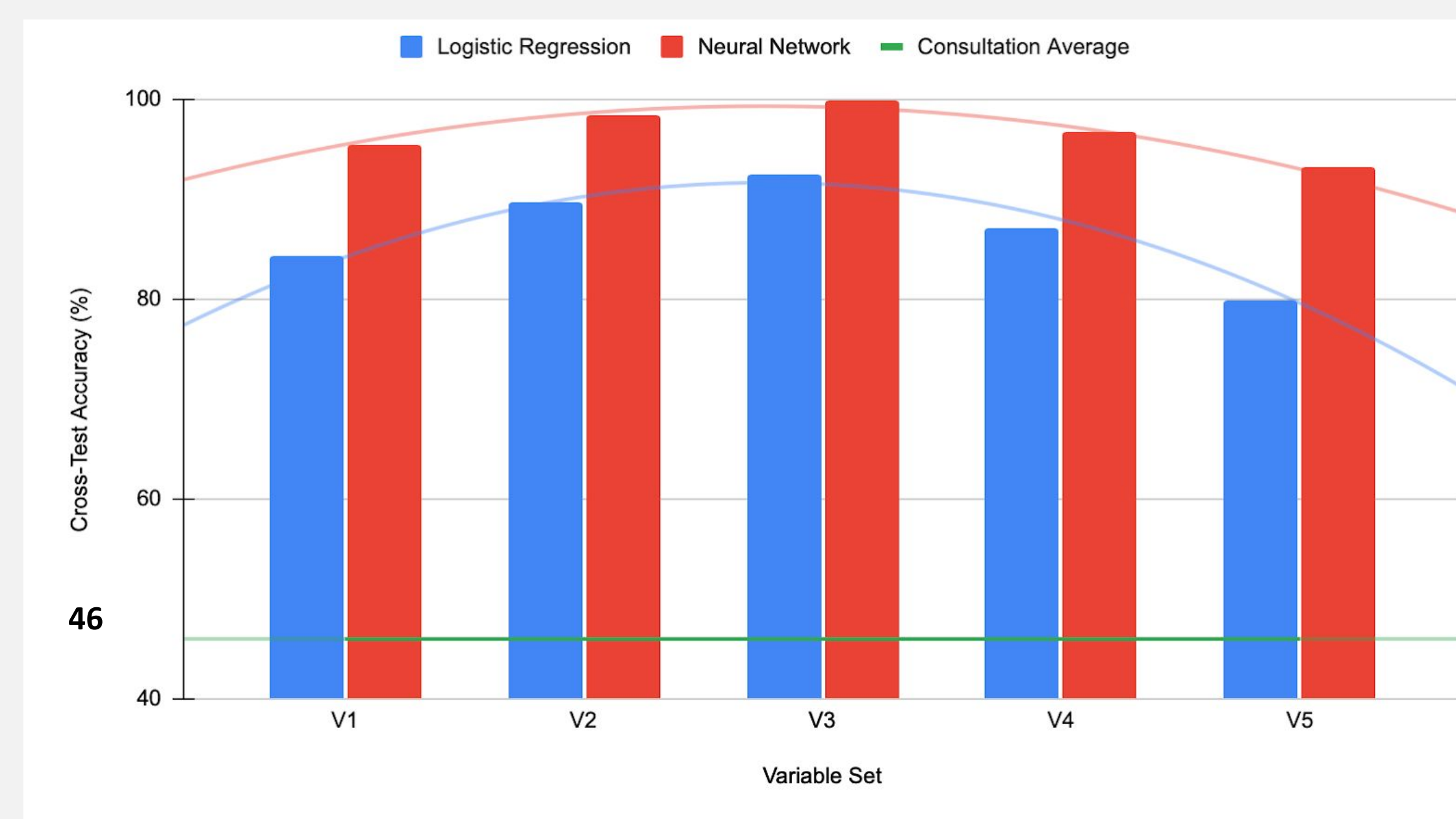
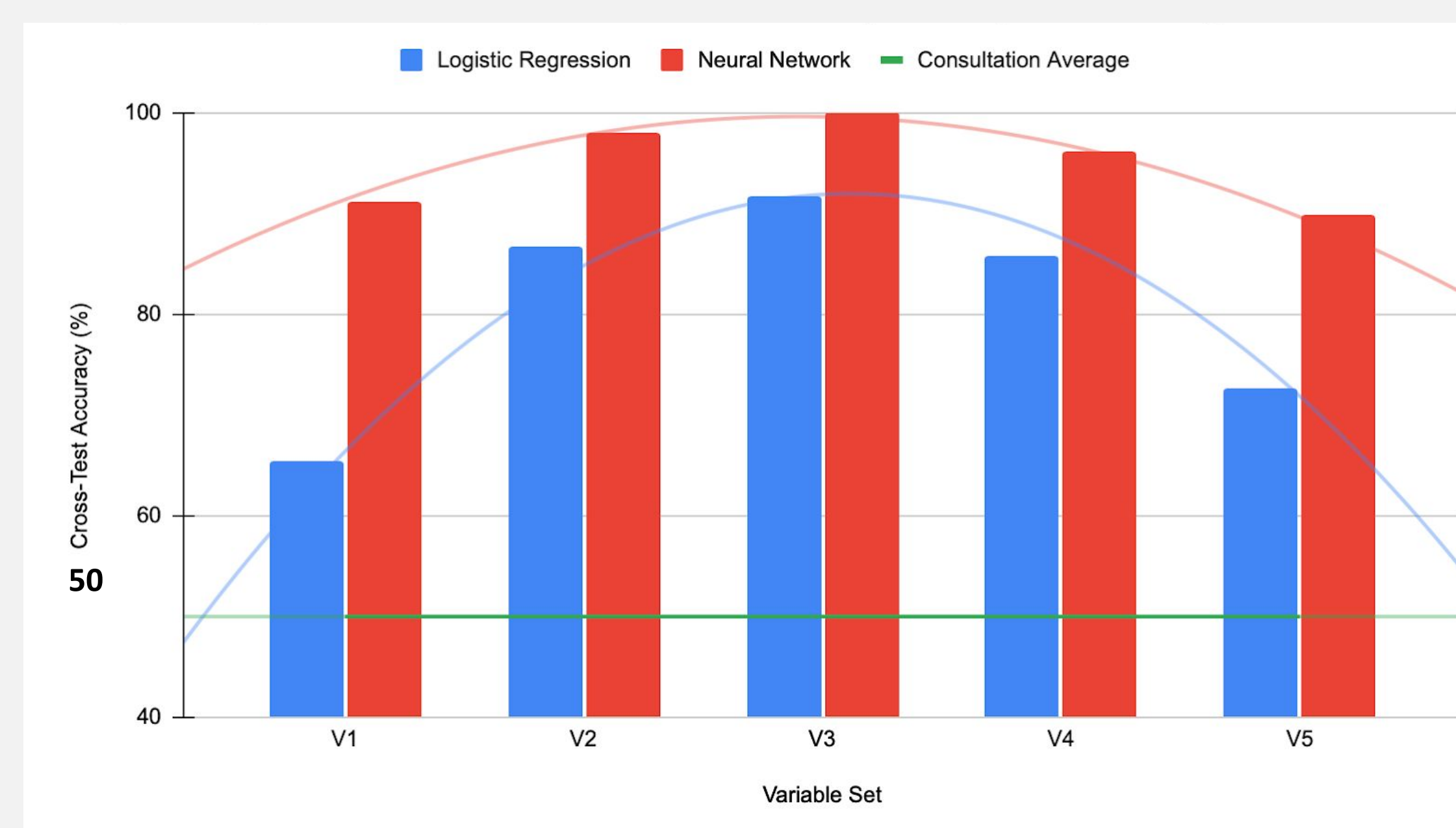


Figure 2. Logistic Regression vs. Neural Network Accuracy for Depression



CONCLUSIONS

- Both **logistic regression (65.5% to 92.5%) and neural network (90% to 99%) models are relatively accurate** in predicting anxiety and depression diagnoses and **tend to be more precise than the average therapist consultation (46% to 50%)**
- **Neural networks are superior to logistic regression** in all cases (see Figure 1 & 2), confirming previous studies that have shown that, unlike logistic regression, the weight systems of neural networks can address the negative effects of redundant variables while not falling susceptible to overfitting (McCaffrey, 2018)
- Neural network accuracy is optimized with **set V3 of 32 variables**
- Indicates that models must **retain specificity** for variables that are highly correlated with mental health outcomes (e.g. demographics, living situation) but **should not fit for inter-correlated** (e.g. race and ethnicity) and **insignificant** (e.g. state of collection) variables
- **This neural network model, with high testing accuracy, can be used by therapists alongside their consultation sessions** to help guide the mental health diagnoses of their patients

FUTURE RESEARCH

- Additional research is needed to analyze the impact of **other specialized variables** not tracked in our data (e.g. neighborhood and school location, exact number of stressor events) on model accuracy
- Moving forward, an important step is **in-field implementation of the proposed approach** and understand accuracy, efficacy, and refinements needed for successful use

REFERENCES

Afifi, M. (2007). Gender differences in mental health. *Singapore Medical Journal*, 48(5), 385-391. <https://pubmed.ncbi.nlm.nih.gov/17453094/>

Al-Huthail, Y. R. (2008). Accuracy of referring psychiatric diagnosis. *International Journal of Health Science*, 2(1), 35-38. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3068718/>

Businelle, M. S., Mills, B. A., Chartier, K. G., Kendzor, D. E., Reingle, J. M., & Shuval, K. (2013). Do stressful events account for the link between socioeconomic status and mental health? *Journal of Public Health*, 36(2), 205-212. <https://doi.org/10.1093/pubmed/fdt060>

CDC. (2021, May 12). *Mental Health*. Centers for Disease Control and Prevention. Retrieved April 7, 2023, from <https://www.cdc.gov/healthyyouth/mental-health/index.htm>

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94-98. <https://doi.org/10.7861%2Ffuturehosp.6-2-94>

Koneru, V. K., Weisman de mamani, A. G., Flynn, P. M., & Betancourt, H. (2007). Acculturation and mental health: Current findings and recommendations for future research. *Applied and Preventive Psychology*, 12(2), 76-96. <https://doi.org/10.1016/j.appsy.2007.07.016>

McCaffrey, J. D. (2018, July 7). *Why a neural network is always better than logistic regression*. Retrieved April 7, 2023, from <https://jamesmccaffrey.wordpress.com/2018/07/07/why-a-neural-network-is-always-better-than-logistic-regression/>

Rowan, K., Mcalpine, D. D., & Blewett, L. A. (2013). Access and cost barriers to mental health care, by insurance status, 1999-2010. *Health Affairs*, 32(10), 1723-1730. <https://doi.org/10.1377%2Fhlthaff.2013.0133>

ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Kaustubh Supekar of Stanford University, for guiding my research as I explored the nuances of the intersection between mental health, psychology, and data science and conducted my literature review and algorithm analysis. I would also like to thank my AAR teacher, Ms. Angell, for providing daily guidance throughout the school year and my fellow AAR students for bringing insightful discussion about a variety of exciting interdisciplinary topics to class and creating a safe, enjoyable environment for research.