A QUANTITATIVE STUDY FOR THE PREDICTION OF ANXIETY, STRESS AND DEPRESSION

Group: 4

Team Line-Up:

1) Brandon Schoener

2) Ishansh Sahni

3) Rajiv Nagesh

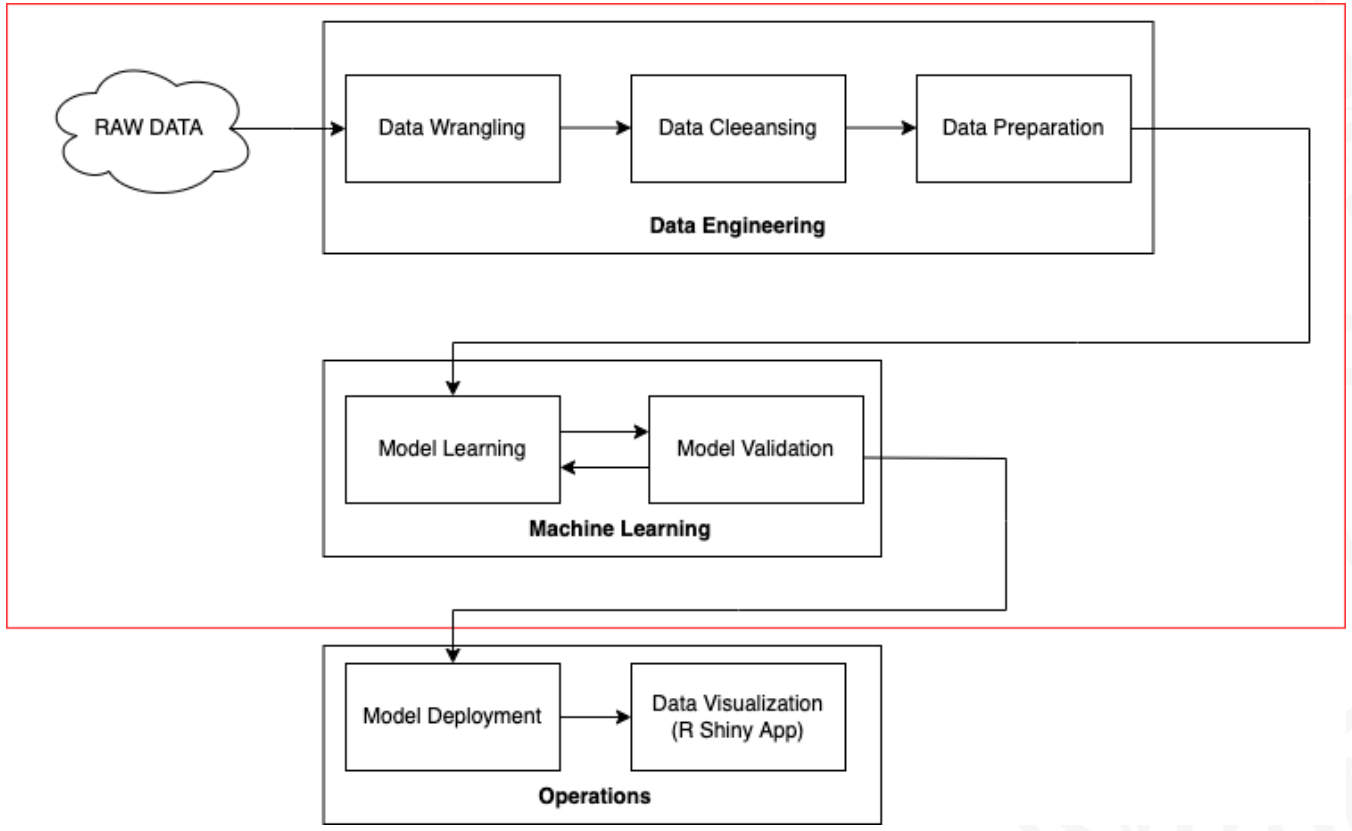
University at Buffalo The State University of New York

University at Buffalo The State University of New York

PHASE-2 DISCUSSIONS

1) Project Pipeline
2) Data Cleaning
3) Feature Engineering
4) Intermediate ML Model (K-Means)
5) Research Questions

Project Pipeline



Data Engineering

1. Data Cleaning
2. Fake Responses (Check)
3. Feature Selection

VCL1	boat
VCL2	incoherent
VCL3	pallid
VCL4	robot
VCL5	audible
VCL6	cuivocal
VCL7	paucity
VCL8	epistemology
VCL9	florted
VCL10	decide
VCL11	pastiche
VCL12	verdid
VCL13	abysmal
VCL14	lucid
VCL15	betray
VCL16	funny

Fig1: Fake Responses Check

A	B	C	D	E	F	G	H	I
Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
4	4	2	4	4	4	4	4	2
4	1	2	3	4	4	3	4	3
3	1	4	1	4	3	1	3	2
2	3	2	1	3	3	4	2	3
2	2	3	4	4	2	4	4	4
1	1	2	1	3	1	1	3	3
1	1	2	3	4	1	3	3	3
1	1	1	1	3	2	2	1	1
4	4	3	4	3	4	4	4	4
3	2	4	1	4	4	3	4	4
3	1	2	1	3	3	1	3	1
3	3	2	2	4	3	1	3	4
1	1	1	1	1	2	1	3	1
1	4	1	2	3	1	1	3	2
1	1	1	1	1	1	1	1	1
3	1	2	2	2	1	1	1	2
3	1	2	2	1	3	1	3	4
3	1	2	1	2	3	1	2	2
3	2	2	1	3	1	1	2	1

Fig2: data snippet

AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG
Q42	country	testelapse	TIPI1	TIPI2	TIPI3	TIPI4	TIPI5	TIPI6	TIPI7	TIPI8	TIPI9	TIPI10	education	gender	age	married	familysize
4	IN	167	1	5	7	7	7	7	7	5	1	1	2	2	16	1	2
2	US	193	6	5	4	7	5	4	7	7	1	5	2	2	16	1	4
4	PL	271	2	5	2	2	5	6	5	5	3	2	2	2	17	1	3
2	US	261	1	1	7	4	6	4	6	1	6	1	1	2	13	1	5
3	MY	164	2	5	3	6	5	5	5	6	3	3	3	2	19	1	4
2	US	349	2	1	6	1	7	7	7	2	6	7	2	2	20	1	4
3	MX	45459	2	5	6	5	3	2	6	3	5	5	2	2	17	1	4
2	GB	232	7	6	4	5	3	2	6	3	5	2	4	2	29	1	2
4	US	195	1	4	5	7	5	7	6	7	1	4	2	1	16	1	4
4	DE	120	1	7	5	7	5	7	1	2	1	7	1	2	18	1	3
2	US	162	5	3	6	6	3	4	4	7	5	7	1	1	15	1	1
2	US	511	6	5	6	6	6	2	5	3	3	3	2	2	18	1	2
2	US	113	5	1	4	6	5	5	7	6	2	1	3	2	20	1	2
3	EE	175	5	2	5	5	2	6	7	6	6	2	4	1	31	1	5
1	US	143	3	5	6	1	6	5	3	2	7	2	3	1	34	3	2
3	US	155	3	5	4	5	6	5	4	3	4	4	2	2	17	1	1
3	US	148	1	7	5	7	1	6	5	4	1	3	2	2	19	1	2
1	GB	99	1	4	4	0	7	5	4	4	3	1	3	2	18	1	5
4	US	128	2	7	7	1	6	5	4	7	6	2	2	1	19	1	1
4	US	280	2	2	5	4	6	7	4	6	3	6	2	2	19	1	2
2	US	182	7	1	3	6	6	1	7	6	1	1	1	2	15	1	5
3	US	236	1	6	5	7	2	6	2	7	2	7	1	1	15	1	1
3	CA	118	3	3	5	5	5	5	5	3	1	1	2	2	22	1	1
4	EC	223	1	5	5	6	3	6	5	3	3	5	2	2	19	1	3
3	AU	254	2	4	3	5	5	6	5	7	3	4	2	1	20	1	3
4	US	229	1	6	5	7	7	7	1	6	1	2	1	2	16	1	2

Fig3: data snippet

Feature Engineering

1. **Computed ASD Score**
2. **Aggregated Time Elapsed (Each Survey)**
3. **Standardizing Numerical Features**
4. **Encoding Categorical Features**

ASD (Anxiety, Stress, Disorder) Score: The score computed on the basis of responses of 42 questions to each individual with their agreement/disagreement. Ranges between 0 and 1 where 0 means not stressed and 1 means showing prominent signs of anxiety, stress or disorder.

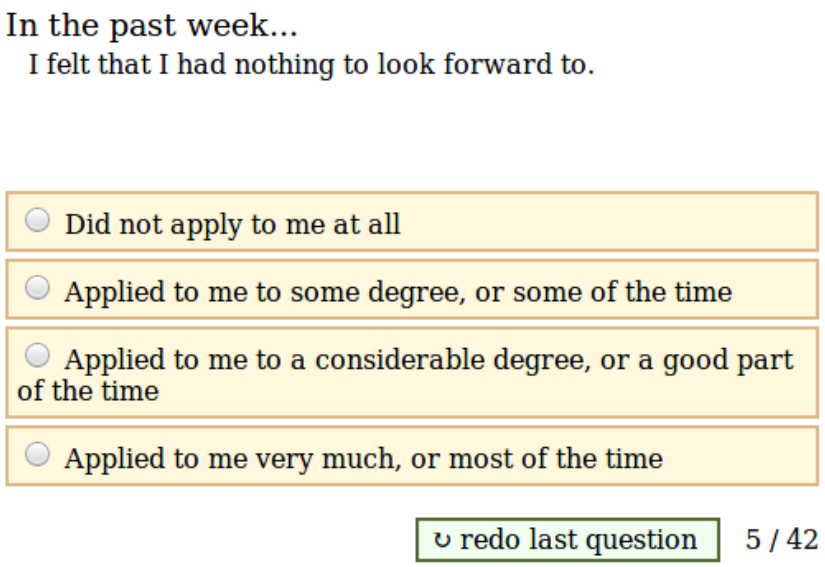
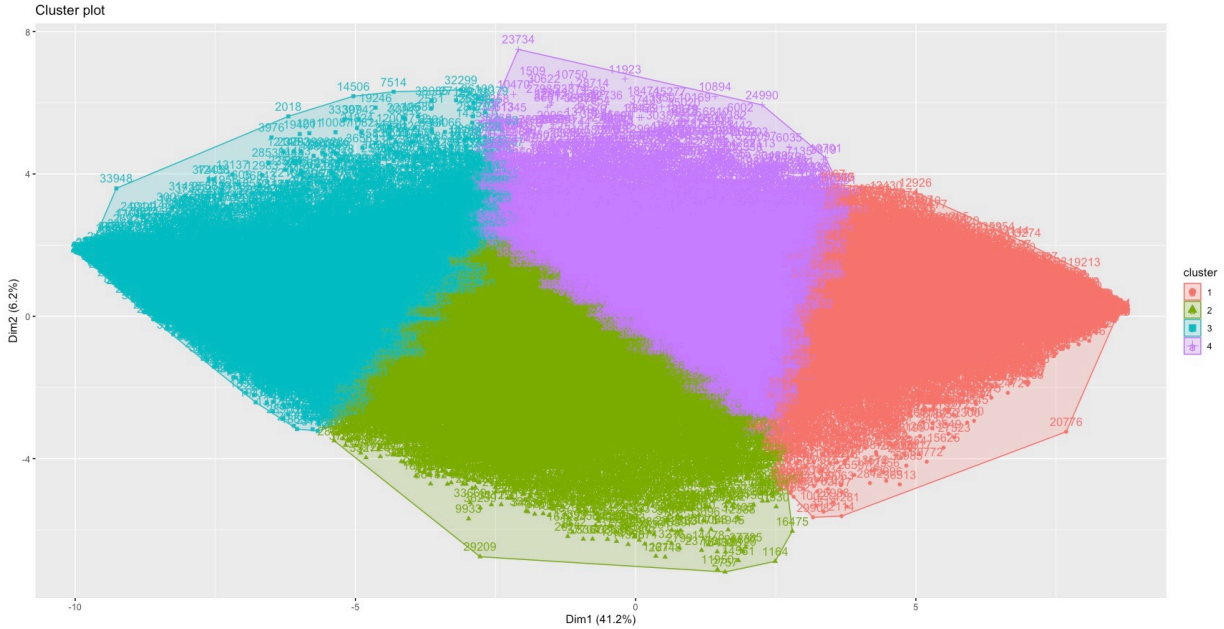


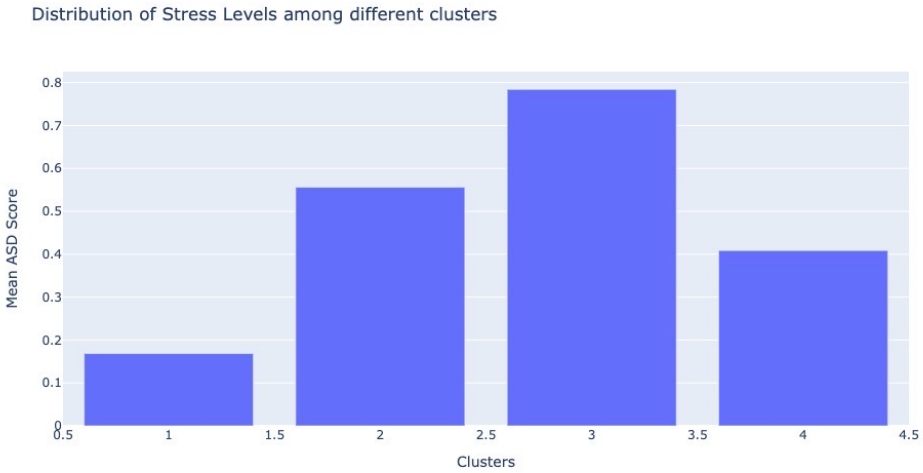
Fig4: Sample Question

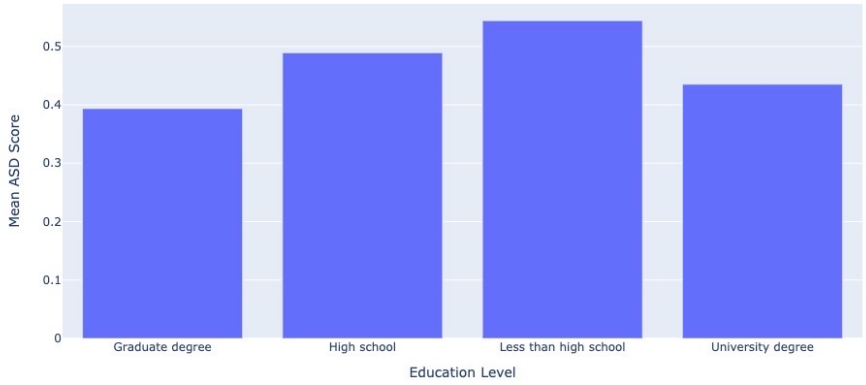
Total_Score	ASD_Score
143	0.8015873
110	0.53968254
110	0.53968254
91	0.38888889
143	0.8015873
73	0.24603175
106	0.50793651
56	0.11111111
149	0.84920635
146	0.82539683
82	0.31746032
108	0.52380952
89	0.37301587

Fig5: ASD Score

Intermediate Machine Learning Model (K-Means)







Research Questions and Future Scope

1) Using ASD score as a metric in the everyday working scenario in organizations.
2) Creating an interactive R Shiny Dashboard with different metrics like Age, Gender, Country to classify which class and sector of people tend to be more/least depressed.
3) In the mental health department, psychiatrists can use the ASD score and treat an individual and use the model analysis to work on a particular behavior of an individual.

Challenges

1) Cannot totally rely on data. (For ex., a person's mental state can change on a given day).
2) Lack of clinical data to verify clusters obtained.
3) Lack of terminology for the ASD Score.