MINDGAME INSIGHTS

A DEEP DIVE INTO GAMING AND PSYCHOLOGICAL WELL-BEING RELATIONSHIPS

COURSE: INST 737 - INTRODUCTION TO DATA SCIENCE

MILESTONE 2

TEAM:

RAJEEVAN MADABUSHI

PRANAV ADIRAJU

ASMITA SAMANTA

RESEARCH QUESTION

- **Dependent Variable**

Generalized Anxiety Disorder ***(GAD_T)***

- **Independent Variables**

Narcissism, Social Phobia Inventory ***(SPIN_T)****, Hours, Satisfaction With Life* ***(SWL_T)****, Reasons to Play* ***(whyplay_clean)****, Status of Work* ***(Work)****, Style of Play* ***(Playstyle_clean)***

DATA CLEANING EFFORTS

MILESTONE 1

- Independent variable **whyplay** had **132** records under **Other** category
- Independent variable **playstyle** had **155** records under **Other** category

MILESTONE 2

- Independent variable **whyplay** had **31** records under **Other** category
- Independent variable **playstyle** had **14** records under **Other** category

LINEAR REGRESSION

01

COEFFICIENTS

Feature	Coefficient
Intercept	6.257917418814023
whyplay_clean_freq_encoded	-2.9472759792159877
Work_freq_encoded	0.6648876016362912
Narcissism	0.3376492930253228
Playstyle_clean_freq_encoded	-0.21394735131423487
SWL_T	-0.17866427544801958
Hours	0.12419096613795266
SPIN_T	0.006705640455700115

02

PREDICTIVE FEATURES

All independent variables are significant except Style of Play

03

MOST PREDICTIVE FEATURES

Reasons to Play and Status of Work are most predictive features

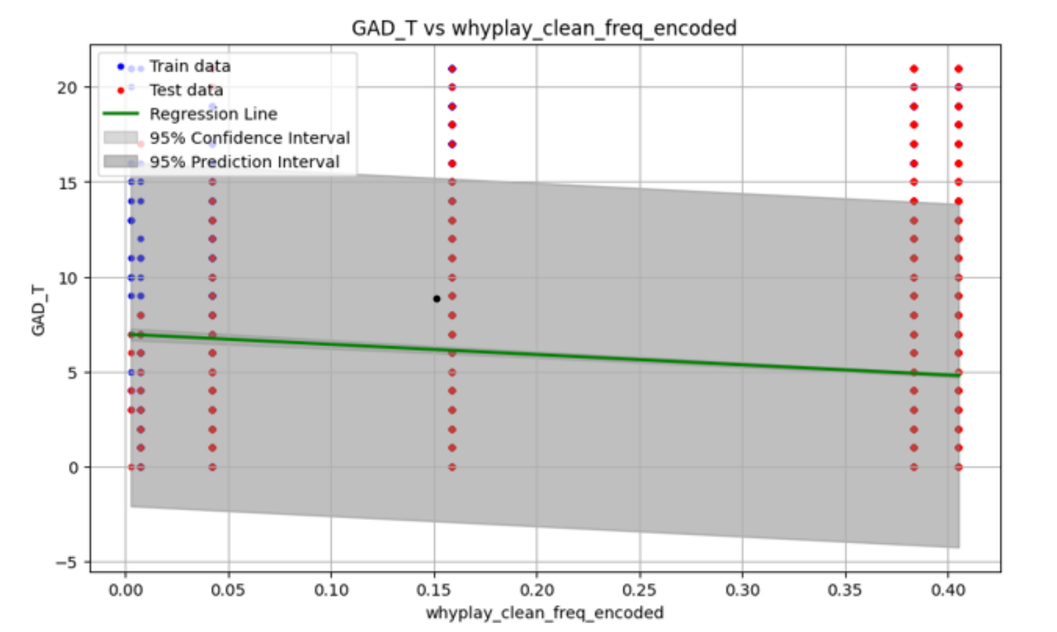
04

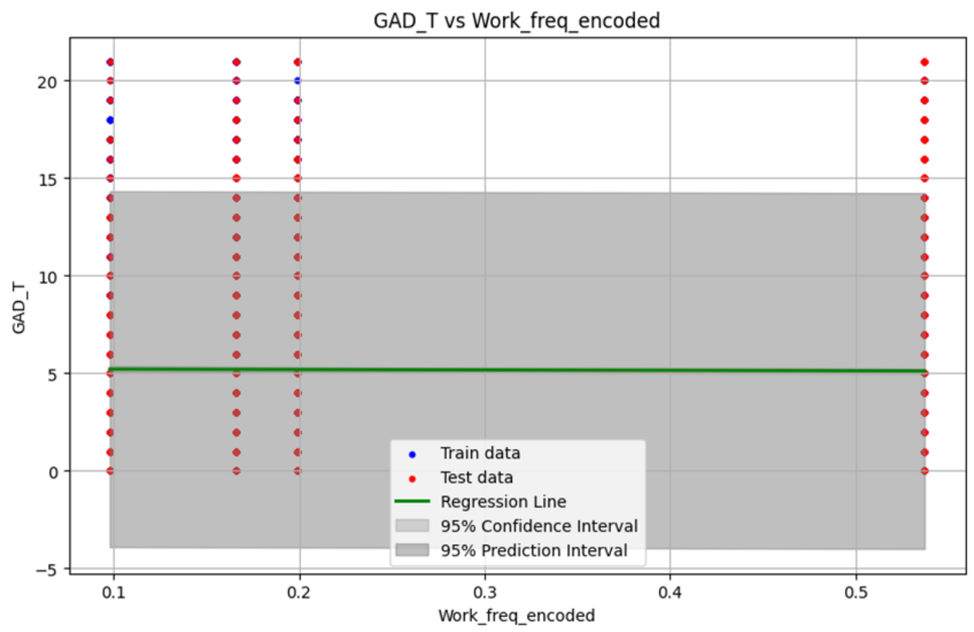
APPLICABILITY OF LINEAR REGRESSION

Only SPIN_T has linear regression applicability with R2 above 20% and resuduls absolute mean below 10

LINEAR REGRESSION

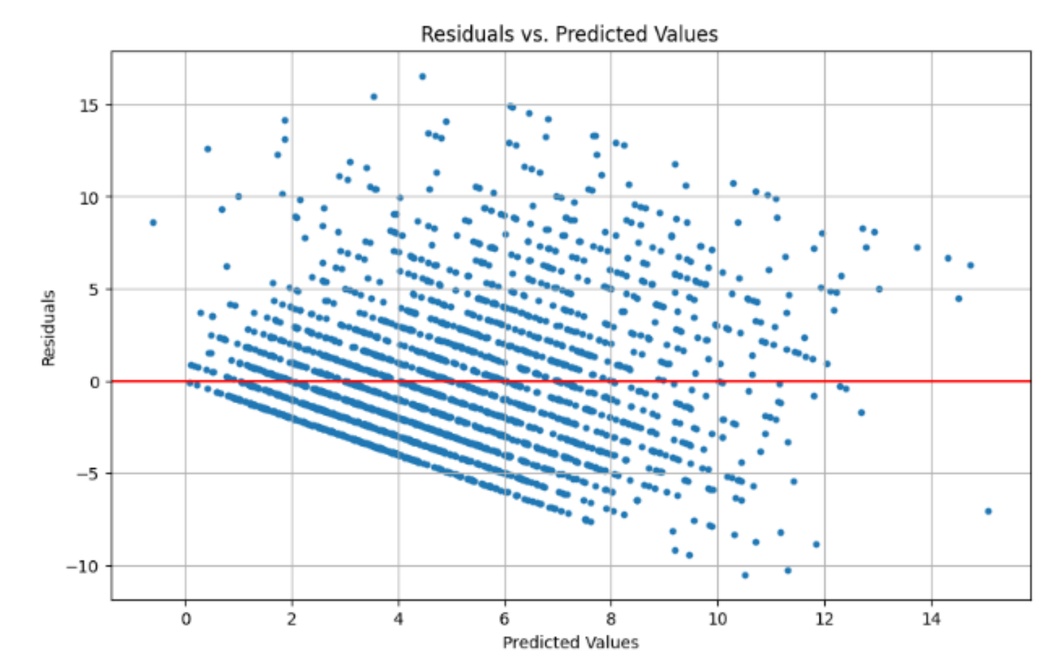
RESULTS WITH CONFIDENCE AND PREDICTION BANDS

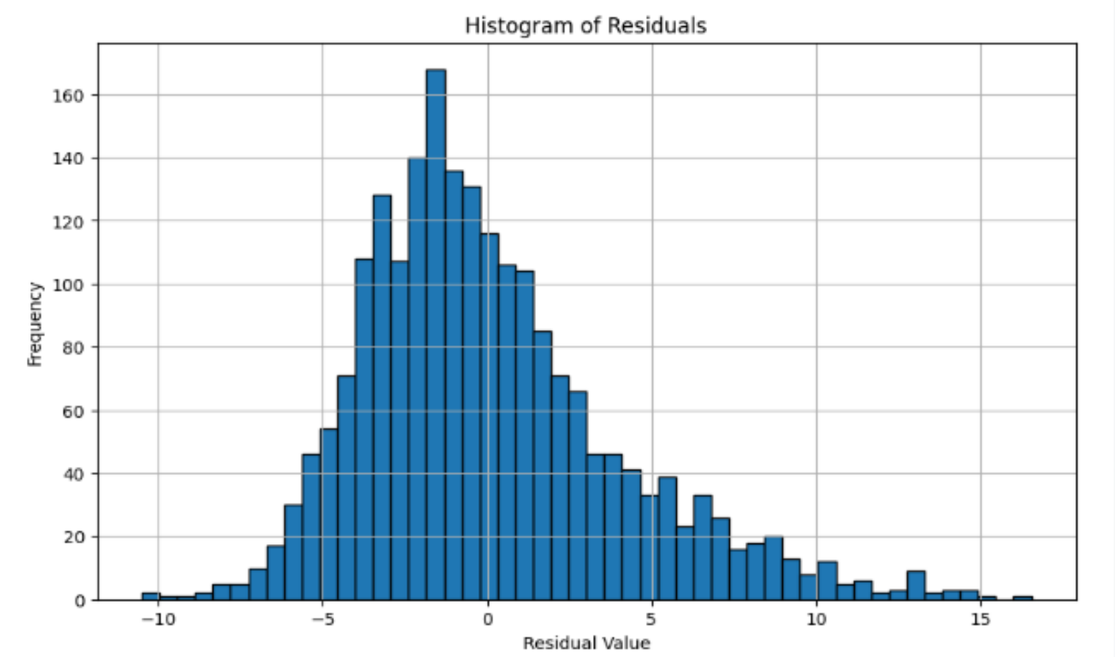




LINEAR REGRESSION

RESIDUALS





LINEAR REGRESSION

PREDICTION ACCURACY OF MODEL

Feature Name	Correlation	MSE	R2	Linear Regression Applicabl
Narcissism	0.033	22.549	-0.002	No
SPIN_T	0.482	17.309	0.231	Yes
Hours	0.096	22.307	0.009	No
SWL_T	0.405	18.828	0.163	No
whyplay_clean_freq_encoded	0.119	22.197	0.014	No
Work_freq_encoded	0.062	22.490	0.001	No
Playstyle_clean_freq_encoded	0.048	22.479	0.001	No

MULTIVARIATE REGRESSION

01

BEST COMBINATION OF FEATURES

Social Phobia Inventory (SPIN_T), Satisfaction With Life (SWL_T), Reasons to Play (whyplay_clean)

02

COEFFICIENT FOR MOST PREDICTIVE FEATURES

Feature	Coefficient
whyplay_clean_freq_encoded	-3.2539532000859888
SWL_T	-0.1789557702324269
SPIN_T	0.12446524210259205

03

PREDICTION ACCURACY USING CORRELATION

Correlation between the predicted and real values for best combination is 0.549

04

PREDICTION ACCURACY USING MSE

mean square error between the predicted and real values for best combination is 15.732, R2 value: 30.09%

REGULARIZATION

UNIVARIATE REGRESSION

Feature Name	Model	Correlation	MSE	R2
whyplay_clean_freq_encoded	Ridge	0.119	22.197	0.014
whyplay_clean_freq_encoded	Lasso	0.000	22.512	-0.000
SPIN_T	Ridge	0.482	17.309	0.231
SPIN_T	Lasso	0.482	17.344	0.229
SWL_T	Ridge	0.405	18.828	0.163
SWL_T	Lasso	0.405	18.870	0.162

MULTIVARIATE REGRESSION

Ridge	0.548	15.761	0.300
Lasso	0.546	15.850	0.296

SIMILARITIES AND DIFFERENCES ACROSS MULTIPLE RUNS OF UNIVARIATE, MULTIVARIATE REGRESSION AND REGULARIZATION

Category	Feature/Model	Correlation	MSE	R2
Univariate	whyplay_clean_freq_encoded	0.126	21.976	0.016
Univariate	SPIN_T	0.477	17.275	0.226
Univariate	SWL_T	0.434	18.200	0.185
Multivariate		0.561	15.326	0.314
Regularization	Ridge	0.561	15.325	0.314
Regularization	Lasso	0.556	15.526	0.305

LOGISTIC REGRESSION

01

DATA PREPARATION EFFORTS

Categorized the dependent variable General Anxiety Disorder (GAD_T) using the Median

02

INTERCEPT, COEFFICIENTS & STATISTICAL SIGNIFICANCE

Intercept for the model is 0.671

Variable	Coefficient	p-value
const	0.691	0.0
SPIN_T	0.055	0.0
SWL_T	-0.079	0.0
yplay_clean_freq_encoded	-1.369	0.0

03

LOG ODDS & ODDS RATIO

Predictor	Log Odds	Odds Ratios
SPIN_T	0.055	1.057
SWL_T	-0.079	0.924
whyplay_clean_freq_encoded	-1.369	0.254

04

MOST PREDICTIVE FEATURES

Reasons to play (whyplay_clean) is the most predictive feature

LOGISTIC REGRESSION

CONFUSION MATRIX

	Predicted: Yes	Predicted: No
Actual: Yes	924 True Positive	249 False Negative
Actual: No	388 False Positive	558 True Negative

PREDICTION METRICS

- *ACCURACY: 0.699*
- *Precision: 0.691*
- *Recall: 0.589*
- *F1_score: 0.636*

NAIVE BAYES

PREDICTION ON TRAINING DATA

	Predicted: Yes	Predicted: No
Actual: Yes	863 True Positive	310 False Negative
Actual: No	321 False Positive	625 True Negative

- ***Accuracy: 0.702***
- ***Precision: 0.668***
- ***Recall: 0.660***
- ***F1_score: 0.664***

PREDICTION ON TESTING DATA USING LAPLACE SMOOTHING

	Predicted: Yes	Predicted: No
Actual: Yes	863 True Positive	310 False Negative
Actual: No	321 False Positive	625 True Negative

- ***Accuracy: 0.702***
- ***Precision: 0.668***
- ***Recall: 0.660***
- ***F1_score: 0.664***

DECISION TREES AND RANDOM FORESTS

TRAINING SPLIT

Category	Distribution
0	0.353
1	0.205
2	0.237
3	0.205

TESTING SPLIT

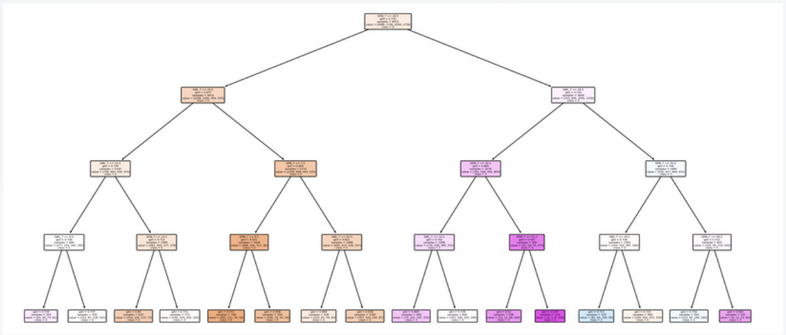
Category	Distribution
0	0.359
1	0.195
2	0.237
3	0.21

ORIGINAL SPLIT

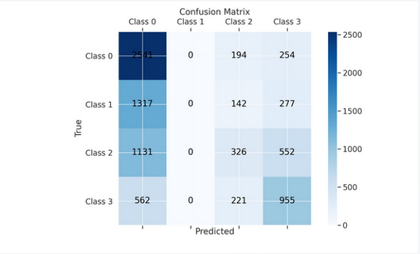
Category	Distribution
0	0.354
1	0.203
2	0.237
3	0.206

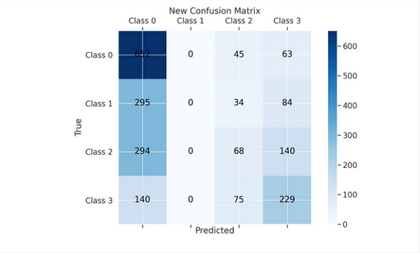
DECISION TREES AND RANDOM FORESTS

DECISION TREE



CONFUSION METRICES



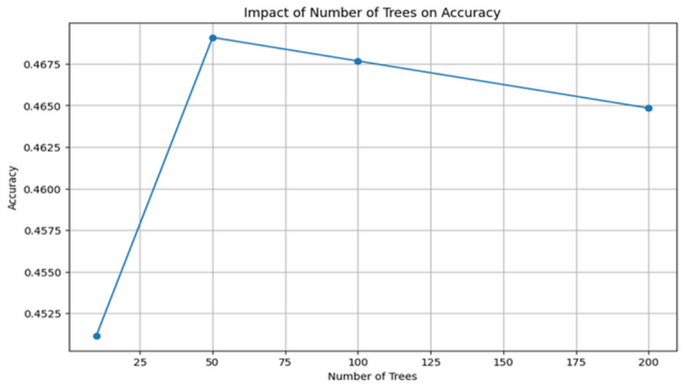


ORIGINAL SPLIT

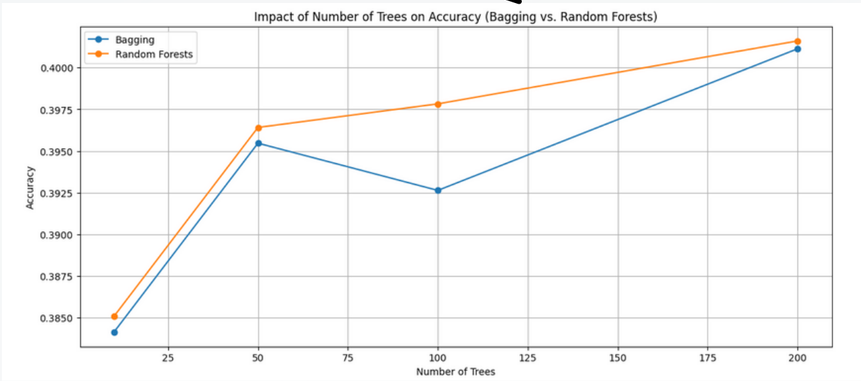
- Correctly Classified Samples for Train Data: 0.451
- Incorrectly Classified Samples for Train Data:0.548
- Correctly Classified Samples for Test Data: 0.447
- Incorrectly Classified Samples for Test Data: 0.552

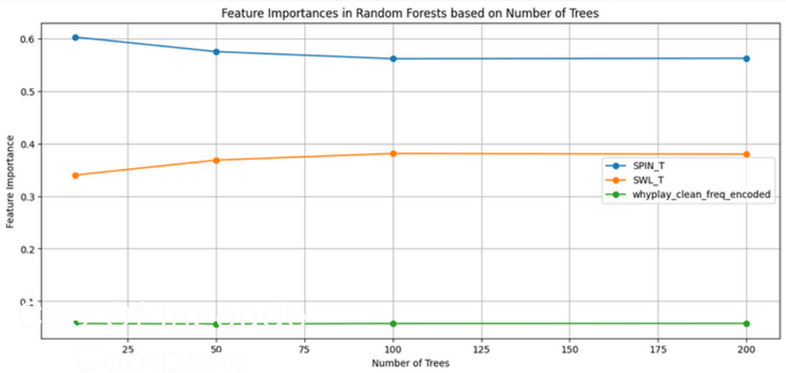
DECISION TREES AND RANDOM FORESTS

Gradient Boosted Decision Trees



Bagging and Random forests





COMPARATIVE ANALYSIS

Model	Accuracy
Linear Regression With Ridge for multivariate	31.38%
Logistic Regression	69.93%
Naive Bayes with Laplace	70.22%
Decision Trees with Boosting with 50 trees	46.9%
Bagging & Random Forests with 200 trees	40.1%

Naive Bayes with Laplace smoothing is top choice for this research question

NEXT STEPS

01 Support Vector Machines

02 Neural Networks

03 Clustering

04 Comparative Analysis of SVMs, Neural networks and clustering

05 Application of feature selection: filter, wrapper or embedding

THANK YOU