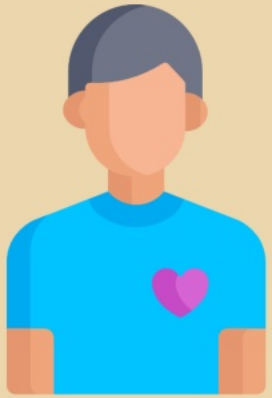
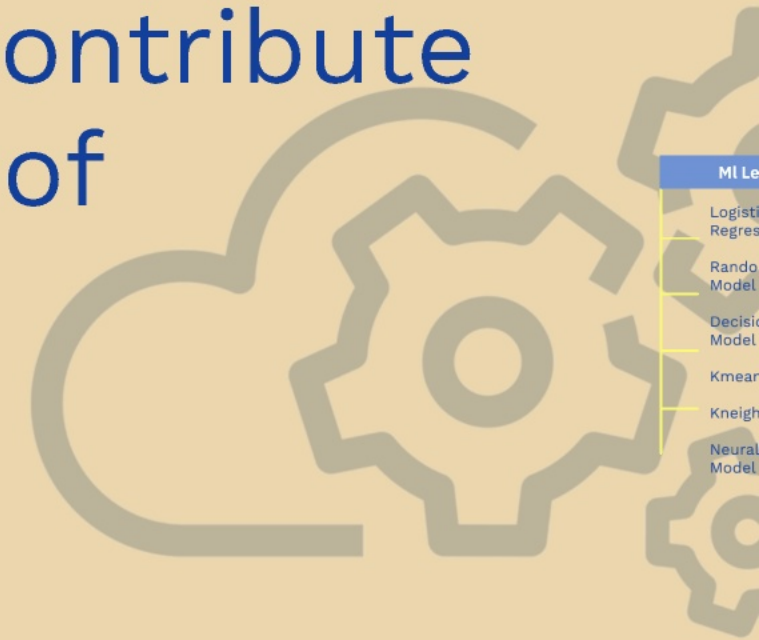
Alzheimer's Disease

Can We Predict It?

This project aims to develop a tool to identify the key factors that contribute to the onset and progression of Alzhiemer's disease





Data

https://

Tools and Technology

Ml Learning	ML Libraries	Additional Tools
Logistic Regression Model	sklearn	numpy
Random Forest Model	seaborn	pandas
Decision Trees Model	tensorflow	matplotlib
Kmeans Model	imblearn	Tableau
Kneighbors Model		PostgreSQL
Neural Networks Model		Jupyter Notebook

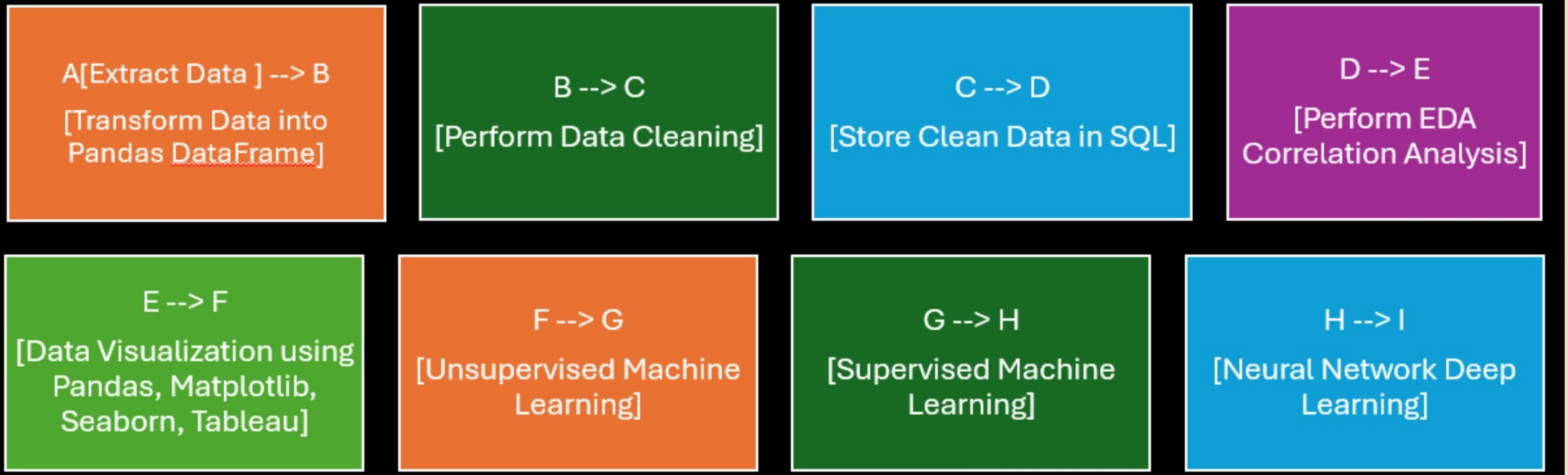
Neural Networks Model

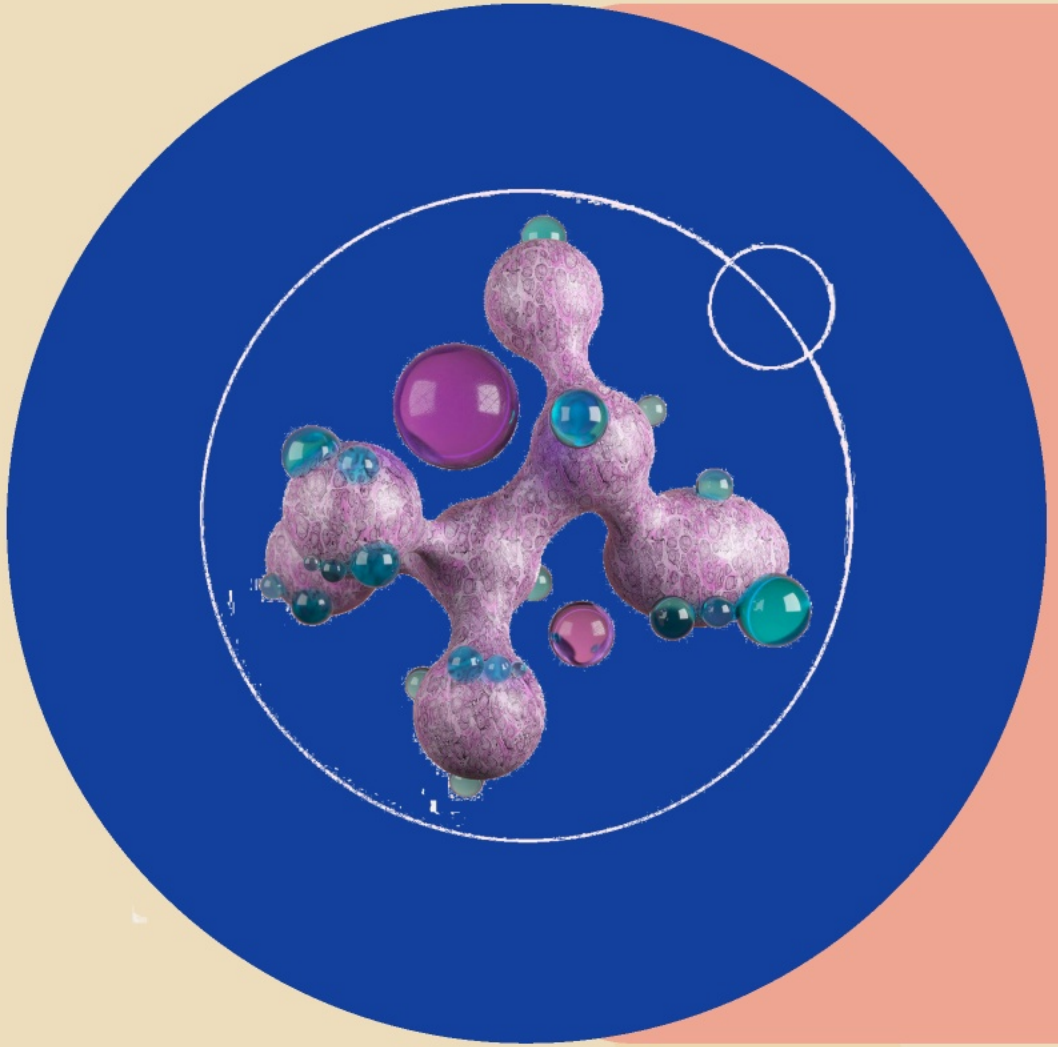
Dataset:

https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset/data

PatientID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption	PhysicalActivity	DietQuality	SleepQuality	FamilyHistoryAlzheimer	CardiovascularDisease	Diabetes	Depression	HeadInjury	Hypertension	SystolicBP	DiastolicBP	CholesterolTotal	CholesterolLDL	Cholestero
4751	73	0	0	2	22.92774923	0	13.29721773	6.327112474	1.347214306	9.025678666	0	0	1	1	0	0	142	72	242.3668397	56.15089696	33.682
4752	89	0	0	0	26.82768119	0	4.542523818	7.61988454	0.518767139	7.151292743	0	0	0	0	0	0	115	64	231.162595	193.4079955	79.0284
4753	73	0	3	1	17.79588244	0	19.55508453	7.844987791	1.826334665	9.673574158	1	0	0	0	0	0	99	116	284.1818578	153.3227622	69.7722
4754	74	1	0	1	33.80081704	1	12.20926555	8.42800135	7.43560414	8.392553685	0	0	0	0	0	0	118	115	159.5822396	65.36663684	68.4574
4755	89	0	0	0	20.71697383	0	18.45435609	6.310460689	0.795497509	5.597237678	0	0	0	0	0	0	94	117	237.6021836	92.86969988	56.8743
4756	86	1	1	1	30.62688555	0	4.140143784	0.211061631	1.584922011	7.261952505	0	0	1	0	0	0	168	62	280.7125387	198.3346285	79.0805
4757	68	0	3	2	38.38762186	1	0.646047271	9.25769491	5.897387927	5.477685594	0	0	0	0	1	0	143	88	263.734149	52.47066963	66.5333
4758	75	0	0	1	18.77600941	0	13.72382571	4.649450668	8.341903192	4.213209925	0	0	0	0	0	0	117	63	151.3831368	69.62351041	77.3468
4759	72	1	1	0	27.83318838	0	12.16784763	1.531359788	6.736882044	5.748223869	0	0	0	0	0	1	117	119	233.6057552	144.0457396	43.0758
4760	87	0	0	0	35.45630173	1	16.02868824	6.440772687	8.086019121	7.551773444	0	1	0	0	0	0	130	78	281.6300502	130.4975804	74.2912
4761	89	0	3	1	39.46303422	0	9.811292129	8.819950351	0.434020278	7.6440973	0	0	0	0	0	0	131	86	224.5264374	160.7564751	68.3283
4762	78	0	0	2	22.46338265	1	19.30018298	3.834639382	8.279189504	8.312325537	0	0	1	0	0	1	165	97	254.5865603	132.9600115	39.009
4763	84	1	0	1	26.770946	0	10.97802164	3.978078872	7.024417381	8.253004551	1	0	0	1	0	0	145	63	254.4353861	148.2914229	98.5855
4764	78	1	0	1	28.87065239	1	10.1947063	0.631280727	1.653281417	7.333235623	1	0	0	0	0	0	137	82	221.3053385	194.6003789	26.3339
4765	64	1	0	2	27.94286273	0	2.17577965	9.714565829	5.317231744	9.087141195	0	1	1	1	0	0	94	98	247.5846143	193.9084814	94.8117
4766	69	0	0	1	18.04591747	0	8.116831616	2.956494729	7.570632794	6.736796642	0	0	0	0	0	1	124	72	204.6707593	97.75564858	99.2863
4767	63	1	1	2	22.82289624	1	4.433961006	7.182894555	7.929485772	4.65482804	0	1	0	0	0	0	148	116	200.5001232	181.0235604	51.8671
4768	65	1	0	1	16.33328275	1	4.161794911	1.306320412	2.888935528	5.436422671	0	0	0	0	0	1	154	61	183.1123335	101.2581964	39.2296
4769	72	0	0	2	37.93246903	0	9.385602765	7.127939893	3.314982664	6.790196088	0	0	0	1	0	0	132	107	214.9084055	132.1273691	20.8879
4770	68	0	0	3	20.04140036	0	18.42636447	4.060713917	3.361536133	7.393126211	0	0	0	0	0	0	144	98	195.2046348	181.9779289	50.5978
4771	82	1	0	0	36.2230988	0	4.19289551	6.381502407	7.971127069	9.521026999	0	1	0	0	0	0	120	93	271.8414494	84.47888882	76.3703
4772	65	0	1	2	37.54394317	1	12.06395931	9.126038218	3.531206847	9.574004795	0	0	0	0	0	0	178	89	210.1505183	94.12843564	68.1956
4773	82	1	0	3	37.58387706	0	5.83642077	7.676580404	9.767579025	9.250389465	0	0	0	0	0	0	120	71	225.6582355	87.32844007	43.2241
4774	82	1	2	1	21.96939009	0	3.289650198	6.934726371	3.242343132	4.412595501	0	0	0	0	0	0	106	95	201.6216146	110.1518657	26.6350

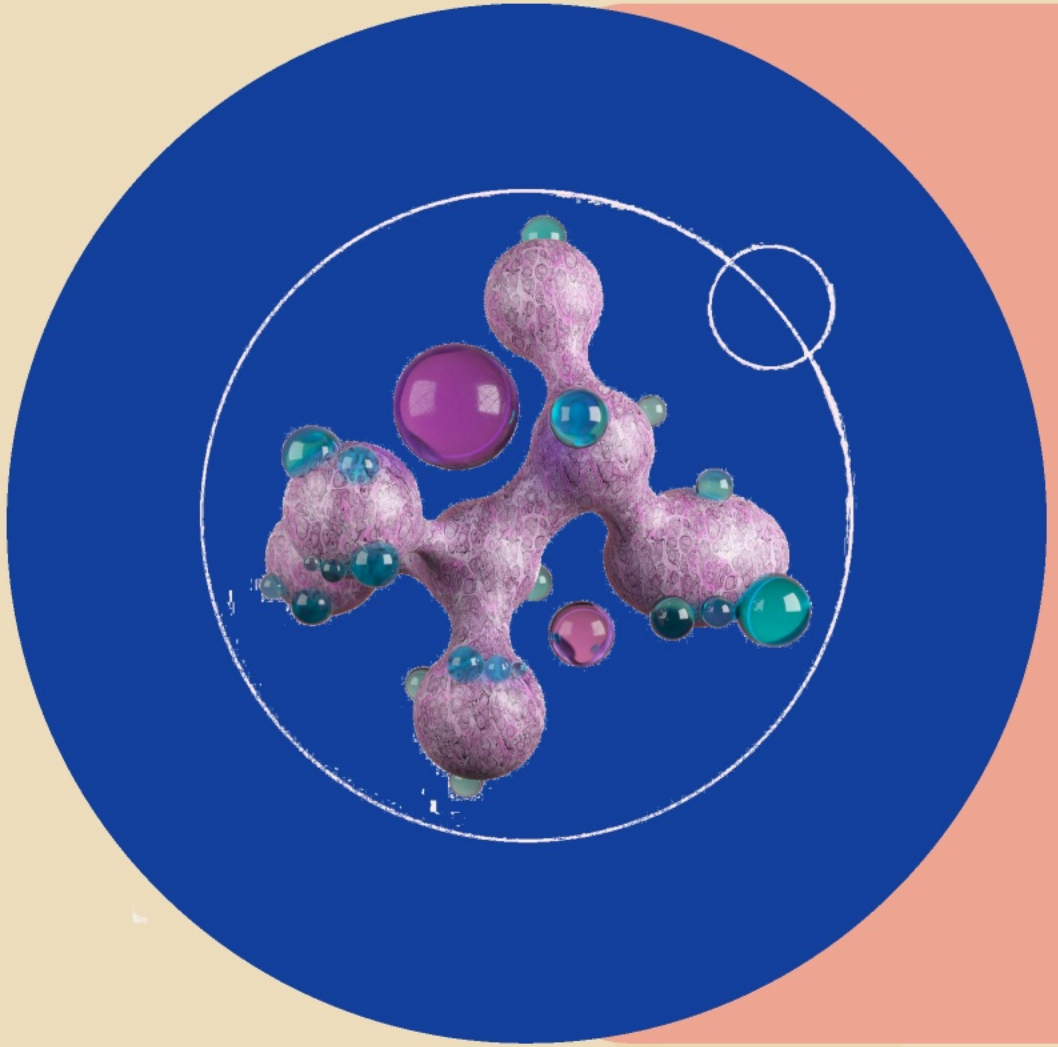
ETL Data Processing Flow Chart



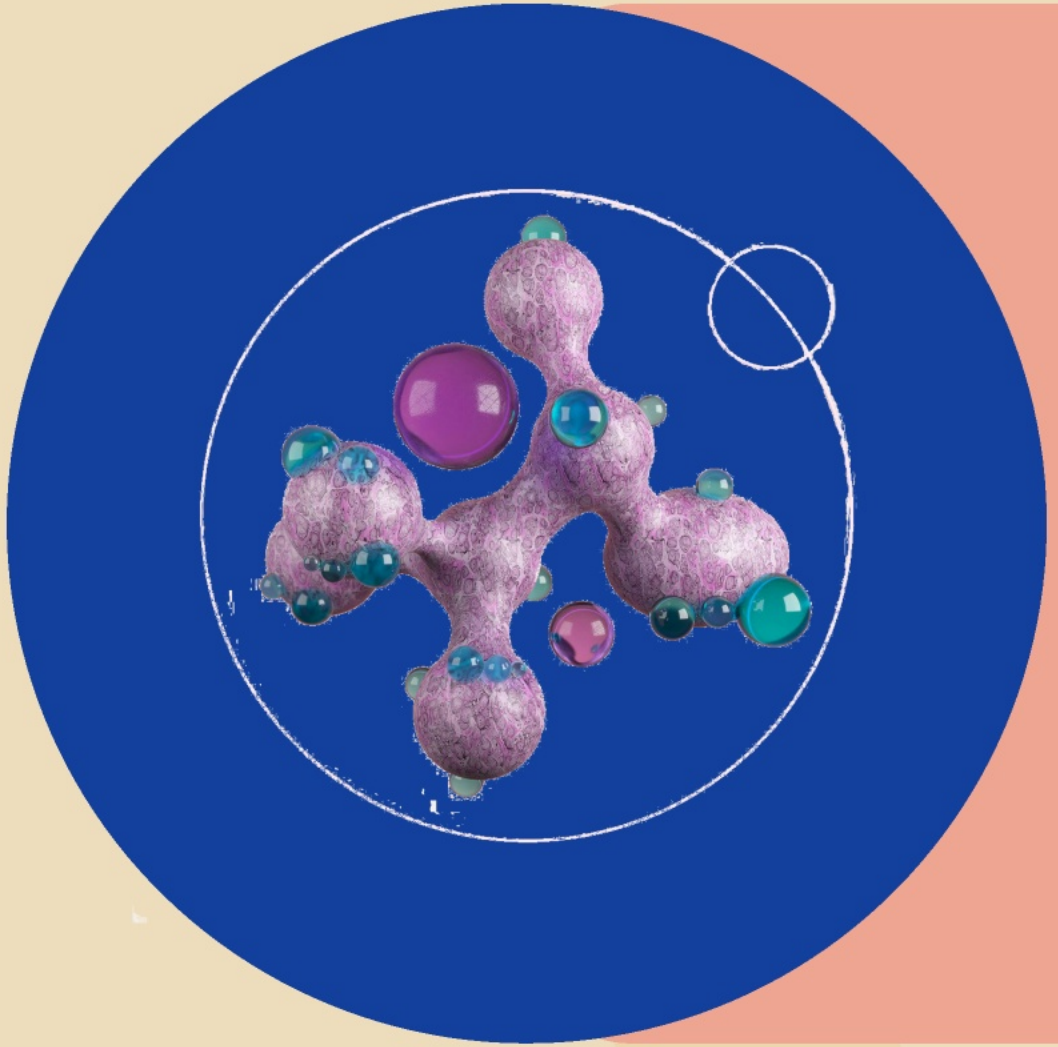


What are modifiable and nonmodifiable metrics that most accurately predict the onset of Alzheimer's Disease?

Alzheimer's is a progressive neurological disease and a type of dementia that affects the memory, thinking and behaviour of an individual [alz.org]



What are modifiable and nonmodifiable metrics that most accurately predict the onset of Alzheimer's Disease?

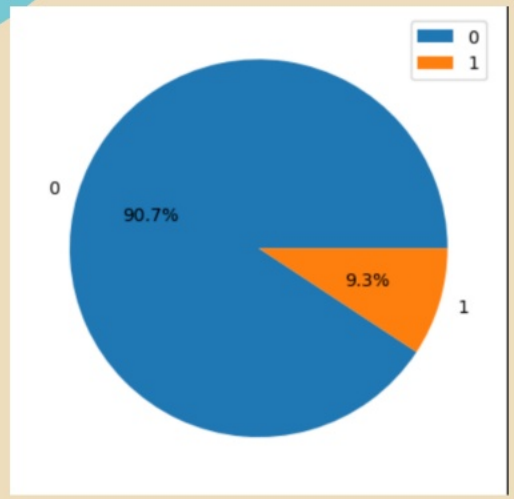


What are modifiable and nonmodifiable metrics that most accurately predict the onset of Alzheimer's Disease?

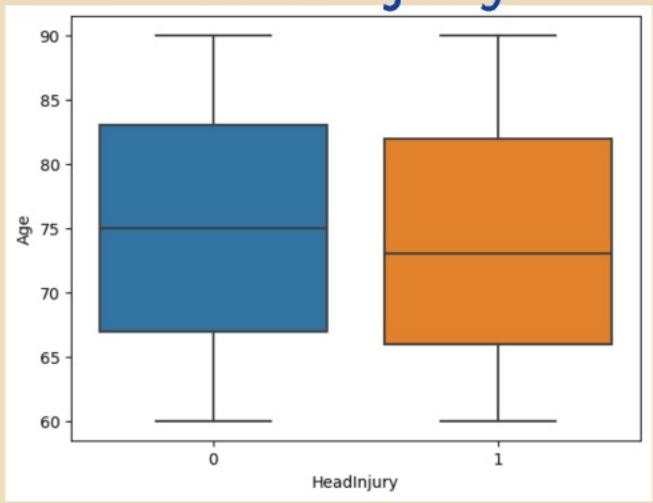
There is no one metric that can predict Alzheimer's disease

Medical History

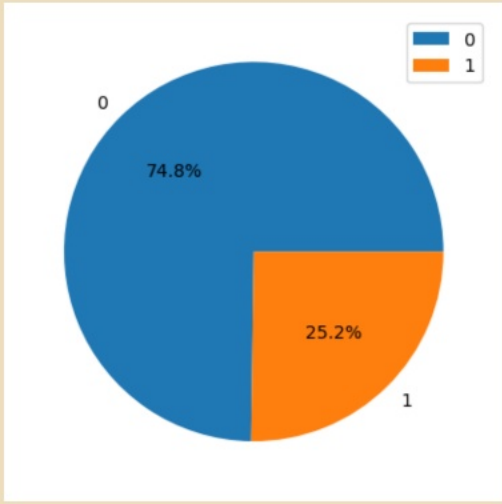
Hypertension



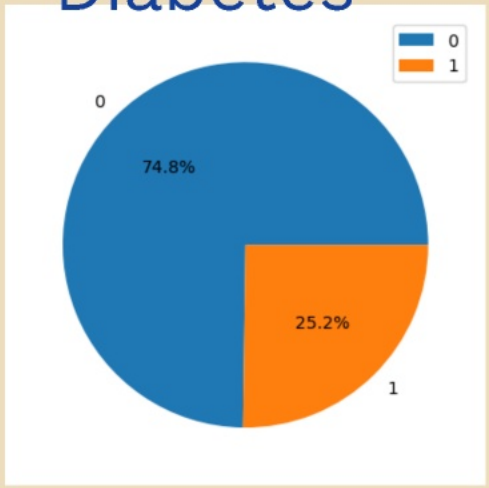
Head Injury



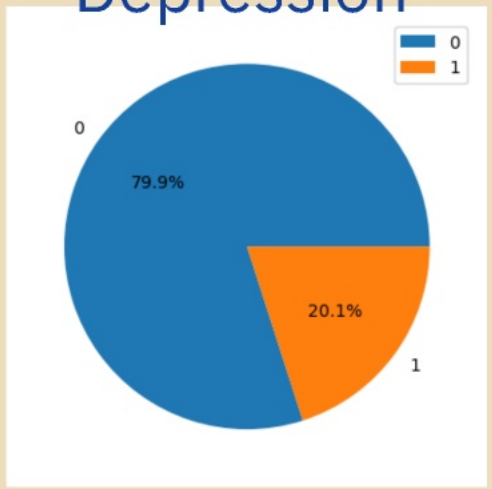
Family History of Alzheimer's



Diabetes



Depression



Decision Tree

confusion_matrix

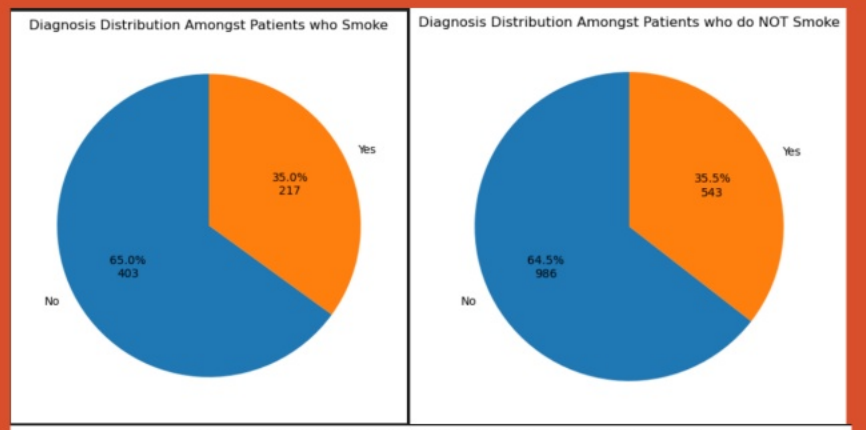
	Predicted 0	Predicted 1
Actual 0	277	70
Actual 1	55	136

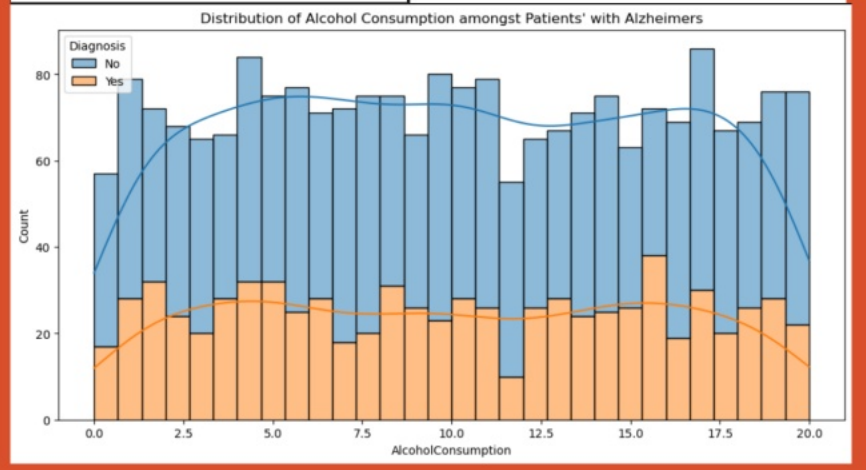
Accuracy Score: 0.7676579925650557
classification Report

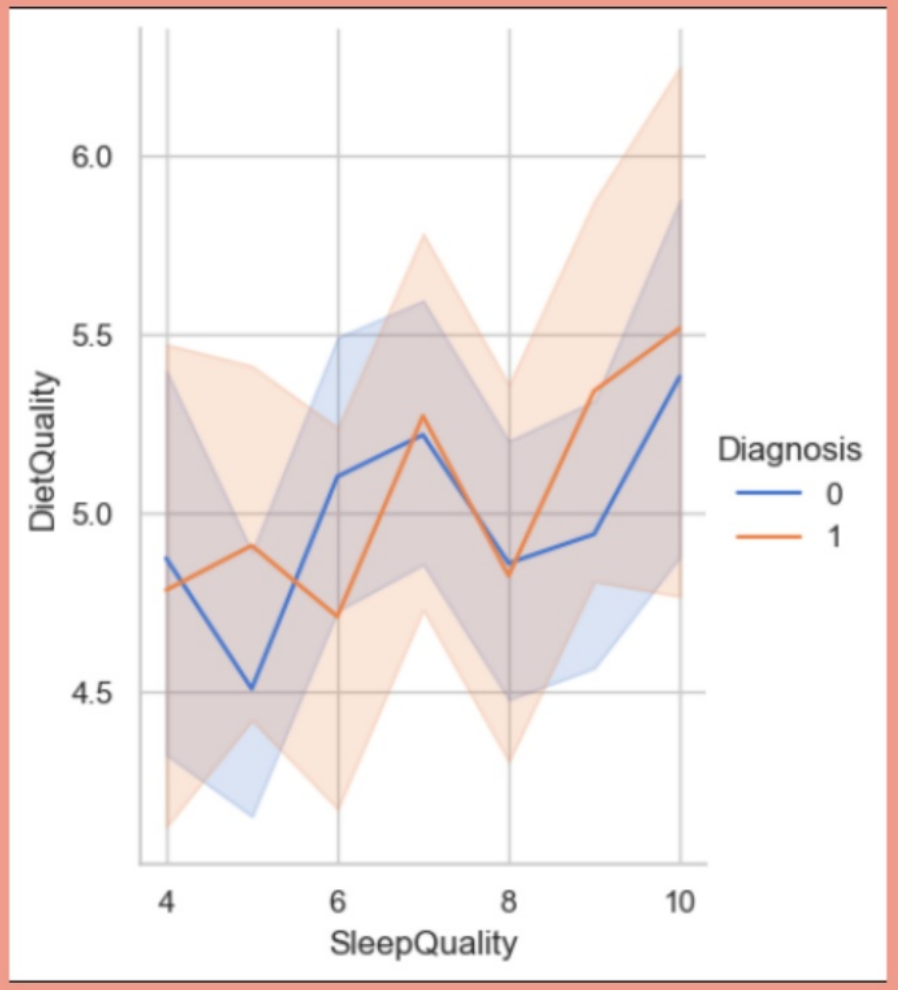
	precision	recall	f1-score	support
0	0.83	0.80	0.82	347
1	0.66	0.71	0.69	191
accuracy			0.77	538
macro avg	0.75	0.76	0.75	538
weighted avg	0.77	0.77	0.77	538

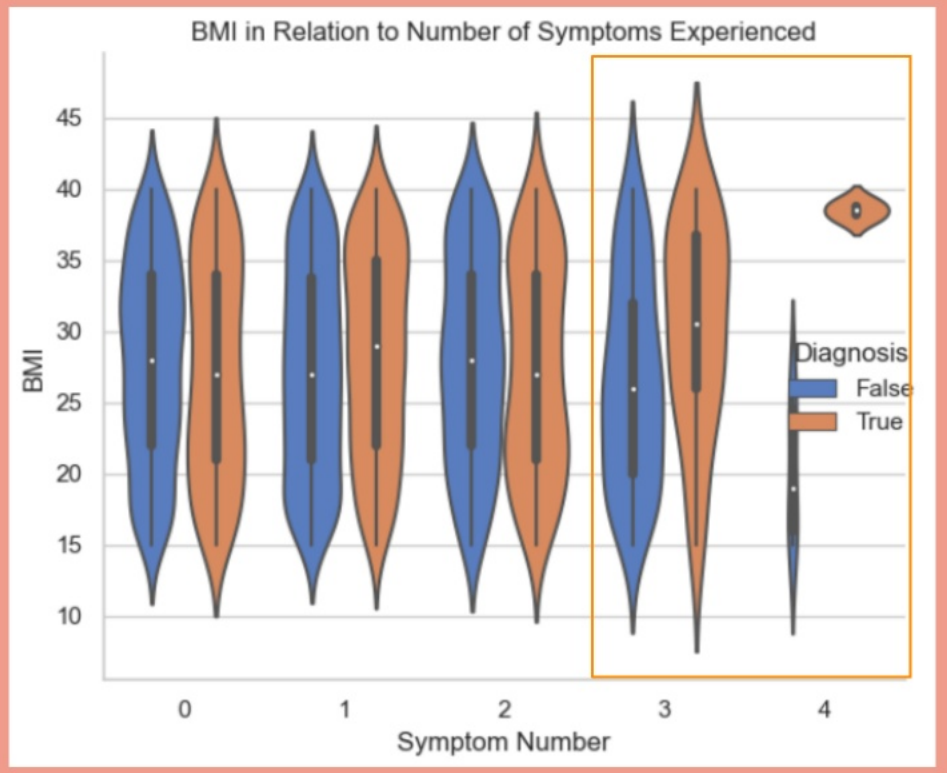
LifeStyle

BMI
Smoking
AlcoholConsumption
PhysicalActivity
DietQuality
SleepQuality






6.0
5.5
5.0
4.5
DietQuality
4
6
8
10
SleepQuality
Diagnosis
0
1


BMI in Relation to Number of Symptoms Experienced
45
40
35
30
25
20
15
10
BMI
0
1
2
3
4
Symptom Number
Diagnosis
False
True

Supervised Learning

```
alzheimer_df.columns
```

0.0s

```
dex(['BMI', 'Smoking', 'AlcoholConsumption', 'PhysicalActivity',
       'DietQuality', 'SleepQuality', 'CardiovascularDisease', 'Diabetes',
       'SystolicBP', 'DiastolicBP', 'CholesterolTotal', 'CholesterolLDL',
       'CholesterolHDL', 'CholesterolTriglycerides', 'MMSE',
       'FunctionalAssessment', 'MemoryComplaints', 'BehavioralProblems', 'ADL',
       'Confusion', 'Disorientation', 'PersonalityChanges',
       'DifficultyCompletingTasks', 'Forgetfulness', 'Diagnosis'],
      dtype='object')
```

Supervised Learning

Supervised Learning

RandomForest

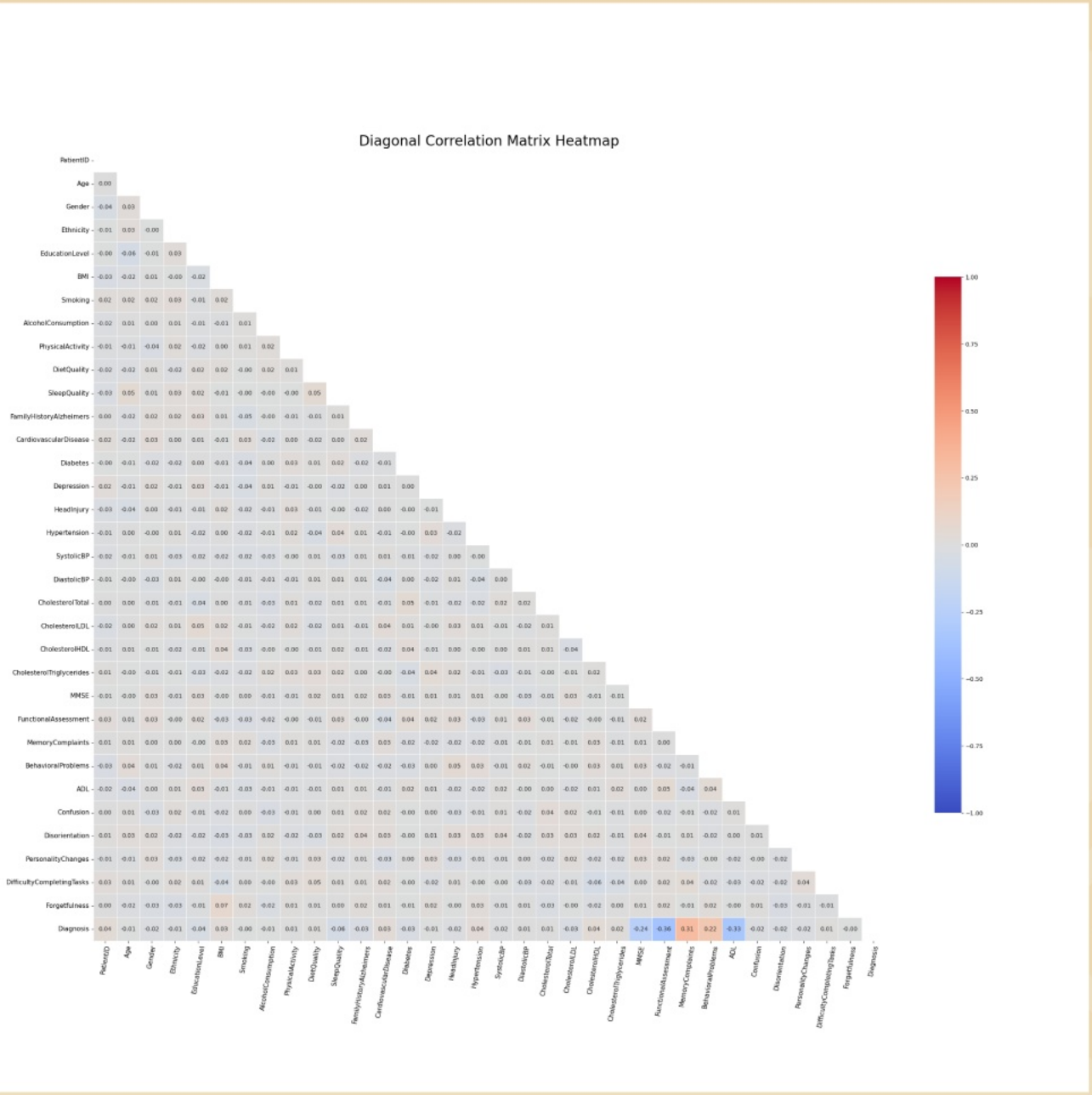
Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	335	13
Actual 1	23	167

Accuracy Score : 0.9330855018587361

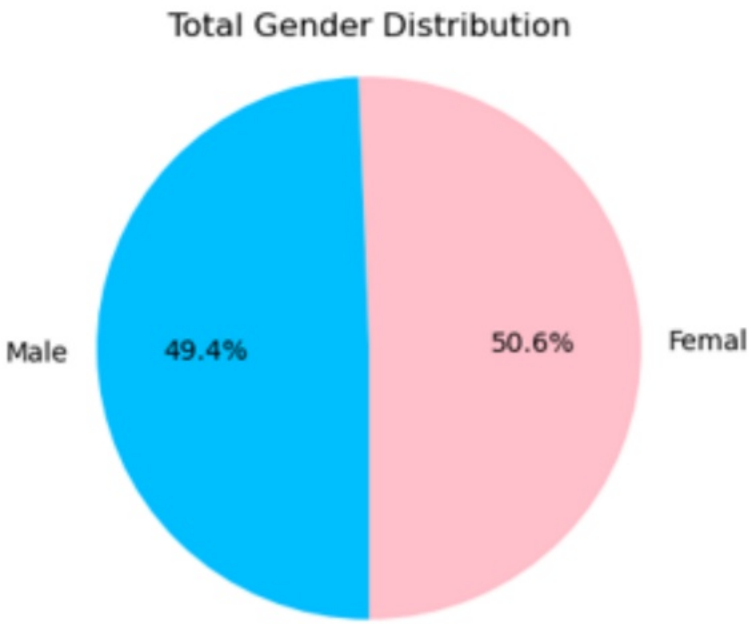
Classification Report

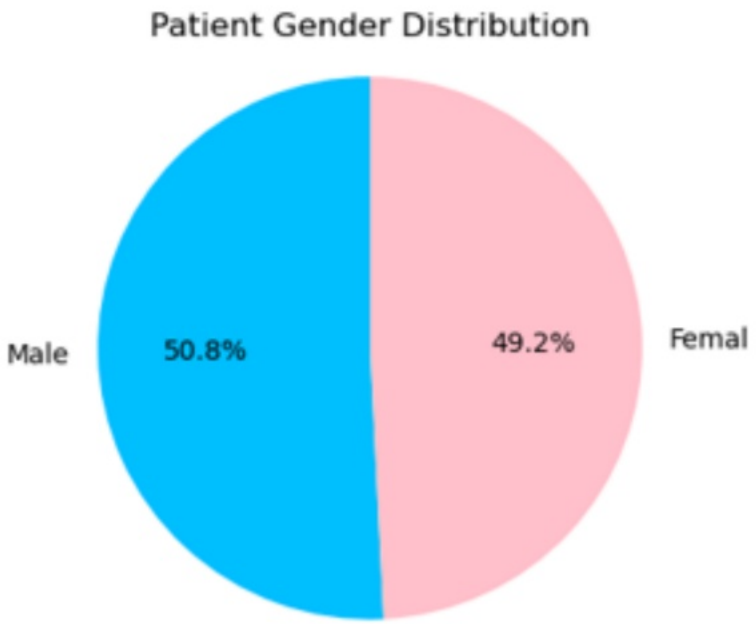
	precision	recall	f1-score	support
0	0.94	0.96	0.95	348
1	0.93	0.88	0.90	190
accuracy			0.93	538
macro avg	0.93	0.92	0.93	538
weighted avg	0.93	0.93	0.93	538

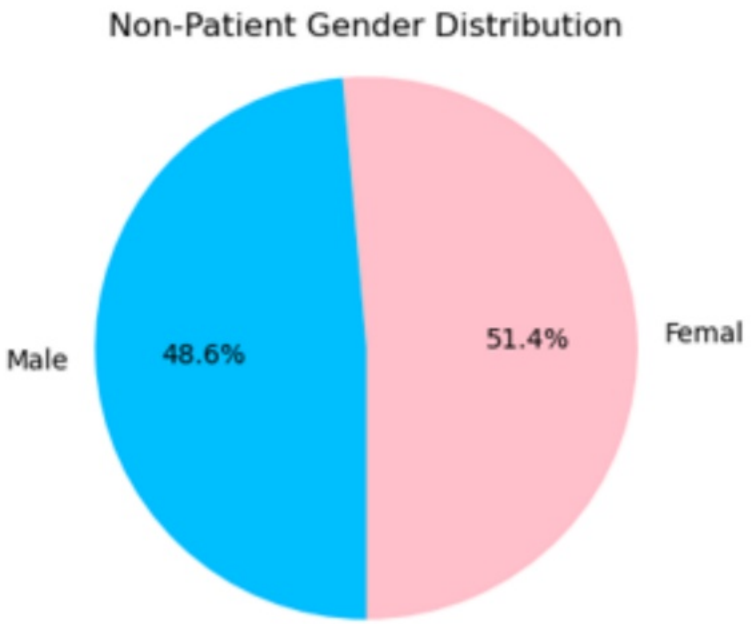


Correlation Matrix Heatmap

Five Cognitive and Functional Assessments features features demonstrate the correlation with dignosis



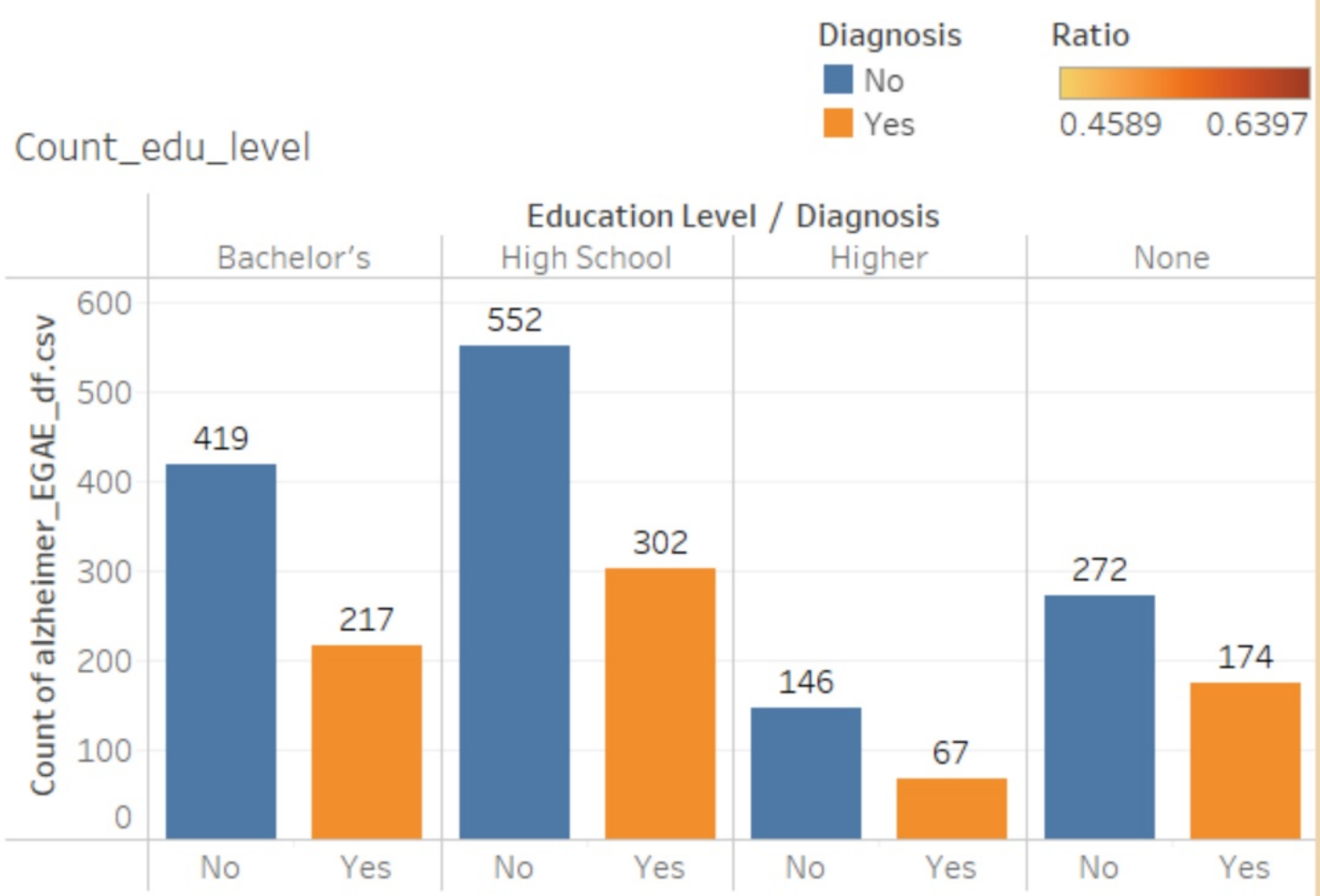


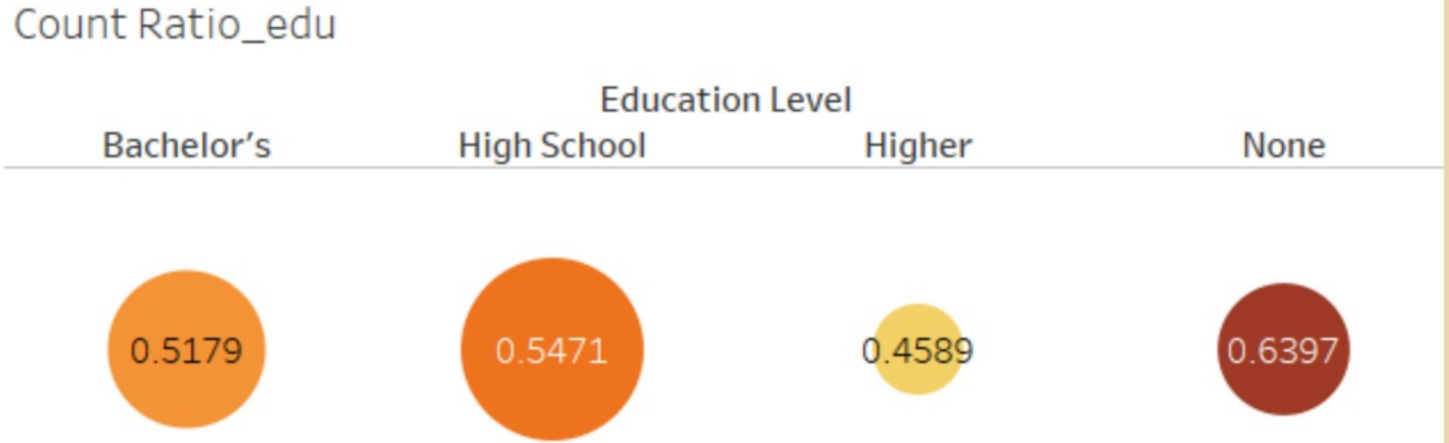


Total participants -The proportion of female participants is slightly higher than that of male participants(1.2%)

Patients - The proportion of male patients is slightly higher than that of female patients(1.6%)

Non-patients - The portion of female non-patients is slightly higher than that of male non-patients(3.2%)



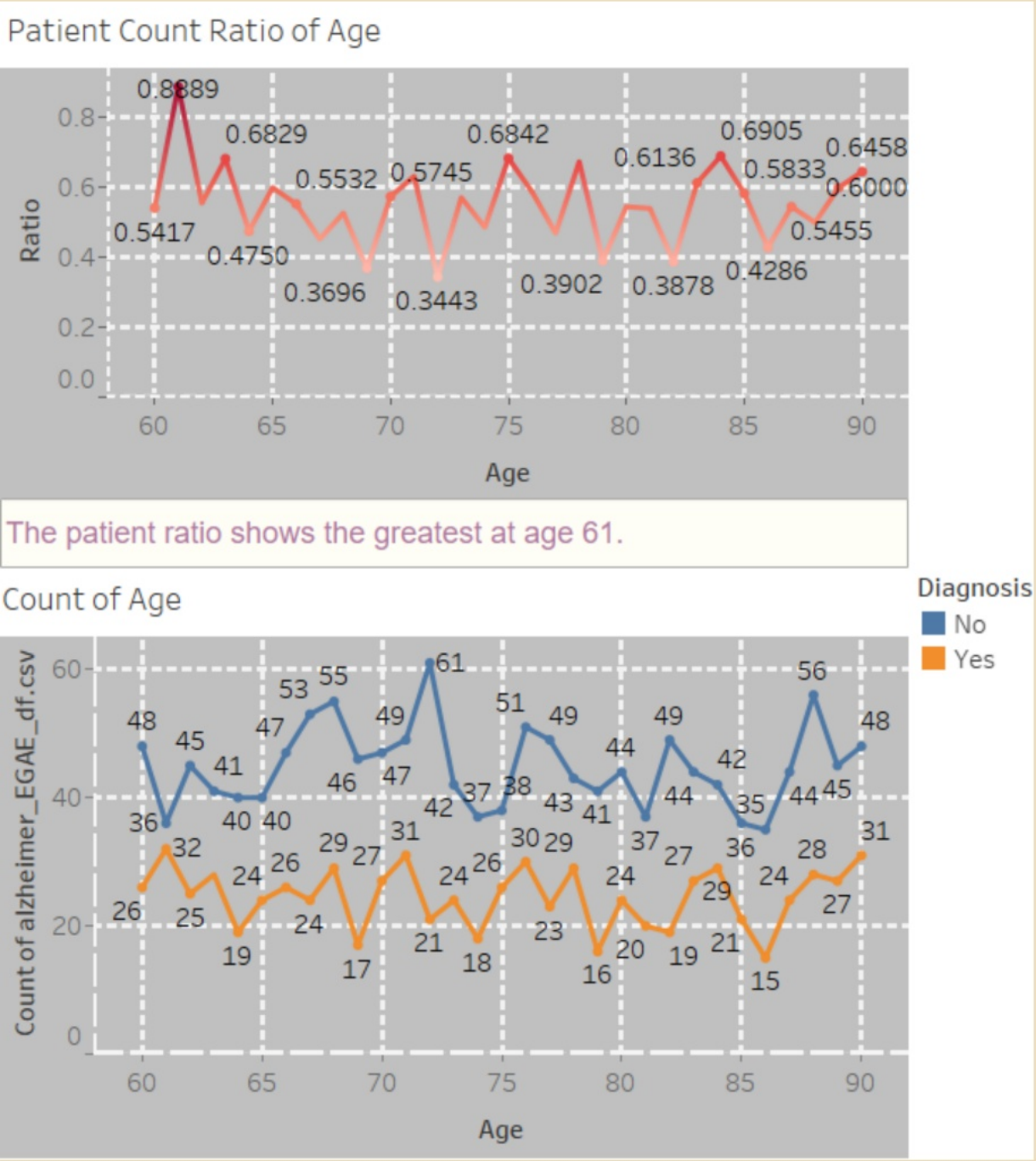


Education Level Analysis

1.Patient ratio in non education group is the highest among all education levels.

2.Patient ratio in high school and Bachlor's degree are similar.

3.Partient ratio in the higher education group is the lowest.




```
# Perform a t-test to see if the difference is statistically significant from zero
t_statistic, p_value = stats.ttest_1samp(diff, 0)
```

```
# Print the results
print(f"Mean difference: {mean_difference:.4f}")
print(f"Median difference: {median_difference:.4f}")
print(f"Standard deviation of difference: {std_difference:.4f}")
print(f"T-statistic: {t_statistic:.4f}")
print(f"P-value: {p_value:.4f}")
```

```
Mean difference: 0.0059
Median difference: 0.0060
Standard deviation of difference: 0.1114
T-statistic: 0.2886
P-value: 0.7749
```

```
print("There is no significant difference due to p-value is greater than 0.05")
```

```
There is no significant difference due to p-value is greater than 0.05
```

No significant difference of patient ratio among age groups from 60-90

Feature Selection

```
Ranked Features based on Spearman's Correlation Coefficients:
Diagnosis                    1.000000
FunctionalAssessment         0.366687
ADL                          0.330450
MemoryComplaints             0.306742
MMSE                         0.236271
BehavioralProblems           0.224350
SleepQuality                 0.056069
EducationLevel               0.043325
CholesterolHDL               0.042542
Hypertension                 0.035080
FamilyHistoryAlzheimers      0.032900
CholesterolLDL               0.032010
Diabetes                     0.031508
CardiovascularDisease        0.031490
BMI                          0.026402
Disorientation               0.024648
CholesterolTriglycerides     0.023072
HeadInjury                   0.021411
Gender                       0.020975
PersonalityChanges           0.020627
```

6 Machine Learning by Top Features

5 **Feature Selection by Top Correlation >0.2**

4 Machine Learning by Category

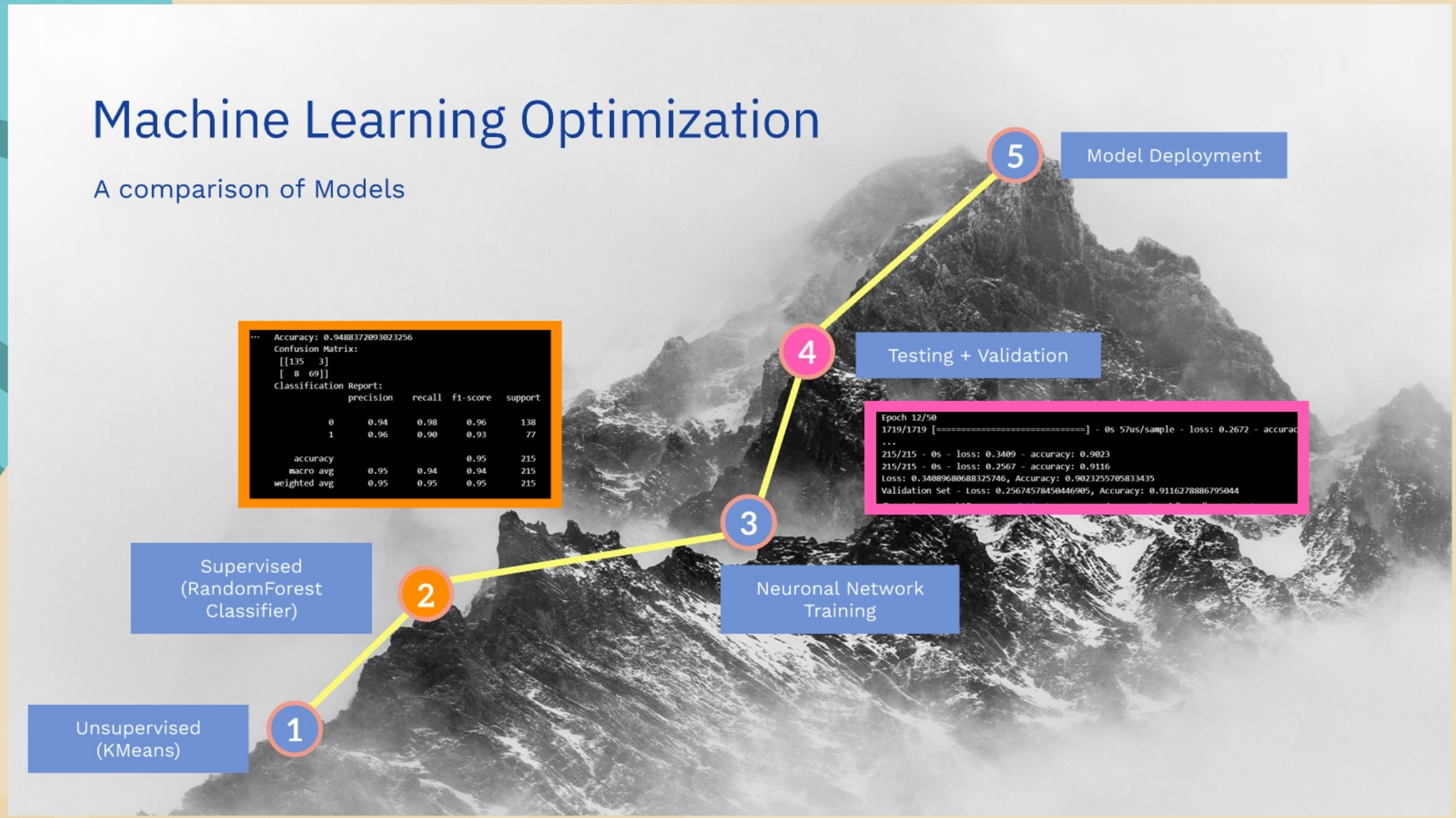
3 Pattern Visualization by Category

2 Feature by Major Category

1 Data Cleaning

Machine Learning Optimization

A comparison of Models



The Predictive Capacity of an Alzheimer Machine Learning Model

Target:

- Diagnosis (of Alzheimer's)

Features:

- Age
- Gender
- Ethnicity
- Education Level
- BMI
- Smoking
- Alcohol Consumption
- Physical Activity
- Diet Quality
- Sleep Quality
- Family History of Alzheimer's
- Cardiovascular Disease
- Diabetes
- Depression
- Head Injury
- Hypertension
- Systolic BP
- Diastolic BP
- Cholesterol Total
- Cholesterol LDL
- Cholesterol HDL
- Cholesterol Triglycerides
- MMSE
- Functional Assessment
- Memory Complaints
- Behavioral Problems
- ADL
- Confusion
- Disorientation
- Personality Changes
- Difficulty Completing Tasks
- Forgetfulness

Limitations and Future Directions

- Skewed dataset
- Confounding factors
- Feature ambiguity

- Increase data pool
- Extended feature combinations
- Clarify feature collection criteria