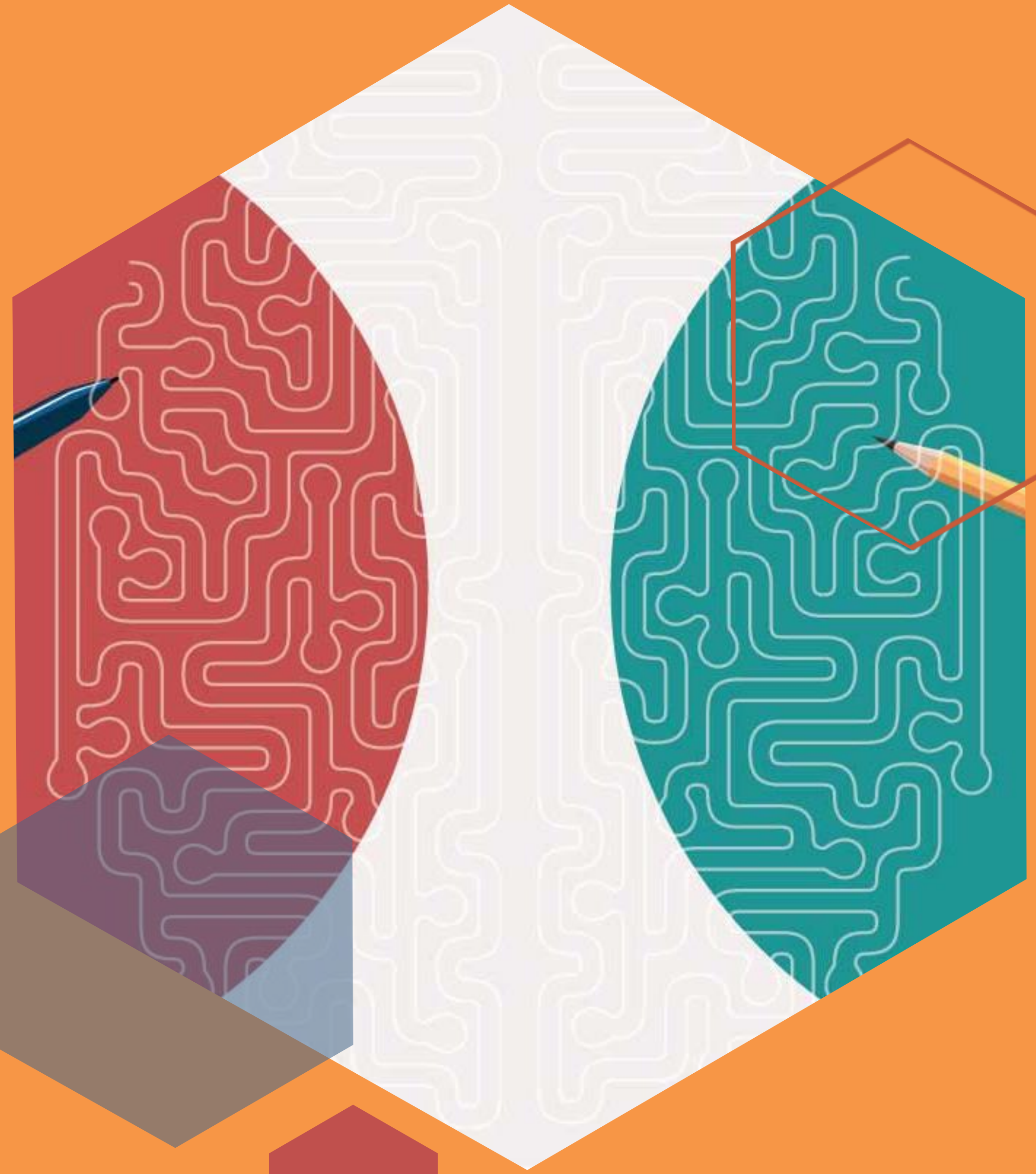
EEG: Brain & Handedness

Team 9

Josiah Tsang
Lisa Skelton
Martina Chiesa
Muyang Li
Sofia Caltabiano
Simon Cotterill



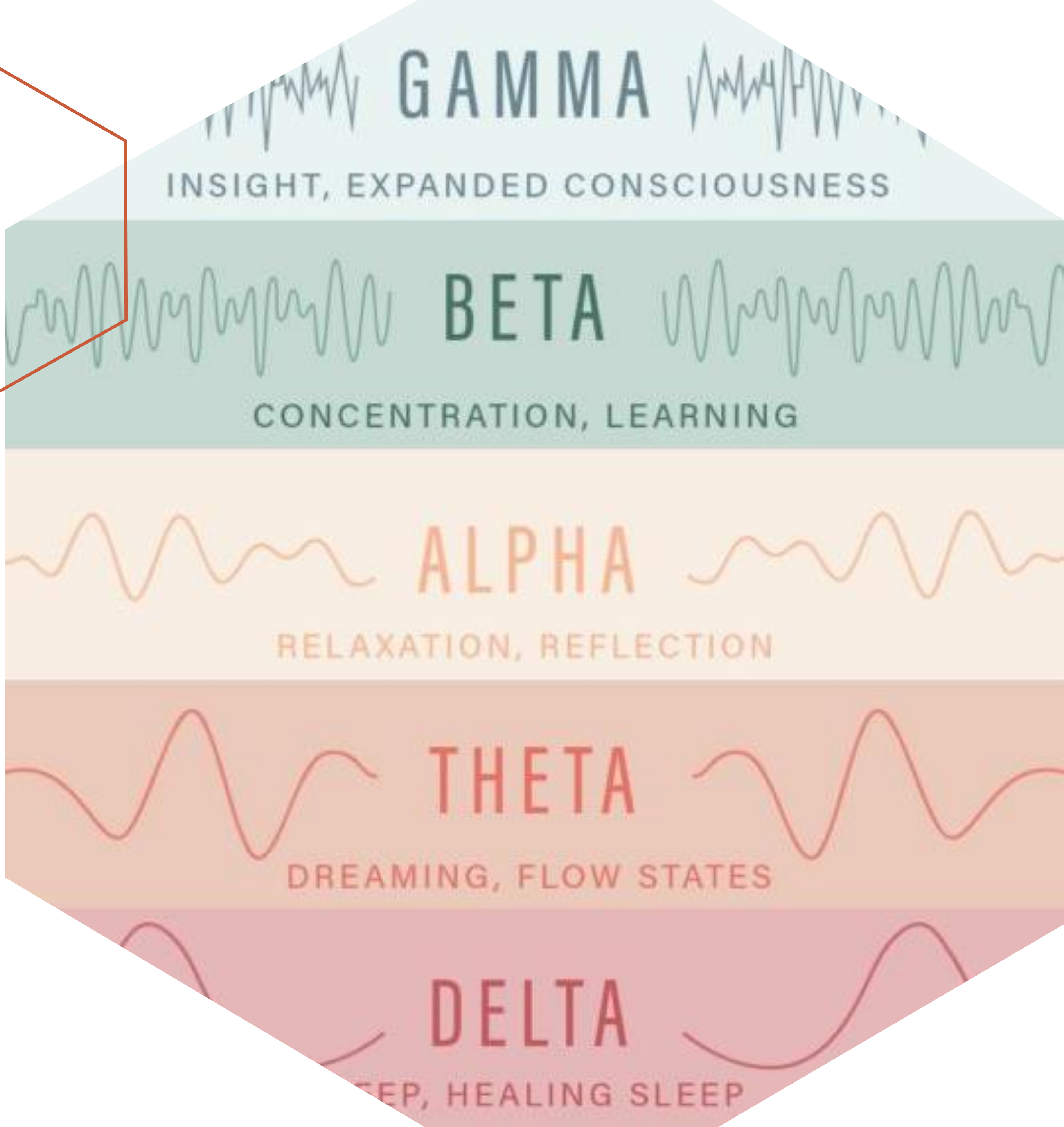
Overview

- Introduction
- Data Collecting
- Data Processing
- Machine Learning
- Results

Introduction

Handedness is inversely correlated to the brain's dominant hemisphere.

Our project sought to determine a person's dominant hand from their brainwaves.



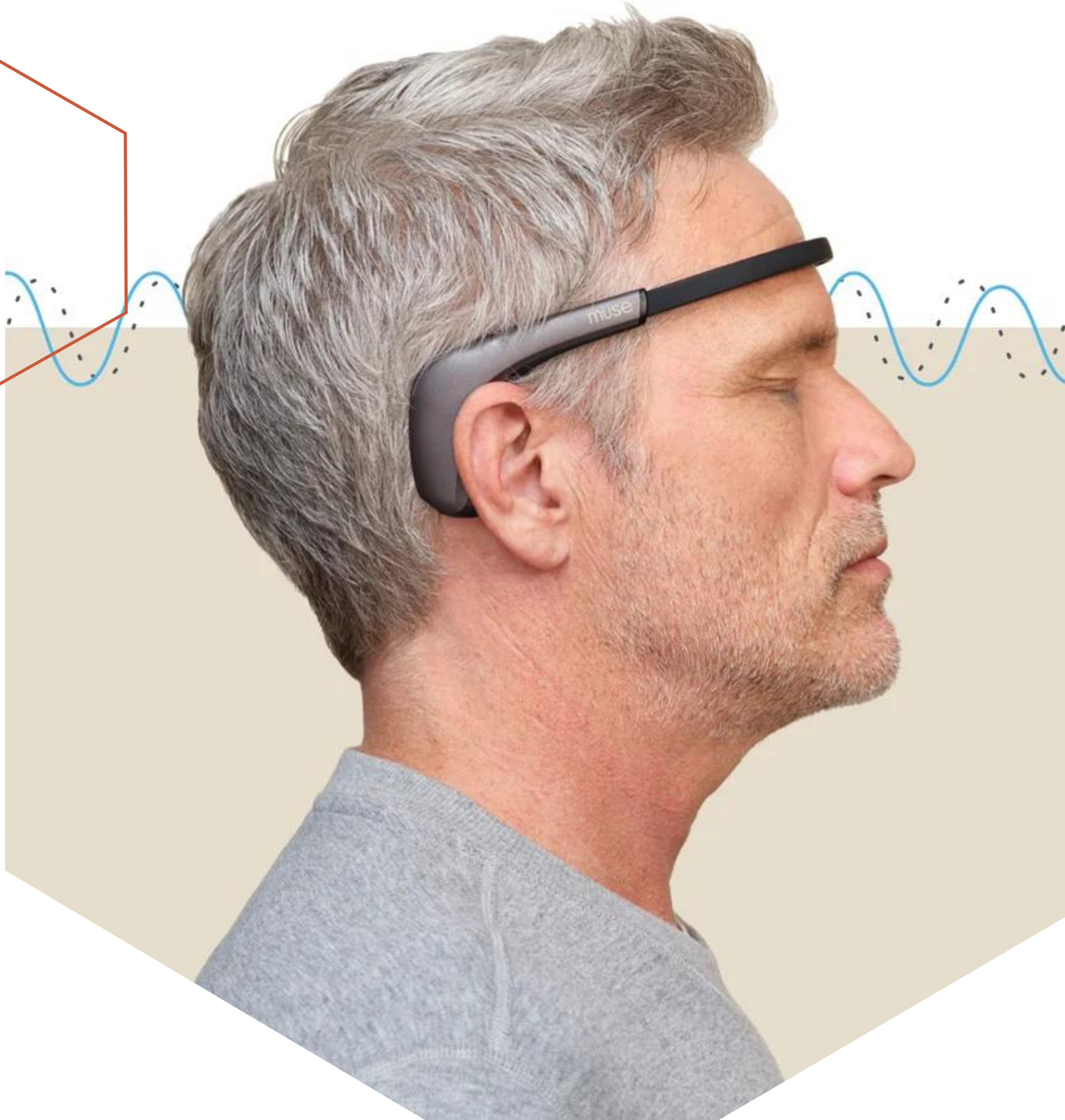
Muse

The muse is an EEG device that reads your brainwaves. It is designed as a neurofeedback tool to help a person improve their ability to meditate.

In this project we repurposed the muse device to collect raw EEG data during our testing using an aftermarket app- **Mind Monitor**

Objective

- Use machine learning to predict a person's dominant hand.
- Distinguish between a random hand motion (scribble) and a deliberate action (legible writing).



Data Collecting

Testing

Manu 465 EEG Group 4 10/21/2022

LEFT HAND SAMPLES	RIGHT HAND SAMPLES
Print “The quick brown fox jumped over the lazy dog.” in this box	Print “The quick brown fox jumped over the lazy dog.” in this box

Scribble inside the circle 30 Seconds

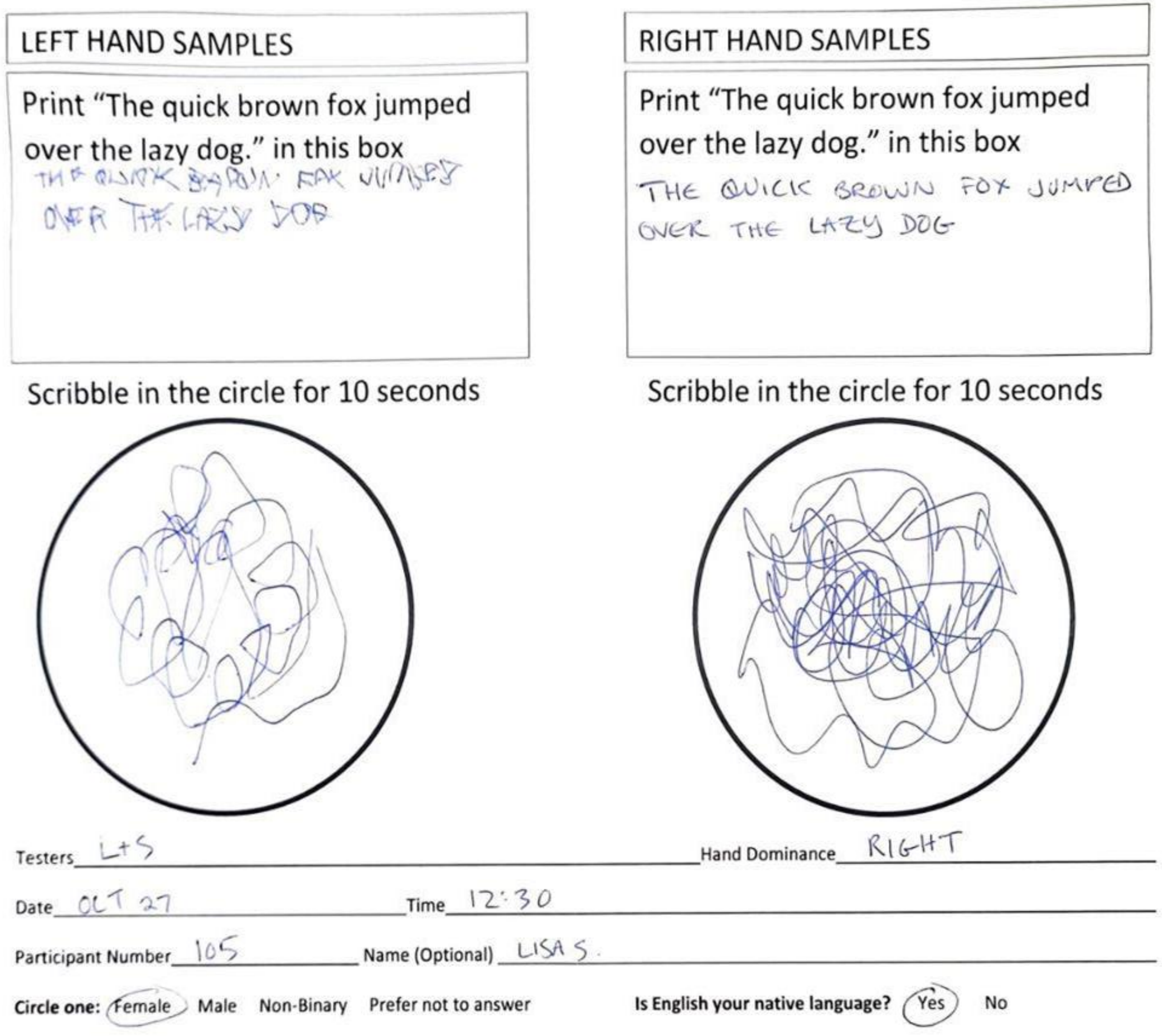
Scribble inside the circle 30 Seconds

Testers____________________ Hand Dominance____________________

Date____________________ Time____________________

Participant Number____________________ Name (Optional) ____________________

Testing



LEFT HAND SAMPLES

Print "The quick brown fox jumped over the lazy dog." in this box

THE QUICK BROWN FOX JUMPED OVER THE LAZY DOG

Scribble in the circle for 10 seconds

RIGHT HAND SAMPLES

Print "The quick brown fox jumped over the lazy dog." in this box

THE QUICK BROWN FOX JUMPED OVER THE LAZY DOG

Scribble in the circle for 10 seconds

Testers L+S

Hand Dominance RIGHT

Date OCT 27

Time 12:30

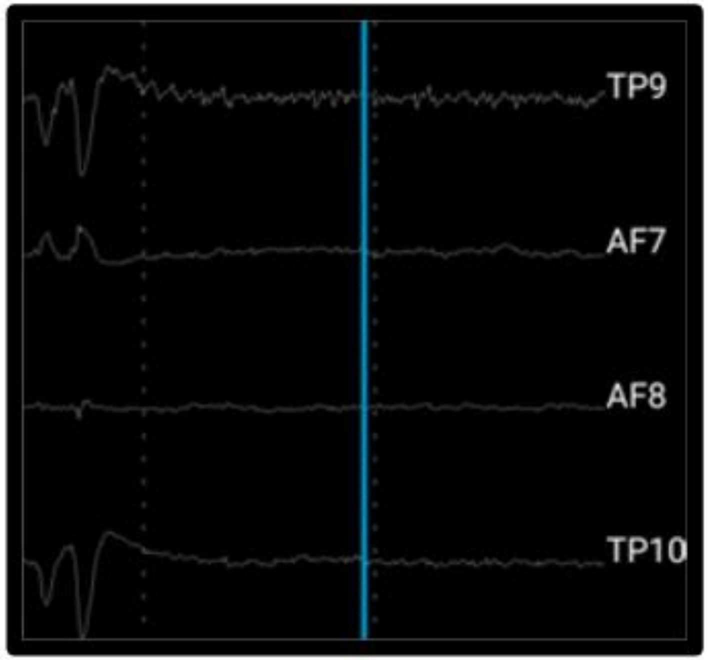
Participant Number 105

Name (Optional) LISA S.

Circle one: Female Male Non-Binary Prefer not to answer

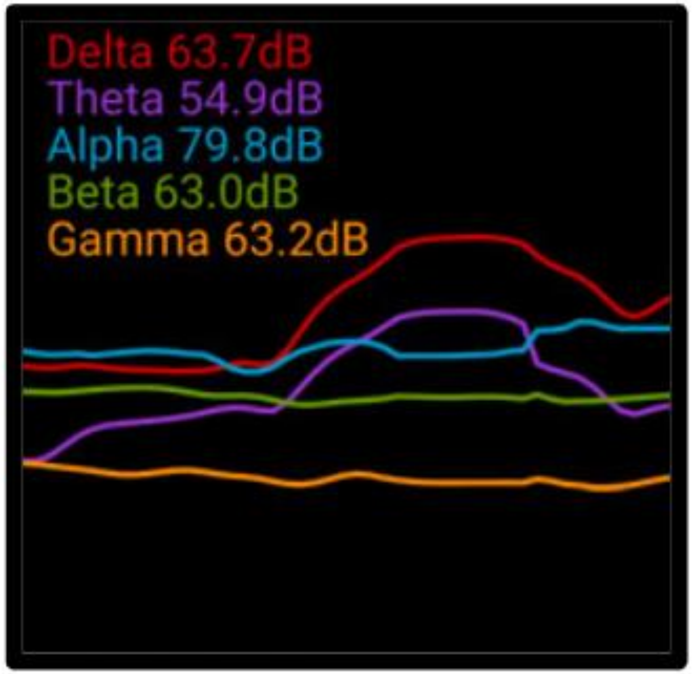
Is English your native language? Yes No

Testing: Muse UI



Raw EEG Values

Sensors



Brain Waves

Brain wave values

TimeStamp	Delta_TP9	Delta_AF7	Delta_AF8	Delta_TP10	Theta_TP9	Theta_AF7	T
43:07.5							
43:07.5	0	0.955974	0.63044	1.039788	0	0.231531	
43:08.5	0	0.003765	0.072892	0.994671	0	-0.25384	
43:09.6	0	-0.3396	-0.26912	0.863073	0	-0.39025	
43:10.6	0	-0.14214	-0.44765	0.863073	0	-0.14308	
43:11.6	0	-0.37619	-0.33304	0.863073	0	-0.35361	
43:12.6	0	-0.26752	-0.22857	1.275989	0	-0.05308	
43:13.6	0	-0.12623	0.002218	1.258686	0	-0.0689	
43:14.6	0	0.449572	0.18539	1.258686	0	0.170049	
43:15.6	0	0.304969	0.06374	1.258686	0	-0.07405	
43:16.6	0	-0.04776	-0.10239	0.675701	0	-0.03897	
43:17.6	0	0.413807	0.166191	0.969467	0	0.240169	
43:18.6	1.649748	0.433776	0.180397	1.19384	1.880369	-0.07054	
43:19.6	1.628041	0.223583	-0.29801	0.52118	1.092966	-0.16998	
43:20.6	1.643404	0.205837	0.25472	0.764221	0.940859	0.058531	
43:21.7	1.643404	0.184698	0.316756	0.859777	0.940859	-0.15362	
43:22.7	1.564125	0.280636	0.00663	0.915244	1.486856	-0.08998	
43:23.7	1.638804	0.15045	-0.05948	0.939615	1.391405	0.030641	

Raw Values

RAW_TP9	RAW_AF7	RAW_AF8	RAW_TP10
811.0989	772.8205	791.7582	790.9524
793.3699	775.6411	796.9963	826.0073
846.9597	785.7143	800.2198	859.0476
795.3846	778.4616	790.5494	803.4432
780.8792	770.403	788.9377	772.8205
875.5678	776.044	798.2051	845.348
794.5787	772.0147	787.7289	800.6227
689.011	769.1942	786.1172	782.0879
803.4432	776.044	788.1319	807.8755
816.337	785.7143	790.5494	808.6813
851.7949	778.4616	794.5787	840.1099
799.4139	776.8498	792.967	835.6777
809.0842	770.403	788.1319	823.5897
792.1612	777.2528	792.1612	808.2784
795.7875	769.5971	795.3846	811.5018
802.6374	773.2235	789.7436	806.2637
782.0879	774.8351	789.3406	810.293

Participant	Test	Gender	English	Dominance
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right
127	RHC	Male	Yes	Right

Whittaker–Nyquist–Shannon

Twice the highest frequency:

4Hz x 2 = 88Hz

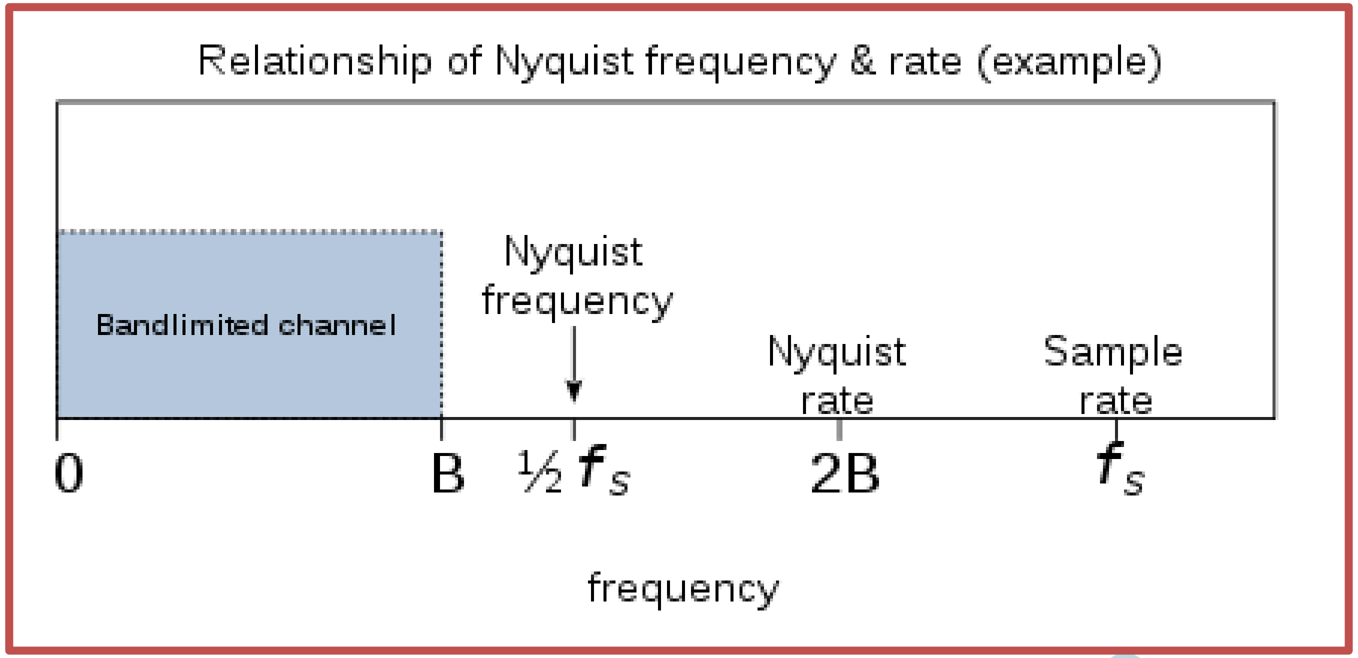
Required Sampling Rate:

Frequency = 1/Time

1/88 ≈ 0.011s

Comparing Rates:

Every 1s << Every 0.011s

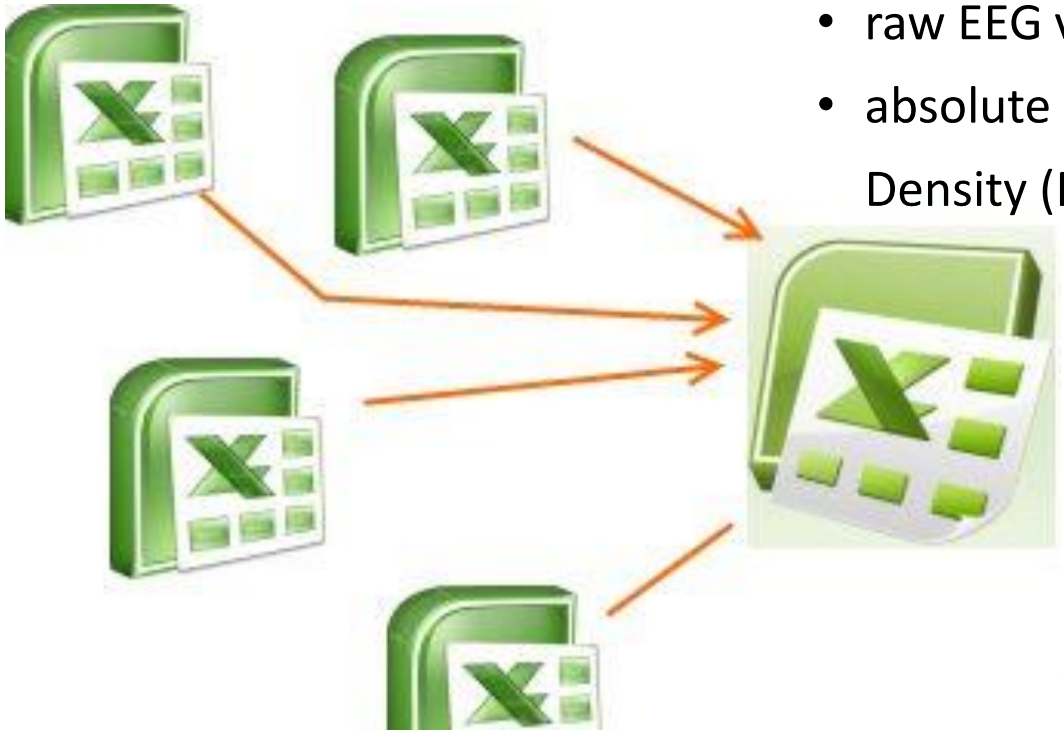


Data Processing

Data Processing

Data Collected:

- Each participant generated 4 .csv files
- Each csv file contains two different types of values for each waves, collected in time domain:
  - raw EEG values show each sensor raw data in microvolts
  - absolute band powers (logarithmic) of the Power Spectral Density (PSD) of the EEG data per channel



Target Output file:

- A single csv for all participants (one data frame)
- Each row represents one test

Feature Generation:

11 Functions were performed on the combined csv file to generate new features

Features generation

In order to condensate each csv file in only one row, we need to use summary metrics.

For each wave, which corresponds to 4 columns in each file, we want one value of mean, one value of std, and so on.
a unique row that represents one csv file.

In [171…

```
def mean(x):
    return np.mean(x, axis=0)
def std(x):
    return np.std(x, axis=0)
def ptp(x):
    return np.ptp(x, axis=0)
def var(x):
    return np.var(x, axis=0)
def minim(x):
    return np.min(x, axis=0)
def maxim(x):
    return np.max(x, axis=0)
def argminim(x):
    return np.argmin(x, axis=0)
def argmaxim(x):
    return np.argmax(x, axis=0)
def rms(x):
    return np.sqrt(np.mean(x**2, axis=0))
def skewness(x):
    return stats.skew(x,axis=0)
def kurtosis(x):
    return stats.kurtosis(x,axis=0)

def concatenate_features(x):
    '''''''''
    this function apply several functions defined above.
    It takes as input a numpy array.
    It outputs a vector with the value of each function: mean, std, ...
    '''''''''
    return mean(x),std(x),ptp(x),var(x),minim(x),maxim(x),argminim(x),argmaxim(x),rms
```

Dataset Creation:

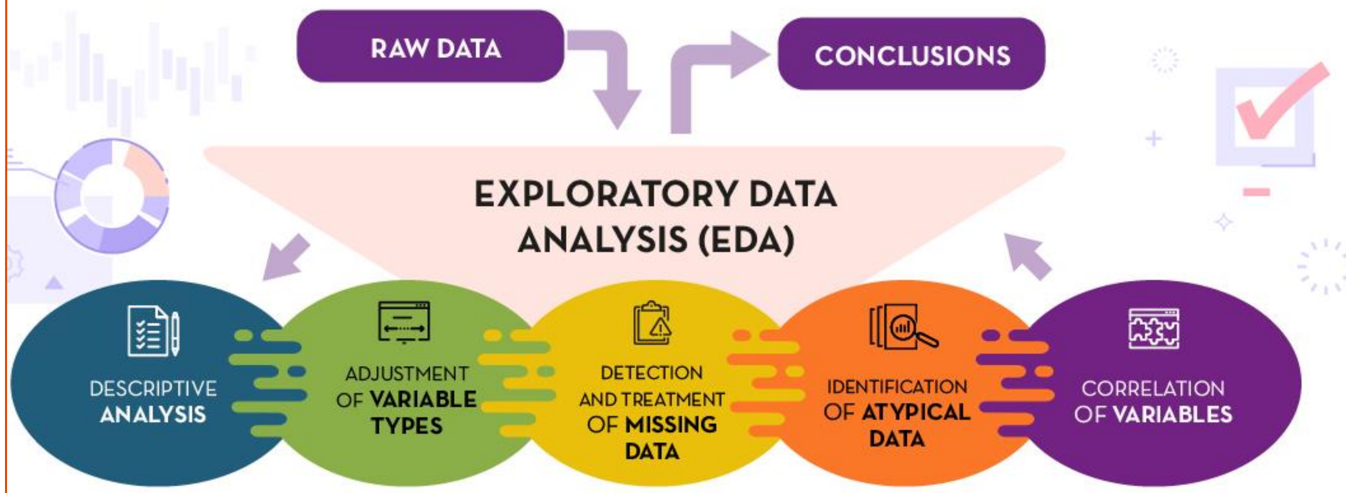
- Added keys to each row (Participant number, Test) and personal information (Gender, English)
- Coding response variable in Dominant and Non-Dominant
- Values from feature generation

Participant	Dominance	Delta_mean	Delta_std	Delta_ptp	Delta_var
101.0	Dominant	0.000000	0.000000	0.000000	0.000000
102.0	Dominant	0.526442	0.294538	1.336331	0.086752
103.0	NonDominant	0.340579	0.413665	1.775144	0.171119
139.0	Dominant	0.933517	0.513017	1.899951	0.263186
106.0	NonDominant	0.567189	0.504123	2.265507	0.254140
...	...	...	...	...	...
322.0	NonDominant	0.582510	0.434822	1.406913	0.189070
323.0	Dominant	0.504330	0.501552	1.719051	0.251554
324.0	NonDominant	0.204568	0.381588	1.164488	0.145609
325.0	Dominant	0.472270	0.334002	0.989795	0.111557
326.0	Dominant	0.679960	0.580307	1.580237	0.336757

EDA

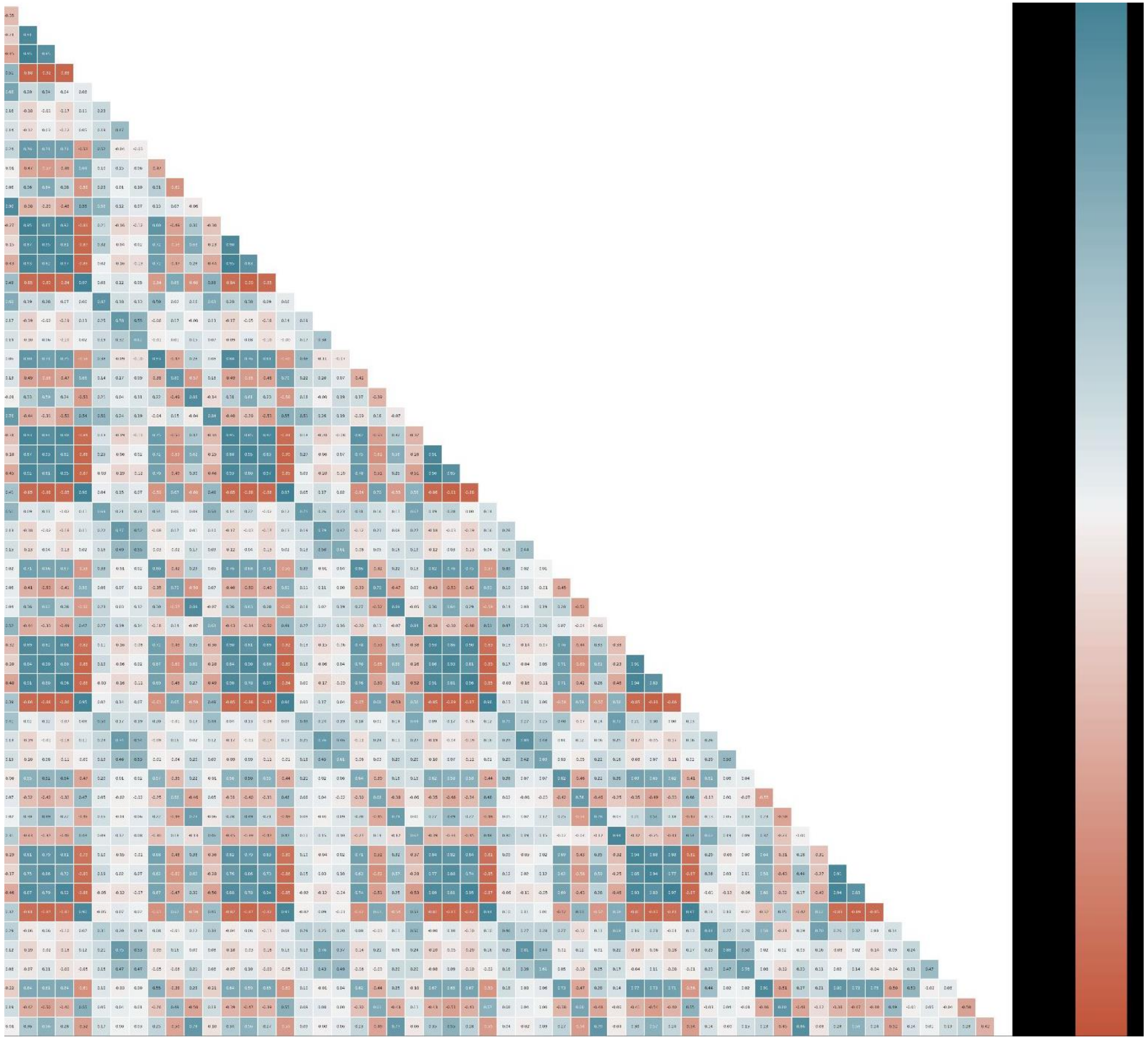
Before applying any machine learning technique, it is important to do some exploratory data analysis (EDA).

EDA is the first step to investigate the data and allow to understand the data.



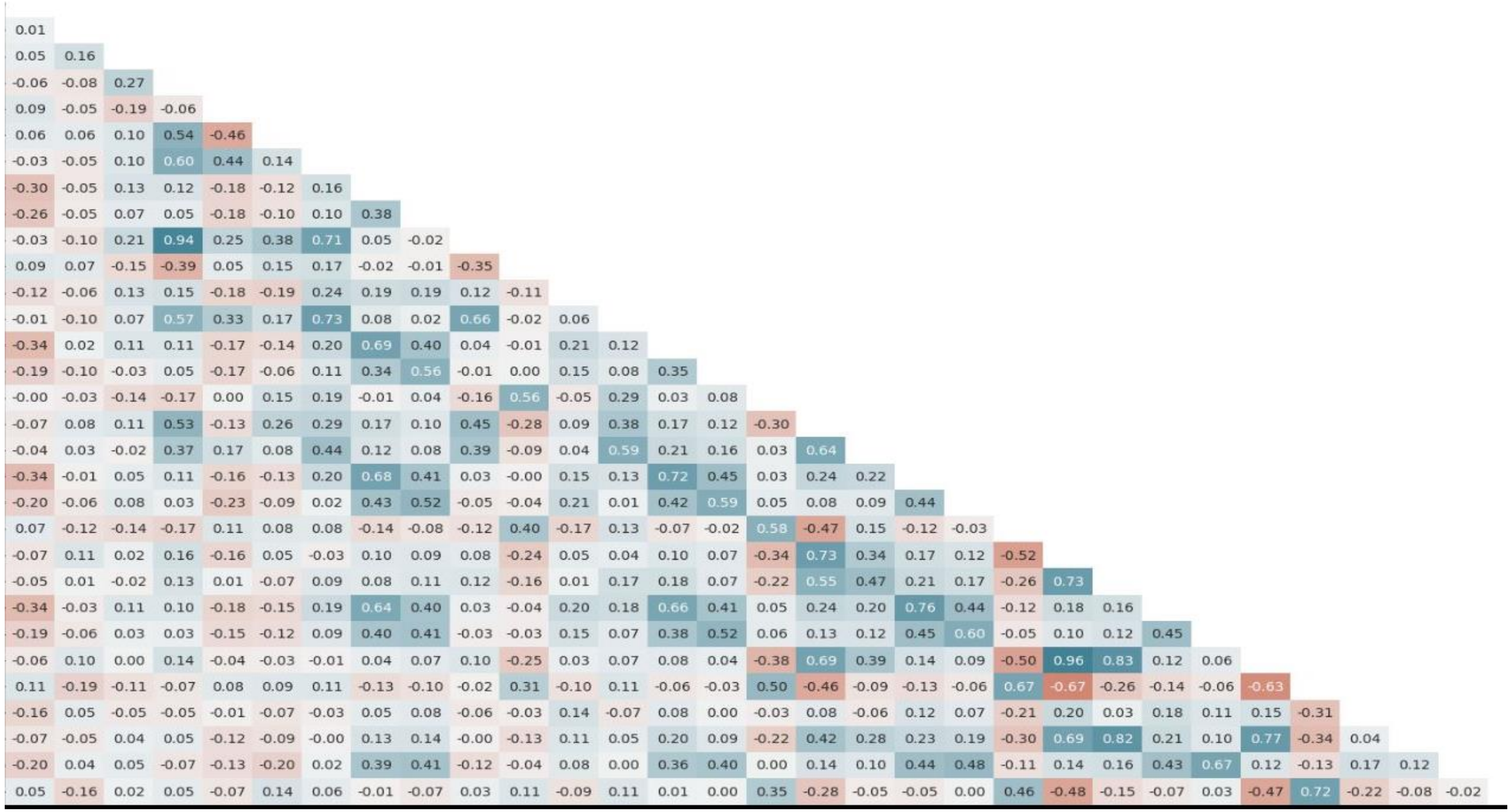
Data Visualization

Correlation Matrix



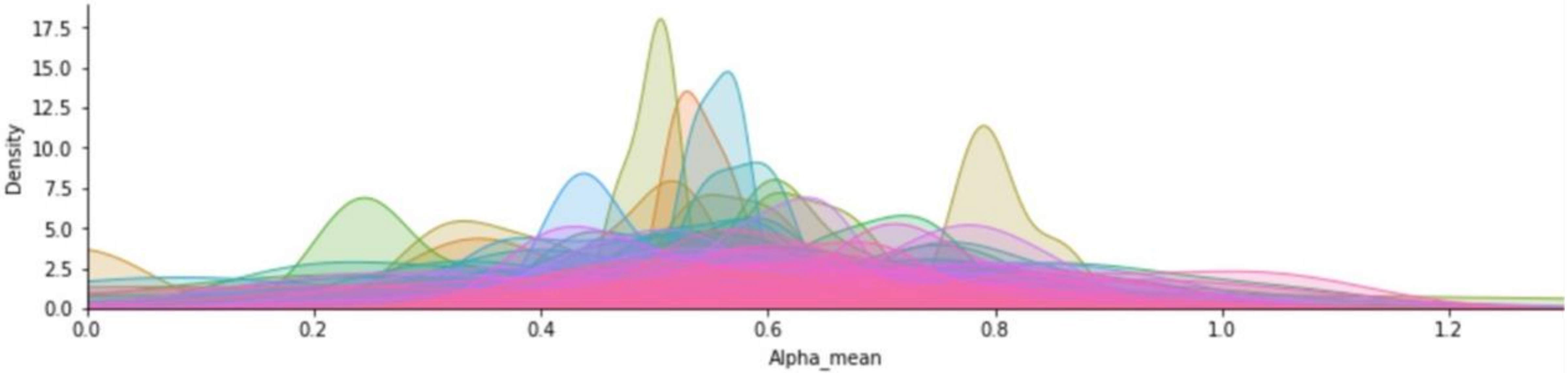
Data Visualization

Applying PCA and removing highly correlated features results in this updated Correlation Matrix



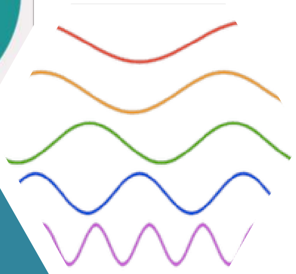
Data Visualization

Distribution of raw sensor data (Alpha)



Observations from dataset:

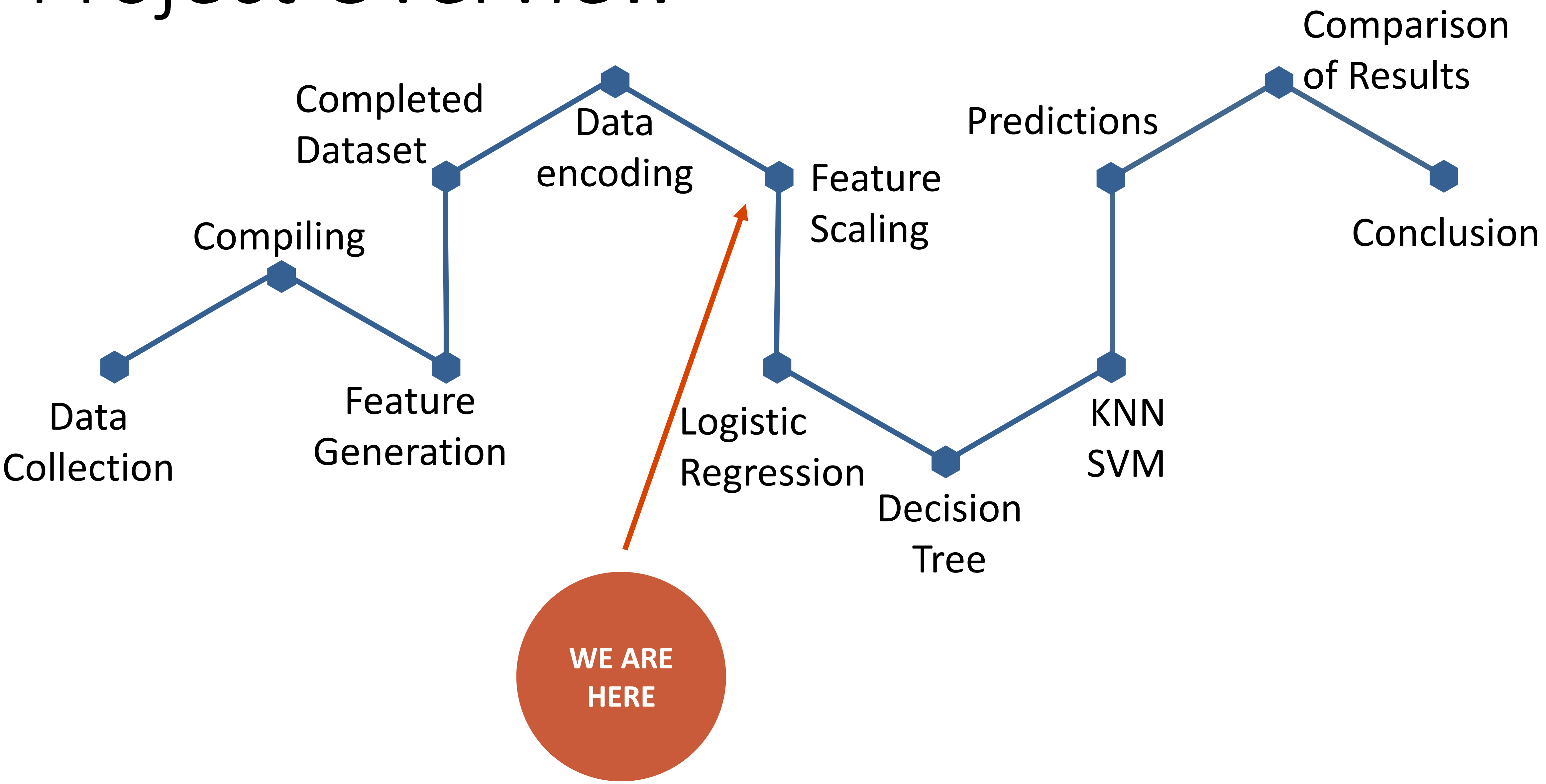




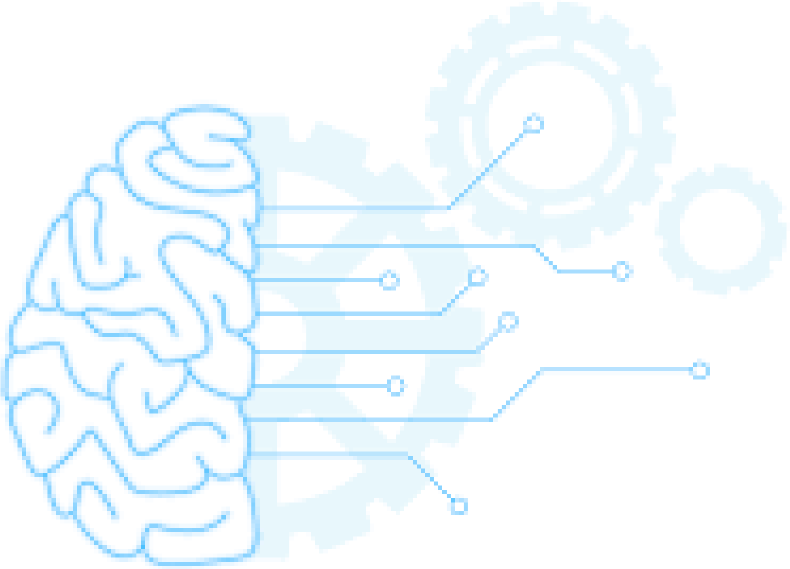
Some variables are highly correlated
Different scale → standardization

Types:
- mostly are numeric (float)
- few are categorical

Project Overview



Machine Learning



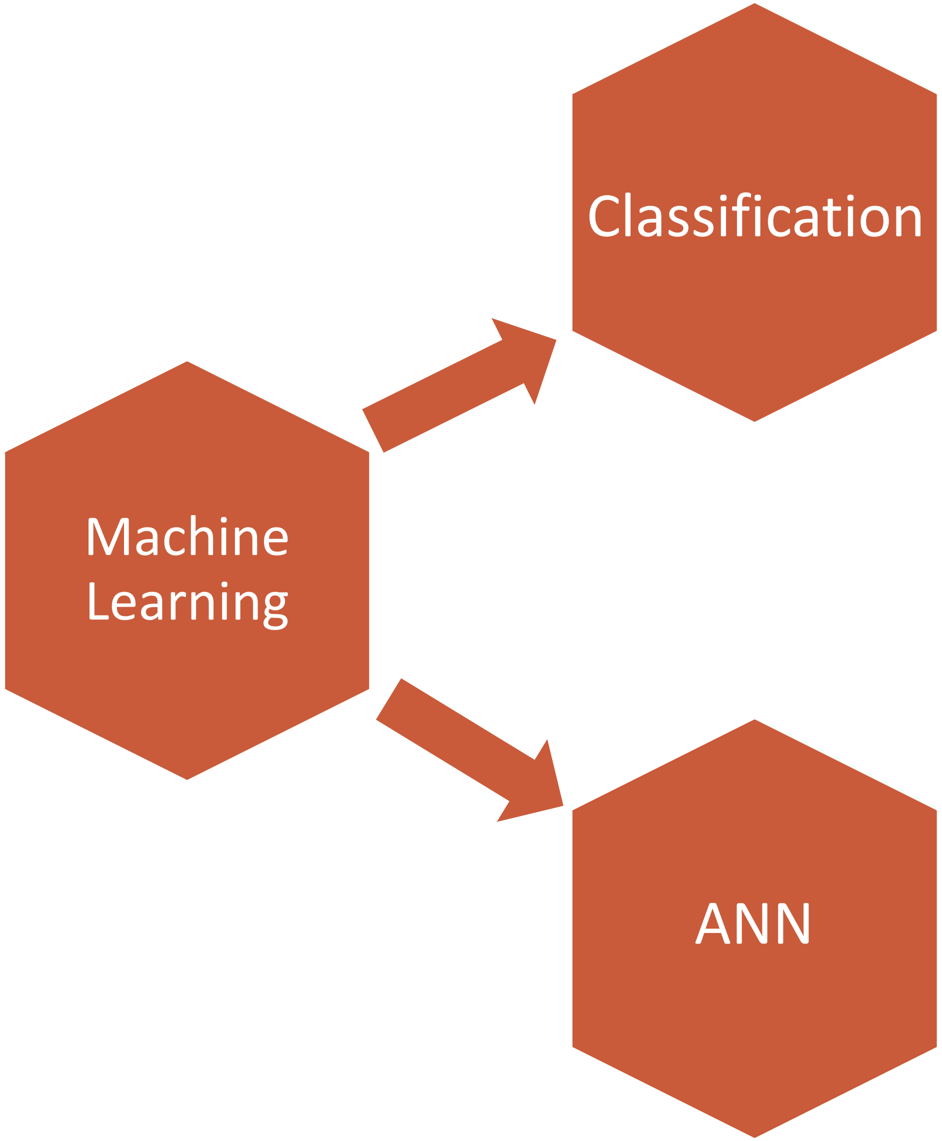
Introduction

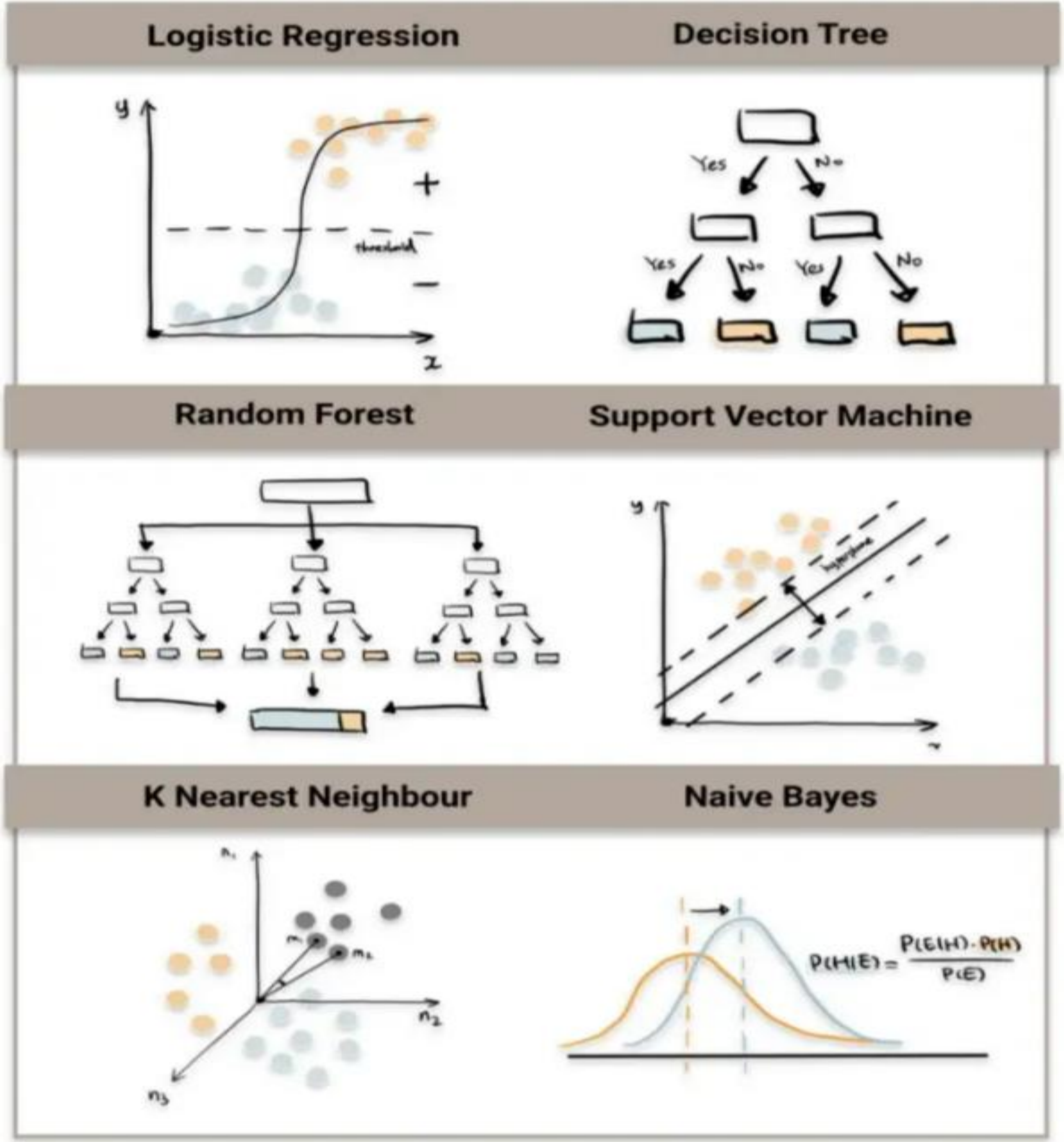
Data Collecting

Data Processing

Machine Learning

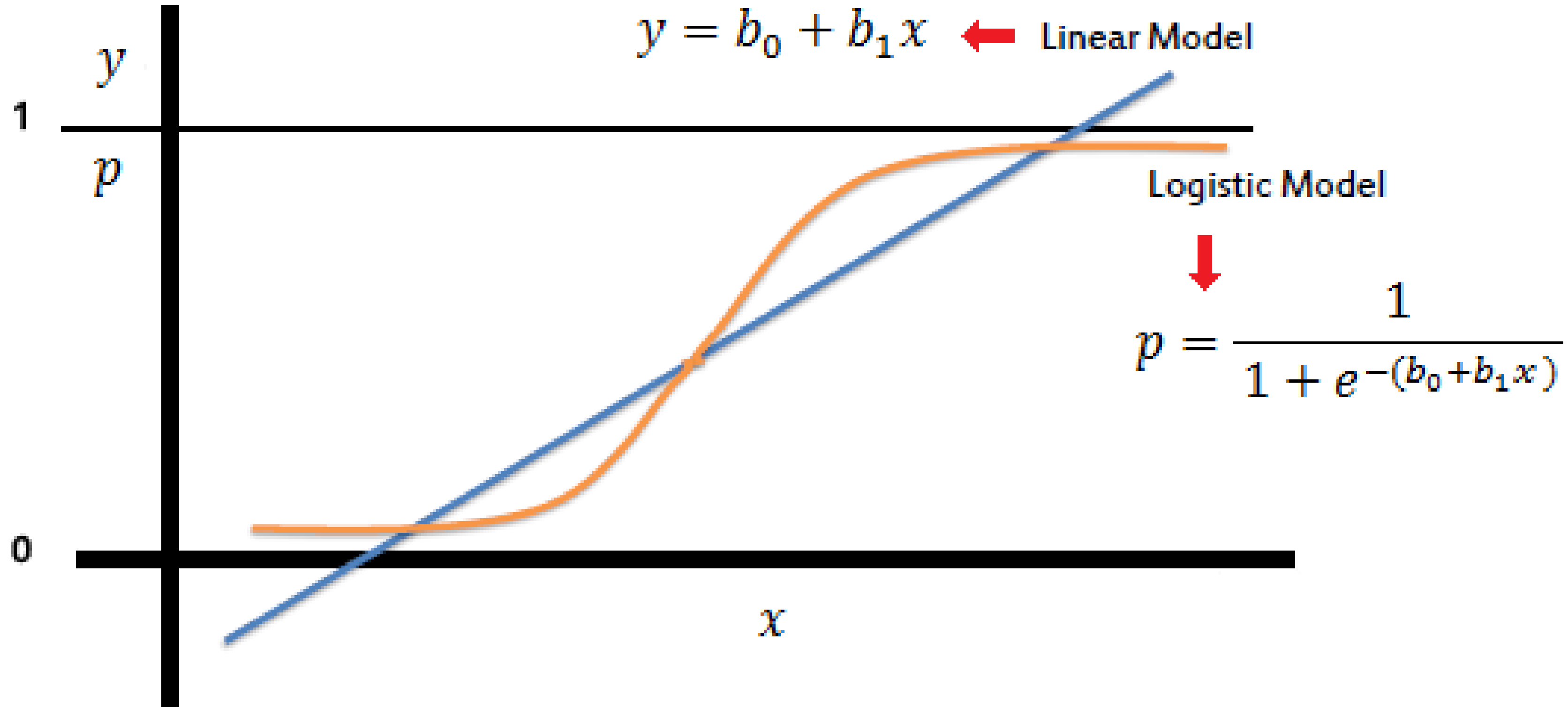
Results


Classification
Machine Learning
ANN

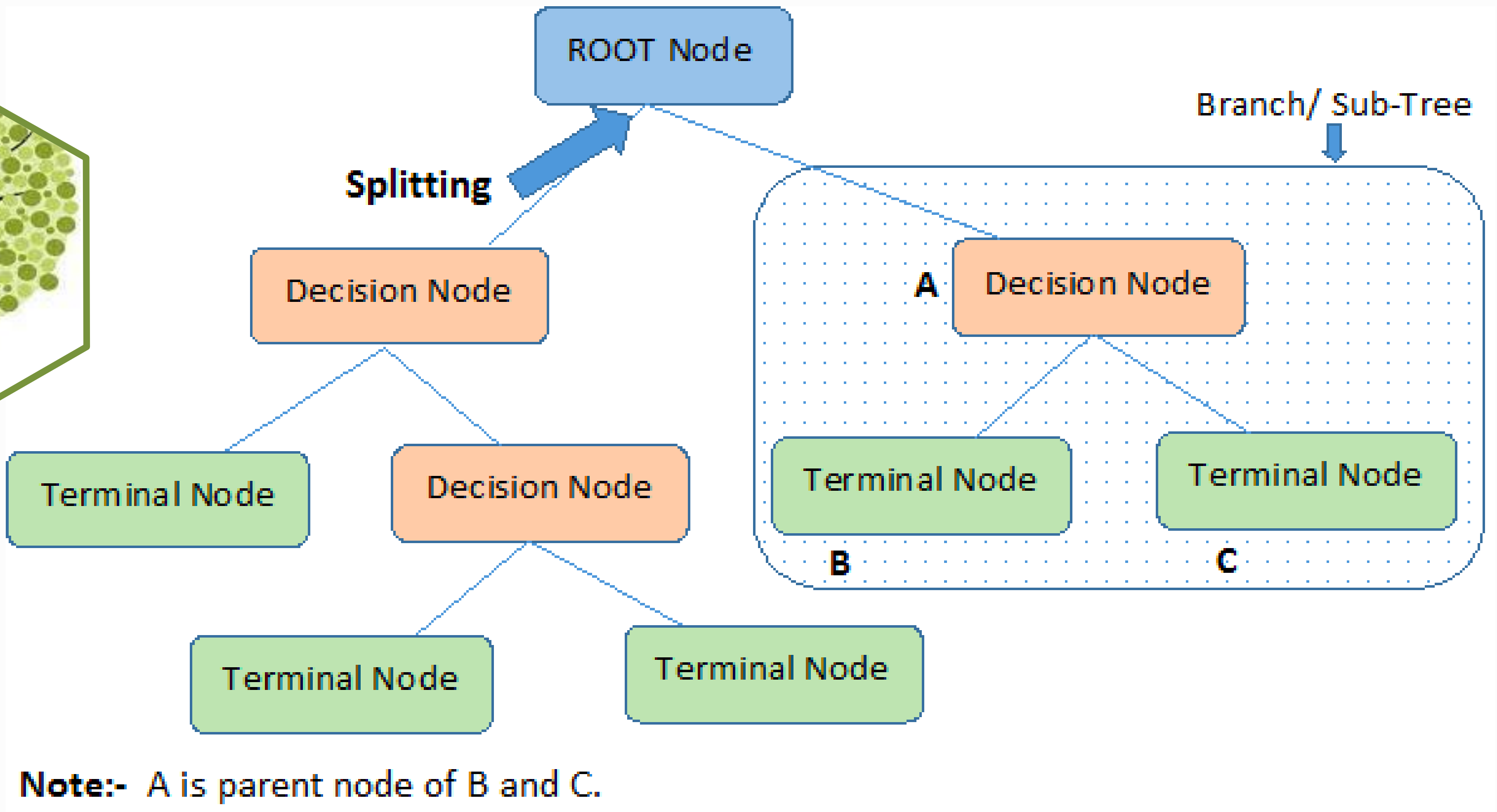

Logistic Regression
Decision Tree
Yes
No
Random Forest
Support Vector Machine
K Nearest Neighbour
Naive Bayes
P(H|E) = P(E|H)·P(H) / P(E)

Binary Logistic Regression: 80.82% test accuracy

Logistic regression is a process of estimating the parameters of a logistic model, a type of sigmoid function that can take any real-valued number and map it into a value between 0 and 1.



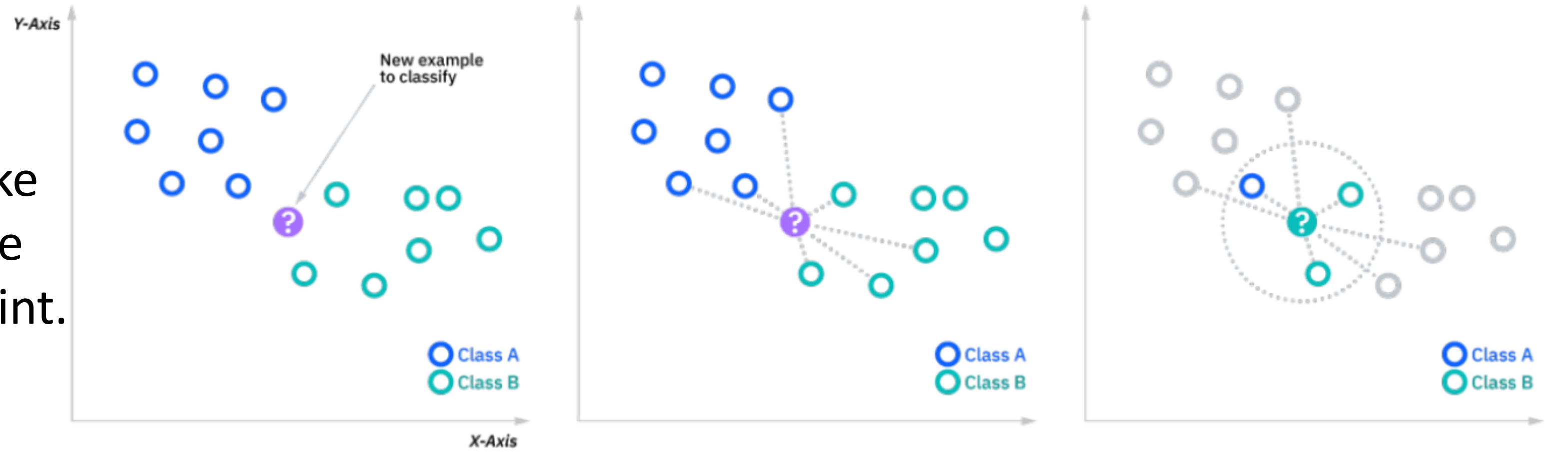
Decision Tree: 79.45% model accuracy



Other Machine Learning Models

KNN Classifier: 71.23%

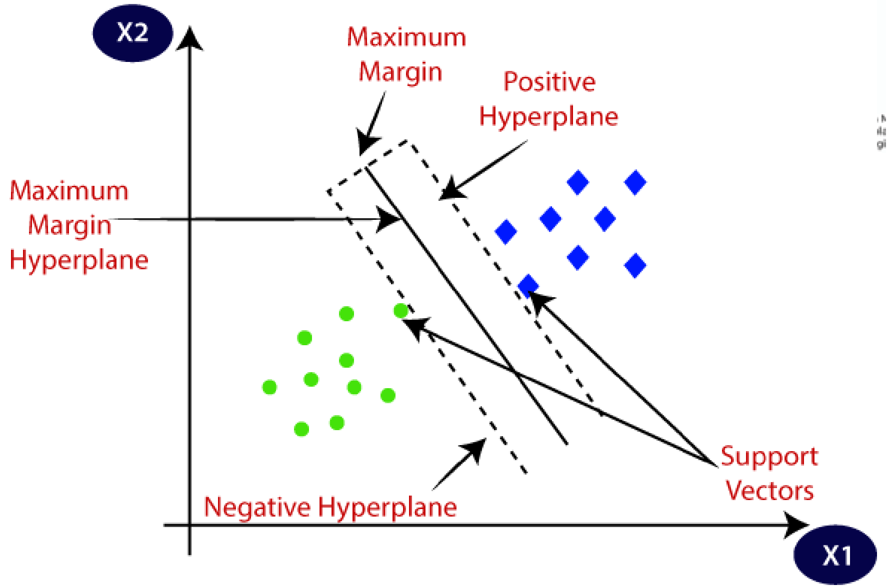
Non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.



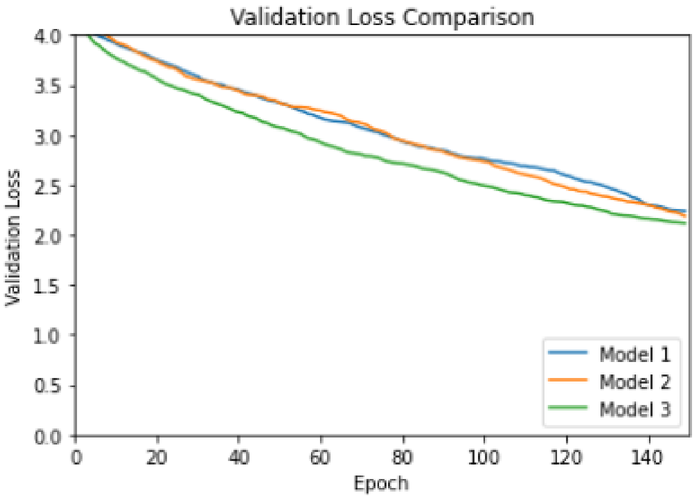
Inputs:
1. Distance metrics
2. K (number of neighbors)

SVM Classifier: 78.08%

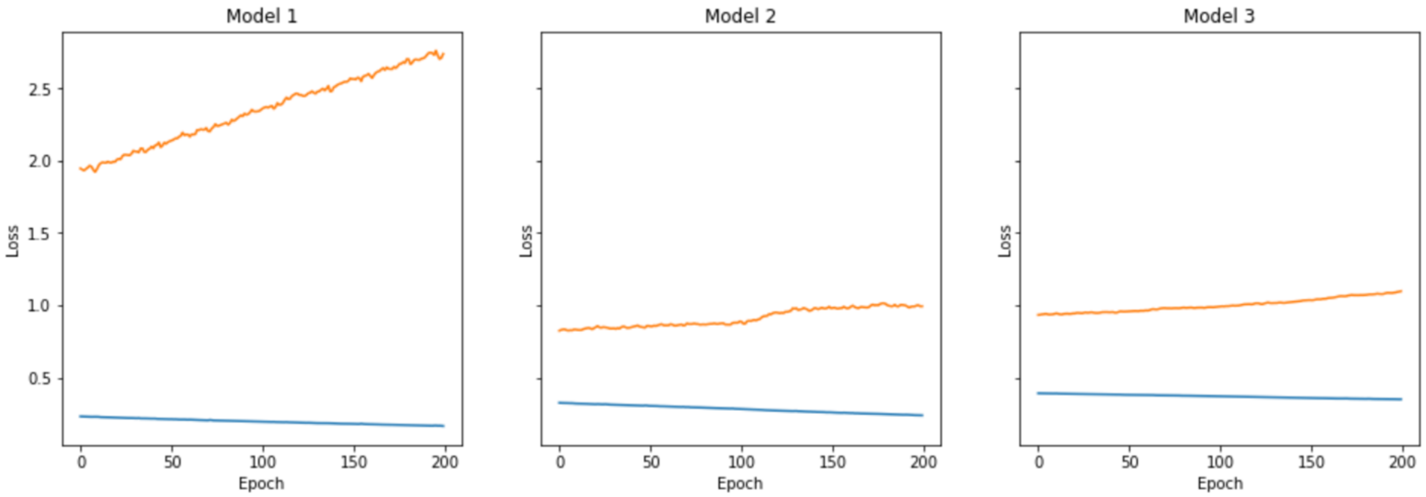
creates a line or a hyperplane (decision boundary) that separates the data into classes.



Artificial Neural Network (ANN)



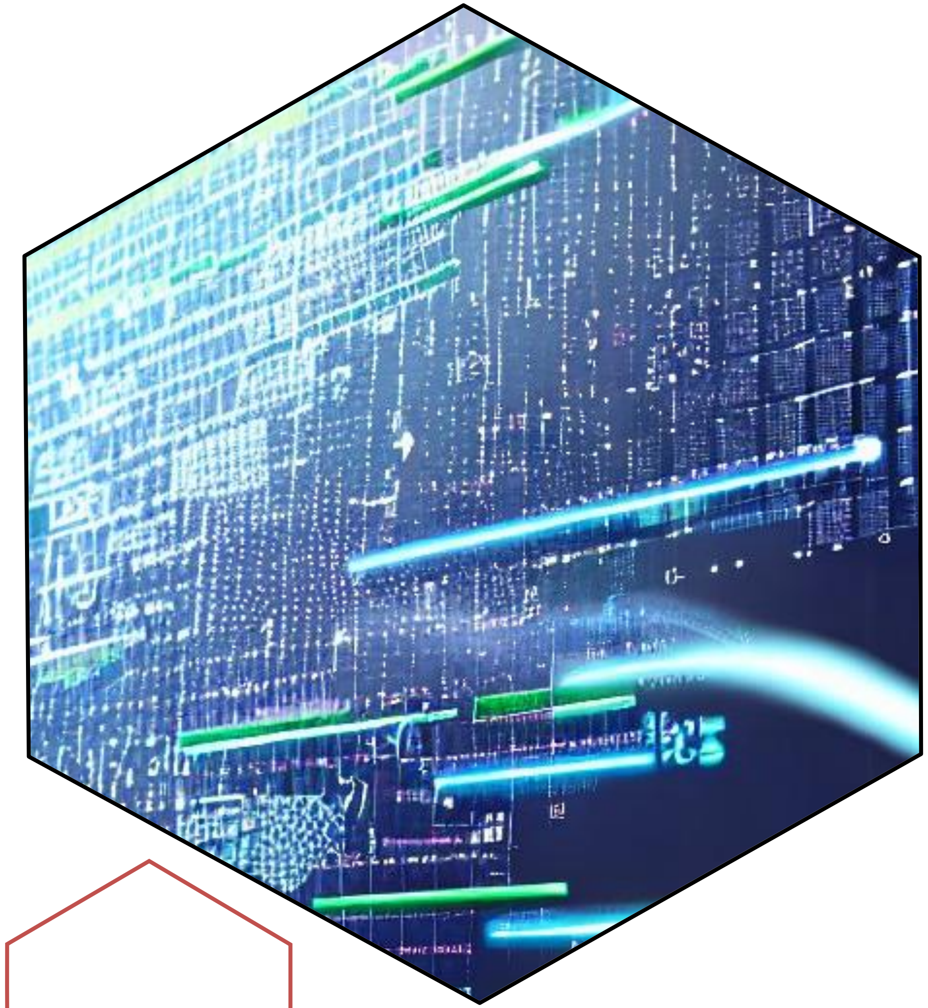
Loss Function:



	Model 1	Model 2	Model 3
Number of hidden layers	1	1	1
Hidden layers input features	unit=64	unit=22	unit=6
Model accuracy	52.05%	52.05%	56.16%
Cross validation	69.14%	71.20%	66.11%

Summary of Algorithms

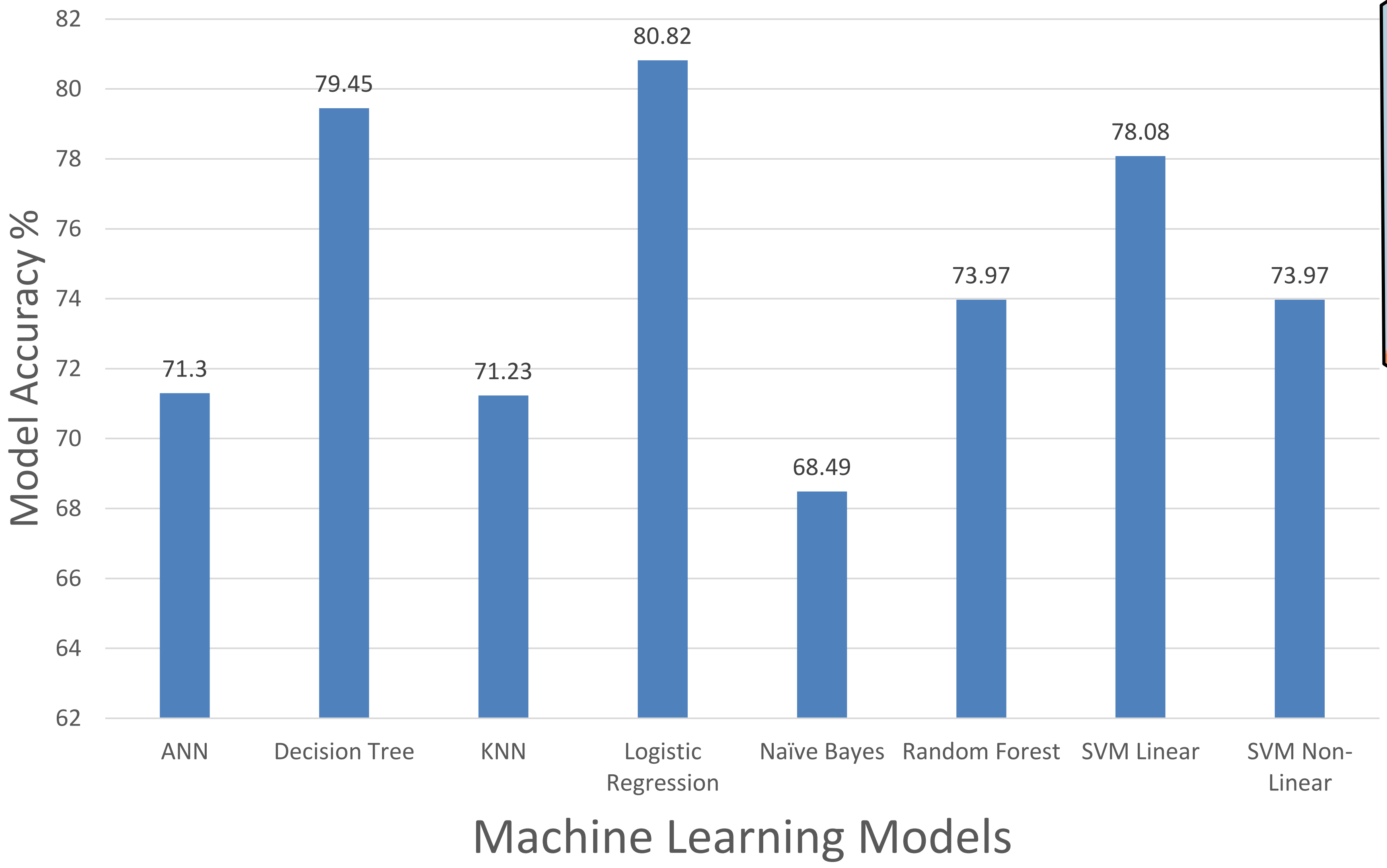
- K Nearest Neighbor
- Support Vector Machine Linear
- Support Vector Machine Non-Linear
- Naive Bayes Classifier
- Decision Tree Classifier
- Random Forest
- Logistic Regression Classifier



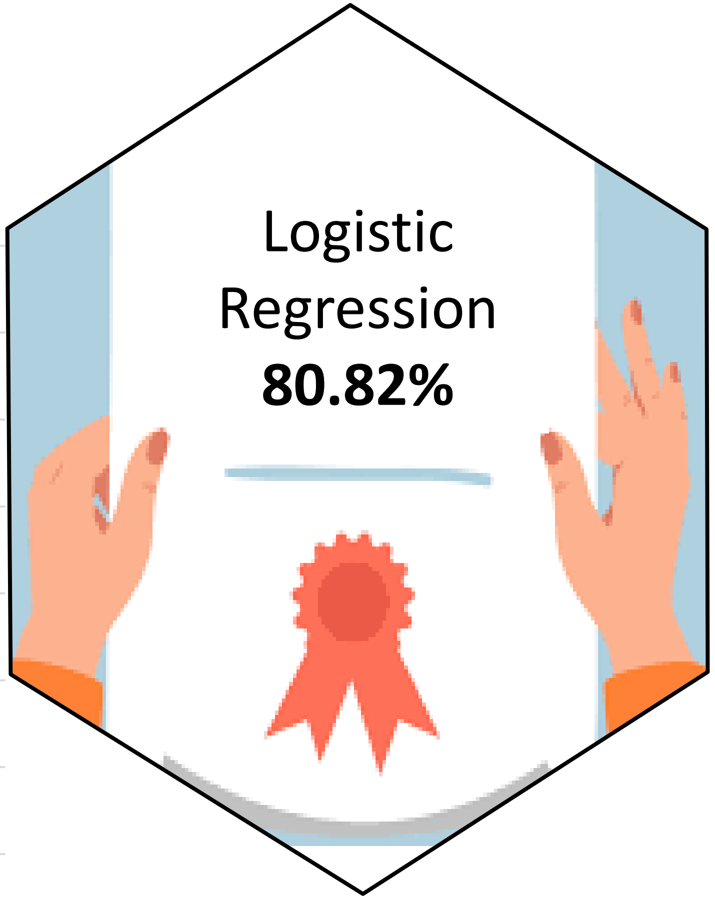
Results


Introduction
Data Collecting
Data Processing
Machine Learning
Results

Results



Model Accuracy %
82
80
78
76
74
72
70
68
66
64
62
71.3
79.45
71.23
80.82
68.49
73.97
78.08
73.97
ANN
Decision Tree
KNN
Logistic Regression
Naïve Bayes
Random Forest
SVM Linear
SVM Non-Linear
Machine Learning Models



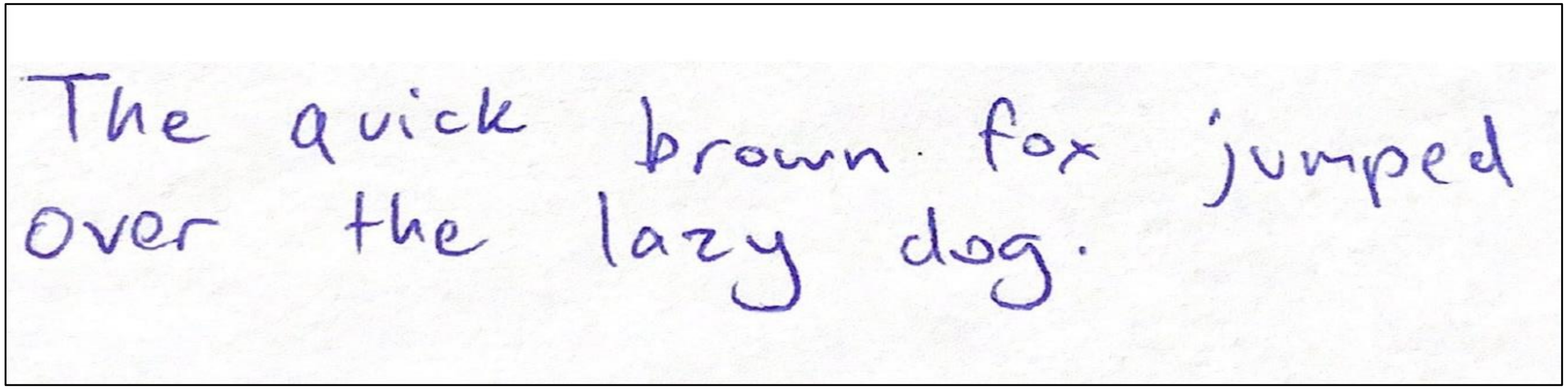
Logistic
Regression
80.82%

Results

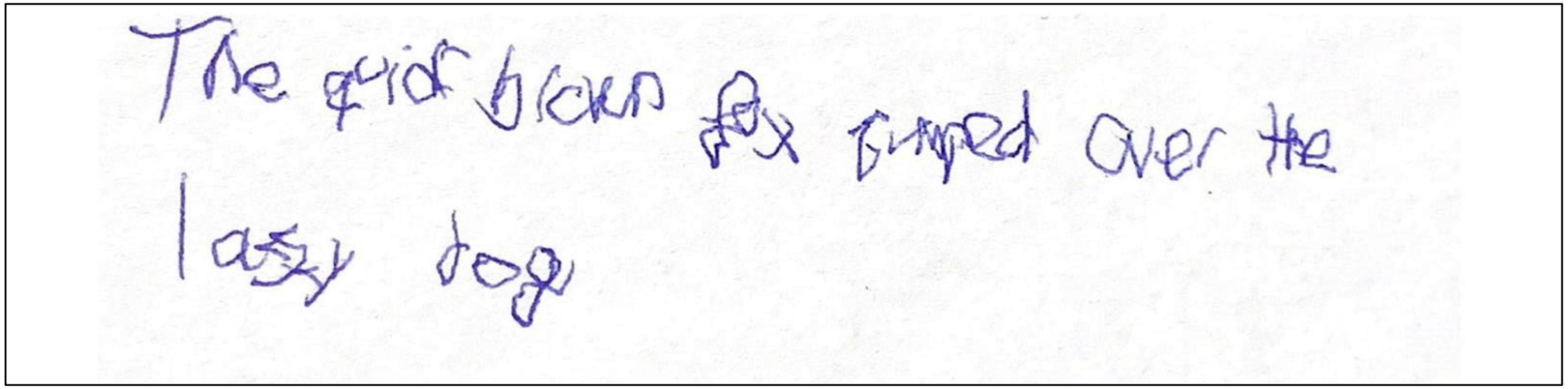
Model	Sensitivity	F1_score
K-Nearest-Neighbor	51.72 %	61.11 %
SVM-Linear	72.41 %	71.12 %
SVM Non-Linear	75.86 %	69.44 %
Naive Bayes Classifier	51.72 %	64.51 %
Random Forest	68.97 %	80.57 %
Decision Tree	68.97 %	66.78 %
Logistic Regression	68.97 %	69.60 %

CNN

A

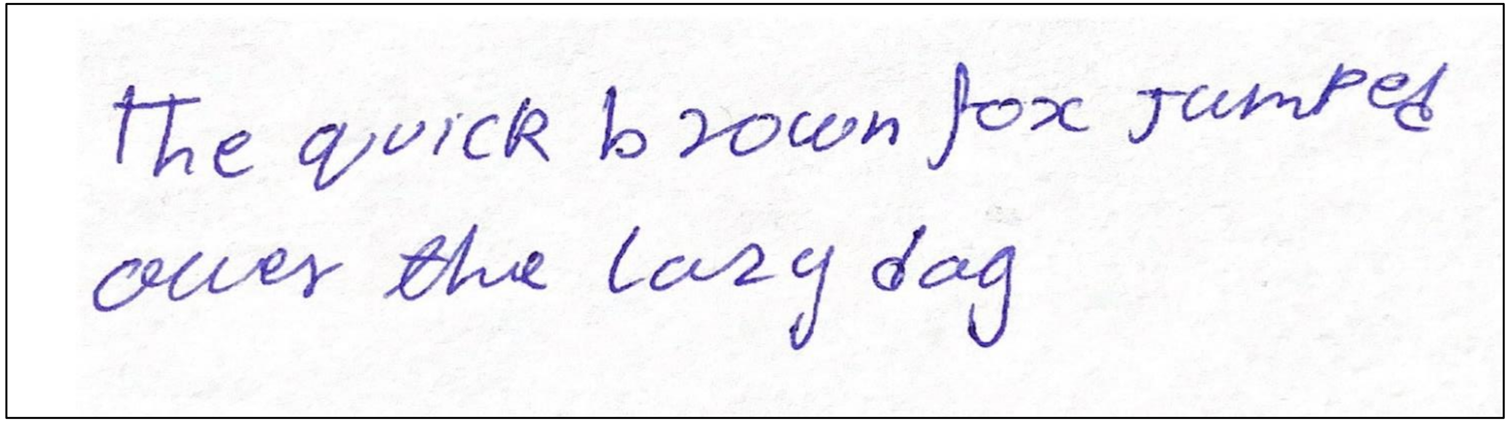

The quick brown fox jumped over the lazy dog.

B

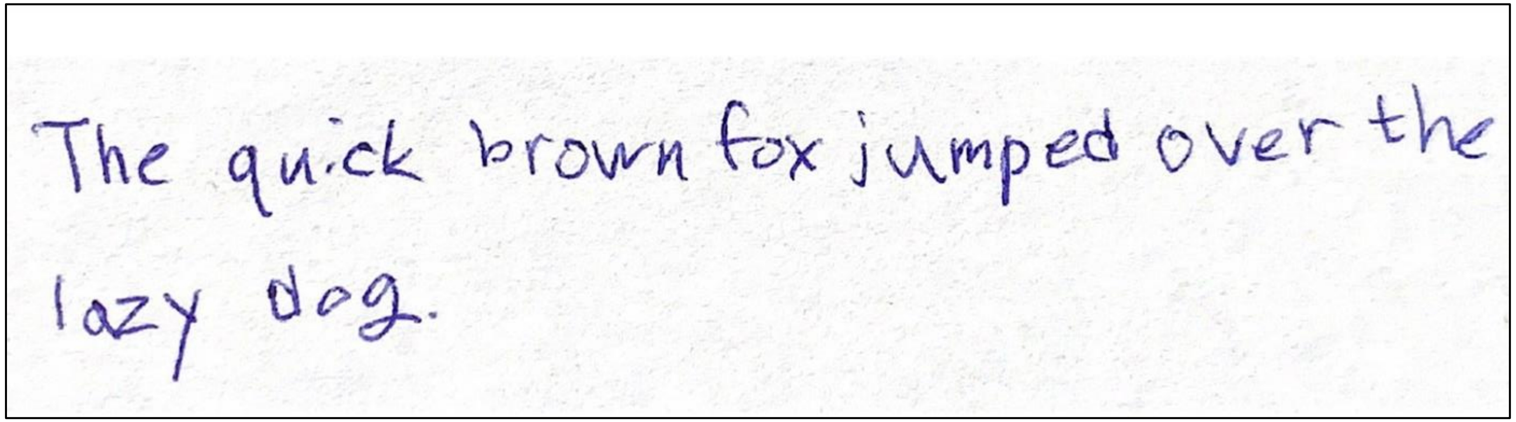

The quick brown fox jumped over the lazy dog

CNN

A



B



Survey:

Show of hands, who thinks ‘A’ was written with the dominant hand?

Future Improvements

- Collect more data
- Increase sampling rate of muse or
- Better fidelity EEG device
- Build a dataset more efficiently
- Implementing a SOM based Algorithm

Q & A

Thank you

Josiah Tsang

Lisa Skelton

Martina Chiesa

Muyang Li

Sofia Caltabiano

Simon Cotterill