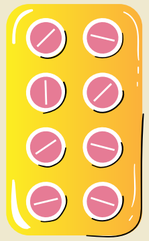
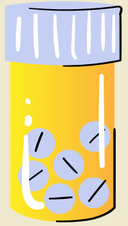
Alzheimer's Prediction Model

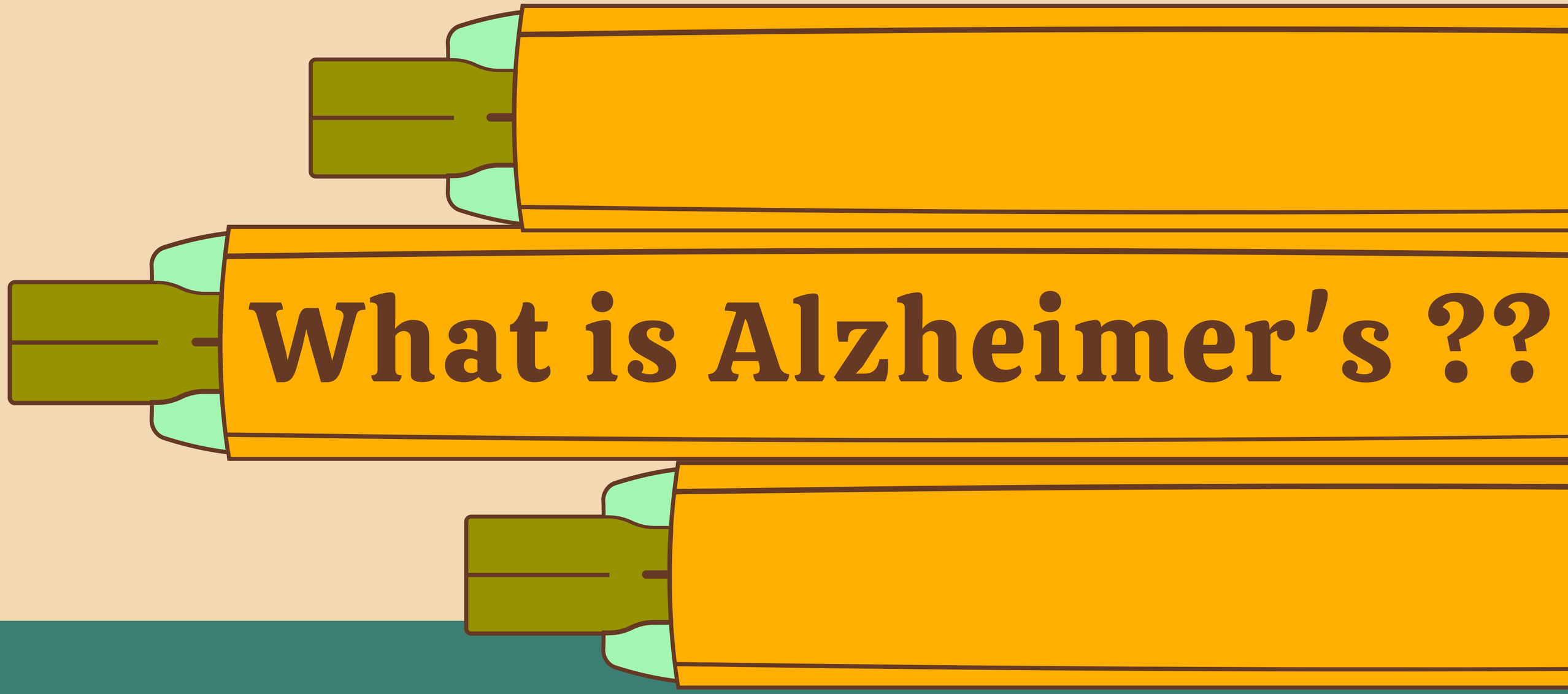
MADE BY:

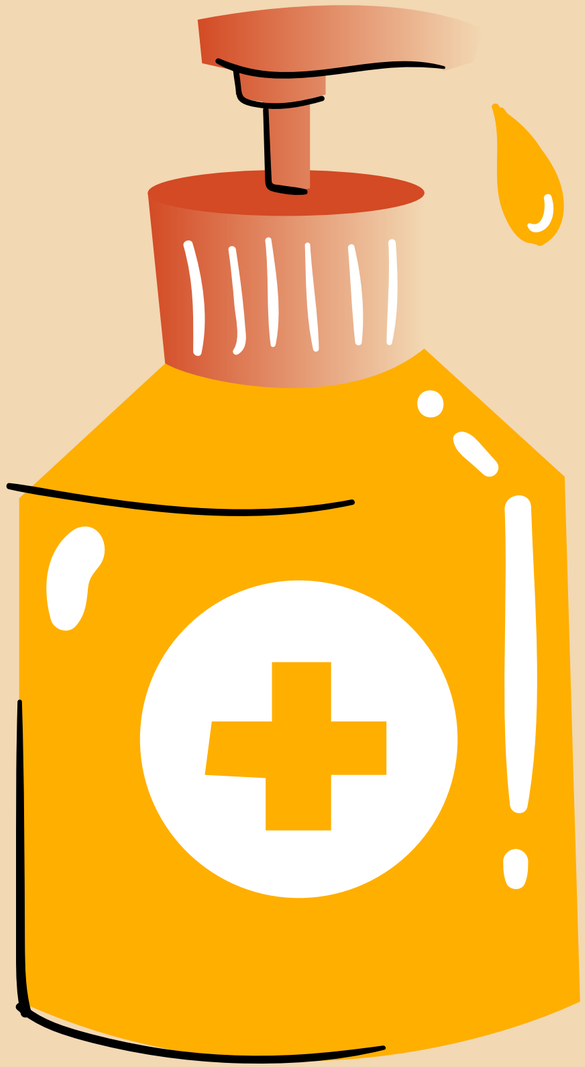
1)ATHARVA SINGH
2)KOMAL GOEL
3)ISHITA UNIYAL

Index






What is Alzheimer's ??



Alzheimer's is a disease that affects the brain, causing memory loss and difficulties with thinking and behavior. It usually starts slowly but gets worse over time, making it hard for people to remember things, carry on conversations, and manage daily tasks. It is the most common cause of dementia, a general term for a decline in mental ability severe enough to interfere with daily life.One of the common names used is dementia

Rate of Alzheimer's Disease Over Time

DATASET

	PatientID	Age	Gender	Ethnicity	EducationLevel	BMI	Smoking	AlcoholConsumption	PhysicalActivity	DietQuality	...	MemoryComplaints	BehavioralProblems
0	4751	73	0	0	2	22.927749	0	13.297218	6.327112	1.347214	...	0	0
1	4752	89	0	0	0	26.827681	0	4.542524	7.619885	0.518767	...	0	0
2	4753	73	0	3	1	17.795882	0	19.555085	7.844988	1.826335	...	0	0
3	4754	74	1	0	1	33.800817	1	12.209266	8.428001	7.435604	...	0	1
4	4755	89	0	0	0	20.716974	0	18.454356	6.310461	0.795498	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2144	6895	61	0	0	1	39.121757	0	1.561126	4.049964	6.555306	...	0	0
2145	6896	75	0	0	2	17.857903	0	18.767261	1.360667	2.904662	...	0	1
2146	6897	77	0	0	1	15.476479	0	4.594670	9.886002	8.120025	...	0	0
2147	6898	78	1	3	1	15.299911	0	8.674505	6.354282	1.263427	...	0	0
2148	6899	72	0	0	2	33.289738	0	7.890703	6.570993	7.941404	...	0	1

2149 rows × 35 columns

Algorithms used

Support Vector Machine (SVM)

- SVM finds the optimal boundary (called a hyperplane) that best separates the data points of different classes. The goal is to maximize the margin, which is the distance between the hyperplane and the nearest data points from either class, called support vectors. By maximizing this margin, SVM aims to improve the model's ability to generalize to new data.

ACCURACY

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, x_test_prediciton)
print(cm)
accuracy_score(y_test, x_test_prediciton)
```

```
[[316  28]
 [ 57 137]]

0.8420074349442379
```

K-Nearest Neighbors (KNN)

- KNN is a simple, instance-based learning algorithm used for classification. It works by finding the 'k' closest data points (neighbors) to a new data point and assigns the most common class among these neighbors to the new data point. The value of 'k' is a positive integer, and the distance between data points is typically measured using metrics like Euclidean distance.

ACCURACY

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)
```

```
[[338  26]
 [123  51]]

0.7230483271375465
```

Logistic Regression

- Logistic regression is a statistical method used for binary classification. It models the probability that a given input belongs to a particular class using a logistic function. The algorithm outputs a value between 0 and 1, which is then used to classify the input into one of two categories based on a threshold, typically 0.5. Despite its name, logistic regression is a classification algorithm, not a regression algorithm.

ACCURACY

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm= confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test,y_pred)
```

[26] ✓ 0.0s

```
[[128  15]
 [ 13  59]]
```

```
0.8697674418604651
```

Decision Tree

- A decision tree is a supervised learning algorithm used for classification. It works by splitting the data into subsets based on the value of input features, creating a tree-like model of decisions. Each node in the tree represents a feature, each branch represents a decision rule, and each leaf node represents an outcome or class label. The goal is to create a model that predicts the class of the target variable by learning simple decision rules inferred from the data features.

ACCURACY

```
[112]
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)
```

```
[[338   6]
 [ 12 182]]

0.966542750929368
```

Random Forest

- Random forest is an ensemble learning method used for classification. It operates by constructing multiple decision trees during training and outputting the mode of the classes (for classification) of the individual trees. Each tree in the forest is trained on a different subset of the data, and the diversity of the trees helps improve the overall performance and robustness of the model. This reduces the risk of overfitting compared to a single decision tree.

ACCURACY

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)
```

```
[[336   8]
 [ 45 149]]

0.9014869888475836
```

Naive Bayes Classification

- Naive Bayes is a probabilistic machine learning algorithm used for classification. It is based on Bayes' Theorem and assumes that the features are independent of each other (hence "naive"). The algorithm calculates the probability of each class given the input features and classifies the data point into the class with the highest probability. Despite its simplifying assumption, Naive Bayes often performs well in practice, especially for text classification tasks like spam detection.

ACCURACY

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)
```

```
[[305  39]
 [ 64 130]]

0.8085501858736059
```

Conclusion

- In conclusion, while various classification algorithms offer unique strengths, decision trees stand out due to their interpretability and ability to handle both numerical and categorical data effectively. Their straightforward structure allows for easy visualization and understanding of the decision-making process. Moreover, decision trees often achieve better accuracy, especially in scenarios where the relationship between features and the target variable is complex and non-linear. This makes them a preferred choice in many practical applications where accuracy and transparency are critical.

Thank you for listening!

Don't hesitate to ask any questions!