Project Initialization and Planning Phase

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	1. Individuals with ASD: Children and adults with Autistic Spectrum Disorder. 2. Clinical features: Behavioral observations, medical history, and symptom profiles. 3. Neuroimaging data: MRI, fMRI, EEG, and other neuroimaging modalities to study brain structure and function. 4. Genetic data: Genetic mutations, variants, and expression profiles. 5. Behavioral data: Observations of social interactions, communication patterns, and repetitive behaviors.
Scope	1. Diagnosis: Accurate detection and classification of ASD. 2. Phenotyping: Characterization of ASD subtypes and severity levels. 3. Biomarker discovery: Identification of reliable biomarkers for ASD diagnosis and monitoring. 4. Personalized interventions: Development of tailored treatment plans based on individual characteristics. 5. Prognostic modeling: Prediction of treatment outcomes and long-term prognosis.
Problem Statement	
Description	- Accurate detection and classification of ASD using machine learning algorithms and neural networks. - Development of personalized diagnostic models incorporating clinical, behavioral, and neuroimaging features.

	- Identification of novel biomarkers and risk factors for ASD.
Impact	Improved diagnostic accuracy and earlier intervention for individuals with ASD. - Enhanced personalized treatment plans and better treatment outcomes. - Increased understanding of ASD's neural mechanisms and underlying causes.
Proposed Solution	
Approach	1. Machine Learning: Using algorithms to analyze behavioral, clinical, and neuroimaging data to detect patterns and predict diagnoses. 2. Deep Learning: Utilizing neural networks to learn complex representations of ASD features from large datasets. 3. Natural Language Processing: Analyzing speech and language patterns to identify potential indicators of ASD.
Key Features	Age,results,symtomps

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware Requirements:		
Computing Resources	CPU/GPU specifications, number of cores	T4 GPU
Memory	RAM specifications	16 GB
Storage	Disk space for data, models, and logs	512 SSD
Software Requirements:		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	Scikit-learn, pandas, NumPy, Seaborn, matplotlib
Development Environment	IDE, version control	Google colab, VS code

Data		
Data	Source, size, format	Kaggle, dataset, csv

Initial Project Planning Template

Date	15 JULY 2024
Team ID	740075
Project Name	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	4 Marks

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create a product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members	Sprint Start Date	Sprint End Date (Planned)
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, and password, and confirming my password.	2	High	Jayanth	2024/03/15	2024/03/22
Sprint-1		USN-2	As a user, I will receive a confirmation email once I have registered for the application	1	High	Jayanth	2024/03/22	2024/03/29
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	Low	Nagaraju	2024/03/22	2024/03/29
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	Nagarani	2024/03/29	2024/04/05
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	Sathwika	2024/03/29	2024/04/05

Project Initialization and Planning Phase

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	1. Individuals with ASD: Children and adults with Autistic Spectrum Disorder. 2. Clinical features: Behavioral observations, medical history, and symptom profiles. 3. Neuroimaging data: MRI, fMRI, EEG, and other neuroimaging modalities to study brain structure and function. 4. Genetic data: Genetic mutations, variants, and expression profiles. 5. Behavioral data: Observations of social interactions, communication patterns, and repetitive behaviors.
Scope	1. Diagnosis: Accurate detection and classification of ASD. 2. Phenotyping: Characterization of ASD subtypes and severity levels. 3. Biomarker discovery: Identification of reliable biomarkers for ASD diagnosis and monitoring. 4. Personalized interventions: Development of tailored treatment plans based on individual characteristics. 5. Prognostic modeling: Prediction of treatment outcomes and long-term prognosis.
Problem Statement	
Description	- Accurate detection and classification of ASD using machine learning algorithms and neural networks. - Development of personalized diagnostic models incorporating clinical, behavioral, and neuroimaging features.

	- Identification of novel biomarkers and risk factors for ASD.
Impact	Improved diagnostic accuracy and earlier intervention for individuals with ASD. - Enhanced personalized treatment plans and better treatment outcomes. - Increased understanding of ASD's neural mechanisms and underlying causes.
Proposed Solution	
Approach	1. Machine Learning: Using algorithms to analyze behavioral, clinical, and neuroimaging data to detect patterns and predict diagnoses. 2. Deep Learning: Utilizing neural networks to learn complex representations of ASD features from large datasets. 3. Natural Language Processing: Analyzing speech and language patterns to identify potential indicators of ASD.
Key Features	Age,results,symtomps

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware Requirements:		
Computing Resources	CPU/GPU specifications, number of cores	T4 GPU
Memory	RAM specifications	16 GB
Storage	Disk space for data, models, and logs	512 SSD
Software Requirements:		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	Scikit-learn, pandas, NumPy, Seaborn, matplotlib
Development Environment	IDE, version control	Google colab, VS code

Data		
Data	Source, size, format	Kaggle, dataset, csv

Data Collection and Preprocessing Phase

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset	There is no correct features in the dataset.	Medium	We check the dataset features correctly and we changed the features to get correct output.

Model Optimization and Tuning Phase Template

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

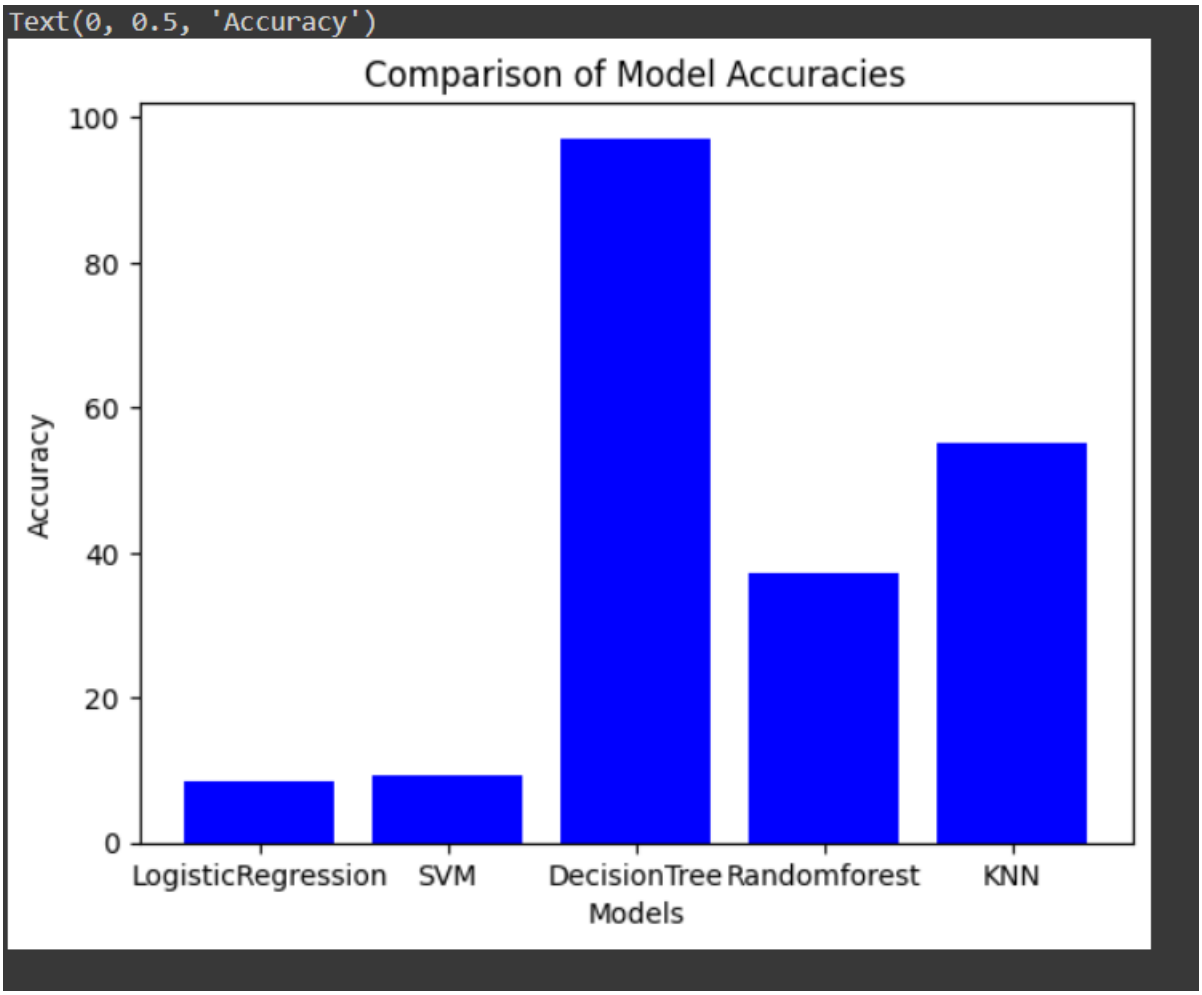
The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Performance Metrics Comparison Report (2 Marks):

```
accuracy_df = pd.DataFrame({
'Model': ['LogisticRegression', 'SVM', 'DecisionTree', 'Randomforest', 'KNN'],
'Accuracy': [accuracy_LR*100, accuracy_SVC*100, accuracy_dt*100, accuracy_RF*100, accuracy_KNN*100]})
print(accuracy_df)
```

```
                Model   Accuracy
0  LogisticRegression   8.490566
1                 SVM   9.433962
2        DecisionTree  97.169811
3        Randomforest  37.264151
4                 KNN  55.188679
```

```
models = ['LogisticRegression', 'SVM', 'DecisionTree', 'Randomforest', 'KNN']
accuracies = [accuracy_LR*100, accuracy_SVC*100, accuracy_dt*100, accuracy_RF*100, accuracy_KNN*100]
plt.bar(models, accuracies, color='blue')
# Add title and axis Labels
plt.title('Comparison of Model Accuracies')
plt.xlabel('Models')
plt.ylabel('Accuracy')
```



Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Decision tree	The model Decision tree usually provides high accuracy due to combining the predictions of multiple decision trees. Its ability to handle complex relationships, minimize overfitting. It can handle both classification and regression justifying its selection as the final model.

Model Development Phase Template

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	6 Marks

Model Selection Report

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

Model Selection Report:

Model	Description		Performance Metric (e.g., Accuracy, F1 Score)
1. K Nearest Neighbors Model	A variable is created with name knn which has KNeighborsClassifier() algorithm initialised in it.The knn model is trained using the .fit() function.The model is trained on the X_train and y_train data that is the training features and target variables.		`Accuracy_KNN: 55.188679245283026`

2. SVM Model	A variable is created with name svm which has SVC() algorithm initialised in it.The svm model is trained using the .fit() function.The model is trained on the X_train and y_train data that is the training features and training target variables.		`Accuracy_SVM: 9.433962264150944`
3. Decision Tree Model	A variable is created with name dt classifier which has Decision Tree Classifier() algorithm initialised in it with a parameter max_depth set to 7. The dtclassifier model is trained using the .fit() function. The model is trained on the X_train and y_train data that is the training features and training target variables.		`Accuracy_DT: 97.16981132075472`

4. Random Forest Model	Random Forest Classifier is a Bagging model which utilises multiple decision trees and takes their aggregate to give a prediction. A variable is created with name rfclassifier which has Random ForestClassifier() algorithm initialised in it.The rfclassifier model is trained using the .fit() function.The model is trained on the X_train and y_train data that is the training features and training target variables.		`Accuracy_RF: 97.16981132075472`

Data Collection and Preprocessing Phase

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	#Structure of the data:   #Descriptive Statistical: Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features.

<table>
<tr>
<td></td>
<td>

```
data.describe()
```

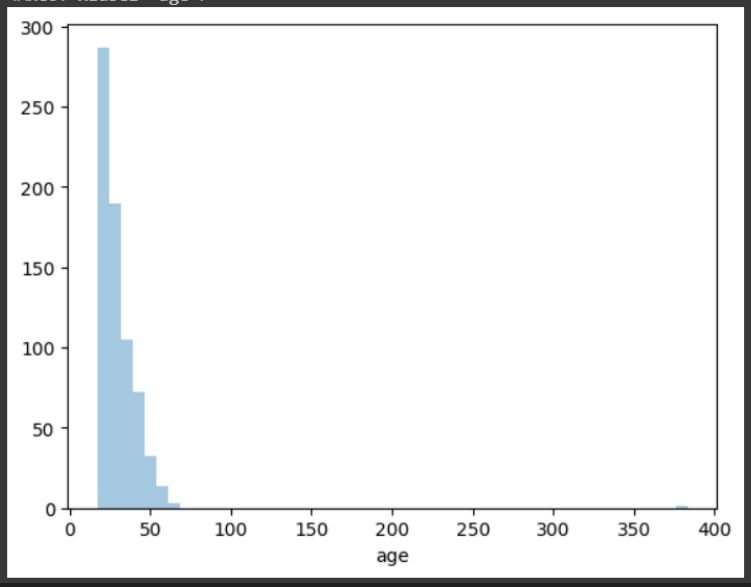
	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	A10_Score	result
count	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000	704.000000
mean	0.721591	0.453125	0.457386	0.495739	0.498580	0.284091	0.417614	0.649148	0.323864	0.573864	4.875000
std	0.448535	0.498152	0.498535	0.500337	0.500353	0.451301	0.493516	0.477576	0.468281	0.494866	2.501493
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	3.000000
50%	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	1.000000	4.000000
75%	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	7.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	10.000000

</td>
</tr>
<tr>
<td>Univariate Analysis</td>
<td>
Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand data. It is a way to quickly identify patterns, trends, and outliers in the data, which can help to gain insights and make informed decisions.
Univariate Analysis:
In simple words, univariate analysis is understanding the data with a single feature. We have displayed three different types of graphs and plots.
For simple visualizations we can use the matplotlib. pyplot library. Here the plt. figure() command is used to determine the size of the plot.
We have histogram for all features of the dataset which include phosphorus, humidity, temperature as well . The histogram shows the distribution of nitrogen fertilizers for crop.

```
Code cell output actions
[ ]  sns.distplot(data['age'],bins=50,kde=False)
```

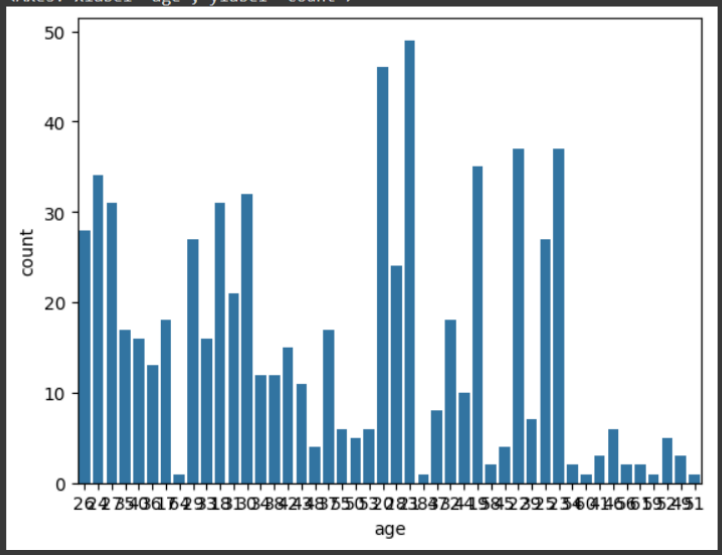
</td>
</tr>
</table>

```
sns.distplot(data['age'],bins=50,kde=False)
<Axes: xlabel='age'>
```



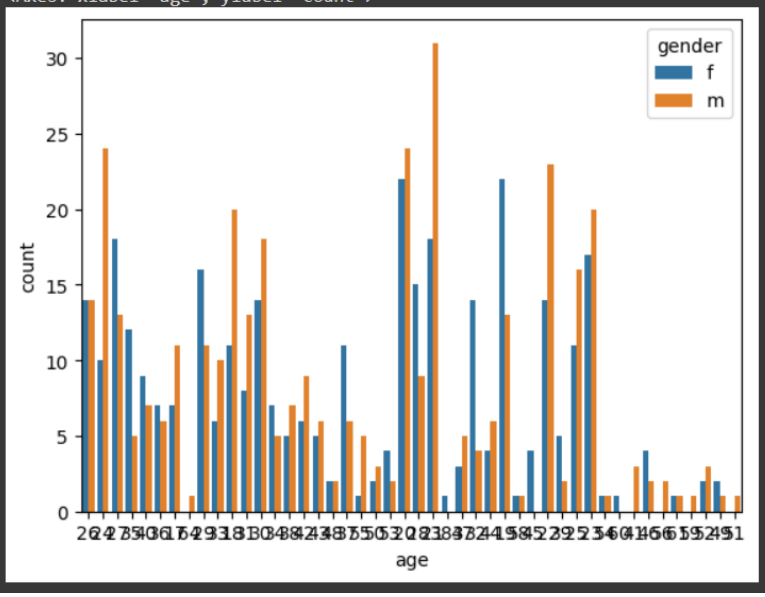

```
[ ] sns.countplot(x='age',data=data)
```

```
<Axes: xlabel='age', ylabel='count'>
```



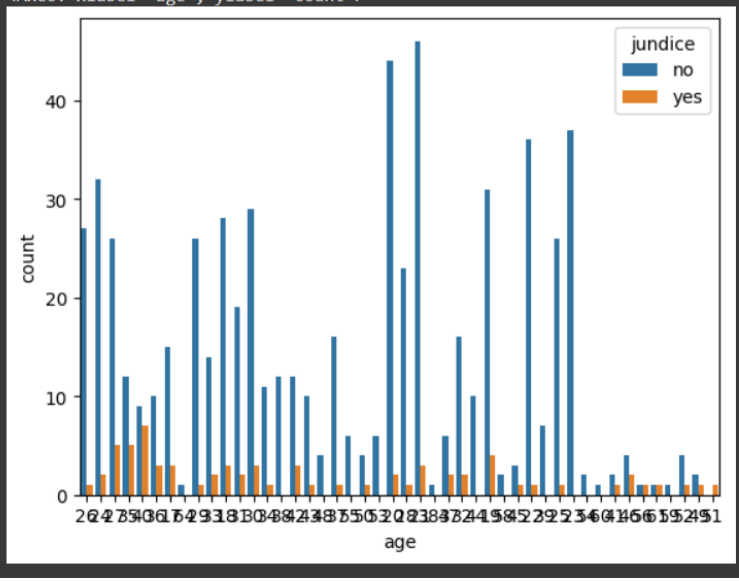
```
[ ] sns.countplot(x='age',hue='gender',data=data)
```

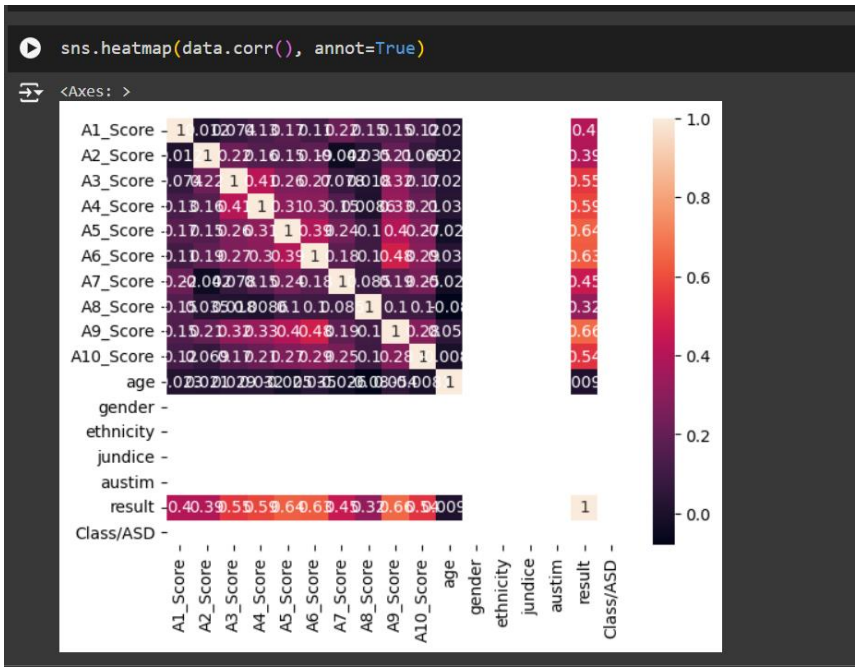
```
<Axes: xlabel='age', ylabel='count'>
```




```
sns.countplot(x='age',hue='jundice',data=data)
```

```
<Axes: xlabel='age', ylabel='count'>
```



	
Outliers and Anomalies	There is no Outliers in our project.
Outliers and Anomalies	There is no Outliers in our project.
Data Preprocessing Code Screenshots	
Loading Data	#Loading the data

<table>
<tr>
<td></td>
<td>

```
data = pd.read_csv('/content/Autism_Data.arff')
data
```

	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	A10_Score	...	gender	ethnicity	jundice	austim	contry_of_res	used_app_before	result
0	1	1	1	1	0	0	1	1	0	0	...	f	White-European	no	no	'United States'	no	6
1	1	1	0	1	0	0	0	1	0	1	...	m	Latino	no	yes	Brazil	no	5
2	1	1	0	1	1	0	1	1	1	1	...	m	Latino	yes	yes	Spain	no	8
3	1	1	0	1	0	0	1	1	0	1	...	f	White-European	no	yes	'United States'	no	6
4	1	0	0	0	0	0	0	1	0	0	...	f	?	no	no	Egypt	no	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
699	0	1	0	1	1	0	1	1	1	1	...	f	White-European	no	no	Russia	no	7
700	1	0	0	0	0	0	0	1	0	1	...	m	Hispanic	no	no	Mexico	no	3
701	1	0	1	1	1	0	1	1	0	1	...	f	?	no	no	Russia	no	7
702	1	0	0	1	1	0	1	0	1	1	...	m	'South Asian'	no	no	Pakistan	no	6

</td>
</tr>
<tr>
<td>Handling Missing Data</td>
<td>

```
[ ] data.shape
```

```
(704, 21)
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 704 entries, 0 to 703
Data columns (total 21 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   A1_Score         704 non-null    int64
 1   A2_Score         704 non-null    int64
 2   A3_Score         704 non-null    int64
 3   A4_Score         704 non-null    int64
 4   A5_Score         704 non-null    int64
 5   A6_Score         704 non-null    int64
 6   A7_Score         704 non-null    int64
 7   A8_Score         704 non-null    int64
 8   A9_Score         704 non-null    int64
 9   A10_Score        704 non-null    int64
 10  age              704 non-null    object
 11  gender           704 non-null    object
 12  ethnicity        704 non-null    object
 13  jundice          704 non-null    object
 14  austim           704 non-null    object
 15  contry_of_res    704 non-null    object
 16  used_app_before  704 non-null    object
 17  result           704 non-null    int64
 18  age_desc         704 non-null    object
 19  relation         704 non-null    object
 20  Class/ASD        704 non-null    object
dtypes: int64(11), object(10)
memory usage: 115.6+ KB
```

</td>
</tr>
</table>

	 For checking the null values . isnull() function is used. To sum those null values we use. sum() function. From the below image we found that there are no null values present in our dataset. So we can skip handling the missing values step.

Data Collection and Preprocessing Phase

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	Develop a machine learning model to detect and classify Autistic Spectrum Disorder (ASD) using neuroimaging and clinical data. Improve diagnostic accuracy and enable earlier intervention for individuals with ASD. Identify novel biomarkers and risk factors for ASD. Enhance understanding of ASD's neural mechanisms and underlying causes.
Data Collection Plan	Kaggle

Raw Data Sources Identified	1. Clinical data: - Electronic Health Records (EHRs) - Clinical interviews and assessments - Behavioral observations - Medical history 2. Neuroimaging data: - Magnetic Resonance Imaging (MRI) - Functional MRI (fMRI) - Electroencephalography (EEG) - Magnetoencephalography (MEG) 3. Behavioral data: - Parent-reported behavioral questionnaires - Teacher-reported behavioral questionnaires - Direct behavioral observations - Wearable sensors (e.g., actigraphy, GPS)

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions

Dataset 1	The dataset consists single CSV file.This dataset is mainly concerning Indian climatic conditions.There are seven input features only one output features.	`/content/Autism_Data.Csv`	CSV	146KB	Public

Model Development Phase Template

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	5 Marks

Feature Selection Report Template

In the forthcoming update, each feature will be accompanied by a brief description. Users will indicate whether it's selected or not, providing reasoning for their decision. This process will streamline decision-making and enhance transparency in feature selection.

Feature	Description	Selected (Yes/No)	Reasoning
Clinical Features	. Age of Diagnosis: The age at which ASD was diagnosed.	YES	Symptom Severity: A score indicating the severity of ASD symptoms. Behavioral Observations: Scores from behavioral assessments, such as the Autism Behavior Checklist (ABC).
Neuroimaging Features	Brain Structure Volumes: Volumes of brain structures, such as the amygdala	Yes	Functional Connectivity Metrics: Measures of brain connectivity, such as default mode network (DMN) activity.

	and hippocampus		Brain Activity Patterns: EEG frequency bands, such as alpha, beta, and theta waves.
Behavioral Features	**Social Skills Assessments: Scores from social skills assessments, such as the Social Responsiveness Scale (SRS).**	Yes	**Communication Skills Assessments: Scores from communication skills assessments, such as the Communication and Symbolic Behavior Scales (CSBS).** **Repetitive Behavior Assessments: Scores from repetitive behavior assessments, such as the Repetitive Behavior Scale (RBS).**
Genetic Features	Genetic Variant Frequencies: Frequencies of specific genetic variants, such as SNPs and CNVs	Yes	Gene Expression Levels: Levels of gene expression, measured through RNA sequencing.
Social Media and Online Features	Social Media Usage Patterns: Frequency and duration of social media use.	Yes	Online Behavior Patterns: Search queries, browsing history, and other online behaviors.
Environmental Features	Exposure to Environmental Toxins: Exposure to pesticides, heavy metals, and other toxins	Yes	Nutrition and Dietary Patterns: Food preferences, nutrient intake, and other dietary metrics.

Wearable Device Features	Physical Activity Levels: Step count, exercise intensity, and other physical activity metrics		Sleep Patterns: Sleep duration, quality, and other sleep metrics.

Model Development Phase Template

Date	15 JULY 2024
Team ID	740075
Project Title	Detection Of Autistic Spectrum Disorder: Classification
Maximum Marks	4 Marks

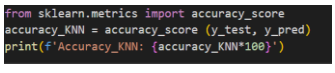
Initial Model Training Code, Model Validation and Evaluation Report

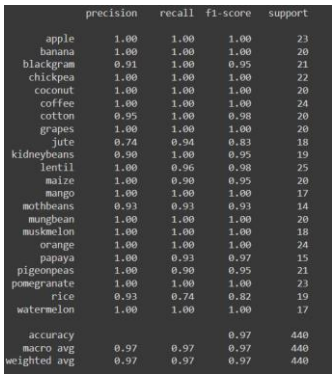
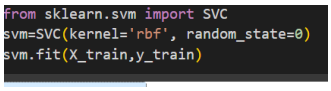
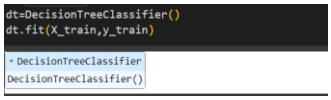
The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include classification reports, accuracy, and confusion matrices for multiple models, presented through respective screenshots.

Initial Model Training Code:

Paste the screenshot of the model training code

Model Validation and Evaluation Report:

Model	Classification Report	Accuracy	Confusion Matrix
1. K Nearest Neighbors Model		**55.18867**	

			
2. SVM Model		**9.433**	
3.Decision Tree Model		97.166	
4. Random Forest Model		97.1669