1. ***2 ASSISTANTCOMMUNICATION PROBLEM***

Dataset:

The provided dataset contains pairs of original and disfluent questions. Each pair is identified by a unique question ID. The dataset includes the following columns:

question\_id: Unique identifier for each question pair.

original: The original, fluent version of the question.

disfluent: The disfluent version of the question.

Objective:

Participants are required to develop an AI/ML model that can accurately transform disfluent questions into their original, fluent versions. The model should be able to handle various types of disfluencies and produce coherent and accurate outputs.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/AI+for+Social+Good/Conversation+dataset/2assistants/2assistantcommunication.csv

1. ***JIGSAW PUZZLE***

Dataset:

The provided dataset contains user comments along with labels indicating whether each comment is toxic or non-toxic. The dataset includes the following columns:

id: Unique identifier for each comment.

comment\_text: The text of the user comment.

toxic: Binary label indicating whether the comment is toxic (1) or non-toxic (0).

severe\_toxic: Binary label indicating whether the comment is severely toxic (1) or not (0).

obscene: Binary label indicating whether the comment is obscene (1) or not (0).

threat: Binary label indicating whether the comment contains threats (1) or not (0).

insult: Binary label indicating whether the comment contains insults (1) or not (0).

identity\_hate: Binary label indicating whether the comment contains identity hate (1) or not (0).

Objective:

Participants are required to develop an AI/ML model that can accurately classify comments as toxic or non-toxic and identify specific types of toxicity. The model should be able to handle various types of toxic language and produce reliable predictions.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/AI+for+Social+Good/Jigsaw+toxic+comment/jigsaw-multilingual-toxic-comment-classification/jigsaw-toxic-comment-train.csv

1. ***SENTIMENT 140***

Dataset:

The provided dataset contains social media posts along with labels indicating the sentiment of each post. The dataset includes the following columns:

target: Binary label indicating whether the sentiment is positive (4) or negative (0).

id: Unique identifier for each post.

date: The date and time when the post was created.

flag: A flag indicating whether there was a query.

user: The username of the person who created the post.

text: The text content of the post.

Objective:

Participants are required to develop an AI/ML model that can accurately classify social media posts as having positive or negative sentiment. The model should be able to handle various types of text data and produce reliable predictions.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/AI+for+Social+Good/Sentiment+analysis+/sentiment+140/Sentiments.csv

1. ***SKIN DISEASES***

Dataset:

The dataset consists of images of skin samples along with their corresponding labels indicating the presence of various skin diseases. The dataset is provided in a CSV file with the following columns:

image: The name of the image file.

MEL: Melanoma (1 if present, 0 otherwise).

NV: Melanocytic nevus (1 if present, 0 otherwise).

BCC: Basal cell carcinoma (1 if present, 0 otherwise).

AK: Actinic keratosis (1 if present, 0 otherwise).

BKL: Benign keratosis (1 if present, 0 otherwise).

DF: Dermatofibroma (1 if present, 0 otherwise).

VASC: Vascular lesion (1 if present, 0 otherwise).

SCC: Squamous cell carcinoma (1 if present, 0 otherwise).

UNK: Unknown (1 if present, 0 otherwise).

Task:

Participants are required to build a machine learning model that can accurately classify the images into one of the nine categories mentioned above. The model should be trained on the provided dataset and evaluated on a separate test set

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Bio-Medical+Analysis/Skin+diseas/Skin\_Problems.zip

***5.DRONES***

Dataset:

The provided dataset contains images captured by drones along with bounding box coordinates and labels for objects detected in the images. The dataset includes the following columns:

image: The name of the image file.

xmin: The x-coordinate of the top-left corner of the bounding box.

ymin: The y-coordinate of the top-left corner of the bounding box.

xmax: The x-coordinate of the bottom-right corner of the bounding box.

ymax: The y-coordinate of the bottom-right corner of the bounding box.

label: The label indicating the type of object detected (e.g., stand).

Task:

Participants are required to develop an AI/ML model that can accurately detect and classify objects in drone images based on the provided bounding box coordinates and labels.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/Drones+dataset/drones.zip

***6. PIPELINE DEFECT DATASET***

Dataset:

The provided dataset contains images of pipelines with visible defects along with bounding box coordinates and labels for the detected defects. The dataset includes the following columns:

image: The name of the image file.

xmin: The x-coordinate of the top-left corner of the bounding box.

ymin: The y-coordinate of the top-left corner of the bounding box.

xmax: The x-coordinate of the bottom-right corner of the bounding box.

ymax: The y-coordinate of the bottom-right corner of the bounding box.

label: The label indicating the type of defect detected (e.g., crack).

Additionally, the dataset includes information about the classes of defects:

nc: 6 (number of classes)

names:

0: Deformation

1: Obstacle

2: Rupture

3: Disconnect

4: Misalignment

5: Deposition

Objective:

Participants are required to develop an AI/ML model that can accurately detect and classify defects in pipeline images based on the provided bounding box coordinates and labels.

<https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Robot/Pipeline+Defect+Dataset+(2).zip>

***7. CHEST X-RAY***

Dataset:

The provided dataset contains chest X-ray images categorized into two labels: normal and pneumonia. The dataset is organized into three main folders: train, test, and val, each containing subfolders named normal and pneumonia. These subfolders represent the labels for the images.

Objective:

Participants are required to develop an AI/ML model that can accurately classify chest X-ray images into one of the two categories: normal or pneumonia.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Bio-Medical+Analysis/Chest+X-Ray+Images+(Pneumonia)/chest\_xray.zip

***8. QUORA DATASET***

Dataset:

The provided dataset contains questions along with their corresponding labels indicating whether the question is relevant (1) or not relevant (0). The dataset includes the following columns:

qid: Unique identifier for each question.

question\_text: The text of the question.

target: Binary label indicating whether the question is relevant (1) or not relevant (0).

Objective:

Participants are required to develop an AI/ML model that can accurately classify questions as relevant or not relevant based on the provided text.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/NLP/Quora\_Dataset/quora\_dataset\_grafest.csv

***9. NEWYORK TAXI FARES***

Dataset:

The provided dataset contains information about NYC taxi trips, including details such as pickup and dropoff times, locations, passenger count, trip distance, and fare amounts. The dataset includes the following columns:

VendorID: A code indicating the provider associated with the trip record.

tpep\_pickup\_datetime: The date and time when the meter was engaged.

tpep\_dropoff\_datetime: The date and time when the meter was disengaged.

passenger\_count: The number of passengers in the vehicle.

trip\_distance: The elapsed trip distance in miles.

pickup\_longitude: The longitude where the meter was engaged.

pickup\_latitude: The latitude where the meter was engaged.

RatecodeID: The final rate code in effect at the end of the trip.

store\_and\_fwd\_flag: This flag indicates whether the trip record was held in vehicle memory before sending to the vendor.

dropoff\_longitude: The longitude where the meter was disengaged.

dropoff\_latitude: The latitude where the meter was disengaged.

payment\_type: The payment method (e.g., cash, credit card).

fare\_amount: The fare amount in dollars.

extra: Miscellaneous extras and surcharges.

mta\_tax: The MTA tax that is automatically triggered based on the metered rate in use.

tip\_amount: The tip amount in dollars.

tolls\_amount: The amount paid for tolls.

improvement\_surcharge: The improvement surcharge in dollars.

total\_amount: The total amount paid by the passenger.

Objective:

Participants are required to develop an AI/ML model that can accurately predict the total fare amount for a taxi trip based on the provided trip attributes.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/AI+for+Social+Good/Newyork\_taxi\_fare/New\_York\_Taxi\_Fare\_With\_Nan.csv

***10. NEWSARTICLES***

Dataset:

The provided dataset contains newspaper articles with labels indicating their category, title, and body. The dataset includes the following columns:

category: The category of the article (e.g., politics, sports, technology).

title: The title of the article.

body: The full text of the article.

Objective:

Participants are required to develop an AI/ML model that can accurately classify newspaper articles into their respective categories based on the provided title and body text.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/NLP/News\_Articles/new\_articles.csv

***11. MENTAL HEALTH***

Dataset Description:

The provided dataset contains user queries related to mental health issues and corresponding responses from mental health professionals. The dataset includes the following columns:

* Context: The user's query or statement regarding their mental health concerns.
* Response: The response provided by a mental health professional to address the user's query.

Objective:

The goal of this competition is to develop an AI/ML model that can provide mental health support by analyzing user queries and providing appropriate responses. The model should be able to understand the context of the user's query and generate empathetic and helpful responses.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/NLP/Mental+health+conversation/mental+health+convo/train\_augmented.csv

***12. FLICKERED IMAGE COMMENTS DATASET***

**Case Study: Clustering Flickr Image Comments**

**Dataset Description:**

The provided dataset contains comments associated with Flickr images. Each comment describes the content of an image in natural language. The dataset consists of a single column comment which contains textual descriptions.

**Objective:**

The goal of this competition is to develop a clustering algorithm that groups similar comments together. The clustering should be based on the semantic similarity of the comments, meaning that comments describing similar scenes or objects should be clustered together.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/NLP/flicker+comment/output.csv

***13. MENTAL HEALTH SURVEY***

Dataset:

The provided dataset contains information about employees in tech companies and their responses to a mental health survey. The columns in the dataset are as follows:

Timestamp: The time when the survey was taken

Age: The age of the employee

Gender: The gender of the employee

Country: The country where the employee is located

State: The state where the employee is located (if applicable)

Self\_employed: Whether the employee is self-employed

Family\_history: Whether the employee has a family history of mental health issues

Treatment: Whether the employee has sought treatment for mental health issues

Work\_interfere: How often mental health issues interfere with work

No\_employees: The number of employees in the company

Remote\_work: Whether the employee works remotely

Tech\_company: Whether the company is a tech company

Benefits: Whether the company provides mental health benefits

Care\_options: Whether the company provides options for mental health care

Wellness\_program: Whether the company has a wellness program

Seek\_help: Whether the company encourages employees to seek help for mental health issues

Anonymity: Whether anonymity is protected when seeking help for mental health issues

Leave: How easy it is to take leave for mental health issues

Mental\_health\_consequence: Whether there are negative consequences for discussing mental health issues at work

Phys\_health\_consequence: Whether there are negative consequences for discussing physical health issues at work

Coworkers: How supportive coworkers are regarding mental health issues

Supervisor: How supportive supervisors are regarding mental health issues

Mental\_health\_interview: Whether discussing mental health issues during an interview would have negative consequences

Phys\_health\_interview: Whether discussing physical health issues during an interview would have negative consequences

Mental\_vs\_physical: Whether mental health is considered as important as physical health at work

Obs\_consequence: Whether there are observed consequences for discussing mental health issues at work

Comments: Additional comments from employees

Objective:

Participants are required to develop an AI/ML model that can accurately predict whether an employee will seek treatment for mental health issues based on the provided factors.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/AI+for+Social+Good/Mental\_Health\_Survey/mental\_health\_survey\_tech\_company.csv

***14. WHO SUICIDE SURVEY***

Dataset:

The dataset contains the following columns:

country: The name of the country

year: The year of the data

sex: The sex of the individuals (male/female)

age: The age group of the individuals

suicides\_no: The number of suicides

population: The population of the age group

Objective:

The objective of this competition is to develop a machine learning model that can accurately predict the number of suicides in a given country, year, sex, and age group based on the provided demographic data.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/AI+for+Social+Good/suicided\_count/who\_suicide\_statistics.csv

***15. PREDICTING GLOBAL LAND TEMPERATURES BY COUNTRIES***

Dataset:

The dataset contains the following columns:

dt: The date of the temperature recording

AverageTemperature: The average temperature recorded on that date

AverageTemperatureUncertainty: The uncertainty in the recorded average temperature

Country: The name of the country

Objective:

The objective of this competition is to develop a machine learning model that can accurately predict the average temperature for a given country and date based on the provided historical data.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/AI+for+Social+Good/Global\_land\_Temperatures\_by\_Countries/GlobalLandTemperaturesByCountry.csv

***16. BRAIN***

Dataset:

The dataset is organized into the following structure:

brain/: The main folder containing all the images.

dirty\_test/: Subfolder containing test images.

glioma/: Subfolder containing images of glioma tumors.

meningioma/: Subfolder containing images of meningioma tumors.

notumar/: Subfolder containing images of notumar tumors.

pituitary/: Subfolder containing images of pituitary tumors.

train/: Subfolder containing training images.

glioma/: Subfolder containing images of glioma tumors.

meningioma/: Subfolder containing images of meningioma tumors.

notumar/: Subfolder containing images of notumar tumors.

pituitary/: Subfolder containing images of pituitary tumors.

Objective:

The objective of this competition is to develop a machine learning model that can accurately classify MRI images of brain tumors into one of the following categories: glioma, meningioma, notumar, and pituitary.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Bio-Medical+Analysis/Brain.zip

***17.LINKEDIN***

Dataset:

The dataset is organized into the following structure:

companies/: Folder containing company-related data.

companies.csv: CSV file containing company information.

company\_industries.csv: CSV file containing company industry information.

companies\_speciality.csv: CSV file containing company specialty information.

employee\_count.csv: CSV file containing employee count information.

jobs/: Folder containing job-related data.

benefits.csv: CSV file containing job benefits information.

job\_industries.csv: CSV file containing job industry information.

job\_skills.csv: CSV file containing job skills information.

salaries.csv: CSV file containing salary information.

mappings/: Folder containing mapping-related data.

industries.csv: CSV file containing industry mappings.

skills.csv: CSV file containing skill mappings.

Objective:

The objective of this competition is to develop a machine learning model that can cluster LinkedIn data into meaningful groups based on the provided information. Participants will use clustering techniques to identify patterns and relationships within the data.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/NLP/linkedin+dataset\_NLP.zip

***18.Helmet Classification***

**Dataset Description:**

The provided dataset contains images categorized into two classes: "helmet" and "no helmet." The dataset is organized into three subfolders: train, test, and validation. Each subfolder contains two additional subfolders named helmet and no helmet, which hold images corresponding to their respective labels.

**Objective:**

The objective of this competition is to develop an AI/ML model that can accurately classify images as either "helmet" or "no helmet." The model should be capable of distinguishing between images of individuals wearing helmets and those not wearing helmets. Participants will preprocess the dataset, develop and evaluate a machine learning model, and ensure it performs well across different image conditions and variations. The goal is to enhance safety measures by accurately identifying helmet usage in various scenarios.

<https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/helmet.zip>

***19.Kidney Stone***

Dataset Description:

The provided dataset contains images related to kidney stones, organized into three main subfolders: train, test, and validation. Each of these subfolders contains two additional subfolders: image and label. The image subfolder holds the images, while the label file contains the corresponding labels for each image, indicating the presence or absence of kidney stones.

Objective:

The objective of this competition is to develop an AI/ML model that can accurately detect kidney stones from medical images. The model should be capable of distinguishing between images with and without kidney stones. Participants will preprocess the dataset, develop and evaluate a machine learning model, and ensure it performs well across different image conditions and variations. The goal is to enhance diagnostic accuracy and support medical professionals in identifying kidney stones effectively.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Bio-Medical+Analysis/Kidney+stone+dataset/kidney+stone.zip

***20. Traffic***

**Dataset Description:**

The provided dataset contains images related to traffic scenarios, organized into two main subfolders: train and test. Each of these subfolders contains two additional subfolders: image and label. The image subfolder holds the traffic images, while the label file contains the corresponding labels for each image, indicating various traffic conditions or objects.

**Objective:**

The objective of this competition is to develop an AI/ML model that can accurately classify traffic images based on their labels. The model should be capable of distinguishing between different traffic conditions or objects present in the images. Participants will preprocess the dataset, develop and evaluate a machine learning model, and ensure it performs well across different image conditions and variations. The goal is to enhance traffic management and safety by accurately identifying and classifying traffic scenarios.

<https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/new+traffic.zip>

***21. Number Plate Classification***

**Dataset Description:**

The provided dataset contains images of vehicle number plates categorized by Regional Transport Offices (RTOs) across India. The dataset is organized into subfolders named after the RTO codes (e.g., DL, UP, HR, etc.). Each subfolder contains images of number plates corresponding to the respective RTO.

**Objective:**

The objective of this competition is to develop an AI/ML model that can accurately classify vehicle number plates based on their RTO codes. The model should be capable of distinguishing between number plates from different regions in India. Participants will preprocess the dataset, develop and evaluate a machine learning model, and ensure it performs well across different image conditions and variations. The goal is to enhance vehicle identification and management by accurately classifying number plates according to their RTO codes.

<https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/num+plate.zip>

***22. Sports Classification***

**Dataset Description:**

The provided dataset contains images related to various sports, organized into three main subfolders: train, test, and validation. Each of these subfolders contains additional subfolders named after different sports (e.g., basketball, soccer, tennis, etc.). Each sports-specific subfolder holds images corresponding to that particular sport.

**Objective:**

The objective of this competition is to develop an AI/ML model that can accurately classify images based on the sport depicted. The model should be capable of distinguishing between different sports categories. Participants will preprocess the dataset, develop and evaluate a machine learning model, and ensure it performs well across different image conditions and variations. The goal is to enhance sports image classification accuracy, which can be useful for organizing sports media, enhancing sports analytics, and improving user experiences in sports-related applications.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/sports.zip

***23. Bird Eye***

**Dataset Description:**

The provided dataset contains images of bird eyes, organized into two main subfolders: train and test. Each of these subfolders contains two additional subfolders: image and label. The image subfolder holds the bird eye images, while the label file contains the corresponding labels for each image, indicating specific characteristics or species of the birds.

**Objective:**

The objective of this competition is to develop an AI/ML model that can accurately classify bird eye images based on their labels. The model should be capable of distinguishing between different bird species or characteristics from the images of their eyes. Participants will preprocess the dataset, develop and evaluate a machine learning model, and ensure it performs well across different image conditions and variations. The goal is to enhance the accuracy of bird species identification and contribute to ornithological research and conservation efforts.

<https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/Bird+eye+view+dataset/top+view.zip>

***24.Yoga***

**Dataset Description:**

The provided dataset contains images of various yoga postures, organized into subfolders named after different yoga poses (e.g., Balasana, Savasana, etc.). Each subfolder contains images corresponding to the specific yoga posture.

**Objective:**

The objective of this competition is to develop an AI/ML model that can accurately classify images based on the yoga posture depicted. The model should be capable of distinguishing between different yoga poses. Participants will preprocess the dataset, develop and evaluate a machine learning model, and ensure it performs well across different image conditions and variations. The goal is to enhance the accuracy of yoga posture classification, which can be useful for yoga practice apps, fitness tracking, and improving user experiences in yoga-related applications.

https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/yoga.zip

***25. Handwriting Signature Dataset***

Dataset:

The dataset is organized into the following structure:

handwritten\_signature\_verification/: The main folder containing all the images.

forged/: Subfolder containing images of forged signatures.

real/: Subfolder containing images of real signatures.

Objective:

The objective of this competition is to develop a machine learning model that can accurately classify handwritten signature images as either real or forged.

<https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/Handwritten+signature+verification.zip>

***26.Mobile Segmentation Dataset***

Dataset:

The dataset is organized into the following structure:

Mobile\_Segmentation\_Dataset/: The main folder containing all the images.

good/: Subfolder containing images of surfaces without defects.

ground\_truth\_1/: Subfolder containing images with ground truth annotations for defect type 1.

ground\_truth\_2/: Subfolder containing images with ground truth annotations for defect type 2.

oil/: Subfolder containing images of surfaces with oil stains.

scratch/: Subfolder containing images of surfaces with scratches.

stain/: Subfolder containing images of surfaces with other types of stains.

Objective:

The objective of this competition is to develop a machine learning model that can accurately segment and classify images of mobile surfaces into one of the following categories: good, ground\_truth\_1, ground\_truth\_2, oil, scratch, and stain.

<https://grafest-bucket-neural-nexus-ctf.s3.ap-south-1.amazonaws.com/Neural+Nexus+Dataset/Image+classification+and+CV/Mobile+Segmentation+Dataset+Raw.zip>