Github Links **2**
Session 0: what and why of ML use cases in industry **7**
Group 0 7
Group 1 8
Group 2 8
Group 3 8
Group 4 8
Group 5 8
Group 6 8
Group 7 8
Group 8 8
Group 9 8
Session 1: model, action, and metrics **8**
Group 0 10
Group 1 11
Group 2 12
Group 3 12
Group 4 13
Group 5 15
Group 6 16
Group 7 17
Group 8 17
Group 9 18
Session 2: deep dive into statistical learning process **20**
Group 0 20
Group 1 20
Group 2 20
Group 3 20
Group 4 20
Group 5 20
Group 6 21
Group 7 21
Group 8 21
Group 9 21
Session 3: ML on cloud **21**
Session 4: what and why of Deep Learning use cases in industry **23**
Group 0 23
Group 1 23
Group 2 23
Group 3 23
Group 4 23
Group 5 24
Group 6 24

Group 7 24
Group 8 24
Group 9 24
Session 5: deep dive into neural networks **24**
Group 0 24
Group 2 24
Group 3 25
Group 4 25
Group 5 25
Group 6 25
Group 7 25
Group 8 25
Group 9 25
Final Project Groups **25**
Questions **32**

Github Links

Please do the following:

- create a GitHub account
- create an **AppliedMachineLearning** repository
- paste the link to the repository below together with your roll number, name and email

roll number	name	email	github
MDS202340	Rohit Roy	rohitatcmi@gmail.com	https://github.com/scarroy-02/AppliedMachineLearning
MDS202312	Aryansh Raj Saxena	aryansh.mds2023@cmi.ac.in	https://github.com/Ars-Gilfoyle-99
MCS202304	Aritra Majumder	aritram.mcs2023@cmi.ac.in	https://github.com/Aritra8438/AppliedMachineLearning
MDS202335	Nandini Jaiswal	nandini.mds2023@cmi.ac.in	https://github.com/Nandini-Jaiswal/AppliedMachineLearning

MDS202324	Esha Bhattacharya	esha.mds2023@cmi.ac.in	https://github.com/EBH2002/AppliedMachineLearning
MCS202311	Sai Sasank Y	sai.sasank.yadati@gmail.com	https://github.com/SasankYadati/AppliedMachineLearning
MDS202352	Utpalraj Kemprai	utpalraj.mds2023@cmi.ac.in	https://github.com/U1Kemp/AppliedMachineLearning
MDS202320	Deepanshu Mittal	deepanshu.mds2023@cmi.ac.in	Deepanshuharsh2209/AppliedMachineLearning
MDS202326	Hiba AP	hiba.mds2023@cmi.ac.in	https://github.com/HibaAp/Applied-Machine-Learning
MCS202409	J Shankar Narayanan	shajaganbpgc@gmail.com	https://github.com/shankar-n/AppliedMachineLearning
MDS202437	Srishti Lakhotia	srishtilakhotia1@gmail.com	https://github.com/creationisme/AppliedMachineLearning
MDS202315	Ayush Yadav	ayushy.mds2023@cmi.ac.in	https://github.com/yadavayush7028/AppliedMachineLearning
MDS202330	Kalyani Gohokar	kalyani.mds2023@cmi.ac.in	https://github.com/Kvgohokar/AppliedMachineLearning
MDS202334	Mayank Nagar	mayank.mds2023@cmi.ac.in	https://github.com/mkreman/AppliedMachineLearning
22f3001413	Dhruv R	22f3001413@ds.study.iitm.ac.in	https://github.com/dhruv1108git/AppliedMachineLearning
21f3002975 (IIT Madras, BS)	A.J.R.Vasu	21f3002975@ds.study.iitm.ac.in	https://github.com/21f3002975/AppliedMachineLearning
MDS202339	Pritam Padhan	pritamp.mds2023@cmi.ac.in	https://github.com/PritamPadhan/AML-Pritam-Padhan
MDS202333	Malde Dharmi	dharmi.mds2023@cmi.ac.in	https://github.com/DharmiMalde/AppliedMa

			chineLearning
MDS202301	Aalekhya Mukhopadhyay	aalekhya.mds2023@cmi.ac.in	https://github.com/M-Aalekhya/AppliedMachineLearning
MDS202341	Salokya Deb	salokya.mds2023@cmi.ac.in	https://github.com/Salokya35/Applied-Machine-Learning-2025
MDS202342	Samaroha Chatterjee	Samaroha.mds2023@cmi.ac.in	https://github.com/Samshoww/-AppliedMachineLearning
MDS202318	Chandranath Bhattacharya	chandranath.mds2023@cmi.ac.in	https://github.com/CHANDRANATHBHATTACHARYA/AppliedMachineLearning
MDS202331	Keshev Kumar	keshev.mds202331@cmi.ac.in	https://github.com/Big-Marvel/AppliedMachineLearning
NA: 0	Nikita Kumari	nikitakumari@study.iitm.ac.in	https://github.com/nikita-12-com/AppliedMachineLearning
MDS202344	Satyaki Mullick	satyaki.mds2023@cmi.ac.in	https://github.com/satyaki-02/Applied-Machine-Learning
MDS202350	Sreeja Choudhury	sreeja.mds2023@cmi.ac.in	https://github.com/SreejaChoudhury/Applied-Machine-Learning
MDS202347	Siddhesh Maheshwari	siddheshm.mds2023@cmi.ac.in	https://github.com/msiddhesh/AppliedMachineLearning
MDS202337	Nooh Ali	nooh.mds2023@cmi.ac.in	noohalibengla/AppliedMachineLearning
MDS202307	Ananya Sinha	ananyas.mds2023@cmi.ac.in	github.com/ananyasinha01/AppliedMachineLearning
MDS202308	Aniket Tiwari	anikett.mds2023@cmi.ac.in	https://github.com/ani98622/AppliedMachineLearning

MDS202332	Kironmoy Roy	kironmoy.mds2023@cmi.ac.in	https://github.com/kiru2001/AppliedMachineLearning
MDS202314	Atul Anant	atula.mds2023@cmi.ac.in	https://github.com/AtulAnant/AppliedMachineLearning
MDS202325	Gauranga Kumar Baishya	gauranga.mds2023@cmi.ac.in	https://github.com/Gauranga2022/CMI/tree/main/Sem4/AppliedML
MDS202317	Bibek Paul	bibek.mds2023@cmi.ac.in	https://github.com/X-Warrior361/AppliedMachineLearning
MCS202103	Bijayan Ray	bijayan@cmi.ac.in	https://github.com/BijayanRay/Applied-Machine-Learning-2025-CMI-course
MDS202310	Aparna C	aparnac.mds2023@cmi.ac.in	https://github.com/Aparnamaya/Applied-ML
MDS202329	Jerin Biju	jerin.mds2023@cmi.ac.in	https://github.com/JerinKallakulam/Applied-ML
MDS202311	Arka Roy	arkar.mds2023@cmi.ac.in	https://github.com/ArkaRoy1998/Applied-Machine-Learning
MDS202406	Aman Raj	amanr.mds2024@cmi.ac.in	
MDS202322	Divyanshi Kumari	divyanshi.mds2023@cmi.ac.in	https://github.com/divK12/Applied-Machine-Learning
MDS202345	Sayantani Saha	sayantanis.mds2023@cmi.ac.in	https://github.com/Sayantani0503/AppliedMachineLearning
MDS202327	Himanshu	himanshu680.mds2023@cmi.ac.in	https://github.com/waitasecant/CMI-HW/tree/main/AppliedML
MDS202351	Suneha Sen	suneha.mds2023@cmi.ac.in	https://github.com/SunehaSen/AppliedMachineLearning

MDS202305	Aman	aman.mds2023@cmi.ac.in	https://github.com/Aman-exp/AppliedMachineLearning
MDS202336	Narendra C	narendra.mds2023@cmi.ac.in	https://github.com/Naren221/Applied-Machine-Learning
BMC202230	Krishanu Bandyopadhyay	krishanu@cmi.ac.in	krish-baner/AppliedMachineLearning: Applied Machine Learning Course in CMI (Sem VI)
MDS202319	Chenna Sai Sandeep	saisandeep.mds2023@cmi.ac.in	https://github.com/chennasaisandeep/Applied-Machine-Learning
MDS202309	Ankita	ankita.mds2023@cmi.ac.in	https://github.com/ankita-rajpal/Applied-Machine-Learning
MDS202323	Eada Surya Deva	surya@cmi.ac.in	https://github.com/Infinitysurya/AppliedMachineLearning
MDS202348	Soumya Dasgupta	soumya@cmi.ac.in	https://github.com/Soumya86177/Applied-Machine-learning-Soumya
MDS202353	Vikas Chaudhary	vikas.mds2023@cmi.ac.in	https://github.com/Vikas9758/AppliedMachineLearning
MDS202316	Bhuvnesh Sahu	bhuvnesh.mds2023@cmi.ac.in	https://github.com/bhuvneshsahu01/Applied-Machine-Learning.git
MDS202354	Vishal Maurya	vishal.mds2023@cmi.ac.in	https://github.com/Weshallfly/AppliedMachineLearning
MDS202313	Ashish Bisht	ashishb.mds2023@cmi.ac.in	https://github.com/Ash-park/Applied-Machine-Learning.git
MDS202328	Himanshu	himanshu801.mds2023@cmi.ac.in	https://github.com/22Himanshu/AppliedMachineLearning

MDS202349	Soumyajoy Kundu	soumyajoy.mds2023@cmi.ac.in	https://github.com/soumyajoykundu/Applied-Machine-Learning-2025
MDS202343	Sattik Biswas	sattik.mds2023@cmi.ac.in	https://github.com/sattik19
MCS202415	Shruti Patil	shruti@cmi.ac.in	https://github.com/shrutisharadpatil/Applied-Machine-Learning
MCS202302	Akhoury Shauryam	akhoury@cmi.ac.in	https://github.com/Satan-Claws/Applied-Machine-Learning
BMC202239	Nirjhar Nath	nirjhar@cmi.ac.in	https://github.com/combinoob/Applied-Machine-Learning-2025

Session 0: what and why of ML use cases in industry

Groups: last digit of your roll number
(e.g. MCS202304 -> Group 4, 21f3002975 -> Group 5)

- You will use the comment option for each group member to propose the use case.
- Each person will propose at least one use case.
- If a use case is already proposed, you will have to propose a different one

Goal:
Understand the ML use cases (**what and why**) at different well-known companies.
Please give a succinct description of use case as following “what/why” format:
“X to improve Y” (e.g. email spam filter to improve user engagement) by specifying both X, the what part and Y, the why part. Also provide the link to the use case.
You may use: https://www.evidentlyai.com/ml-system-design

Group 0

Apple

Group 1

Google

Group 2

Facebook

Group 3

Uber

Group 4

LinkedIn.

Group 5

Amazon

Group 6

Netflix

Group 7

Spotify

Group 8

Swiggy

Group 9

Paypal

Session 1: model, action, and metrics

Groups: last digit of your roll number

(e.g. MCS202304 -> Group 4, 21f3002975 -> Group 5)

- The person with **the smallest "second last digit"** value is moderator by default, in case of collision go to "last two digits" (or last 3, 4, 5 digits etc) for resolution.
- You will use the comment option for each group member to propose the answer.
- The moderator will synthesize and select the final answers from the comments.
- You may also use the direct chat option in zoom.

Goal:
Understand the causal flow of ML applications (data=>model=>action=>impact) underlying the real world use cases of the common ML methods using different modality of data (e.g. numerical, text, image, multimodal etc)

For each box below fill in the following:

- what/why: X to improve Y (e.g. email spam filter to improve user engagement)
- link (the url from evidently/ml-system-design
- data: what is the choice of data used to train the model
- model: input->output (e.g. sms_text -> pr(spam))
- action: what action is taken on the model output (e.g. classify as positive above threshold)
- impact: what is the (business) impact (e.g. reduction in spam => better use experience => more engagement)

Do not worry about quantifying the choices.
Also do not worry about the metrics and experiments.
We are limiting this exercise to only the following qualitative understanding:

- using data D we build model M (of input->output)
- using model M we take action A
- using action A we drive impact I

https://www.evidentlyai.com/ml-system-design

data/method	classification	regression	ranking
numerical	- what - why - link - data - model: input->output - action - impact		
text			
image			
audio			

video			
text + image			
text + audio			

You may use separate sections below to note down the answers.
Please use the google doc comment feature.

Steps

1. select a use case link for classification, regression, or ranking method and put the link in appropriate box together with group name. please do not select the same link for two groups. whichever group has pasted the link first in the box keeps it and the second group has to search for another link.
2. identify what
3. identify why
4. identify the choice of data
5. identify the choice of model: input -> output
6. identify the choice of action
7. identify the choice of impact

	choices	metrics
model	What core ML model do we use? What is its input? What is its output? (ML model could also be pretrained model or black-box model)	How do we measure the effectiveness of our ML model? How do we select the best model? (Pretrained or black-box model's published performances on various tasks can be a guide)
action	What would you do with the output of the model?	How do we measure the effectiveness of the action? How do we select the best action?

Group 0

- **Link:** [Etsy Engineering | Efficient Visual Representation Learning And Evaluation]
- **What:** The blog discusses efficient visual representation learning using models like EfficientNet and Vision Transformers (ViT) for image retrieval at Etsy.

- **Why:** To improve search-by-image functionality by learning better visual embeddings while optimizing for computational efficiency and retrieval accuracy.
- **Data:** Etsy's large-scale product image dataset is used to train and evaluate models through contrastive learning and nearest neighbor retrieval tasks.
- **Model:** Combination of EfficientNet and Vision Transformers (ViT) for visual representation learning. EfficientNet provides a computationally efficient convolutional neural network, while ViT captures long-range dependencies in images.
- **Action:** Adopting advanced deep learning architectures, fine-tuning models, and leveraging multitask learning for embedding optimization.
- **Impact**: Enhanced search accuracy, faster retrieval, and improved shopping experience for Etsy users.

	choices	metrics
model	image -> embedding (last layer of a multi-task model)	- multi-task loss used during training of the embedding model
action	- (offline) build vector index using embeding model - (online) vectorize search_image and query the index and retrieve top-k relevant	- recall@k

Group 1

- Link: https://research.google/blog/auto-generated-summaries-in-google-docs/
-
- What: Auto-generated summaries in Google docs
- Why: To improve user experience
- Data: manually-generated summaries from academic abstracts that were consistent with typical use cases
- model (input->output): document words -> summary words
- Action: writers annotate their documents with summaries, and readers comprehend and navigate documents more easily
- Impact: expected impact is increased user engagement with docs

	choices	metrics
model	pretraining:	

	collection of sentences with some masked -> masked sentences token probabilities final: document -> summary	cross entropy BLEU/ROUGE
action	display (on demand)	% accepts (offline manual)

Group 2

- link:https://www.google.com/url?q=https://slack.engineering/how-we-built-slack-ai-to-be-secure-and-private/&sa=D&source=docs&ust=1738071012700545&usg=AOvVaw0GyVNoLZnAJZUHrgAHoTJN
- What: Slack introduced AI features like summarization and search to boost productivity.
- why:Addressed user challenges of information overload and difficulty finding content.
- data:Used only customer data visible to the user, ensuring strict privacy compliance.
- model (input->output):Leveraged off-the-shelf LLMs with RAG for secure, stateless operations.
- action:Built AI architecture within Slack's VPC with enterprise-grade security and compliance.
- impact:Improved productivity, ensured data privacy, and delivered relevant AI-driven insights.

Same as Group 4

Group 3

- **Link**: https://medium.com/airbnb-engineering/prioritizing-home-attributes-based-on-guest-interest-3c49b827e51a
- **What**: Prioritizing home attributes using user interest classification and understanding guest preferences by analyzing unstructured text data.
- **Why**: To improve the relevance of home listings for users, increasing user satisfaction and engagement.
- **Data**: Historical search data, including text descriptions of listings and user interaction data (e.g., clicks, bookings).
- **model (input->output)**: Input: Property features (e.g., size, location, amenities), guest preferences.
- **Output:** Ranked list of attributes based on guest interest.
- **Action:** Use the predicted relevance scores to prioritize and display home attributes that are most likely to align with user interests.

- **Impact:** Enhanced user experience by surfacing relevant attributes, leading to increased engagement, bookings, and user retention.

	choices	metrics
model	posting -> home attributes (via NER + linker) —---------- home_location + property_type + capacity + luxury_level + home_attribute -> importance score for home attribute (frequency of mention of home_attribute in messages, reviews, and customer service tickets)	precision/recall/confision matrix —---------- AUC ROC on manually labelled balanced sample of posting, attribute -> is_relevant
action	rank all possible attributes for the home based on an importance score derived from the predicted frequencies display top-ranked attributes (which are believed to be most important to the potential guests)	(offline) precision@k (basesd on manually labeled relevance labels)

From Session 0

- **Link**: https://github.com/Ivan-Zhou/Uber_Driver_Schedule_Optimization
- **What**: Mixed Integer Programming (MIP) model is needed to optimize trip value.
- **Why**: Scheduling flexibility and heatmaps of customer demand to maximize revenue, but a Mixed Integer Programming (MIP) model is needed to optimize trip value by balancing demand and trip duration.
- **Data**: Uber's "Uber Movement" initiative.
- **Model** (input->output): Mixed Integer Programming (MIP) model.
- **Action**: Optimize trip value using MIP.
- Impact: Optimize trip value balancing demand and trip.

Group 4

- **Link:** https://slack.engineering/how-we-built-slack-ai-to-be-secure-and-private/
- **What**: Slack AI is a secure and private AI system integrated into Slack, designed to provide features like message summarization and intelligent search without compromising user data.
- **Why**: To enhance productivity while maintaining strict data privacy and security, ensuring user trust and compliance with enterprise standards.
- **Data**: Encrypted user messages, metadata, anonymized training data, and security logs.
- **Model (input->output):** Input: Encrypted messages + metadata. Output: AI-generated responses (e.g., summaries, answers). Techniques: Fine-tuned LLMs, on-device processing, zero-data retention
- **Action**: Train on anonymized data, enforce encryption, and ensure no data storage.
- **Impact**: Secure conversation for corporate agencies, preventing adversarial communication.

	choices	metrics
model	text_chunk -> embedding	- black box model performance on various tasks (e.g. squad question answering)
action	- (offline) build vector index on historic text data - (online) embed search_text and query the vector index to retrieve the matches	- NDCG@k (manual evaluation) - Mean Reciprocal Rank (based on search history click data)

Group 4 (alternate)

- **Link:** Improving job matching with machine-learned activity features
- **What**: Improving job matching with machine-learned activity features
- **Why**: To enhance the accuracy and relevance of job recommendations for users.
- **Data**:
  - **Job IDs**: Unique identifiers for jobs interacted with.
  - **Job Embeddings**: Low-dimensional vector representations of jobs, including features like industry, skills, and title.
  - **Actions**: Member actions on jobs (e.g., apply, save, dismiss).
  - **Timestamps**: Time of each job interaction.
  - **Sequence Length**: Truncated or padded to 32 most recent actions per member.
- **Model (input->output):**
  - **Input**: a sequence of Pensieve job embeddings *(j,e,a,t)* concatenated with their three-dimensional one-hot label (APPLY/SAVE/DISMISS)
  - **Output**: Improved job recommendations tailored to user preferences.
  - **Techniques**:

        - Machine-learned activity embedding
        - Geometrically-decaying average embedding
- **Action**:
    - first trained a simple Geometrically-decaying average embedding model that produced three separate activity embeddings per member, by simply averaging each job embedding within each action type.
    - The Machine-learned activity embedding sequence model is trained on the first n-1 job embeddings and their corresponding one-hot action labels as input, generates a final embedding, and uses it with the held-out job embedding to predict if the last action was positive (APPLY/SAVE) or negative (DISMISS) via a Hadamard product, fully-connected layer, and softmax, trained with cross-entropy loss.
    - In the both the cases, output was the job recommendations for users, so they were rolled out for users under the job tab.
- **Impact**: Across the four iterations of implementing this model,, the number of job application increased by over 10% and confirmed hires by 5% during online A/B testing.

Group 5

- **Link**:https://www.amazon.science/latest-news/how-amazon-music-uses-recommendation-system-machine-learning
- **What**: Amazon Music uses ML to enhance its recommendation system for personalized music suggestions through Alexa.
- **Why**: The system provides better recommendations by incorporating real-time context like mood or activity, often missing in traditional methods.
- **Data**: User's music history, user persona profile and real-time contextual data gathered through Alexa conversations.
- **Model** (input->output): Inputs are user preferences and context, processed to generate personalized music recommendations.
- **Action**: Alexa uses Amazon Music's recommendation system to offer personalized song suggestions based on user preferences and past interactions, all through voice commands.
- **Impact**: Boosts user engagement by delivering personalized experiences, measured through satisfaction surveys, higher interaction rates with Alexa for recommendations, and an increase in music listening and platform activity.

	choices	metrics
model	user_context + alexa prompt -> pr(alexa prompt is useful to the user and user ultimately selects a recommended music) (recommendation based on	AUC ROC

	vector DB nearest neighbor retrieval on encoded conversation)	
action	dialogue policy selecting follow-up prompt based on usefulness probability	(offline) cross entropy of pr(follow-up prompt) with ultimate success/failure after using the follow-up prompt (online) #successful_outcomes #conversation_turns

From Session 0

- Link: https://aws.amazon.com/personalize/
- What: Personalized product recommendations.
- Why: To improve customer satisfaction and sales.
- Data: Purchase history, browsing patterns, ratings.
- Model (input->output): Customer behavior → Recommended products.
- Action: Display recommended products on the website or app.
- Impact: Increased sales and better customer retention.

Group 6

- **Link**: https://engineering.grab.com/llm-powered-data-classification
- **What**:LLM-powered data classification to identify sensitive data at Grab.
- **Why**:To automate metadata generation, streamline data access, and improve data governance.
- **Data**:Use schemas and sample records to identify sensitive fields.
- **Model (input->output)**:LLM processes schema info to classify fields into appropriate tags (e.g., PII, business metrics).
- **Action**:Automatically tag sensitive fields using rules and an LLM.
- **Impact**:Efficient, accurate classification reduces manual effort and enhances data security.

	choices	metrics
model	prompt + column_name -> column_tag (black-box model)	- published benchmark performance of the black-box model on different tasks
action	automatically obtain tags (and store in DB)	- confusion matrix - precision

		- recall

Group 7

- **Link:** https://engineering.atspotify.com/2021/11/the-rise-and-lessons-learned-of-ml-models-to-personalize-content-on-home-part-ii/
- **What:** Using Machine Learning models to personalize home page content for users, ensuring relevant recommendations and better engagement.
- **Why:** For better user experience and user retention.
- **Data:** The best albums, playlists, artists, and podcasts are selected for each listener.
- **Model (input->output):** Processes user behavior, listening history, and preferences to generate personalized recommendations.
- **Action:** Dynamically updates home page content for each user based on their unique preferences and activity.
- Impact: Improved user engagement, increased satisfaction, and higher retention through relevant and personalized content delivery.

	choices	metrics
model	user profile, podcast profile, historic interactions, day-time -> pr(user will listen to podcast)	AUC ROC
action	(explore/exploit) rank show top-k	NDCG@k

Group 8

- Link: https://medium.com/foodpanda-data/introduction-optimising-budget-through-data-analysis-030b2f39ad0c
-
- What: The ultimate objective is to create a marketplace environment where the fulfilment of customer orders is not only efficient but also economically sustainable
- Why: This involves creating a robust framework that enables us to allocate resources effectively, ensuring that there is a harmonious match between the influx of customer orders and the capacity of riders available to fulfil them.
- Data: Utilized historical order trends, peak demand periods, and rider availability patterns.

- model (input->output): Employed machine learning algorithms to forecast future demand and optimize rider deployment based on budget inputs.
- Action: **:** Use ML to forecast demand and optimize rider deployment. Balance CARC incentives (drive orders) and rider payments (ensure supply).Guide Allocations via Dashboard priortising profit and CPO.
- Impact: Improved operational efficiency, maximized profitability, and maintained rider satisfaction, ensuring long-term platform viability.

	choices	metrics
model	demand, supply, *customer_acquisition_budget*, rider_incentive_budget -> GMV (daily/weekly)	- MAPE (growth curve, profitability curve for sanity check with domain intuition)
action	choose budget that maximizes growth while maintaining certain profitability levels and rider availability constraints	- estimated increase in n_orders - estimated rider fill rate

Group 9

- Link: https://careersatdoordash.com/blog/improving-etas-with-multi-task-models-deep-learning-and-probabilistic-forecasts/
- What: Enhancing Estimated Time of Arrival (ETA) predictions using multi-task models.
- Why: To provide customers with accurate and reliable ETAs across a broad spectrum of delivery types and ETA scenarios, each delivery's variability in timing, geography, and conditions.
- Data:
    - **Choices:**
        - **Signals:** Order Data (location, time, number of items, type, special instructions), Delivery Person Data (location, rating, experience, average_delay, order_history), Merchant Data (location,history,average_delay, active_orders,type), Geographic Data (distance, traffic),Weather Data (real time weather), Customer Data (uid, order_history), Special Events Calendar.
        - **Labels:** Order fulfilment time in minutes.
    - **Metrics:** Stratification across different Delivery types, geographical regions (city, country, state etc.), special events, driver rating, time of day, weather pattern, historic ETA and arrival time distributions.
    - **Experiments:** data profile by segments, distributions of ETA, errors

- Model (input->output):
    - **Choices:**
        - Real time data -> **Aggregated/ temporal features, Categorical features** -> **Feature Embeddings -> predicted ETA**
    - **Metrics:**
        - Calibration
        - Accuracy:
            - continuous ranked probability score (CRPS)
    - **Experiments:**
        - Benchmarks:
            - directly model ETA for each type
            - residual to static model
        - Features: Train deep neural network for generating feature embeddings.
        - Regression model to get ETA for each bin

- Action:
    - **Choices:** This data is processed via APIs and machine learning models for accurate, dynamic updates.
    - **Metrics:** Latency of ETA service.
    - **Experiments:** Optimizing model for better tradeoff b/w complexity, accuracy and response time.
- Impact
    - **Choices:** Improved ETA prediction system enhances customer satisfaction through accurate delivery times, reduces churn, reduces order cancellations, and increases trust in the platform. For Dashers, it optimizes route planning, reduces idle time, and boosts efficiency, ultimately driving higher engagement and retention across users.
    - **Metrics:** % cancellations due to arrival after pickup ETA, average absolute ETA error, churn rate, dashers mean idle time.
    - **Experiments:** A/B Testing

	choices	metrics
model	order profile + customer profile + restaurant profile + weather condition + driver demand-supply -> pr(ETA <= x)	- CRPS (measures how well the predicted ETA aligns with actual delivery times)
action	- display 95% confidence interval before customer	- accuracy (how often delivery arrives within time based on

	places order (can also be used the ETA estimates for other downward optimization tasks)	model prediction interval) - consistency (variance of actual vs predicted)

Session 2: deep dive into statistical learning process

Groups: last digit of your roll number
(e.g. MCS202304 -> Group 4, 21f3002975 -> Group 5)

- The person with **the smallest "second last digit"** value is moderator by default, in case of collision go to "last two digits" (or last 3, 4, 5 digits etc) for resolution.
- Please use the comment section to **post 3 questions** regarding the topic
- Group number i + 1 (mod 10) will answer the questions for group i

Group 0

train/
test split

Group 1

bias variance tradeoff

(Please recover the questions from Session 2 from All Comments tab and repost in the new comment or Samaroha Chatterjee please reopen the comment)

Group 2

learning curves

Group 3

model sanity checks

Group 4

data preparation

Group 5

exploratory data analysis

Group 6

feature engineering

Group 7

hyperparameter tuning

Group 8

model debugging

Group 9

prediction vs inference

Session 3: ML on cloud

Groups: last digit of your roll number (Group 0 merges into Group 9)
(e.g. MCS202304 -> Group 4, 21f3002975 -> Group 5)

- The person with **the smallest "second last digit"** value is moderator by default, in case of collision go to "last two digits" (or last 3, 4, 5 digits etc) for resolution.
- Please find the **components** and **process flow** to implement each of the following ML solutions on different cloud service providers

	AWS	GCP	Azure
batch scoring	Group 1	Group 2	Group 3
real-time scoring (streaming)	Group 4	Group 5	Group 6
RAG	Group 7	Group 8	Group 9

scribbr.com

AWS
https://www.googleadservices.com/pagead/aclk?sa=L&ai=DChcSEwjJiMGj1euMAxU1KkQIHSeIFhcYABABGgJkeg&ae=2&aspm=1&co=1&ase=5&gclid=Cj0KCQjw_JzABhC2ARIsAPe3ynpFo-BIft6GqM99msFD_9vQEEYTY0LUowBff0DgyZrssEeHloQrQkMaAknjEALw_wcB&ohost=www.google.com&cid=CAESVeD2errMfizsNewrj_2c1bX9poY6uvLR0NVAokuD25uenx3

JfSOhNWM21YVgJqaA_oaZSnKXD6d5tdDrOgGnDc0OFW6kBAEcA5ADWODZTYHs-iFAyT8&sig=AOD64_2ul40tnnFSYt3Sf_K-rsGX7bWgoA&q&adurl&ved=2ahUKEwj56Lyj1euMAxVTD0QIHX8nI3cQ0Qx6BAgWEAE

GCP
https://www.googleadservices.com/pagead/aclk?sa=L&ai=DChcSEwjpl8HD1euMAxXtCe8CHStRLB4YABADGgJkeg&co=1&ase=2&gclid=Cj0KCQjw_JzABhC2ARIsAPe3ynoGOgWOO6J1ALQvLzNCCQk4AwMxqgqJiZwyGxRVuE_bQpdM6ViSvkAaAgF6EALw_wcB&ohost=www.google.com&cid=CAESVeD22cHt1PeSrT-2MSWezmMb01AE9i0aQNUzCLBxIZFNFbN4ScUpWFHVA0zzpTJBY07p1pess2LcNlEdOOQgLcnHvtytTDVsi4GyeC5zav1rkRbwoaY&sig=AOD64_2cenhgp0nDgsx_cWmE2YEiqE_h2A&q&nis=4&adurl&ved=2ahUKEwj-0rzD1euMAxV7HUQIHYBtEbkQ0Qx6BAgLEAE

Azure
https://www.googleadservices.com/pagead/aclk?sa=L&ai=DChcSEwjD0eTX1euMAxUuL0QIHe_gN_0YABACGgJkeg&ae=2&aspm=1&co=1&ase=5&gclid=Cj0KCQjw_JzABhC2ARIsAPe3ynrnScMdmoeXSwGIgI4is6CqN1Wdn70oUKHdSMC0NNBclipUpUIa10EaAhWjEALw_wcB&ohost=www.google.com&cid=CAESVeD29Bzg2CiOrLsl-ml_zhkPeA4ePzxXp5DcQ3UERc9lIw-TlJRJiQ03owhzeHhAw5M70h4OTI56NOjDcd-yaR55aixUMv-6ZAdhpWOALaKgWNPBns4&sig=AOD64_0zgJgQ63vrcNs1fbQx0AVem-TUGQ&q&adurl&ved=2ahUKEwi_xuDX1euMAxVGOUQIHdhIMUUQ0Qx6BAgJEAE

serverless function
serverless container
vm cluster

database server (sql/nosql, row-oriented vs column oriented)
object storage
file storage
volume storage

serverless endpoint
API Gateway
Load Balancer
CDN
Cache
Message Queue

batch job
streaming

Managed ML Platform

Session 4: what and why of Deep Learning use cases in industry

Groups: last digit of your roll number
(e.g. MCS202304 -> Group 4, 21f3002975 -> Group 5)

- You will use the comment option for each group member to propose the use case. specific to deep learning application
- Each person will propose at least one use case.
- If a use case is already proposed, you will have to propose a different one

Goal:
Understand the Deep Learning use cases (**what and why**) at different well-known companies.
Please give a succinct description of use case as following “what/why” format:
“X to improve Y” (e.g. email spam filter to improve user engagement) by specifying both X, the what part and Y, the why part. Also provide the link to the use case.
You may use: https://www.evidentlyai.com/ml-system-design
Together with what and why, please write brief description of the **method** used during deep learning application (e.g. **black-box use, finetuning, transfer learning, training from scratch** etc)

Group 0

Apple

Group 1

Google

Group 2

Facebook

Group 3

Uber

Group 4

LinkedIn

Group 5

Amazon

Group 6

Netflix

Group 7

Spotify

Group 8

Swiggy

Group 9

Paypal

Session 5: deep dive into neural networks

Groups: last digit of your roll number
(e.g. MCS202304 -> Group 4, 21f3002975 -> Group 5)

- The person with **the smallest "second last digit"** value is moderator by default, in case of collision go to "last two digits" (or last 3, 4, 5 digits etc) for resolution.
- Please use the comment section to **post 3 questions** regarding the topic
- Group number i + 1 (mod 10) will answer the questions for group i

Group 0

Neural networks: graph representation

Group 1
Neural networks: function composition representation (nonlinearity, matrix multiplication etc)

Group 2

activation functions

Group 3

backpropagation

Group 4

Regularization for neural networks

Group 5

pytorch autograd and different optimization methods (e.g. adam, momentum etc)

Group 6

cnn

Group 7

rnn

Group 8

lstm

Group 9

transformer

Final Project Groups

- Max 4 people in one group
- Each person has a primary responsibility

For projects, please submit the tentative plan by **20 Mar 2025**

- problem statement
- data (e.g. Kaggle datasets, open source ML datasets, manual curation)
- models
- responsibilities (each person has a primary responsibility)

Final presentation:

- 10min + 5min for question answers

References:

https://www.projectpro.io/projects/data-science-projects

http://cs230.stanford.edu/past-projects/

https://ocw.mit.edu/courses/15-097-prediction-machine-learning-and-statistics-spring-2012/pages/projects/

Group - group name - member 1 - member 2 - member 3 - member 4 preferred slot: **6th/7th/8th/9th 5pm to 7pm (15-20min each group)** on each day we will have presentations of all those people who have preferred the given slot, in the alphabetical order of the name of the group	Project - problem - data - model - responsibilities Github link: - code - presentation

Group name: Breaking RAG
- Aman (MDS202305)
- Himanshu (MDS202327)
- Keshev Kumar (MDS202331)
- Utpalraj Kemprai (MDS202352)
 	
Preferred Slot : 6th May 2025, 5:00 pm
	- Problem: RAG-powered search engine for Research Papers
	- Data: arxiv api, semantic scholar api etc.
	- Model: Mistral 7, Gemini 2.0-flash, Phi-3-mini, etc.
	- Responsibilities:
	  ○ **Himanshu** - Data Extraction from APIs, UI Development, ppt
	  ○ **Keshev** - Data Extraction from pdfs, docs etc, Research possible tools/tech for use
	  ○ **Utpalraj** - End to end integration, UI development
	  ○ **Aman** - model evaluation and comparison
	
	Project (GitHub Link)
	https://github.com/U1Kemp/RAG-powered-search-engine-for-Research/
	
	Presentation(Slides)
	https://github.com/U1Kemp/RAG-powered-search-engine-for-Research/blob/main/Presentation/RAG-Powered%20Chatbot%20for%20Research%20Queries.pdf

Group name : KasuKabe Defence Force - Aniket Tiwari (MDS202308) - Deepanshu Mittal (MDS202320) - Kironmoy Roy (MDS202332)	Project - **Problem**: GenAI application which converts any CV into the company's standard template - **Data** : No specific data used

Preferred Slot : 6th May 2025, 6:30 pm	- **Model**: Gemini or any other open source model - **Responsibilities**: 1-Custom Functions to analyse user CV (Deepanshu Mittal) 2-Custom template development (Kironmoy Roy) 3-Frontend & API Development (Aniket Tiwari) Presentation: https://github.com/ani98622/AML-Project/blob/main/presentation/AML%20Presentation.pdf Project Link(Github): https://github.com/ani98622/AML-Project/tree/main Demo Link(streamlit): https://resume-builder-aml-project.streamlit.app/
Group Name : Very Good Team - Ananya Sinha (MDS202307) - Ankita (MDS202309) - Divyanshi Kumari (MDS202322) - Rohit Roy (MDS202340) Preferred Slot: **8th May**	Project ● Problem: GANs for Human Face Generation and Completion ● Data: The CelebA dataset ● Model: DCGAN ● Responsibilities: ○ Ananya Sinha : Data acquisition and preprocessing, presentation ○ Rohit & Ankita : Model architecture and training ○ Divyanshi : Model testing and fine-tuning Links : ● Code ● Presentation

Group Name: Noise Assassins - Aalekhya Mukhopadhyay (MDS202301) - Ayush Yadav (MDS202315) - Kalyani Gohokar (MDS202330) - Mayank Nagar (MDS202334) **Preferred Slot: 9th May 2025 (5 PM)**	Project - **Problem**: Deep-Learning Based Audio Denoising & Enhancement. - **Data**: Voicebak-DEMAND, Self-made (LibriSpeech + MUSAN) - **Model**: Wave-U-NET, SEGAN - **Responsibilities**: ○ Aalekhya: Dataset curation (download and organize), Preprocessing, DVC. ○ Mayank & Ayush: Models' building, training, testing and logging. ○ Kalyani: MLOps integration and model deployment. ○ All: Presentation and Documentation. **Code:** https://github.com/mkreman/DL-Audio-Denoising-_AML-Project **Presentation:** Noise Assassins Presentation
Group name : Longshot - Dhruv (22f3001413, IIT-Madras) - Vasu (21f3002975, IIT-Madras) - Nikita (IC11107, IIT-Madras) - Gauranga (MDS202325, CMI) Preferred Slot : **9th May**, **5pm**	**Problem Statement** Systematic Comparison of Classical Models and Modern Architectures for Text Analytics using GoodReads Reviews data. **Data:** Goodreads Datasets. In particular, 1. Detailed book graph (~2gb, about 2.3 million books): goodreads_books.json.gz 2. Complete book reviews (~15 million multilingual reviews about ~2 million books and 465k users): goodreads_reviews_dedup.json.gz **Models:** Logistic Regression, XGBoost, ANN, Transformers (BERT, Distilbert) **Experiments across** 1) **Embeddings:** TF-IDF, BoW, Skip-Gram, BERT 2) **Sampling :** Downsampling **Responsibilities:** ○ *Background research, Data pre-processing and Design* - **Nikita** ○ *Code base and ANN Modelling* - **Vasu** ○ *DL/ML Modelling and Experiments* - **Gauranga** ○ *Code base and ANN Modelling* - **Dhruv** **Code/Presentation:** G-drive, Github

Group Name:Trade Cartels - Aparna C (MDS202310) - Hiba AP (MDS202326) - Jerin Biju (MDS202329) - Nooh Ali (MDS202337) Preferred Slot- 9th May 2025 6:30pm	Project ● Problem: Data: historical stock data ● Model: DQN ● Develop an intelligent trading agent capable of making buy, sell, or hold decisions by learning from historical stock price data ● ● Responsibilities: ● Data Collection and Preprocessing ● Environment and Agent Design ● Training and Evaluation: ● Optimization and Deployment Project link: AppliedMachineLearning/prjct at main · noohalibengla/AppliedMachineLearning
Group name: Team Seaslug - Nandini Jaiswal (MDS202335) - Aritra Majumder (MCS202304) - Satyaki Mullick (MDS202344) - Sreeja Choudhary (MDS202350) Preferred slot: 7th May, 6 p.m.	Project - **Problem**: *Braille* to *English* Translator and vice versa using Convolutional Neural networks. (Given an image containing the text in Braille, extract the text and return the English translation and vice versa.) - **Data**: We will use datasets related to *Optical Character Recognition* to train/fine-tune YOLO model. ○ Angelina Braille character Dataset. ○ Kaggle Braille character Dataset. ○ Kaggle standard OCR dataset. - **Model**: Ultralytics Yolo model. - **Responsibilities**: ○ Project initialization and dataset curation: **Satyaki**. ○ Fine-tuning model: **Nandini**, **Sreeja**. ○ Evaluating performance: **All**. ○ Organize and deployment: **Aritra**. ○ Documentation and presentation: **All**. Presentation Link (Google Slides): https://docs.google.com/presentation/d/1anX78Leh4zTzyWojHcP5FC-Nix9eMBsvL3ihCG5IUW0/edit?usp=sharing Project Link (Github): https://github.com/Aritra8438/braille-utils Braille - English Translator App Link: https://aritra8438-braille-utils.hf.space/

<table>
<tr>
<td>Group: Caption Captains
- Vishal Maurya(MDS202354)
- Ashish Bisht(MDS202313)
- Bhuvnesh Sahu(MDS202316)
- Vikas Chaudhary(MDS202353)

Preferred Slot:- 6th May 6:00 pm</td>
<td>Project
- Problem : Image Captionin Generation using CNN and LSTM
- Data : Flickr8k
- Model : VGG16 +LSTM based architecture
- Responsibility :
1. Data: Vishal
2. Architecture and Modeling: (Bhuvnesh)
3. Validation: Backtest and analyze performance(Ashish)
4. Documentation and presentation(Vikas)

Project and Slides:
https://github.com/bhuvneshsahu01/Applied-Machine-Learning/tree/main/Image%20Captioning%20project</td>
</tr>
<tr>
<td>Group : AudioSentinels

- Chandranath Bhattacharya (MDS202318)
- Salokya Deb (MDS2023)
- Soumyajoy Kundu (MDS202349)

Slot :

- May 7, 2025 (5pm)</td>
<td>Project : SonicShield

- Problem: AI-Powered Guardian Against DeepFake Speech
○ The rise of AI-driven voice cloning enables realistic DeepFake speech, posing risks like identity fraud and misinformation. A reliable detection system is needed to distinguish AI-generated voices from real human speech in real-time.

- Data: We will use datasets that consist of real human speech from well-known figures and their AI-generated counterparts, created using Retrieval-based Voice Conversion (RVC) to train/fine-tune various ML/DL models.

- Model: State-of-the art ML/DL models, Wave2Vec, Wave-U-Net

- Responsibilities:
○ Dataset curation (download and organize), Preprocessing, DVC. : Salokya
○ Models’ building, training, testing and logging : Soumyajoy
○ MLOps integration and model deployment : Chandranath
○ Presentation and Documentation : All</td>
</tr>
</table>

Group : ChadGPT - Chenna Sai Sandeep (MDS202319) - Bibek paul (MDS202317) **Preferred date : 8th May 5:00PM**	Project - CMI InfoBot: A Domain-Specific Chatbot with Real-Time Data Retrieval - Data : CMI website, Google Scholar, Linkedin - Model : Pertained models like LLAMA, GPT, Gemma - Responsibilities: a. Web Scraping (Data Engineering) - Bibek b. Model selection and training (ML Engineering) - Sai c. Deployment and Frontend (ML Ops) - Both **Presentation Link -** https://docs.google.com/presentation/d/1grc2jdKXxKx26090E5YLfuzv6WBlAM9l/edit?usp=sharing&ouid=109984858608257599141&rtpof=true&sd=true **Github Link -** **https://github.com/chennasaisandeep/CMI_Info_Bot/tree/scrapper**
Group: NeuraPaint - Bijayan Ray (MCS202403) - Aryansh Saxena (MDS202312) - Krishanu Bandopadhyay (BMC202230) **Preferred Slot: 8th May 2025 (6PM)**	Project - Problem: Diffusion model based image generation and model maintenance - Data: image data - Model: pretrained or scratch - Responsibilities: ○ Bijayan Ray: Model building and dataset preparation ○ Aryansh Saxena: Model building and model maintenance via MLOps ○ Krishanu Bandopadhyay: Model building and dataset preparation **Github Link** https://github.com/BijayanRay/Diffusion-model-implementation-applied-machine-learning-course-project
Group name: NoName 1. Atul Anant MDS202314 2. Soumya Dasgupta 3. Pritam Padhan **Preferred Slot: 9th May 2025 (5.30 PM)**	Project - Problem : Train ML models collaboratively across decentralized devices while preserving data privacy, applicable in medical imaging, vehicular communication, and IoT. - Data : Decentralized datasets like EMNIST/FEMNIST, medical imaging, and IoT sensor data, handled locally on client devices. - Model : Implement FedAvg algorithm – clients train local models, send updates to the server, and the server aggregates updates to update the global model iteratively. - Responsibility : 1. Atul: Implement FedAvg algorithm and manage model aggregation.

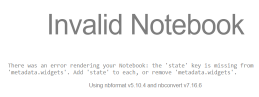
	2. Soumya: Integrate FedAvg with TensorFlow Federated and test with datasets. 3. Pritam: Implement data privacy mechanisms and performance analysis. Code and ppt : FederatedLearning@ppt_code
Group - group name: Catch Me If You Scan ● Arka Roy ● Samaroha Chatterjee **Preferred slot : 9th May , 6 pm**	Project - Problem: License-Plate Recognition from On Road Images of Cars - data: (https://public.roboflow.com/ds/K9IjAYLk5y?key=9th6uXdcDR) - - Model: YOLO v5, Tesseract OCR, regex - Responsibilities: - Data Preprocessing, Train & fine-tune YOLOv5 for plate detection: Arka Roy - Prepare OCR labels; augment plate crops Optimise OCR for character recognition and get presentable results : Samaroha Chatterjee Presentation Link: Project link(GitHub) : https://github.com/Samshoww/-AppliedMachineLearning/tree/main/FINAL%20PROJECT
Group Name : 404 Not Found :) - Malde Dharmi (MDS202343) - Esha Bhattacharya (MDS202324) - Narendra C(MDS202336) - Siddhesh Maheshwari (MDS202347) **Preferred Slot: 8th May 2025 (6:30 PM)**	Project - Problem: Instance Segmentation - Data: BDD100k Image dataset - Model: Mask R-CNN model and similar models. - Responsibilities: EDA & PPT : Siddhesh Maheshwari Data Inspection : Malde Dharmi Model Training : Narendra C Model Evaluation & PPT : Esha Bhattacharya Project link(GitHub) : https://github.com/EBH2002/AML-PROJECT Presentation link: https://onedrive.live.com/:p:/g/personal/DC39E

	0855C566098/EZcBwrkquRFOh9NzJMBLI_AB8vsd9_ERBAdphKWuVrTfWQ?resid=DC39E0855C566098!sb9c20197b92a4e1187d37324c04b23f0&ithint=file%2Cpptx&e=mbQ5P2&migratedtospo=true&redeem=aHR0cHM6Ly8xZHJ2Lm1zL3AvYy9kYzM5ZTA4NTVjNTY2MDk4L0VaY0J3cmtxdVJGT2g5TnpKTUJMSV9BQjh2c2Q5X0VSQkFkcGhLV3VWclRmV1E_ZT1tYlE1UDI
Group Name : EchoVision - Himanshu(MDS202328) - Sattik Biswas - Sayantani Saha - Suneha Sen **Preferred date : 7th May 6PM**	Project - Detection of faces in an image→ and video - Data : Kaggle, Live - Model: YOLO, ResNet-50 (Deep Learning Model) - Responsibilities: 1. Himanshu - Project management, dataset organisation 2. Sayantani - Implementation and train the model using AI/ML techniques. 3. Sattik - Software development, OpenCV implementation. 4. Suneha - Model evaluation, bug fixing, deployment. Presentation Link : AML Presentation.pptx Project link (Github) : AppliedMachineLearning/AML Project at main · Sayantani0503/AppliedMachineLearning
Group name: SANS Overfit - Akhoury Shauryam (MCS202302) - Nirjhar Nath (BMC202239) - Sai Sasank Y (MCS202311) - Shruti Patil (MCS202415) **Preferred Slot : 9th May 2025, 5:00 pm**	**Project** - A CLI tool that can give you the shell command you need from natural language descriptions using llms. **Data** - https://huggingface.co/datasets/westenfelder/NL2SH-ALFA **Model** - llama-3-8b, qwen-2.5-14b **Responsibilities** - ● Model selection and fine-tuning (Sai, Shauryam) ● CLI interface development and Ollama inference integration (Shauryam, Nirjhar, Shruti) ● Model evaluations, testing and error analysis (Nirjhar, Shruti, Sai)

	GitHub: https://github.com/SasankYadati/aish **PPT**: AML_Presentation_2025.pdf
Group WazzUpBeijing - Shankar Narayanan (MCS202409) - Eada Surya Dev (MDS202323) **Date : 9th May, 6.30 PM**	Project - **Title**: Skin Health Companion Application - **Datasets:** ○ DermNet ○ HAM10000 - **Models:** Finetuned CNN, Finetuned Transformer - **Responsibilities:** ○ Project and Data Management: Shared ○ Prototyping and ML Experiments: - End-to-end training: Shankar - EDA, Experiment 2: Surya ○ App Dev. & Deployment: Shankar ○ Presentation: Shared - Github Repo - Presentation
Group: GroupOfOne - Dipanjoy Saha	Project - Identifying individual animals to monitor population and reduce manual errors. - Images of endangered animals with individual IDs, Publicly available datasets (e.g., from wildlife conservation organizations) , Custom datasets collected via camera - Pretrained deep learning models (e.g., ResNet, EfficientNet) for image feature extraction. Fine-tuned CNN for animal face recognition and identification. Ensemble approaches combining detection (YOLO) and identification.

Questions

date	name	question
23 Jan 2025	Vasu	Q1) In case of data imbalance - do we take equal (or almost equal) samples of various classes for

		training OR do we prefer to create synthetic data for the under-represented class during training OR is there any other better way to manage this? Q2) Related, ils there a benchmark on when to consider these methods. For example is 80-20% split of two classes ok? Should we consider these methods if it is less than 10%. Is there a threshold?
4 Feb 2025	Vasu	How do we know which distribution to use when? (e.g. poisson, normal, log normal etc)
	Vasu	**NN/CNN Diagram Utility** https://alexlenail.me/NN-SVG/index.html
15 April 2025	Gauranga	 **Any idea how to mitigate past this problem ?** https://github.com/Gauranga2022/CMI-MSc-Data-Science/blob/main/Sem4/AppliedML/Assignment%205/Sentiment%20Analysis/Sentiment_Analysis.ipynb
22nd April	Aritra	Refer to this blog. It may get messy, but I will tell you the steps. Let's say, the nth (from the last) commit has the unwanted modification, so this command `git reset --soft HEAD~n` will take you back to that place. Then remove the unwanted modification, commit the rest. Then, `git push origin main -f` to force push to github, everything should be fixed. But anyways, that's risky (if you

		haven’t used git that much), so, it’s better to just add the mistake to the readme. @Vasu Thanks so much @Aritra. This was for assignment 3, and now after assignment 4 and 5 reversing it to the right version and then updating both assignments may get tricky but I will try.