Machine Learning

Adult dataset

	Predict whether an individual's income exceeds $50K annually based on demographic and employment data.
Objectives	Analyze factors that significantly contribute to income classification.
Considerations	Data preprocessing steps
	Model building (from a selection of how choices) & hyperparameter tunning
	You can see some baseline model performance in the provided link. Make sure you can achieve at least these numbers.

Dry Bean Data set

	Classify dry beans into their respective types based on physical properties.
Objectives	Determine the most distinguishing features for bean classification.
Considerations	Data preprocessing steps
	Choice of classifiers, take note of its high dimensionality
	Techniques used to address the dimensionality

Rice data set

	Differentiate between Cammeo and Osmancik rice varieties using physical characteristics.
Objectives	Identify key traits that most affect the classification.
Considerations	Data preprocessing steps
	Feature dimension reduction?
	Choice of AI algorithms - supervised learning vs CNN? (CNN: acc=77.5%)

Wine quality

Objectives	Predict wine quality based on physicochemical tests.
	Discover the relationships between wine components and perceived quality.
Considerations	Data preprocessing steps - skewed data?
	Models: Consider side effects of skewed data and how to interpret the results in actual use case

Bank marketing dataset

	Predict whether a client will subscribe to a term deposit following a marketing campaign.
Objectives	Understand which marketing strategies are most effective.
Considerations	Data preprocessing steps
	Feature engineering e.g. time series features
	How do you deal with imbalanced data?
	Models: Classifier for imbalanced dataset?
	Through this exercise, can you figure out the relationship between certain features and campaign outcome?

Abalone dataset

Objectives	Estimate the age of abalone from physical measurements.
	Explore the relationship between physical characteristics and age.
Considerations	Data preprocessing steps
	Relationship between Age and features abalone need not be linear
	Outlier management
	Models: Choice of models and which strategy do you use?
	Understand the metrics and the respective meaning.

Computer Vision

Bird species recognition

Objectives	Identify species of birds from images with high precision.
	Understand which visual features (e.g., beak shape, wing patterns) are most effective for classification.
	Technical Advice:
Considerations	Data exploration
	Data augmentation
	Consider multiple DL models, including transfer learning

Car Model recognition

Objectives	Classify car models from images, focusing on make, model, and year.
	Determine key visual cues that differentiate car models.
Considerations	Data exploration
	Several approaches to attention mechanisms, pros and cons
	Understand feature extraction w.r.t to your dataset

Airplane Model recognition

Objectives	Recognize specific aircraft models from images.
	Highlight features like wing design, tail shape, or engine type that are critical for classification.
Considerations	Data exploration
	DL model for fine grain recognition - possible transfer learning?

Dog breed recognition

Objectives	Classify dog breeds from images.
	Identify which features like fur pattern, ear shape, or body proportions are most indicative of breed.
Considerations	Data augmentation
	Explore transfer learning
	Ensemble learning ?

Scene recogntion

Objectives	Recognize indoor scenes or environments from images.
	Determine which scene elements (texture, layout, objects) are most crucial for scene classification.
Considerations	Data exploration: spatial and contextual
	challenges may include lighting conditions, viewpoints, clutter differences etc.

Face based identity recogntion

Objectives	Identify individuals based on facial features.
	Explore the robustness of facial recognition under varying lighting and expression conditions.
Considerations	Check out some existing models eg Facenet etc.

Facial expression

Objectives	Classify emotions from facial expressions.
	Understand the nuances of emotion detection across different cultures or individuals.
Considerations	Focus on facial areas like eyes, mouth?

Hand gesture recogntion to play game

Objectives	Recognize hand gestures for game control.
	Optimize gesture recognition for real-time, low-latency interaction.
Considerations	Gesture segmentation?

Style transfer for images

Objectives	Transfer artistic styles from one image to another.
	Explore the balance between style and content preservation in the generated images.
Considerations	Neural style transfer

NLP

Spam email detection

Objectives	Classify emails as spam or not spam with high accuracy.
	Identify features or patterns that are most indicative of spam emails.
Considerations	Text preprocessing
	Featur engineering
	ML models: non DL
	Any imbalanced data?

Sentiment analysis

Objectives	Determine the sentiment (positive, negative, neutral) of text data.
	Explore how different aspects of language influence perceived sentiment.
Considerations	Use of TFIDF or word embeddings
	Models: non DL vs DL

POS tagging

Objectives	Automatically label words in text with their corresponding part of speech.
	Achieve high accuracy in tagging, especially for ambiguous cases.
Considerations	Conditional Random Fields (CRF), LSTM
	Use pre-trained models BERT to improve performance

Machine translation

Objectives	Translate text from one language to another with high fidelity.
	Handle nuances like idioms, cultural references, and context in translation.
Consderations	Use Transformer models?
	Metrics BLEU, COMET

Word embedding

Objectives	Represent words as vectors in a continuous vector space where semantic relationships are preserved.
	Enhance downstream NLP tasks by providing better word representations.
Considerations	Compare different embedding models from non-contextual to contextual
	Embedding dimensions?
	Fine tune for domain specific inputs?

Building a chatbot

Objectives	Develop a conversational AI capable of understanding and responding to user queries in a natural way.
	Ensure the chatbot can handle a variety of conversational contexts or intents.
Considerations	Transformer models?
	Prompt engineering and controlled generation via temperature setting etc.
	User interactions - advanced feature will be multi modal

Others

TSP in NSW

Objectives	Solve the Travelling Salesman Problem (TSP) for a set of locations within New South Wales (NSW), aiming to minimize the total distance travelled while visiting each location exactly once and returning to the starting point.
	Compare different algorithmic approaches to understand their efficiency and effectiveness on real-world geographic data.
Considerations	Search methods: Consider Simulated annealing, genetic algo, ant colony optimzation
	Simpler approach: DP?

Understand AlphaGo

Objectives	Gain insight into how AlphaGo uses AI, specifically deep learning and reinforcement learning, to master the game of Go.
	Understand the implications of AlphaGo's achievements for AI research and game theory.
Considerations	Monte Carlo Tree Search? With NN vs without NN
	Limitations and strengths of AlphaGo