5 days Data Science Workshop
For Karakoram International University Students

Outline

Day 1: Introduction to Data Science and AI

1. **Introduction to AI and Data Science**:
    - Definitions: AI, Machine Learning (ML), and Data Science.
    - Key terms: Algorithms, datasets, predictions.
    - Evolution of AI and DS.
2. **Components of Data Science**:
    - Data Collection
    - Data Cleaning
    - Data Analysis
    - Modeling and Prediction
3. **Real-Life Applications**:
    - Healthcare (e.g., disease prediction).
    - Retail (e.g., recommendation engines).
    - Social Media (e.g., content suggestions).
    - Transportation (e.g., self-driving cars).
4. **Tools we need for implementing Data science**:
    - Google Colab
    - Weka
    - Roboflow
    - JenniAI

Practical Activities:

1. **Interactive Poll**: Ask students how they think AI is used in their daily lives.
2. **Group Activity**: Brainstorm one real-life AI use case for each group. Have them present it briefly.
3. **Quick Practical**: Use a simple ChatGPT prompt to show an AI tool in action (e.g., summarize a document).

Day 2: Data Processing, Cleaning, and Visualization

Presentation Content:

1. **Understanding Data**:
    - Types of data: Structured vs. unstructured.
    - Introduction to datasets (e.g., CSV files).
2. **Data Cleaning Basics**:
    - Handling missing values.
    - Removing duplicates.
    - Data transformations (e.g., normalization).
3. **Data Visualization**:
    - Importance of visualization.
    - Tools: Matplotlib, Tableau.
    - Creating graphs: Line, bar, scatter, heatmap.

4. **Case Study**: Analyze a real-world dataset (e.g., sales or weather).

Practical Activities:

1. **Dataset Cleaning**:
    - Provide a messy dataset (e.g., house data with missing values).
    - Guide participants to clean it using Python (pandas library).
2. **Visualization Task**:
    - Use Seaborn/Matplotlib to create bar and scatter plots from a cleaned dataset.
3. **Interactive Q&A**:
    - Share visualizations created by participants and discuss insights.

Day 3: AI Research Tools and Prompt Engineering

Presentation Content:

1. **AI Research Tools**:
    - Platforms: Jenni AI,ChatGPt.
2. **Prompt Engineering**:
    - Writing effective prompts.
    - Dos and don'ts of prompt design.
    - Examples: ChatGPT, DALL·E.
3. **Exploring Pre-Trained Models**:
    - Using Hugging Face to load models.
    - Running inference on small datasets.

Practical Activities:

1. **Prompt Engineering Challenge**:
    - Give participants a scenario (e.g., generate customer service replies).
    - Have them craft prompts and test their effectiveness with ChatGPT.
2. **Kaggle Exploration**:
    - Explore a dataset on Kaggle (e.g., dataset).
    - Discuss how to frame a research question.
3. **Small AI Tool Experiment**:
    - Use Hugging Face to run a sentiment analysis model.

Day 4: Machine Learning – House Price Prediction

Presentation Content:

1. **What is Machine Learning?**:
    - Types: Supervised, unsupervised, reinforcement.
    - ML pipeline: Data preparation, training, evaluation.
2. **Linear Regression and Logistic Regression**
    - Real-life examples: Predicting prices, demand.
3. **Building an ML Model**:

    - Dataset: House price prediction dataset.
    - Using scikit-learn for linear regression.
4. **Model Evaluation**:
    - Metrics: MAE, RMSE, R-squared.
    - Tips for model improvement.

Practical Activities:

1. **Hands-On Coding**:
    - Load and explore the dataset (e.g., housing prices).
    - Train a linear regression model using scikit-learn.
    - Evaluate model performance.
2. **Interactive Task**:
    - Change a parameter (e.g., add/remove features) and observe the impact on predictions.
3. **Discussion**:
    - Group brainstorming on improving model accuracy.

Day 5: Deep Learning for Image-Based Applications

Presentation Content:

1. **Introduction to Deep Learning**:
    - Difference between ML and DL.
    - Neural networks basics: Input, hidden, and output layers.
2. **What are Convolutional Neural Networks (CNNs)?**:
    - Concepts: Filters, pooling, activation functions.
    - Applications: Image classification, object detection.
3. **Building a CNN Model**:
    - Dataset: MNIST or CIFAR-10.
    - Using TensorFlow/Keras for image classification.
4. **Training and Evaluation**:
    - Metrics: Accuracy, precision, recall.
    - Visualizing training performance (loss/accuracy curves).

Practical Activities:

1. **Image Classification Task**:
    - Use TensorFlow/Keras to build a CNN model.
    - Train on MNIST (digits) or CIFAR-10 (objects).
2. **Interactive Debugging**:
    - Identify and resolve issues in model performance (e.g., overfitting).
3. **Group Activity**:
    - Discuss potential applications of CNNs in real-life scenarios (e.g., medical imaging, security).
