OJT Plan for Students _AI

AI project based learning will provide learners with a comprehensive understanding of artificial intelligence principles, techniques, and applications. Participants delve into foundational AI concepts like machine learning, neural networks, deep learning, natural language processing, and computer vision, gaining practical skills in applying AI algorithms to solve complex problems across diverse domains. The course emphasizes hands-on experience, teaching students to develop and deploy AI solutions using programming languages like Python and frameworks.

Tools Required

- Computer / Laptop with good configuration or Web platform like Google Colab Kaggle
- Code Editor (VS Code, Jupyter Notebook etc)
- Programming Languages: Python
- Development Environments: Jupyter Notebooks, Integrated Development Environments (IDEs)
- Libraries and Frameworks: NumPy, MatplotLib, Pandas, SciPy, TensorFlow, PyTorch, Scikit-Learn, Keras
- Self-paced Content

Master Class Sessions: The master class, led by a central trainer, will be conducted online. These sessions are project-focused, covering everything from project initiation to project deployment.

Self-Paced Content: Self-paced content sourced from IBM Skills build platform.

Artificial Intelligence

AI project is to explore and apply artificial intelligence techniques to solve specific problems and tasks. This aims to develop AI models that can analyse data, make predictions, and provide insights relevant to the chosen domain. Through this project, students will learn how to implement machine learning algorithms, use data analysis tools, and evaluate the performance of AI models. Ethical considerations such as fairness, transparency, and privacy will also be emphasized to ensure responsible AI development. Ultimately, they can showcase problem-solving skills, and demonstrate the potential impact of AI applications in addressing real-world challenges.

- A candidate will be able to:
  - Understand the fundamentals and historical context of artificial intelligence.
  - Master various machine learning techniques including supervised, unsupervised, and reinforcement learning.
  - Explore deep learning concepts such as neural networks, CNNs, and RNNs.
  - Gain proficiency in natural language processing (NLP) and computer vision tasks.
  - Learn to deploy machine learning models effectively using containerization and web services

Project Titles

1. Customer Segmentation: Implement a k-means clustering algorithm on customer data to segment them into distinct groups based on their purchasing behavior, allowing businesses to tailor marketing strategies and product offerings to specific customer segments.

2. Sentiment Analysis in Social Media: Implementing machine learning models to analyze social media data and determine the sentiment (positive, negative, neutral) associated with user-generated content, providing insights for brand management, marketing strategies, and customer engagement.

3. Natural Language Processing for Text Classification: Utilizing machine learning techniques in natural language processing (NLP) to classify and categorize text documents, such as customer reviews, support tickets, or news articles, for tasks like sentiment analysis, topic modelling, or spam detection.

4. Image Classification for Medical Diagnosis: Using machine learning techniques to analyse medical images (e.g., X-rays, MRIs) and classify them into different categories, assisting healthcare professionals in diagnosing diseases such as cancer or neurological disorders.

5. Human Activity Recognition: Implement deep learning models to recognize and classify human activities based on sensor data from wearable devices or IoT sensors, enabling applications in health monitoring and fitness tracking.

Weeks	**Milestones**
Weeks 1-5	Problem Understanding and Data Collection
	- Understand problem domain, goals, and requirements.
	- Submit initial research findings and problem statement.
	- Deepen domain-specific knowledge and submit research findings and dataset sources.
	- Collect relevant datasets and understand their structure, features, and characteristics.
	- Clean and preprocess the dataset, handling missing values, outliers, and inconsistencies.
	- Ensure uniformity in non-numeric data and submit a cleaned dataset.
Weeks 6-10	Preprocessing, EDA, and Feature Engineering
	- Continue preprocessing tasks like normalization, scaling, and data augmentation for non-numeric data.
	- Conduct exploratory data analysis (EDA) to identify data distribution, correlations, and patterns.
	- Submit EDA findings and visualizations.
	- Engineer new features or transform existing ones to enhance model performance.
	- Utilize pre-trained models for feature extraction in non-numeric data and submit engineered features.
Weeks 11-15	Model Training, Evaluation, and Deployment
	- Select and train appropriate models based on problem type and dataset characteristics.
	- Evaluate model performance, fine-tune hyperparameters, and submit evaluation results.
	- Deploy the trained model and document the process.

	- Submit the deployed model and documentation.
Weeks 16-20	Testing, Optimization, and Finalization
	- Test and debug the deployed model for accuracy, reliability, and scalability.
	- Optimize models and deployment if time allows, and submit optimized models.
	- Finalize the project, prepare for presentation, and submit final project components.
	- Present project findings, methodology, and results, and deliver a final report.
	- Discuss possible extensions and future work.

Weekly Plan for Students

#	Weekly Learning Outcomes	Weekly Milestone	Self-Paced Courses
Week 1	• **Introduction to Artificial Intelligence** • Definition of AI • Historical context • Milestones in AI development • Applications of Artificial Intelligence • Types of AI (Narrow AI vs General AI) • **Master Class 1: Employability Skills**	• After studying these topics, students will fully understand Artificial Intelligence (AI). • They will learn what AI is, its history, important events, and how it's used in different industries. • Students will know about key advancements like expert systems and neural networks. • They'll understand the difference between Narrow AI (focused on specific tasks) and General AI (more human-like). • Students will see how AI has changed healthcare and finance. • With this knowledge, students will have a basic understanding of AI's principles and its potential to change many areas. • This will encourage them to learn more and contribute to the field of AI.	• https://skills.yourlearning.ibm.com/activity/URL-AD79NYK2KEG • https://skills.yourlearning.ibm.com/activity/URL-A0_LO_GDCFW • https://skills.yourlearning.ibm.com/activity/URL-FWOZMMIUQHG • https://skills.yourlearning.ibm.com/activity/PLAN-E7ED2AACB692 • https://skills.yourlearning.ibm.com/activity/ILB-ZKVJNXVZKMZXEWGN
Week 2	• **Fundamentals of AI Technologies** • Introduction to Machine Learning • Definition and principles of machine learning • Types of machine learning: Supervised, Unsupervised, Reinforcement Learning.	• Students who finish these topics will become good at the basics of AI, especially in machine learning (ML) and neural networks (NN). • They will know ML basics, like supervised, unsupervised, and reinforcement learning.	• https://skills.yourlearning.ibm.com/activity/URL-A0_LO_GDCFW • https://skills.yourlearning.ibm.com/activity/URL-FWOZMMIUQHG

	• Machine learning workflow • Deep Learning and Neural Networks • Basic structure and functioning of artificial neurons • Basic idea on loss function • **Master Class 2: Employability Skills**	• Students will understand how ML works from preparing data to checking how well a model works. • They will also learn NN basics, understanding what artificial neurons are and why loss functions matter in training deep learning models. • This is an important step because it gives students a strong base in AI tech. • With this knowledge, they can easily learn more advanced stuff in AI and use it in real life.	• https://skills.yourlearning.ibm.com/activity/URL-WNQKFPCPK1G • https://skills.yourlearning.ibm.com/activity/URL-968D607D4A12
Week 3	• **Key Concepts in AI** • Feature engineering and representation learning • Loss functions and optimization algorithms • Training and testing data, overfitting, and generalization in machine learning • Introduction to Machine Learning Algorithms • Linear Regression • **Master Class 3 : Introduction to Artificial Intelligence Fundamentals of AI Technologies Key Concepts in AI**	• After mastering these topics, students will fully understand important AI concepts. • They will learn about feature engineering, representation learning, loss functions, and optimization algorithms. • Students will understand how training and testing data work in machine learning. • They'll know about overfitting and generalization, which are important for checking how well a model works. • Students will study machine learning algorithms, especially linear regression, and know how to use it. • This marks a strong start in understanding AI concepts, so students can use what they've learned in real life. • With this knowledge, they can specialize further in AI if they want.	• https://skills.yourlearning.ibm.com/activity/URL-A0_LO_GDCFW • https://skills.yourlearning.ibm.com/activity/URL-WNQKFPCPK1G • https://skills.yourlearning.ibm.com/activity/URL-ZWASDVPPFFW • https://skills.yourlearning.ibm.com/activity/PLAN-7913EE1DB030
Week 4	• **Machine Learning (continued)** • Classification • Logistic Regression • Decision Trees • Splitting criteria(Gini Impurity, Entropy) • Advantages and limitations of decision trees • **Master Class 4: Machine Learning**	• Students will deepen their understanding of machine learning by looking into classification techniques. • They will focus on two main methods: logistic regression and decision trees. • Students will understand how logistic regression works for tasks with two categories (binary classification).	• https://skills.yourlearning.ibm.com/activity/URL-A0_LO_GDCFW • https://skills.yourlearning.ibm.com/activity/URL-FWOZMMIUQHG • https://skills.yourlearning.ibm.com/activity/URL-WNQKFPCPK1G • https://skills.yourlearning.ibm.com/activity/URL-968D607D4A12

		• They'll learn about decision trees' structure and how they decide to split data, using criteria like Gini impurity and entropy. • Students will study the pros and cons of decision trees and when they're useful. • This step shows students have a good grasp of machine learning concepts, so they can use and check how well classification algorithms work. • With this knowledge, they'll be able to implement and evaluate classification algorithms effectively.	• https://skills.yourlearning.ibm.com/activity/PLAN-7913EE1DB030
Week 5	• **Model Evaluation and Implementation** • Model Evaluation Techniques • Confusion matrix and its interpretation • Cross-validation techniques: K-fold cross-validation, Stratified cross validation • Model Selection and Hyperparameter Tuning • Implementation of Simple Machine Learning Algorithms (Hands-on) • **Master Class 5: Model Evaluation and Implementation**	• Students will advance to mastering techniques for evaluating models. • They will learn how to interpret confusion matrices and use cross-validation methods like K-fold and stratified cross-validation to assess models well. • Students will study model selection and ways to adjust hyperparameters to make algorithms work better. • They'll get practical experience by doing exercises where they deploy simple machine learning algorithms. • This step shows students can use methods to check how good models are and put them into action. • With this knowledge, they'll be able to use and adjust machine learning algorithms effectively in real life.	• https://skills.yourlearning.ibm.com/activity/PLAN-7913EE1DB030
Week 6	• **Fundamentals of Deep Learning** • Introduction to Deep Learning • Key differences between traditional machine learning and deep learning • Historical context and milestones in deep learning research	• Students will understand the basics of deep learning after finishing these topics. • They will see the main differences between traditional machine learning and deep learning and know the history and important points in deep learning research. • Students will learn about artificial neural networks (ANNs), including how they're structured, how forward propagation works,	• https://skills.yourlearning.ibm.com/activity/URL-A0_LO_GDCFW • https://skills.yourlearning.ibm.com/activity/URL-WNQKFPCPK1G • https://skills.yourlearning.ibm.com/activity/URL-968D607D4A12 • https://skills.yourlearning.ibm.com/activity/PLAN-7913EE1DB030

	• Artificial Neural Networks (ANNs) • Structure of neural networks • Forward propagation and the role of weights and biases • Backpropagation • Optimization algorithms: Gradient Descent, Stochastic Gradient Descent (SGD), Adam • **Master Class 6: Fundamentals of Deep Learning**	and why weights and biases are important. • They'll also understand backpropagation, which is how neural networks learn, and optimization algorithms like Gradient Descent, Stochastic Gradient Descent (SGD), and Adam. • This gives students a strong base to explore deep learning more deeply. • With this knowledge, they'll be able to use and develop deep learning models effectively.	• https://skills.yourlearning.ibm.com/activity/URL-1AE76ADD4550?channelId=CNL_LCB_1567443069285
Week 7	• **Common Deep Learning Architectures** • Convolutional Neural Networks (CNNs) • Convolutional layers, Pooling layers, Fully connected layers • CNNs and their applications in computer vision • Implementation of Basic CNN model • **Master Class 7: Employability Skills**	• After studying these topics, students will become skilled in common deep learning designs, particularly Convolutional Neural Networks (CNNs). • They will understand the structure of CNNs, including convolutional layers, pooling layers, and fully connected layers. • Students will know how CNNs process visual data, especially for computer vision tasks. • They'll get hands-on practice by making a basic CNN model, which will help them understand how CNNs work and apply them to solve real-world problems in computer vision. • This is a big step forward in students' deep learning skills. • With this knowledge, they'll be able to explore more complex designs and uses in deep learning.	• https://skills.yourlearning.ibm.com/activity/URL-A0_LO_GDCFW • https://skills.yourlearning.ibm.com/activity/URL-WNQKFPCPK1G • https://skills.yourlearning.ibm.com/activity/PLAN-7913EE1DB030
Week 8	• **Recurrent Neural Networks (RNNs)** • Recurrent layers, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) • RNNs and their applications in sequential data processing • Challenges with traditional RNNs and solutions	• After studying these topics, students will become skilled in Recurrent Neural Networks (RNNs) and their types like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). • They will understand how RNNs are built, especially the parts that let them handle sequences of data. • Students will focus on the mechanisms of LSTM and GRU cells, which are special parts of	• https://skills.yourlearning.ibm.com/activity/PLAN-7913EE1DB030

	provided by LSTM and GRU • Implementation of Basic RNN Model • **Master Class 8: Common Deep Learning Architectures**	RNNs for processing sequences effectively. • They'll know how RNNs are used in different areas like natural language processing and predicting time series data. • Students will learn about the problems with traditional RNNs and how LSTM and GRU fix them. • They'll get practice by making a basic RNN model, which will help them understand how RNNs work and put them into action. • This is a big step forward in students' deep learning skills. • With this knowledge, they'll be able to handle tasks involving sequences of data confidently.	
Week 9	• **Introduction to Natural Language Processing (NLP)** • Preprocessing Techniques: Tokenization, Stemming, Lemmatization • Text Representation: • Bag of Words (BoW) model • Term Frequency-Inverse Document Frequency (TF-IDF) representation • Word embeddings: Introduction to word2vec, GloVe, and fastText • **Master Class 9: Employability Skills**	• After learning these topics, students will become skilled in the basics of Natural Language Processing (NLP). • They'll understand how to prepare text data for analysis by using techniques like tokenization, stemming, and lemmatization. • Students will learn about different ways to represent text, such as the Bag of Words (BoW) model and Term Frequency-Inverse Document Frequency (TF-IDF) representation, which help capture the meaning of text. • They'll also get introduced to word embeddings like word2vec, GloVe, and fastText, which give word representations in a way that helps understand the meaning of words better. • This is a basic understanding of NLP concepts. • With this knowledge, students can use these techniques in real-life tasks like analyzing sentiment, classifying text, or translating languages.	• SkillsBuild (ibm.com) • https://skills.yourlearning.ibm.com/activity/URL-FB887C26B96B • https://skills.yourlearning.ibm.com/activity/URL-332C7DE56AB8?channelId=CNL_LCB_1567443069285
Week 10	• **Advanced NLP Techniques** • Sentiment Analysis • Techniques for sentiment analysis: Lexicon-based,	• After studying advanced NLP topics, students will become masters in advanced Natural Language Processing (NLP) techniques.	• SkillsBuild (ibm.com) • https://skills.yourlearning.ibm.com/activity/SN-COURSE-V1:IBMDEVELOPERSKILLSNETWORK+CB0

	Machine Learning-based, Deep Learning-based • Named Entity Recognition (NER) • Text Classification • Classification algorithms: Naive Bayes, Support Vector Machines (SVM), Logistic Regression, Neural Networks • **Master Class 10: Introduction to Natural Language Processing (NLP) Advanced NLP Techniques**	• They'll learn about sentiment analysis, including lexicon-based, machine learning-based, and deep learning-based methods for understanding sentiment in text. • Students will also get into Named Entity Recognition (NER), which helps find and classify names of people, organizations, and places in text. • They'll study text classification using methods like Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Neural Networks. • This is a big step forward in understanding NLP techniques. • With this knowledge, students can handle complex NLP tasks and contribute to new research and applications in the field.	101EN+V1?channelId=CNL_LCB_1567443069285
Week 11	• **Introduction to Computer Vision** • Definition and scope of computer vision • Applications of computer vision in various industries such as healthcare, automotive, and security • Image Formation and Representation • Image formats: JPEG, PNG, BMP, etc. Understanding image metadata: Exif data, IPTC data • Image Preprocessing Techniques • Image resizing, cropping, and rotation • Image denoising and smoothing • Histogram equalization and contrast enhancement	• After studying these topics, students will become skilled in the basics of computer vision. • They'll understand what computer vision is and where it's used, like in healthcare, automotive, and security for tasks such as medical imaging, autonomous driving, and surveillance. • Students will learn about how images are made and shown, including different types like JPEG, PNG, and BMP, and info attached to images like Exif and IPTC data. • They'll also know about techniques to change images, like resizing, cropping, and making them clearer, which are important for analyzing images. • This is a basic understanding of computer vision principles.	• https://skills.yourlearning.ibm.com/activity/PLAN-7913EE1DB030 • https://skills.yourlearning.ibm.com/activity/URL-FB887C26B96B •

	• **Master Class 11: Introduction to Computer Vision**	• With this knowledge, students can look into more advanced techniques and uses in the field.	
Week 12	• **Basic Computer Vision Tasks** • Image Classification • Object Detection • Object detection techniques: Haar cascades, Histogram of Oriented Gradients (HOG), Deep Learning-based approaches • Feature Extraction and Descriptors • Feature extraction techniques: SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), ORB (Oriented FAST and Rotated BRIEF) • **Master Class 12: Basic Computer Vision Tasks & Advanced Computer Vision Techniques**	• After studying these topics, students will become skilled in basic computer vision tasks. • They'll understand image classification, which is sorting images into categories. • Students will learn about object detection methods, like Haar cascades, Histogram of Oriented Gradients (HOG), and Deep Learning approaches. These help find and locate objects accurately in images. • They'll also learn about feature extraction techniques such as SIFT, SURF, and ORB. These methods help identify and describe unique features in images. • This is a big step forward in students' computer vision skills. • With this knowledge, students can explore more in-depth techniques and areas within the field.	• https://skills.yourlearning.ibm.com/activity/URL-FB887C26B96B
Week 13	• **Introduction to Reinforcement Learning** • Definition and key concepts of reinforcement learning (RL) • Applications of reinforcement learning in various domains (e.g., gaming, robotics, finance) • Markov Decision Processes (MDPs)	• After studying these topics, students will have a strong understanding of Reinforcement Learning (RL). • They'll know what RL is and how it works, focusing on how an agent interacts with its environment to get the most rewards over time. • Students will see how RL is used in various fields like	• https://skills.yourlearning.ibm.com/activity/URL-FB887C26B96B • https://skills.yourlearning.ibm.com/activity/URL-ADSZJZB-AX8?channelId=CNL_LCB_1567178224443

	• Components of an MDP: States, Actions, Transition probabilities, Rewards • The Bellman Equation • Bellman Expectation Equation and Bellman Optimality Equation • Value functions and optimal policies • **Master Class 13: Employability Skills**	gaming, robotics, and finance, helping make decisions in uncertain situations. • They'll learn about Markov Decision Processes (MDPs), which include states, actions, probabilities of moving between states, and rewards. • Students will also look into the Bellman Equation and its different forms, such as the Bellman Expectation Equation and Bellman Optimality Equation. They'll understand value functions and optimal policies, which are important for making good RL strategies. • This is a big step in understanding RL principles. • With this knowledge, students can explore more advanced RL algorithms and uses.	
Week 14	• **Reinforcement Learning Algorithms** • Q-Learning • Q-learning algorithm, Q-value iteration, and convergence properties • Policy Gradient Methods • Policy gradients and the REINFORCE algorithm • Deep Reinforcement Learning • Combining deep learning with reinforcement learning • Deep Q-learning algorithm and its extensions (e.g., Double DQN, Dueling DQN) • **Master Class 14: Introduction to Reinforcement Learning**	• After mastering these topics, students will become proficient in various Reinforcement Learning (RL) algorithms. • They'll understand Q-Learning, including its structure, how Q-values are updated, and how it finds the best policies through exploration and exploitation. • Students will also learn about Policy Gradient Methods, focusing on policy gradients and the REINFORCE algorithm. This helps agents optimize policy parameters directly. • Additionally, they'll delve into Deep Reinforcement Learning, where they combine deep learning with RL. • Particularly, they'll understand the Deep Q-learning algorithm and its extensions like Double DQN and Dueling DQN. These use neural networks to	• https://skills.yourlearning.ibm.com/activity/URL-FB887C26B96B • https://skills.yourlearning.ibm.com/activity/URL-332C7DE56AB8?channelId=CNL_LCB_1567443069285

		handle complex problems efficiently. • This marks an advanced level of proficiency in RL algorithms. • With this knowledge, students can effectively solve challenging decision-making problems in different fields.	
Week 15	• **Advanced Topics in Deep Learning:** • Generative Adversarial Networks (GANs) and Autoencoders • GANs and their applications in image generation, style transfer, and data augmentation • Components of GANs: Generator, Discriminator, Adversarial training • Popular GAN architectures: DCGAN, WGAN, CycleGAN • Introduction to Autoencoders • **Master Class 15: Advanced Topics in Deep Learning**	• After finishing these topics, students will master advanced Deep Learning concepts. • They'll learn about Generative Adversarial Networks (GANs) and Autoencoders, which are important for tasks like creating images, changing styles, and making more data. • Students will understand the parts of GANs, like the Generator and Discriminator, and how they work together through adversarial training. • They'll explore different GAN types like DCGAN, WGAN, and CycleGAN, and see where each is used best. • Additionally, students will get introduced to Autoencoders, which help learn features and compress data. • This is a big step forward in students' deep learning skills. • With this knowledge, they'll be able to handle complex tasks in image processing and creation effectively.	• https://skills.yourlearning.ibm.com/activity/ILB-RKEVKEEWNQZVV8JV?channelId=CNL_LCB_1551384685793 • https://skills.yourlearning.ibm.com/activity/URL-332C7DE56AB8?channelId=CNL_LCB_1567443069285 • https://skills.yourlearning.ibm.com/activity/MDL-298?channelId=CNL_LCB_1567443069285
Week 16	• **Transfer Learning and Advanced Deep Learning Models** • Transfer learning and its importance in deep learning • Types of transfer learning: Feature extraction, Fine-tuning • Fine-tuning Pre-trained Models	• Students will become proficient in Transfer Learning and advanced Deep Learning models. • They'll understand how Transfer Learning lets them use pre-trained models to solve new tasks efficiently, saving data and computing power. • Students will learn about different types of Transfer Learning, like Feature	• https://skills.yourlearning.ibm.com/activity/ILB-QPDKNWEKJEPN2QPX?channelId=CNL_LCB_1569538212365 • https://skills.yourlearning.ibm.com/activity/PLAN-C4FCC67D3E76?channelId=CNL_LCB_1551390152858

	• Popular pre-trained models such as VGG, ResNet, and BERT • **Master Class 16: Employability Skills**	Extraction and Fine-tuning, which help them adapt pre-trained models to new areas. • They'll also practice Fine-tuning Pre-trained Models, including popular ones like VGG, ResNet, and BERT, for tasks in computer vision and natural language processing. • This shows advanced skill in using Transfer Learning and advanced Deep Learning models. • With this knowledge, students can confidently tackle complex challenges in AI.	
Week 17	• Introduction to GitHub • Collaborating on GitHub • GitHub Branching Strategies • Continuous Integration with GitHub Actions • **Deploying Machine Learning Models** • Introduction to Model Deployment • Challenges and considerations in model deployment: Scalability, Performance, Security • Model Monitoring and Maintenance • Techniques for model monitoring and maintenance: Logging, Alerting, Versioning • Continuous integration and continuous deployment (CI/CD) pipelines for model updates • **Master Class 17: Transfer Learning and Advanced Deep Learning Models**	• After these topics, students will be skilled in deploying and maintaining machine learning models. • They'll understand the basics of model deployment, including challenges like scalability, performance, and security, which are important for real-world use. • Students will learn techniques for monitoring and maintaining models, like logging, alerting, and versioning, to keep models reliable and performing well over time. • They'll explore continuous integration and continuous deployment (CI/CD) pipelines designed for updating models, making it easy to integrate and deploy improvements. • This is a big step in being able to deploy, monitor, and maintain machine learning models effectively. • With this knowledge, students can ensure their models remain useful and up-to-date in real-world situations.	• https://skills.yourlearning.ibm.com/activity/PLAN-1ACC2ADE86A4 • https://skills.yourlearning.ibm.com/activity/PLAN-C4FCC67D3E76?channelId=CNL_LCB_1551390152858
Week 18	• **Containerization and Web Services**	• After these topics, students will master containerization	• https://skills.yourlearning.ibm.com/activity/PLAN-

	• Introduction to Docker: Containers, Images, Dockerfile, Docker Compose • Deploying Machine Learning Models as Web Services • Web frameworks for deploying machine learning models: Flask, FastAPI, Django • Designing RESTful APIs for model inference endpoints • Scaling and Load Balancing • **Master Class 18: Deploying Machine Learning Models**	and deploying machine learning models as web services. • They'll understand Docker's main ideas, like containers, images, Dockerfile, and Docker Compose, which help package and deploy ML models efficiently. • Students will learn about web frameworks like Flask, FastAPI, and Django, used to make strong APIs for deploying machine learning models as web services. • They'll design RESTful APIs for model inference endpoints, making it easy to interact with deployed models. • Additionally, students will learn about scaling and load balancing to ensure deployed services can handle lots of users and stay reliable. • This shows an advanced level of skill, letting students deploy and manage machine learning models effectively in real-world situations using containers and web services.	C4FCC67D3E76?channelId=CNL_LCB_1551390152858 • https://skills.yourlearning.ibm.com/activity/ILB-QPEKVQEKRYZG2KZ3?channelId=CNL_LCB_1567178224443
Week 19	• Students will prepare deliverables for reporting and presentations • **Masterclass 19: Enhancing the presentation skills** •	By the end of this week, students will be able to incorporate all the required suggestions.	• https://skills.yourlearning.ibm.com/activity/PLAN-C4FCC67D3E76?channelId=CNL_LCB_1551390152858
Week 20	• Showcase of Final project	• After the completion of week, student will be able to showcase their final project and modify according to peer team suggestions.	•

Project based evaluation for Students

Project-based evaluation	Milestones to be tracked by the evaluator
Mid Evaluation (after 10 weeks)	• Understanding & Implementation of the Problem statement (20) • Achievement of weekly milestone (15) • Q n A on coding / Model / Weekly milestone (15)
End Evaluation (after 20 weeks)	• Implementation of the Project (Compete document, source code, ppt etc)(25) • Achievement of weekly milestone(15) • Q n A on coding / Model / Real world Implementation (15)

Problem statement of Projects

Customer Segmentation:

Problem Statement:

The goal of this project is to apply the k-means clustering algorithm to customer data in order to segment customers into distinct groups based on their purchasing behavior. By analyzing patterns in customer purchasing habits, the platform aims to assist businesses in identifying homogeneous customer segments and tailoring marketing strategies and product offerings to better meet the needs and preferences of each segment.

Requirements:

Data Collection: Obtain a dataset containing relevant customer data, including attributes such as purchase history, frequency of purchases, total spending, demographics, and any other pertinent variables.

Data Preprocessing: Clean and preprocess the customer data, handling missing values, outliers, and encoding categorical variables if necessary. Perform feature scaling and normalization to ensure uniformity across features.

Feature Selection: Select relevant features from the customer dataset that are indicative of purchasing behavior and likely to contribute to meaningful segmentation.

K-means Clustering: Implement the k-means clustering algorithm to partition customers into k distinct clusters based on their feature similarities. Experiment with different values of k to determine the optimal number of clusters.

Model Training: Train the k-means clustering model using the preprocessed customer data, optimizing cluster centroids to minimize the within-cluster sum of squared distances.

Cluster Interpretation: Analyze the characteristics of each cluster and interpret the purchasing behavior patterns exhibited by customers within each cluster to gain insights into distinct customer segments.

Visualization: Visualize the clustering results using appropriate plots and visualizations to facilitate interpretation and communication of the identified customer segments.

Business Insights: Provide actionable insights to businesses based on the identified customer segments, including recommendations for targeted marketing campaigns, product recommendations, and customer engagement strategies.

Evaluation Metrics: Evaluate the quality of the clustering solution using internal validation metrics such as the silhouette score or external validation metrics if ground truth labels are available.

Deployment: Deploy the trained clustering model into a production environment where businesses can input customer data and obtain segment assignments for personalized marketing and customer relationship management.

Background:

Customer segmentation plays a crucial role in marketing strategy formulation, enabling businesses to better understand their customer base and tailor their offerings to specific customer segments. By leveraging clustering algorithms like k-means, businesses can identify homogeneous groups of customers with similar purchasing behavior, allowing for more effective targeting and customization of marketing efforts.

Real-world Use Case:

Targeted Marketing: Businesses can use customer segmentation results to tailor marketing messages and promotional offers to specific customer segments, increasing the relevance and effectiveness of marketing campaigns.

Product Recommendations: By understanding the preferences and purchasing habits of different customer segments, businesses can provide personalized product recommendations and suggestions, enhancing the overall customer experience.

Customer Retention: Identifying high-value customer segments enables businesses to prioritize resources and efforts towards retaining these valuable customers through loyalty programs, special incentives, and personalized communications.

Market Expansion: Clustering analysis can reveal potential new market segments or niche markets that businesses can target with tailored products or services, driving expansion and growth opportunities.

Customer Lifetime Value (CLV): Segmenting customers based on their predicted lifetime value allows businesses to allocate resources more efficiently, focusing on acquiring and retaining customers with the highest potential long-term value.

Sentiment Analysis in Social Media:

Problem Statement:

The objective of this project is to develop machine learning models capable of analyzing social media data and accurately determining the sentiment expressed in user-generated content. By leveraging natural language processing (NLP) techniques and sentiment analysis algorithms, the platform aims to provide businesses with valuable insights into customer perceptions, attitudes, and opinions expressed on social media platforms.

Requirements:

Data Collection: Gather a diverse dataset containing user-generated content from various social media platforms, including text posts, comments, reviews, and tweets, along with corresponding sentiment labels (positive, negative, neutral).

Data Preprocessing: Clean and preprocess the social media data, removing noise, special characters, and irrelevant information. Tokenize the text, remove stopwords, and perform lemmatization or stemming to normalize the text data.

Feature Engineering: Extract relevant features from the text data that may be indicative of sentiment, such as word frequencies, n-grams, sentiment lexicon scores, or word embeddings.

Model Selection: Evaluate and select appropriate machine learning models for sentiment analysis, including techniques such as logistic regression, support vector machines (SVM), naive Bayes, or deep learning models like recurrent neural networks (RNNs) or transformers.

Model Training: Train the selected sentiment analysis model using the preprocessed social media data, optimizing model parameters and hyperparameters to maximize predictive performance on sentiment classification tasks.

Model Evaluation: Assess the performance of the trained sentiment analysis model using evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis to evaluate its effectiveness in sentiment classification.

Deployment: Deploy the trained sentiment analysis model into a production environment where it can analyze real-time social media data and classify user-generated content into positive, negative, or neutral sentiment categories.

Integration: Integrate the sentiment analysis system with social media monitoring tools, customer relationship management (CRM) systems, or marketing automation platforms to facilitate automated sentiment analysis and reporting.

Real-time Monitoring: Implement mechanisms for real-time monitoring of social media sentiment, enabling businesses to track changes in customer sentiment and respond promptly to emerging trends or issues.

Actionable Insights: Provide actionable insights and recommendations based on the sentiment analysis results, enabling businesses to adjust their brand management strategies, marketing campaigns, and customer engagement efforts accordingly.

Background:

Social media platforms have become invaluable sources of customer feedback, opinions, and sentiments, providing businesses with a wealth of data to analyze and leverage for brand management and customer relationship management. By employing sentiment analysis techniques, businesses can gain valuable insights into customer sentiment, enabling them to tailor their strategies and communications to better meet customer needs and preferences.

Real-world Use Case:

Brand Reputation Management: Businesses can use sentiment analysis to monitor social media conversations about their brand and products, identifying positive sentiments to amplify and negative sentiments to address proactively.

Marketing Campaign Optimization: Sentiment analysis helps businesses evaluate the effectiveness of marketing campaigns by analyzing customer reactions and sentiment towards promotional content, allowing for adjustments and optimizations in real-time.

Product Feedback Analysis: By analyzing sentiment in user-generated product reviews and feedback, businesses can identify areas for product improvement, feature enhancements, or customer support enhancements.

Crisis Management: Social media sentiment analysis enables businesses to detect and respond to potential crises or negative publicity quickly, minimizing reputational damage and restoring customer trust.

Competitive Analysis: Businesses can compare their brand sentiment with that of competitors, gaining insights into areas of competitive advantage or areas for improvement in brand perception.

Natural Language Processing for Text Classification:

Problem Statement:

This project aims to leverage machine learning algorithms in natural language processing (NLP) to classify and categorize text documents into predefined categories or labels. By analyzing the textual content of documents, the platform seeks to perform tasks such as sentiment analysis, topic modeling, and spam detection, providing valuable insights and automation for various applications across different domains.

Requirements:

Data Acquisition: Gather a diverse dataset containing text documents spanning different categories or labels, such as customer reviews, support tickets, news articles, or email messages, along with corresponding ground truth labels or annotations.

Data Preprocessing: Clean and preprocess the text data, including steps such as tokenization, removing stopwords, punctuation, and special characters, and performing stemming or lemmatization to normalize the text.

Feature Extraction: Extract relevant features from the text data that capture important information for classification tasks, such as word frequencies, n-grams, term frequency-inverse document frequency (TF-IDF) vectors, or word embeddings.

Model Selection: Evaluate and select appropriate machine learning algorithms for text classification tasks, including techniques such as logistic regression, support vector machines (SVM), naive Bayes, decision trees, or deep learning models like convolutional neural networks (CNNs) or recurrent neural networks (RNNs).

Model Training: Train the selected text classification model using the preprocessed text data, optimizing model parameters and hyperparameters to maximize predictive performance on classification tasks.

Model Evaluation: Assess the performance of the trained text classification model using evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis to evaluate its effectiveness in classifying text documents.

Task-specific Applications: Apply the trained text classification model to perform specific tasks such as sentiment analysis, topic modeling, or spam detection, depending on the requirements and objectives of the application.

Deployment: Deploy the trained text classification model into a production environment where it can classify text documents in real-time, providing automated categorization and classification capabilities for various applications.

Integration: Integrate the text classification system with existing software applications, workflows, or data pipelines to facilitate seamless data exchange and interoperability with other systems.

Continuous Improvement: Implement mechanisms for continuous model monitoring, performance tracking, and model retraining to adapt to changing data distributions and improve classification accuracy over time.

Background:

Text document classification is a fundamental task in natural language processing (NLP), enabling automated analysis and categorization of textual data for various applications. By applying machine learning techniques, businesses can gain valuable insights from text documents, automate repetitive tasks, and enhance decision-making processes across different domains.

Real-world Use Case:

Sentiment Analysis: Businesses can analyze customer reviews, social media posts, or survey responses to gauge sentiment towards products, services, or brands, informing marketing strategies and customer engagement efforts.

Topic Modeling: Text classification enables the identification of key topics or themes in large collections of documents, facilitating content organization, information retrieval, and knowledge discovery.

Spam Detection: Automated classification of email messages or online comments helps identify and filter out spam or malicious content, improving user experience and security in communication platforms.

Customer Support Ticket Routing: Text classification can categorize support tickets or inquiries based on their content, enabling efficient routing to the appropriate departments or agents for timely resolution.

News Article Categorization: Media organizations can classify news articles into different topics or sections (e.g., politics, sports, technology) to organize content for readers and personalize news recommendations.

Image Classification for Medical Diagnosis:

Problem Statement:

The project aims to develop machine learning models capable of analyzing medical images, such as X-rays and MRIs, to classify them into different categories based on the presence or absence of diseases or abnormalities. By leveraging image processing techniques and deep learning algorithms, the platform seeks to assist healthcare professionals in diagnosing diseases such as cancer, neurological disorders, or other medical conditions.

Requirements:

Data Acquisition: Gather a comprehensive dataset containing medical images, including X-rays, MRIs, CT scans, or other modalities, along with corresponding ground truth labels indicating the presence or absence of diseases or abnormalities.

Data Preprocessing: Clean and preprocess the medical images, including steps such as resizing, normalization, and augmentation to enhance data quality and facilitate model training.

Feature Extraction: Extract relevant features from the medical images using techniques such as convolutional neural networks (CNNs) or pre-trained models like ResNet, VGG, or DenseNet to capture important patterns and characteristics for classification tasks.

Model Selection: Evaluate and select appropriate machine learning algorithms for medical image classification, including CNN architectures optimized for image recognition tasks, transfer learning techniques, or ensemble models for improved performance.

Model Training: Train the selected image classification model using the preprocessed medical image data, optimizing model parameters and hyperparameters to maximize predictive performance on classification tasks.

Model Evaluation: Assess the performance of the trained image classification model using evaluation metrics such as accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curve analysis to evaluate its effectiveness in disease diagnosis.

Deployment: Deploy the trained image classification model into a production environment where it can analyze medical images in real-time, providing automated classification and diagnostic support for healthcare professionals.

Integration: Integrate the image classification system with existing medical imaging software, picture archiving and communication systems (PACS), or electronic health record (EHR) systems to facilitate seamless integration into clinical workflows.

Interpretability: Provide explanations or visualizations of model predictions to help healthcare professionals understand and interpret the classification results, aiding in diagnostic decision-making.

Regulatory Compliance: Ensure that the image classification system complies with regulatory requirements and standards governing medical devices, data privacy, and patient confidentiality, such as HIPAA regulations in the United States.

Background:

Medical image analysis plays a crucial role in disease diagnosis and treatment planning, enabling healthcare professionals to visualize and interpret anatomical structures and abnormalities in patients. By leveraging machine learning techniques, such as deep learning, medical image classification models can assist healthcare professionals in diagnosing diseases and conditions accurately and efficiently.

Real-world Use Case:

Cancer Diagnosis: Image classification models can analyze mammograms, CT scans, or histopathology images to detect and classify cancerous lesions, aiding oncologists and radiologists in early detection and treatment planning.

Neurological Disorder Detection: MRI and CT images can be analyzed to identify neurological disorders such as Alzheimer's disease, multiple sclerosis, or brain tumors, facilitating timely intervention and patient management.

Fracture Detection: X-ray images can be classified to detect and localize fractures in bones, assisting orthopedic surgeons and emergency physicians in diagnosis and treatment planning for musculoskeletal injuries.

Cardiac Imaging: Echocardiograms and cardiac MRI images can be analyzed to classify cardiac abnormalities such as myocardial infarction, heart failure, or congenital heart defects, supporting cardiologists in patient evaluation and management.

Lung Disease Diagnosis: Chest X-rays and CT scans can be classified to detect and classify pulmonary diseases such as pneumonia, tuberculosis, or lung cancer, aiding pulmonologists in diagnosis and treatment planning.

Human Activity Recognition:

Problem Statement:

The project aims to develop deep learning models capable of recognizing and classifying human activities based on sensor data collected from wearable devices or IoT sensors. By leveraging deep learning architectures such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs), the platform seeks to accurately identify various physical activities performed by individuals, enabling applications in health monitoring, fitness tracking, and activity analysis.

Requirements:

Data Collection: Gather a diverse dataset containing sensor data from wearable devices or IoT sensors, including accelerometer, gyroscope, and magnetometer readings, along with corresponding labels indicating the type of human activity being performed.

Data Preprocessing: Clean and preprocess the collected sensor data, including steps such as normalization, feature extraction, and segmentation to prepare the data for training deep learning models.

Feature Engineering: Extract relevant features from the sensor data that capture important patterns and characteristics indicative of different human activities, such as time-domain features, frequency-domain features, or statistical features.

Model Selection: Evaluate and select appropriate deep learning architectures for human activity recognition tasks, including RNNs (e.g., LSTM, GRU), CNNs, or hybrid models combining both architectures, based on the complexity and characteristics of the sensor data.

Model Training: Train the selected deep learning model using the preprocessed sensor data, optimizing model parameters and hyperparameters to maximize predictive performance on activity classification tasks.

Model Evaluation: Assess the performance of the trained activity recognition model using evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis to evaluate its effectiveness in classifying human activities.

Deployment: Deploy the trained deep learning model into a production environment where it can analyze real-time sensor data from wearable devices or IoT sensors, providing automated activity recognition capabilities for health monitoring and fitness tracking applications.

Integration: Integrate the activity recognition system with existing health and fitness applications, wearable devices, or IoT platforms to facilitate seamless data exchange and interoperability with other systems.

Real-time Monitoring: Implement mechanisms for real-time monitoring and visualization of recognized human activities, providing users with feedback and insights into their activity patterns and behaviors.

Privacy and Security: Ensure that the activity recognition system complies with privacy regulations and standards governing the collection and processing of personal health data, implementing encryption and access controls to protect sensitive information.

Background:

Human activity recognition plays a vital role in health monitoring and fitness tracking applications, enabling individuals to monitor their physical activities and exercise routines accurately. By leveraging deep learning techniques, such as RNNs and CNNs, activity recognition models can analyze sensor data from wearable devices or IoT sensors and provide insights into users' activity patterns and behaviors.

Real-world Use Case:

Fitness Tracking: Activity recognition models can track and classify various physical activities such as walking, running, cycling, or swimming, providing users with detailed insights into their exercise habits and performance metrics.

Health Monitoring: By analyzing activity patterns and behaviors, activity recognition systems can detect anomalies or changes in users' activity levels, providing early warnings for potential health issues or irregularities.

Rehabilitation Support: Activity recognition models can assist in monitoring and assessing patients' physical activities during rehabilitation exercises, providing feedback and guidance to support recovery and rehabilitation efforts.

Elderly Care: Activity recognition systems can monitor the daily activities of elderly individuals living alone or in assisted living facilities, providing caregivers with insights into their well-being and detecting any signs of unusual behavior or emergencies.

Sports Performance Analysis: Coaches and athletes can use activity recognition models to analyze and optimize training routines, track performance metrics, and identify areas for improvement in sports activities and techniques.

Guidelines for Selecting Custom Project

Follow the below guidelines, if you are selecting your own topic for project

Students need to choose from two technology tracks**: Artificial Intelligence (AI) and Web Development.** Your task is to create innovative solutions linked to **specific social challenges** through your problem statements.

Social Challenges

Your project should aim to address one of the following social challenges:

1. Eradicating Poverty and Hunger

- Solutions promoting health care, sanitation, preventive health care, and safe drinking water.
- Awareness to initiatives promoting sanitation, clean water, best agricultural practices.

2. Improving Education

- Enhancing education, vocational skills for children, women, elderly, differently-abled individuals.
- Livelihood enhancement projects.

3. Promoting Gender Equality

- Solutions for empowering women, setting up homes and hostels for women and orphans.
- Facilities for senior citizens and reducing inequalities for socially and economically backward groups.

4. Environmental Sustainability

- Safeguarding ecological balance, protection of flora and fauna, and conservation of natural resources.
- Initiatives for maintaining soil, air, and water quality, including river rejuvenation projects.

5. Protecting National Heritage and Culture

- Restoration of historical sites and works of art.
- Promotion and development of traditional arts and handicrafts.

6. Supporting Armed Forces and Their Families

- Benefits for veterans, war widows, their dependents, and Central Armed Police Forces personnel.

7. Promoting Sports

- Training and support for rural, nationally recognized, Paralympic, and Olympic sports.

8. Rural Development Projects

- Solutions aimed at improving rural infrastructure and living conditions.

9. Slum Area Development

- Projects focused on the development of areas declared as slums by government authorities.

10. Disaster Management

- Initiatives for relief, rehabilitation, and reconstru
- cted areas.

Guidelines for Creating Problem Statement

Your problem statement should clearly outline the following:

1. The Social Challenge

- Identify which social challenge from the list above your project will address.

2. The Technology Track

- Specify whether your solution will use AI or Web Development technologies.

3. The Solution

- Describe the project you plan to create, including its goals, functionalities, and how it leverages the chosen technology to address the social challenge.

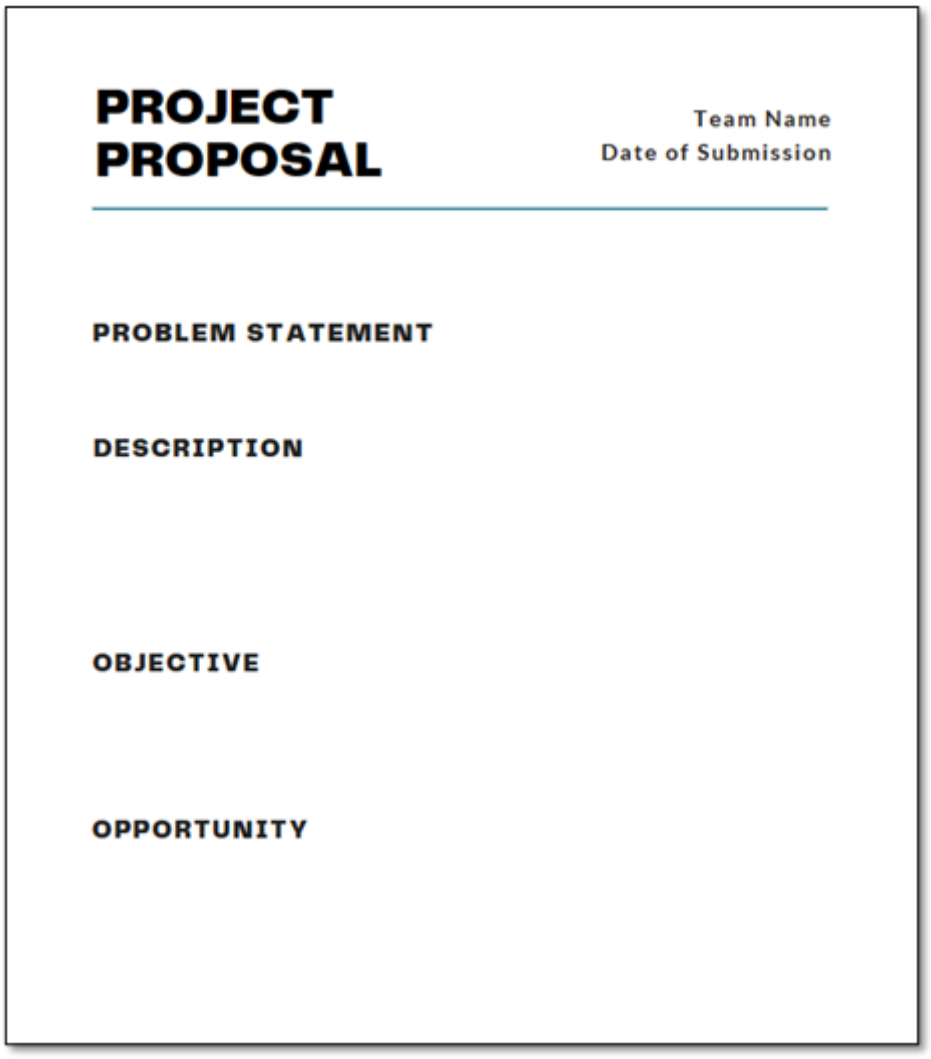
4. Impact

- Explain the potential impact of your solution on the community or target audience.

Guidelines for Submitting Problem Statement

- ✓ Prepare it in word document
- ✓ Mentioned the below details in proposal document
  - Project title (should be concise and clear)
  - Description (should not exceed 500 words)
  - Objective (should not exceed 200 words)
  - Opportunity (should not exceed 200 words)
- ✓ Submit your problem Statement by 22nd of May

Reference document:



PROJECT PROPOSAL

Team Name
Date of Submission

PROBLEM STATEMENT

DESCRIPTION

OBJECTIVE

OPPORTUNITY