Six Months Online Certificate Course: "**Data Science and Gen AI LLMs"**

Subject 1: **Python Programming**

S. No.	Module Name	Topic
Module 1	Introduction to Python	What is AI ML DL NLP
		Basics of Algorithm/Pseudocode, Program, Kinds of Programming Languages, Compilers, and Interpreters
		Introduction to Python, Types of IDE (Anaconda)
		Identifiers, Variables, Operators, Data Types,
		Conditions, Loops
Module 2	Data Structures using Python	Strings: Introduction, functions, and operations on Strings, Application Programs on Strings.
		List: Introduction, functions and operations on List, Application Programs on Lists. Tuple: Introduction, functions, and operations on Tuple
		Dictionaries: Introduction, functions and operations on Dictionaries, Application Programs on Dictionaries.
		Sets: Introduction, functions, and operations on Sets, Applications on Sets, Frozensets
		List Comprehension, Dictionary Comprehension
		Functions Defining and Invoking functions, Scope, Parameter types
Module 3	Functions, Modules and Collections	Recursive functions
		Built in Functions such as enumeration, zip, sorted, map, filter and Applications
		Modules in Python, creating custom modules and calling them
		Lambda functions
		Collections, Iterators, Generators, Decorators, OrderedDict, defaultdictetc
		File I/O operations: Reading and Writing data from various formats
Module 4	Working with Databases and Text	Regular Expressions, Identifiers, Quantifiers. Application Programs on Regular expressions
		Working with Databases: Databases and Data Science, SQLite database and Insert, Update, Delete, Retrieve operations,
		Exception Handling: Need for Exception handling, Raising exceptions,
		Need for Object Orientation, OOPS basics, Principles of OOPS
Module 5	Object Oriented Programming using Python	Classes, Objects, Pass by reference, Self, Collection of objects, Constructors
		Need for Encapsulation and Abstraction, Private Attributes, Getter, and Setter Methods– Python Implementation
		Inheritance: Need for Inheritance, Kinds of Inheritance
		Polymorphism Abstract methods, Overloading and Overriding
		Statistics for Data Science
Module 6	Mathematical modelling for Data Science	Mathematical Computing with NumPy
		Statistics and Probability using Numpy
		Data Manipulation and Analysis
Module 7	Exploratory Data Analysis and Data Visualization	Exploratory Data Analysis with Pandas
		Matplotlib and Seaborn libraries for Visualization
		Web Programming using Flask
Module 8	GUI Programming	GUI programming TCL TK
		Web Scraping using Python

*****

Six Months Online Certificate Course: "**Data Science and Gen AI LLMs"**

Subject 2: **Machine Learning Deep Learning**

S. No.	Module Name	Topic
Module 1	Data Preprocessing Techniques	• Data Imputation • Data Encoding • Data Integration • Data Normalization • Outlier detection Techniques • Dimensionality reduction • Feature Engineering
		Exploratory Data Analysis (EDA) • Univariate Analysis • Multivariate Analysis Case studies on Titanic / Wine Quality / Loan Prediction / Diabetic datasets from Kaggle
Module 2	Machine Learning (Supervised Learning)	Introduction to Machine Learning •Types of Machine Learning • Supervised Learning • Unsupervised Learning • Reinforce Learning
		Regression Analysis • Simple Linear Regression • Multilinear Regression • Metrics for Regression • Lasso, Ridge Regression Regularization Techniques • Case Studies on House Price Prediction/Car Price Prediction … • Deployment of Regression using Flask

		Classification Techniques
		• KNN classifier
		• Logistic Regression classifier
		• Decision Tree classifier
		• Naïve Bayes Classifier
		• SVM classifier
		• Random Forest classifier
		• Ensemble methods
		  ○ Boosting algorithms
		  ○ Bagging algorithms
		  ○ Stacking algorithms
		• Case Studies on IRIS / Churn modeling/ Credit card Fraud detection etc from Kaggle
		• Deployment of Classification using Flask

Module 3	Bias, Variance and Optimization Techniques	– Model Selection and Evaluation • Train/Validation/Test split, K-Fold Cross Validation • The Problem of Over-fitting and Underfitting (Bias-Variance tread- off)
		Learning Best Practices for Model Evaluation • ML Pipeline techniques • Parameter Tuning mechanisms (Grid Search/ Random Search) • Debugging algorithms with learning and validation curves

Six Months Online Certificate Course: "**Data Science and Gen AI LLMs"**

Module 4	Machine Learning (Unsupervised Learning)	Association Analysis • Association Rules & Interesting measures • Apriori Algorithm, FP-Grouth algorithm • Case Studies on Retail Analysis
		Clustering Analysis • Similarity distance measures • K-means Clustering, Hierarchical Clustering. DB Scan Clustering • Case Studies on Customer Segmentation
Module 5	Introduction to Deep Learning	Introduction to Deep Learning
		Image Processing Operations using Numpy
		Neural Network Basics, Activation Functions
		Multilayer Perceptron Algorithm, Gradient Descent Algorithm
		Introduction to TensorFlow Basic programming using TensorFlow
Module 6	Deep Learning Algorithms	Artificial Neural Networks (ANN) Case Study on Churn Modeling using ANN
		Convolution Neural Networks (CNN) Case Study on MNIST Digit Classification
		Recurrence Neural Networks (RNN-LSTM): Case Study on Stock Market Prediction, Image Captioning etc.
		Deployment of Deep Learning Models
❖ **Natural Language Processing:**		
Module 7	Introduction to NLP	What is NLP, Various levels of NLP. • Morphological, • Lexical Analysis, • Syntactic analysis, • Semantic analysis, • Discourse level, • Pragmatic • Applications of NLP
	Text Processing using NLTK, Blob, Spacy	– **Introduction to Text Processing** Working with • Text Files, • HTML files, Web scraping • XML files, • JSON files and • PDF files • Working with Regular Expressions
		Text Processing • Tokenization, • Stemming, • Lemmatization,

		• Removal of Stop Words, • POS tagging and • Named Entity recognition, • Text Preprocessing, • Phrase Matching
		Text Feature Extraction using SciKit-Learn • Vector Space Model representation, • Term Frequency, • Document Frequency, • TF_IDF frequency, • Count Vectorizer, • TF-IDF Transformer, • TF-IDF Vectorizer • Text Similarity • Word Embedding Layer using Deep Learning, • Word2Vec and Doc2Vec
Module 8	Applications of NLP	• Text Classification, • Text Clustering • Text Summarization • Topic Modelling • Recommendation Systems - Collaborative filtering • Case Studies and Application development
	Sentiment Analysis	• Introduction to Sentiment Analysis • Creating NLP Pipeline for Text Mining (Social Media data/Web data) • Using Bag Of Words representation, using TF-IDF representations Data Set : IMDB dataset using Scikit-Learn
	Natural Language Understanding	• Parts of Speech Tagging (POS), • Dependency Parsing, • Named entity recognition (using Spacy module)
	Web Mining and Generative AI	• Web Scraping • Textual data sources and formats, social media, web scrapping APIs (example: scrapy) • NLP Chatbot & Voice bots

*****

Subject 3: **Generative AI and Large Language Models**

Six Months Online Certificate Course: "**Data Science and Gen AI LLMs"**

S. No.	Module Name	Topic
Module 1	Generative AI and its Industry Applications Topics:	• Generative AI Principles • Types of Generative Models • Applications of Generative Models • Machine Learning Algorithms with GenAI • Applications of Generative AI • Generative AI: Advantages and Disadvantages • Ethical Considerations
Module 2	Language Models and Transformer-based Generative Models	• Exploring Language Models • Types of Language Models • Applications of Language Models • The Transformer Architecture: Attention Mechanism • Advanced Transformer Models • GPT • BERT • Applications of Transformer-based Models
Module 3	Prompt Engineering	• Prompt Engineering Principles • What is Prompt Engineering? • Importance and Applications • Prompt Design Strategies • Types of Prompting • Crafting Effective Prompts • Parameter Tuning
Module 4	Generative AI with LLMs	• LLMs and Generative AI Project Lifecycle • LLM Pre-Training and Scaling • Fine-Tuning LLMs with Specific Instructions • Efficient Fine-Tuning of Parameters • Reinforcement Learning from Human Response
Module 5	LLMs for Search, Prediction, and Generation	• Search Query Completion • Next Word Prediction • Word Embeddings • Transformers • Generating Text • Stacking Attention Layers
Module 6	LangChain for LLM Application Development	• LangChain Foundations • Benefits of using LangChain • Using LangChain to Develop LLM Applications • Value Propositions of LangChain • Components of LangChain

Six Months Online Certificate Course: "**Data Science and Gen AI LLMs"**

		• Off-the-Shelf Chains in LangChain • Build and Deploy LLM-Powered Applications using LangChain
Module 7	Interacting with Data Using LangChain and RAG	• Understanding Retrieval-Augmented Generation (RAG) • Document Loading and Splitting • Vector Stores and Embeddings • Retrieval • Question Answering with Chatbots • Building RAG Models using LangChain
Module 8	Generative AI on Cloud	• Cloud Computing Foundations • AWS S3 • Amazon EC2 Trn1n • Amazon EC2 Inf2 • Amazon Sagemaker • Amazon CodeWhisperer • Amazon Bedrock • Azure OpenAI
Module 9	Working with ChatGPT	• Introduction to ChatGPT • Leveraging ChatGPT for Productivity • Mastering Excel through ChatGPT • Becoming a Data Scientist using ChatGPT • Data Analysis in PowerBI with ChatGPT • Creating a Content Marketing Plan • Social Media Marketing using ChatGPT • Keyword Search and SEO using ChatGPT • Generating Content using ChatGPT • Implementing ChatGPT for Customer Service • ChatGPT for Developers • ChatGPT for Creating Programs • ChatGPT for Debugging • ChatGPT for Integrating New Features • ChatGPT for Testing • Documenting the Code with ChatGPT

Six Months Online Certificate Course: "**Data Science and Gen AI LLMs"**

Module 10	Python with Generative AI	• Python Code Generation with Generative AI • Gen AI Tools for Coding • Advanced Code Optimization with ChatGPT Gen AI Tool • Coding with ChatGPT • Building an Application in Python with ChatGPT
Module 11	Evaluating LLM Performance	• LLM Performance Comparison • Perplexity • BLEU Score • Human Evaluation • Choosing the Right Metrics • Interpreting the Results
Module 12	Industry Case Studies	• Project: AI-Powered Text and Image Generator • Case Study: Generative AI for Personalized Learning • Case Study: Generative AI for Creative Content Generation • Case Study: Generative AI for Business

*****