DATA SCIENCE
CURRICULUM

Leveraging Data Science for Social Good

SABUDH

Table of Content

About Sabudh

Sabudh Foundation was founded by Dr. Sarabjot Singh Anand in 2018 along with Dr. Sukhjit Sehra. It empower students to apply Data Science to address real-world challenges and collaborate with citizens, governments and NGOs to drive tangible societal impact.

Recognizing the fact that real Data Scientists need not only to be delivering on customer requirements but also need to feed their hunger to grow their knowledge base in the fast-evolving field.

Data Science

[6 Months]
[48 hrs/week]

Program Learning Objectives

The program aims to provide learners with a comprehensive skill set across various domains. The Python course covers a broad spectrum of programming skills, from fundamentals like control structures to advanced topics such as web scraping and GUI implementation, encompassing software testing and debugging practices. Participants will also acquire proficiency in data acquisition, cleaning, analysis, and visualization, fostering data-driven decision-making. Additionally, they'll gain expertise in machine learning, deep learning concepts, and practical application, including advanced algorithms and training strategies. The Natural Language Processing & Recommender System course focuses on understanding interactions between computers and human language, alongside proficiency in data preprocessing, feature engineering, and recommendation systems.

The Dataiku certification course guides IT professionals through progressive mastery of the platform, covering machine learning models, data pipelines, code integrations, and MLOps practices. The Data Structure & Algorithm course aims to understand data representation concepts and efficient computation methods. Finally, the Passion Project equips students with research methodology skills to define problems, conduct literature reviews, apply research methods, analyze data, and effectively communicate findings.

01 Python Programming [6 weeks]

This course covers a solid foundation in Python programming, emphasizing practical skills essential for software development and data analysis. The learning objectives of this course will be accomplished through the following topics:

1. Python Installation and Setting up IDE
2. Python Basics: Problem-solving, Formatted Output, Comments & Docstrings, Identifiers, Keywords
3. Python Data Types: Including Type Casts and Operators
4. Python Functions: Function Calls, Different forms of Arguments, Scope and Lifetime of Variables, Use of the Global Keyword, Types of Functions (built-in and user-defined)
5. Control Flow Statements: if/else, for Loops, while Loops, Transfer Statements (break, continue, pass)
6. Python Data Structures: Overview of Built-in and User-defined Data Structures
7. File Handling: Basic File Operations in Python.
8. NumPy: Arrays, Basic Operators, Universal Functions, Shape Manipulation, Stacking Arrays, Linear Algebra Operations.
9. Pandas: Data Manipulation, Analysis, Merging Data, Basic Visualization with Pandas
10. Matplotlib: Introduction to the Library, State-based vs Object-oriented Interfaces, Creating Various Types of Plots.
11. Seaborn: Advanced Data Visualization Techniques
12. Web Scraping: Using BeautifulSoup and Selenium for Scraping Data
13. Object-Oriented Programming: Basic and Advanced OOP concepts in Python
14. Sklearn: Introduction to Machine Learning with Scikit-Learn
15. Scipy: Overview, Stats Module, Algebra Functions, Handling Sparse Data, Image Processing with SciPy.
16. APIs: Working with REST APIs, JSON Parsing, Security and Authentication in APIs
17. Flask: Building and Deploying Web Applications with Flask
18. Dockerization: Basics of Containerizing Python Applications with Docker
19. Asynchronous Tasks: Using Celery and Redis for Handling Background Tasks

20. Working with SQL Basics: Integrating SQL Databases with Python
21. GUI Programming: Introduction to GUI Development with Streamlit and Gradio
22. Testing and Debugging: Techniques and Best Practices for Testing and Debugging Python code

02 Machine Learning   [12 weeks]

This course covers a comprehensive introduction to data science, core machine learning techniques, and the mathematical foundations essential for understanding and implementing machine learning algorithms. Learning objectives of this course will be accomplished through the following topics:

1. Introduction to data Science and Machine Learning
2. Probability and Statistics for Machine Learning
3. Supervised Learning: Linear Regression, Logistic Regression, Naive Bayes. Support Vector Machines, Decision Trees , k-Nearest Neighbor (k-NN) Algorithm, Ensemble learning Methods: Bagging, Boosting and Random Forests
4. Model Selection, Building and Evaluation
5. Feature Engineering and Pre-processing
6. Model Optimization: Hyperparameter Tuning, Regularization, Grid Search Cross-Validation
7. Unsupervised Learning: Clustering (K-Means, Hierarchical, DB Scan, Expectation-Maximization algorithm (EM) and Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH))
8. Dimensionality Reduction: Principal Component Analysis, Singular Value Decomposition, Matrix factorization, Multidimensional Scaling, T-distributed Stochastic Neighbor Embedding (t-SNE)
9. Time Series Forecasting

03 Data Structures and Algorithms [20 weeks]

This course progresses from fundamental concepts to more advanced topics in DSA, ensuring a smooth learning curve. Each topic is accompanied by practice problems and exercises to reinforce learning and problem-solving skills. Learning objectives of this course will be accomplished through the following topics:

1. Introduction to Data Structures and Algorithms
2. Basics of Algorithm Analysis: Time Complexity Analysis (Big O notation, time complexity classes), Space Complexity Analysis
3. Introduction to Arrays ; Basics of Arrays and their Implementation, Basic Operations on Arrays
4. Introduction to Lists; linked lists, Singly linked lists, Doubly Linked Lists and Circular Linked Lists
5. Stacks & Queues; Implementation using Arrays and Linked Lists
6. Recursion
7. Trees: Binary Trees and Binary Search Trees (BST), Operations on Binary Search Trees (insertion, deletion, searching)
8. Tree Traversal Algorithms; Depth-first Traversal (pre-order, in-order, post-order), Breadth-first Traversal (level-order)
9. Graphs; Graph Representation (adjacency matrix, adjacency list), Graph Traversal Algorithms (depth-first search, breadth-first search)
10. Graph Theory; Graph Properties (connectedness, cycles, degrees), Graph Algorithms (shortest path, minimum spanning tree)
11. Tries; Introduction to Tries (prefix trees), Operations on Tries, Applications of Tries

04 Dataiku [6 weeks]

Dataiku certifications are highly regarded in the data science and analytics industry. This course will cover the fundamental concepts and functionalities of Dataiku DSS and guide you through the process

of designing and building end-to-end data pipelines. Learning objectives of this course will be accomplished through the following certificates:

1. Core Designer Certificate: Dataiku Datasets and Visual Recipes
2. ML Practitioner Certificate: Visual Machine Learning and Interactive Statistics
3. Advanced Designer Certificate: Variables, Data Pipelines, and Scenarios
4. Developer Certificate: Code Recipes, Webapps, and the APIs.

05 Deep Learning [12 weeks]

This module provides a comprehensive journey through deep learning, beginning with foundational concepts to advanced techniques. This will also cover practical applications to gain hands-on experience. Learning objectives of this course will be accomplished through the following topics:

1. Introduction to Deep Learning: From Logistic to Neural Networks, Basics of Neural Networks, Perceptron Model and Activation Functions, Forward Propagation and Backpropagation, Introduction to TensorFlow for Gradient Computation
2. Multilayer Perceptron (MLPs); Structure and Architecture of MLPs, Training MLPs using TensorFlow, Activation Functions for MLPs, Applications of MLPs
3. Convolutional Neural Networks (CNN) and Digital Image Analysis; CNN Architectures (e.g. VGG, ResNet),
4. Training CNNs for Image Classification, Digital Image Analysis Tasks (e.g., object detection, segmentation)
5. Object Detection; Overview, Evolution and Strategies, Single Shot, Few Shot Learning
6. Recurrent Neural Networks (RNN); RNN Architectures (e.g. LSTM, GRU), Applications of RNNs (e.g., sequence prediction, natural language processing)

7. Generative Models; Generative Adversarial Networks (GANs), Vector Autoregressive (VAR) models, Attention Models and Transformers
8. Sequence to Sequence Models: Encoder-Decoder Models
9. Autoencoders
10. Speech Data Processing

06 Natural Language Processing & Recommender systems [13 weeks]

This course provides a thorough journey through the fundamentals and applications of Natural Language Processing (NLP), spanning from basic text processing techniques to advanced topic modeling methods. Additionally, it covers the practical insights into content recommendation systems and dimensionality reduction techniques for enhancing the understanding of real-world NLP challenges and solutions. Learning objectives of this course will be accomplished through the following topics:

1. Natural Language Processing (NLP): Fundamentals of NLP, Basic Text Processing Techniques, Tokenization, Lemmatization, and Stemming, Part-of-Speech Tagging and Named Entity Recognition
2. Syntactic Vectorization
3. Latent Semantic Analysis
4. Multivariate Bernoulli and Multinomial Naive Bayes for Text Classification
5. Latent Variables in NLP
6. Advanced Topic Modeling Techniques: Latent Dirichlet Allocation, Probabilistic Modeling
7. Text Vectorization
8. Content Recommendation: Introduction to Recommender Systems, Content-Based Recommendation Techniques, TF-IDF and Cosine Similarity for Content-Based Recommendation
9. Documents as Vectors
10. PCA and Singular Value Decomposition

07 Passion Project

We provide opportunities to delve into passion projects with a focus on creating positive social change. Through these projects, learners will not only enhance their technical skills but also gain valuable experience in applying data science methodologies to address real-world social issues.

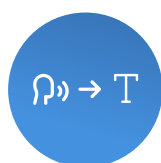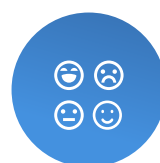|				
Bird Call	Sat Sri Akal	Chatbot	Smart Glasses	Medical NLP
Depression Detection	Bus Route Optimization	Hateful Meme Detector	Solar Power Prediction	Road Scene Analysis
Long Document Keyphrase Generation	AI for Videos Analysis	AI for Book Analysis	Intelligent Document Processing	AI Smart Search Documents
ML Approached for Quality Assessment of the Map Data	Speech To Text Model for Indian Languages	Socio-Economic & Political Analysis through Bollywood Songs	The ECG Monitoring & Decision Support System	Text-to-Speech (Punjabi Language)

PII Information Extraction from Text

Raag Identification & Understanding

Medical Review/ Diagnosis Prediction

Sentiment Analysis for News Covered in Print Media

Object Detection & Classification using Satellite Imagery

Structure Prediction in Proteins using Aminoacid Properties

Parsing & Information Retrieval from Document Using LLMs

Using Deep Learning for Image Processing in Pathology & Radiology

Document Analysis using Large Language Models (LLMs)

Detecting Pronunciation Errors for Automatic Correcting of Speech Based Answers

Program Outcomes

Upon completion of this program, students will achieve:

1. Mastery of Python programming skills, with proficiency in software testing and debugging.
2. Ability to acquire, clean, analyze, and visualize data, enabling informed decision-making through data-driven insights.
3. Expertise in machine learning and deep learning concepts, including the practical application of advanced algorithms and training strategies.
4. Understanding of natural language processing and recommender systems, with skills in data preprocessing, feature engineering, and recommendation algorithms.
5. Proficiency in utilizing the Dataiku platform for machine learning model development, data pipelines, code integrations, and MLOps practices.

6. Understanding of efficient computation methods through the Data Structure & Algorithm course.
7. Research methodology skills, including problem definition, literature review, research methods application, data analysis, and effective communication of findings, developed through the Passion Project.

Certifications

Upon successful completion of Sabudh's Internship Program, you will be awarded a **"Certificate of Completion"** certifying you as JOB-READY in Data Science. This certificate will validate your proficiency in the field and enhance your employability.

Additionally, participants have the opportunity to enroll for individual courses. For the passion project component, eligibility is limited to students enrolled in the full-time 6-month program only.

Overall, this portfolio will serve as a showcase of your skills and capabilities to prospective employers. Beyond demonstrating your technical knowledge, it will highlight your critical thinking skills and your ability to apply data science methodologies to solve real-world problems.

THANK YOU

SABUDH

To know more visit- https://sabudh.org/

Contact Number- 8837662054

Facebook   LinkedIn   YouTube   Instagram

Or Scan