



## THE NELSON MANDELA AFRICAN INSTITUTION OF SCIENCE AND TECHNOLOGY (NM-AIST)



### INFORMATION AND COMMUNICATION TECHNOLOGY RESOURCE CENTER (ICT- RC)

## PROFESSIONAL TRAINING ON DATA SCIENCE AND MACHINE LEARNING FOR THE TOURISM AND CONSERVATION COMPANIES

The ICT-RC in collaboration with Centre for Development of Advanced Computing, Pune (India) has prepared a three weeks professional training to Tourism and Conservation stakeholders for a practical training on using data science and machine learning techniques in working with real-world datasets. Through the training, participants will delve into the world of data science and machine learning as appropriate state of the art approaches in handling complex datasets, analytics and predictions tasks. Specifically, participants will be strengthened on:

- Data preprocessing and exploration
- Statistical foundations in data science and machine learning
- Data visualization
- Machine learning techniques
- Advanced Machine Learning techniques for Tourism and Conservation applications
- Deployment of predictive models

If you are interested please request for invitation letter by writing to: [devotha.nyambo@nm-aist.ac.tz](mailto:devotha.nyambo@nm-aist.ac.tz). Mob: 0752905156

### Facilitation Details:

- i. Name of the course Facilitator: NM-AIST ICT-RC team and CDAC, India
- ii. Facilitators Contacts/Coordinator – [devotha.nyambo@nm-aist.ac.tz](mailto:devotha.nyambo@nm-aist.ac.tz) and [m DOE.mwamnyange@nm-aist.ac.tz](mailto:m DOE.mwamnyange@nm-aist.ac.tz)
- iii. Duration: Three Weeks (17<sup>th</sup> February – 07<sup>th</sup> March 2025)

## **Week 1: Foundations of Data Science and Machine Learning**

**Goal:** Introduce core data science concepts, tools, and techniques with a focus on insurance applications.

### **Module 1: Introduction to Data Science for Tourism and Conservation**

- Overview of data science and its relevance to Tourism, hospitality and conservation.
- Key data challenges in Tourism and conservation (e.g., fraud detection, trends analysis, risk prediction).
- Overview of Python/R for data analysis.

### **Module 2: Data Preprocessing and Exploration**

- Data cleaning and preprocessing for Tourism datasets.
- Exploratory Data Analysis (EDA):
  - Identifying patterns and trends in Tourism datasets.
  - Detecting outliers and anomalies.
- Practical: EDA on a Tourism dataset.

### **Module 3: Statistical Foundations**

- Descriptive and inferential statistics in the tourism industry.
- Correlation, regression, and hypothesis testing.
- Practical: Analysing the relationship between tourists profiles and return visits.

### **Module 4: Data Visualization**

- Creating impactful visualizations for tourism data using tools like Matplotlib, Seaborn, and Power BI.
- Practical: Visualizing clients segmentations.

---

## **Week 2: Machine Learning Techniques for Insurance**

**Goal:** Equip participants with machine learning skills for predictive modeling and decision-making in tourism and conservation domains.

### **Module 5: Introduction to Machine Learning for Insurance**

- Overview of machine learning concepts and workflow.
- Supervised vs. unsupervised learning applications.

## **Module 6: Supervised Learning Models**

- Linear regression for pricing.
- Logistic regression for return prediction.
- Decision trees and Random Forests for risk assessment.
- Practical: Building models to predict pricing.

## **Module 7: Unsupervised Learning Models**

- Clustering for customer segmentation.
- Anomaly detection for fraud identification.
- Practical: Applying k-means clustering to group clients.

## **Module 8: Model Evaluation and Optimization**

- Cross-validation, accuracy, precision, recall, and F1-score.
  - Hyperparameter tuning with Grid Search and Random Search.
  - Practical: Optimizing a predictive model for claim classification.
- 

## **Week 3: Advanced Techniques and Real-World Applications**

**Goal:** Focus on advanced topics and real-world projects tailored to the tourism and conservation industry.

## **Module 9: Time Series Analysis**

- Introduction to time series for clients trends.
- Forecasting applications for weather as related to seasonalities.
- Practical: Building time series models using ARIMA and Prophet models.

## **Module 10: Advanced Applications in Conservation**

- Fire detection algorithms.
- Climate variability and predictions.
- Practical: Implementing a models and mobile-based approaches to monitoring

## **Module 11: Deployment and Integration**

- Introduction to model deployment using Flask/Django.
  - Integration techniques with existing systems.
  - Practical: Deploying a claim prediction API.
-