



## 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 INSURANCE COMPANIES

The ICT-RC in collaboration with Centre for Development of Advanced Computing, Pune (India) has prepared a three weeks professional training to Insurance stakeholders for practical training on using data science and machine learning techniques in working with insurance 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 Insurance applications
- Applications of Deep Learning in Insurance
- 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 [mdoe.mwamnyange@nm-aist.ac.tz](mailto:mdoe.mwamnyange@nm-aist.ac.tz)
- iii. Duration: Three Weeks

## **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 Insurance**

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

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

- Data cleaning and preprocessing for insurance datasets.
- Exploratory Data Analysis (EDA):
  - Identifying patterns in claims and customer data.
  - Detecting outliers and anomalies.
- Practical: EDA on a sample insurance claims dataset.

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

- Descriptive and inferential statistics in insurance contexts.
- Correlation, regression, and hypothesis testing.
- Practical: Analyzing relationships between policy features and claim likelihood.

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

- Creating impactful visualizations for insurance data using tools like Matplotlib, Seaborn, and Power BI.
- Practical: Visualizing customer segmentation and claim trends.

---

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

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

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

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

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

- Linear regression for premium pricing.
- Logistic regression for claim prediction.
- Decision trees and Random Forests for risk assessment.
- Practical: Building models to predict insurance claims and customer churn.

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

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

## **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 insurance industry.

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

- Introduction to time series for insurance.
- Forecasting claim frequency and revenue trends.
- Practical: Building time series models using ARIMA and Prophet.

## **Module 10: Deep Learning in Insurance**

- Neural networks for insurance claim predictions.
- Application of NLP in analyzing insurance documents (e.g., policy terms, claims notes).
- Practical: Sentiment analysis on customer feedback or chatbot logs.

## **Module 11: Advanced Applications in Insurance**

- Fraud detection using anomaly detection algorithms.
- Risk scoring and premium pricing using ensemble models.
- Practical: Implementing a fraud detection pipeline.

## **Module 12: Deployment and Integration**

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

---

### Capstone Project (End of Week 3)

- Participants will work in groups to apply the concepts learned.
- Example Projects:
  1. Build a customer churn prediction model for policyholders.
  2. Create a risk scoring engine for premium pricing.
  3. Design a fraud detection pipeline for claims processing.