THE 2023 WCO BACUDA SCHOLARSHIP PROGRAMME (DATA ANALYTICS)

Offering a five-month offline programme to Customs officials of Member Customs administrations for in-depth study and training in Customs-related data analytics.

I. OVERVIEW OF THE PROGRAMME

Following the COVID-19 pandemic, the role of Customs in international trade and investment has become increasingly important to the economic strength of countries in terms of effective controls, trade facilitation and attracting business and investment. However, Customs must adapt to a changing world if it is to remain relevant and be a positive force within governments.

II. FRAMEWORK OF THE PROGRAMME

This is a five-month, non-degree programme: five months face-to-face. The aim of this programme is to enhance the capacity of Customs officials in the area of data analysis. Data analysis is a tool used by companies and governments for years in their business, to drive priority-setting, decision-making, performance measurement, budget planning and forecasting, and operations. Customs has a substantial amount of data at its disposal, beginning with data submitted for the Customs clearance process. Thanks to the development of digital technology, Customs can tap into data from other government agencies, commercially available databases, and open-source information platforms such as digitized global public records and multilingual news sources.

The framework of the programme includes:

- Fundamentals of Big Data Analytics
- Machine Learning Fundamentals
- Python Programming
- Applied Advanced Analytics
- Data Technology
- WCO BACUDA Algorithm
- WCO Topics, including Data Strategy and Policy for Members.

This second BACUDA Scholarship Programme incorporates a new training component, based on positive feedback from the first cohort of BACUDA Scholars; this new component will last more than one month, and will involve conducting a practicum project as follows:

- Data cleansing with synthetic (or raw) data, for three weeks
- Group assignments to develop analytic models, for two weeks or more

III. SELECTION OF PARTICIPANTS

1. The number of participants in the programme will be twelve (12). Given the objectives of the programme, they should be working-level officials of Customs administrations in developing countries.
2. The candidates should be in good health and possess good communication skills.
3. Officials who apply for the programme must meet all of the following requirements: (1) Certification of completion of the Data Analytics - Beginners Course on WCO CLiKC!; (2) At least two years of experience working in Customs Data Analytics; (3) Proof of English proficiency (both speaking and listening).

IV. FUNDING

The programme will be financed through CCF-Korea, funded by the Korea Customs Service. The fund will cover travel costs, admission fees, tuition fees, institutional costs, accommodation costs and other approved incidental expenses, to enable the participants to complete the programme.

V. PROGRAMME SCHEDULE AND DESCRIPTION

1. Course Description

The programme consists of offline courses/activities, divided into the following modules:

1) Fundamentals of Big Data Analytics
2) Machine Learning Fundamentals
3) Machine Learning Applications
4) Python Programming
5) Analytical Skills for AI and Data Science
6) Network Analysis in Customs
7) Data Technology
8) WCO BACUDA Algorithms (Online/Offline)
9) WCO Topics (Online/Offline)

2. Curricular Modules

At the end of the programme, each participant must submit to the University a research paper on a Customs topic concerning analytic model development, making full and best use of the knowledge and insights gained from the programme. The research paper must be validated by the participant's tutor and must demonstrate the participant's analytical skills, understanding of the topics and discussions from the programme, and his/her strategic ability and skills. Below is an outline of the modules (subject to possible changes which will be made known to the participants in due course).

Module 1	Fundamentals of Big Data Analytics
1	Unlocking the Potential of Customs Data
2	Overview of the Hadoop Ecosystem
3	MapReduce and HDFS Fundamentals
4	NoSQL Databases
5	Real-time Processing
6	Analytics
7	Inductive and Deductive Investigation
8	Introduction to Pig and Hive
9	Communicating your Findings
10	Onsite Deployment

Module 2	Machine Learning Fundamentals
1	Machine Learning Overview
2	Model Performance Metrics
3	Feature Selection and Engineering
4	Machine Learning Algorithm
5	Regression Analysis - Defining information needs and conducting regression analysis
6	Regression Analysis - Structuring problems data and features
7	Regression Analysis - Evaluating regression models and presenting the results
8	Classification Analysis - Conducting classification analysis
9	Classification Analysis - Structuring problems data and features
10	Classification Analysis - Evaluating classification models
11	Clustering Analysis - Conducting clustering analysis
12	Clustering Analysis - Structuring problems data and features
13	Clustering Analysis - Evaluating clustering models

Module 3	Machine Learning Applications
1	Types of Machine Learning Techniques
2	Supervised Learning – Classification, Regression
3	Unsupervised Learning – Clustering, Association
4	Semi-Supervised Learning – Classification, Clustering
5	Reinforcement Learning
6	ML Algorithms: Linear Regression
7	ML Algorithms: Logistic Regression
8	ML Algorithms: Decision Tree
9	ML Algorithms: SVM
10	ML Algorithms: Naïve Bayes
11	ML Algorithms: k-NN
12	ML Algorithms: Random Forest
13	ML Algorithms: Dimensionality Reduction Algorithms
14	ML Algorithms: GBM, XGBoost, LightGBM, CatBoost
15	**Practice ML Problems**

Module 4	Python Programming
1	Python - Variables and Operations
2	Python - Data Structure
3	Python - Control Flow Statement
4	Python - Functions and Packages
5	Python - Classes and Class Inheritance
6	NumPy Basic / Structure and Creation
7	NumPy Array Indexing / Broadcasting
8	NumPy Methods / Data Structures in Pandas
9	Pandas Data Frame / Methods on Pandas Data Frame
10	Reshaping Data Frames / Text Files (.txt, .csv) - Input & Output
11	JSON Files – Input & Output
12	XML -_HTML / Database (SQLITE)
13	Introduction to Data Visualization with Matplotlib
14	Advanced Data Visualization with Matplotlib
15	Visualization with Pandas
16	**Practice: Data Visualization with Pandas and Matplotlib**

Module 5	Analytical Skills for AI and Data Science
1	Introduction to Evidence-based Decision Making: Examples in Economic Development
2	Analytical Thinking and Methods (I): How to Define Research Question and Introduction to Statistical Approaches to Draw Inference
3	Machine Learning Applications for the Public Sector
4	Discovering and Collecting Data: Practical Advice for Government Managers
5	Analytical Thinking and Methods (II): Correlation, Causation and Decision-Making
6	Barriers to Building a Data Practice in Government
7	Data Collaboratives: Creating Public Value By Exchanging Data
8	Strengthening a Data Analytics Culture
9	Analytical Thinking and Methods (III): Optimization
10	Practical Guide to Analytic Selection

Module 6	Network Analysis in Customs
1	What are Networks?
2	Network Analysis as a Method
3	Social Network Analysis
4	Visualizing Network Models
5	Foundational Network Measures
6	Introduction to Big-FINDER Models
7	Air Passenger Intelligence
8	Customs Supply Chain Analysis
9	Delinquent Taxpayer Analysis
10	Practice – CASOS ORA

Module 7	Data Technology
1	What is NoSQL?
2	MongoDB Installation and Environment Setting
3	Data Entry/Modification/Deletion
4	Data Manipulation Method
5	Creating Reports Using NoSQL
6	Report Visualization

Module 8	BACUDA Algorithms & WCO Topics
1	AI HS Recommendation
2	DATE Algorithm
3	Synthetic data generation
4	Image recognition
5	WCO Data Strategy and Policy
6	Data Analytics Framework
7	BACUDA, WCO Data Analytics Project
8	Members' Use Cases and Lessons Learnt (provisional)
9	WCO working on Disruptive Technologies (provisional)

3. Programme Schedule

- **All schedules are provisional and subject to change** *(Please see detailed programme in Excel file)*
- **Best Practice on Data Analytics will be presented by each participant during the face-to-face classes, to share ideas and receive insights from classmates and professors.**
- **Classes will be organized by SKKU in cooperation with the KCS and the private sector.**

VI. Main responsibilities of participants

To improve their analytic capacities and develop analytic models throughout the course, the participants selected are expected to work together with some of their fellow participants in separate groups.

They will be invited to attend an orientation session by the training academy at the beginning of the course, to select a topic for each Group Project, and will undertake study on an assigned topic jointly agreed with participants in cooperation with tutors from the academy and the SKKU.

- A report to evaluate their performance and the outcome of the course should be written by individual trainees (or by each group) and submitted to the SKKU and the WCO.
- Reporting should take place twice over the whole course, i.e., a mid-term progress report and a final report. The necessary template and detailed instructions for this task will be provided by the SKKU during the orientation session.
- The mid-term report should be submitted **within two months after the start of the course**, and the final report should be submitted **before the closing ceremony**. The reports should be uploaded onto the BACUDA Website, https://bacuda.wcoomd.org where appropriate.

Annex 1. BACUDA Scholarship Syllabus

Subject	Hours	Head Professor	Affiliation
• Welcoming remarks	1	TBD	WCO/SKKU
• WCO Topics	8	TBD	*WCO*
• Organizational Behaviour	4	Youngshin Chun	SKKU
• Basic Data Structures	8	(Yun Joo An)	SKKU
• Fundamentals of Big Data Analytics	16	Yunbae Kim	SKKU
• Machine Learning Fundamentals	18	YoungjuNilson	SKKU
• Analytical Skills for AI and Data Science	12	Kiheung Park	SKKU
• Block-chain Technology	10	JeMin Lee	SKKU
• Python Programming - Advanced Course	18	Shin Sunwook	SKKU
• Special Lectures	10	TBD	SKKU
• Machine Learning Applications - Regression, Classification, Clustering	20	Simon Woo	SKKU
• Network Analysis - Graph Theory & Network representations	18	Janghyun Kim	SKKU
• Big Data	20	TBD	SKKU
• BACUDA algorithms and relevant topics	15 (offline) 6 (online)	AP Other regions	BACUDA Scholars
• Practicum Project - 1	70	UNI-PASS system developer	
• Practicum Project - 2	51	Private training academy	
Total	185+70+51=302		