2022 BRICS Skills Competition
(BRICS Future Skills Challenge)

TECHNICAL DESCRIPTION

Machine Learning and Big Data (Online)

CONTENT

1 INTRODUCTION

The development of machine learning (ML) and big data has revolutionized the way businesses operate and make decisions. In the past, businesses relied on humans to make decisions based on their experience and intuition. However, with the advent of ML and big data, businesses can now rely on machines to make better decisions faster and more accurately.

ML is a subset of artificial intelligence (AI) that allows machines to learn from data and improve their performance over time. ML algorithms are able to automatically improve given more data. This is in contrast to traditional AI, which requires humans to write rules for the machine to follow.

Big data is a term used to describe the large volume of data that is generated by businesses and organizations on a daily basis. This data can come from a variety of sources, including social media, website click-streams, sensors, and transactional data.

The combination of ML and big data is providing businesses with the ability to make better decisions, faster. Here are some examples of how ML and big data are being used by businesses:

1. Automated Customer Service: Businesses are using ML to automate customer service tasks, such as answering common questions and routing customer requests. This is possible because ML algorithms can be trained to understand natural language.

2. Predictive Maintenance: Businesses are using ML to predict when equipment is likely to fail so that they can take preventive measures. This is possible because ML algorithms can be trained to identify patterns in data that indicate an impending failure.

3. Fraud Detection: Businesses are using ML to detect fraud before it occurs. This is possible because ML algorithms can be trained to identify patterns in data that indicate fraudulent activity.

4. Personalized Marketing: Businesses are using ML to create personalized marketing campaigns for their customers. This is possible because ML algorithms can be trained to understand the preferences of individual customers.

The development of ML and big data is providing businesses with the ability to make better decisions, faster. This is just the beginning – the potential applications of ML and

big data are endless. As businesses continue to adopt these technologies, we will see even more transformative changes in the way businesses operate.

Since this document contains only a common description for **MLBD Skill Challenge** and Task-specific information, it must be used in association with the following:

- Guidelines and additional common requirements of 20222022 BRICS Skills Competition (BRICS Future Skills Challenge) Organising Committee

- BRICS International Standards Specification for **Machine Learning and Big Data** Competence

- Health, Safety, and Environment Policy and Regulations

- Code of Conduct

2   PARTICIPATION REQUIREMENTS

According to the relevant decision of the 2022 BRICS Skills Competition (BRICS Future Skills Challenge) Organizing Committee led by China, any country can participate in the skills challenge.

Also, registration is up to 10 teams from each country.

For the MLBD Skill the following rules will be implied:

- Each participating country at the online MLBD Skill Challenge must delegate **one** Expert (hereinafter Expert-Compatriot), who will represent the country at MLBD Skill. All communication concerning the online MLBD Skill Challenge, technical and organizational aspects, task, assessment, etc. will go through the Expert-Compatriot.

- If the participated country is a BRICS member, the local BRICS Skills Development Working Group manage the relationship with the BRICS Skills Development Working Group of China and select the Expert-Compatriot for the MLBD Skill.

- If the country is not a BRICS member, the first applicant will be the Expert-Compatriot.

- Expert-Compatriot is responsible for all activities within his/her country of origin in preparation, organization, and participation in the online MLBD Skill Challenge. The Expert-Compatriot can decide the number of own country's Experts participating in the Skill Challenge and the number of Competitors (participating Students).

- Expert-Compatriot is responsible for organizing the proper technical environment for each Competitor and following the Skill Challenge procedures (e.g. Test task distribution, ensuring fulfillment of the Code of Conduct by each Competitor, following the requirements of the time limit for the task accomplishment, etc.), for the assessment of Competitor's projects in their own country, and compliance with Health, Safety, and Environment (HSE) rules.

As a result, Expert-Compatriot will get comprehensive training in organizing their own country MLBD Skill, to run the competition and will be ready to represent their own country at MLBD Skill at the 2022 BRICS Skills Competition (BRICS Future Skills Challenge).

3 SCHEDULE

As the participants of the online MLBD Skill Challenge located across the world with different timezones (from Beijing/China at UTC+8 to San-Paulo/Brazil at UTC-3), the MLBD Skill Team is proposing the distributed and synchronous Skill Challenge form, which allows:

- To participate from any country of the world independently from the Time Zone.

- Subject to the time of the host country.

- Reduce the daily workload for the participants due to splitting the Task Modules over Skill Challenge time and limiting the duration of one Module to 3 hours.

- Enable the participation of an unlimited number of participants from each country: participant gathering/selection, management, assessment, task distribution will be done by the appointed Country's Expert-Compatriot.

- To get familiar with the assessment and with the marking scheme by Expert-Compatriots from all participating countries who are lead by the Chief Expert and the Deputy Chief Expert assessing all participants's tasks.

However, as some common activities are necessary for the Skill Challenge organization and conduction, the Skill Challenge Organizing Committee will select the time which will be appropriate for most countries to do the live sessions. These sessions will be recorded and will be available to participants from other regions who were not able to participate, with additional time follow-up Q&A sessions by the Skill Team.

The Competitors will need three hours a day involvement, the Experts will need additional time for the assessment procedures and follow-ups.

The common events such as introduction and non-competition activities will need less than two hours.

The online MLBD Skill Challenge will have the following schedule:

Day	Topic	Description
C-2	Competition Day 1	Exerts will jointly assess one of the participant tasks to understand the marking and the assessment criteria. They do the assessment of all participants under the leadership of the Chief Expert and Deputy Chief Expert.
C-1	Competition Day 2	
C1	Challenge Opening & Competition Day 3	
C2	Competition Day 4	
C3	Challenge Closing	

That form will enable all participants to utilize 100% of the online MLBD Skill Challenge, get familiar with the MLBD Skill requirements, tasks, procedures and be able to initiate the MLBD Skill implementation on the national level.

4   SKILL SPECIFICATION

SECTION		RELATIVE IMPORTANCE (%)
1	**Work organization and management**	**5**
	The Competitor **must know and understand**: • principles and skills that ensure productive teamwork; • principles and behavior of systems; • aspects of systems that enhance the stability and environmental safety of products; • analysis and assessment of information obtained from various sources. The Competitor **must be able to**: • plan the production schedule for each day in accordance with the available time and take into account time constraints and deadlines; • apply research technologies and skills to understand the latest industry recommendations; • analyze the results of their own activities in comparison with the expectations and needs of the client and the organization.	
2	**Communication and interpersonal relationships**	**5**
	The Competitor **must know and understand**: • importance of documenting decisions; • importance of resolving misunderstandings; • importance of written and verbal communication skills; • importance of thoroughly documenting developed solutions; • manifestations of professionalism in the preparation of documentation. The Competitor **must be able to**: *Use literacy skills to:* • following the documented instructions in the provided manual; • understanding the instructions for the organization of the workplace and other technical documentation; • interpretation and understanding of requirements; • awareness of the latest industry recommendations. *Use oral communication skills to:* • discussing and proposing data; • presentation of the proposed and final decision. Use written communication skills to: • development of user documentation; • work with technical documentation in English; • compilation of interactive reports on data analysis in the environment of Jupyter Notebook or similar. *Use project management skills in:*	

SECTION	RELATIVE IMPORTANCE (%)
• prioritization and scheduling of tasks; • allocation of resources between tasks.	
3 **Problem Solving, Innovation, Creativity**	**10**
The Competitor **must know and understand**: • general types of problems that may arise when developing data analysis solutions; • general types of problems that may arise in a commercial organization; • diagnostic approaches to solving problems; • industry trends and developments, including new technologies, methods, languages, conventions and technical skills. The Competitor **must be able to**: *Use analytical skills to*: • analysis and synthesis of complex or heterogeneous information; • define trivial and nontrivial data dependencies. *Use research and training skills for:* • independently investigate problems. *Independently solve the problems that were encountered in the process of work:* • timely identify and solve problems; • correctly collect and analyze information; • develop alternatives for making decisions, choose the most	

SECTION		RELATIVE IMPORTANCE (%)
	appropriate options and implement the necessary solution.	
4	**Data Analysis and Processing**	25
	The Competitor **must know and understand**: • varieties of data analysis; • data analysis techniques; • put and test hypotheses about data dependence; • methods of data collection; • principles of work with graphic data; • principles of working with text data; • principles of working with audio data; • principles of working with video data. The Competitor **must be able to**: • analyze time series; • identify anomalies and deviations in the data; • structure data; • carry out normalization and preparation of data; • highlight features, properties and characteristics of objects in the data; • clustering data; • search for data correlation; • carry out operations with big data; • identify patterns in the data; • conduct a visual analysis of the data; • create patterns of patterns; • use software tools for research and data processing.	

SECTION		RELATIVE IMPORTANCE (%)
5	**Building Machine Learning Models**	30

	The Competitor **must know and understand**:
	 • various methods and algorithms of machine learning;
	 • quality criteria for machine learning models;
	 • the sequence of development of machine learning models;
	 • what methods of machine learning to apply depending on the source data;
	 • how to work with various data samples;
	 • how to use various software tools to develop and improve models.
	The Competitor **must be able** to apply the classic machine learning algorithms:
	 • unsupervised learning (reduction of dimensionality, search for rules, clustering);
	 • supervised learning (regression, classification); apply reinforcement learning algorithms;
	 • genetic algorithms;
	 • Q-Learning Algorithms; apply ensemble methods of machine learning:
	     ○ stacking;
	     ○ begging;
	     ○ boosting;
	• deep learning methods and neural networks:
	     ○ perceptrons;
	     ○ convolutional neural networks;
	     ○ recursive networks;
	     ○ auto encoders.	

6	**Application Development**	25

SECTION	RELATIVE IMPORTANCE (%)
The Competitor **must know and understand**: • approaches to the development of applied solutions; • architecture of applied solutions; • ways to develop applied solutions; • principles of building human-machine interaction. The Competitor **must be able to**: • own high-level programming languages; • interpret the results of the decision; • interpretation of big data and big images; • carry out the development of applied solutions using machine learning models.	
Total	**100**

5   HEALTH, SAFETY, AND ENVIRONMENT (HSE)

All participants must comply with the Health, Safety, and Environment legislation.

Each Participating Country Technical Delegate is responsible for ensuring that their compatriot Competitors and Experts have been briefed on the correct conduct and the Skills Camp Health, Safety, and Environment policies and requirements before the training begins, so that they understand and adhere to these requirements.

Chief Expert/Technical Experts of the Host Country are responsible for compliance with the requirements at the Skill HQ and at the remote venue as well as for health and safety training sessions for experts and competitors, including compliance with sanitary and epidemiological rules, recommendations for preventing the spread of a new coronavirus infection (COVID-19) and other recommendations of Executive authorities.

Each Expert-Compatriot is responsible for preparing safe Work stations for him/herself and his/her competitors as well as for health and safety training sessions for conducting their competitors.

All competitors are responsible for their state of health and shall confirm that they have no diseases that are a contraindication to participate in the competitions.

HSE documentation shall be included in the Skills Description for each skill and posted on the 2022 BRICS Future Skills Challenge website a week before its start.

6 TASK DESCRIPTION

Within the framework of the competition task, based on a large dataset, the competitor needs to prepare a big data environment required for the task, extract data from multiple sources, transform data into a format and load data into the data warehouse, analysis and visualize data in the data warehouse, then build a machine learning model to make predictions.

The competition task is secret and is available for review by participants 15 minutes before the start of each session by e-mail or through the competition platform which is based on public cloud.

The competition task has several modules that are executed sequentially during 4 competitive days. The modules and time are summarized in table 1.

#	Module	Lead time, hours
1	Preparation for Big Data Environment	3
2	Build a Data Warehouse	3
3	Data Analysis and Visualization	3
4	Machine Learning	3
	Total	**12**

This Task structure is adapted for the specifics of the online form of the challenge.

7   INSTRUCTIONS TO THE COMPETITOR

The competitors will be provided with all the necessary tools. The workplace of the Competitors must be equipped with:

- Personal computer with two 2K monitors.

- Peripherals, e.g. keyboard, mouse.

- Webcam with a resolution of at least 720p HD.

- Google Chrome with the latest version is installed on the PC.

- Internet connection speed of at least 50 Mbps, delay not more than 200 ms.

The following distribution of the modules by competition days will be used (Table 2).

#	Module	Competition Days			
		C-2	C-1	C1	C2
1	Preparation for Big Data Environment	+			
2	Build a Data Warehouse		+		
3	Big Data Analysis and Visualization			+	
4	Machine Learning				+

8 EQUIPMENT, MACHINERY, INSTALLATIONS, AND MATERIALS REQUIRED

The following table defines the version of used software.

Operating System	Windows, MacOS, Linux
BRICS Competition Client	1.0.0 or above
Google Chrome	99.0.4844.84 or above
Tencent Meeting	3.6.2 or above
Foxit Reader	11.1.126.51346 or above
Office Package	2019 or above

A computer camera and headset (or external microphone and speakers) are also required.

During the competition, it is **allowed to use the Internet**. Not allowed to use teas and social networks. The communication platform will be **Tencent Meeting**.

Additional required software for online communication, tracking, etc. will be provided additionally.

9   MARKING SCHEME

The Task is be accompanied by an associated marking scheme matching the assessment criteria as given in the Technical Description (Marking Summary). For each of these criteria, a detailed list of aspects to be assessed is defined.

	Criteria	Points		
		Subject	**Object**	**Total**
A	Preparation for Big Data Environment		**20**	**20**
B	Build a Data Warehouse		**20**	**20**
C	Data Analysis and Visualization	**15**	**15**	**30**
D	Machine Learning		**30**	**30**
	Total	**15**	**85**	**100**