EQUIP UK★VN
Enabling an environment of QUality for International HE Partnerships between the UK and Viet Nam

BRITISH COUNCIL

OBJECTIVE 2 – QUALIFICATION BENCHMARKING REPORT

MSc Data Science

Hanoi University of Science and Technology (HUST)

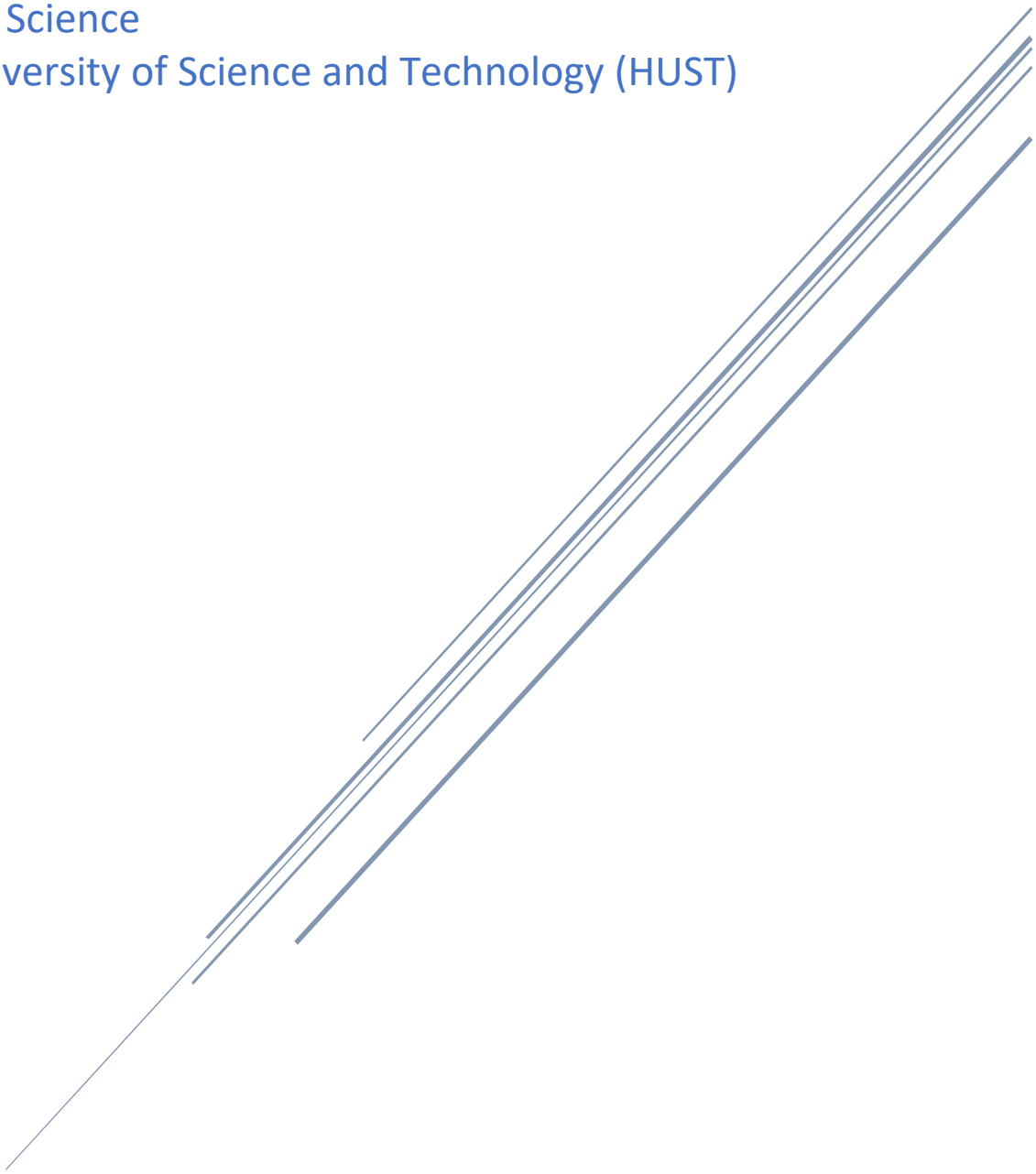

Table of Contents

Introduction

The MSc Data Science programme at Hanoi University of Science and Technology (HUST) addresses the needs of local employers as well as the challenges of the digital transformation ambition of Vietnamese industries and government.

For this benchmarking exercise we have developed a scoring matrix where we identified 5 themes (programming, knowledge management, knowledge abstraction, knowledge representation/communication and research/soft skills). **Programming** theme entails criteria related to design and development of not only software but also other artefacts like algorithms, network, IoT framework etc. The theme also includes the evaluation process and collaborative management of the artefacts. **Knowledge management** primarily focuses on processes and techniques of warehousing different types of data. The theme also includes security and privacy issues related to data management. **Knowledge abstraction** theme focuses on different data analytics and machine learning techniques applied to different types of data. **Knowledge representation/communication** theme includes different visualisation techniques used to represent the results (from database query through to data analytics to algorithm) to a wide range of stakeholders. **Research/Soft skills** theme focuses on the understanding and practice of research methods along with the ability to undertake team work and present results to a wider audience.

Within each theme, we have a set list of criteria against which each course is scored. The score is within the range of 50 – 100. 90 – 100 (fully meets the criteria); 75 – 89 (mostly meets the criteria); 60 – 74 (partially meets the criteria); 50 – 59 (barely meets the criteria). The marks are indeed subjective and therefore debatable. However, the pattern that emerges as result of the scoring of each module/course provides a holistic view of the programme and clearly identifies the areas of strengths and improvements.

Design of the programme

1. The programme is an integrated BSc-MSc Data Science programme that provides a solid foundation on data science along with theoretical knowledge and technical skill to undertake wide range of data science related application and research work. For this benchmarking exercise we only focused on the Master's section of the programme.
2. The Master's part of the programme focuses on specialised skill in knowledge management, knowledge abstraction with advanced machine learning algorithms and knowledge representation with common visualisation tools.
3. The master's programme is 48 credits.

Mode of Delivery

4. All courses are delivered in English and have theory, practice and self-study credit hours. Except for the thesis, self-study hours are 2 hours/week/credit.
5. Only few courses have some degree of soft skill development elements. Acquiring such soft skills is pivotal in the ICT sector. From the syllabus it is evident that most of the courses are exam based and courses have projects through which students experience

teamwork. Setting marks (10-20%) to these informal projects would formalise the work/collaboration undertaken in these projects.

6. The courses are delivered through a mixture of lectures, discussions, and case studies. To further increase student engagement, flipped classroom or peer assessment approaches can be introduced.
7. For practical sessions, information regarding available resources is necessary to evaluate the effectiveness of the practical sessions.

Learning and Teaching

8. The programme in general focuses on machine learning algorithms and data management. From the syllabus it is deemed that the data management pipeline, data collection through to data visualisation are covered. The data management and visualisation course cover almost all steps of the data management pipeline. This is based on foundational knowledge students gained from their UG programme.
9. The programme is for students who already have strong scripting as well as object-oriented programming skills from their UG courses. Continuation and/or upskilling of programming skills is primarily self-directed. Signposting of web-based resources in this regard would be beneficial.
10. For algorithm focused courses, the algorithm evaluation aspect needs more focus even though it is covered during the UG programme.
11. Although critical understanding about code sharing (github etc.) is covered during the UG level, it would be beneficial for the students if examples/case studies of forking, code modulation type of advanced topics were covered during this PG level. Also, basic understanding about different types of code sharing licences (e.g. GNU, GPL) were covered.
12. The syllabus did not provide information regarding reference materials like books or online materials used in teaching.
13. Communication of the results and/or artifacts are an important part of any data science programme. Including a presentation part with most of the model enables students to develop their communication skills.
14. Courses focusing on algorithms (knowledge abstraction) can be improved by reducing the amount of theoretical knowledge and incorporating more case studies and/or application of the algorithm in real life scenarios. In these approaches students would gain more experience of application of the appropriate algorithm in the right context.

Assessment and Feedback

15. No information was available with regard to assessment or exam samples.
16. No information was provided as to how student feedback is captured, evaluated and utilised for the improvement of the courses.

Conclusion and Recommendations

The programme is designed for students with a strong programming background who have completed their undergraduate in this institution. Such prerequisites can limit student recruitment numbers as well as scope of the programme itself.

17. Teaching modality
    a. More discussion-based teaching approach including flip classroom type teaching model can be introduced to increase student engagement and self-directed study.
    b. Project-based learning approach can be implemented to get more knowledge about different real-life projects, their short comings etc.
18. Teaching content
    a. Low code/No code based programming are becoming popular (10.3390/electronics10101192) in universities with the rise of online education and as a result of COVID-19. Adaptation with new trend will help students to develop new applications/algorithms more easily. This impacts not only skill development but confidence also.
    b. In this regard API based programming e.g. GPT-3 like language model (from OpenAI etc.) to any software/app would benefit students with high quality trained dataset/model integration.
    c. Analysis of real-life data from different domains (finance, healthcare, social media etc.) is essential to get understanding about different data sources and types.
    d. Engagement with stakeholders and requirement capture is pivotal. Therefore, with different types of programming/machine learning courses these aspects need to be included.
    e. Use of online content/course can introduce students to new topics and choice of learning sources (in contrast to recommended book). This diversity of content and modality of delivery not only helps students to be in line with current trends but also initiate peer learning.
    f. Skill development on code sharing (through github etc.) and open licence needs to added to the course curriculum along with collaborative code development (e.g. Google Colab, AWS).
    g. Cyber security aspect of software/algorithm design and development can be improved by incorporating some topics from cyber security, particularly access control to source code and sensitive data (e.g. health data).
    h. Basic understanding of how to protect intellectual property rights related to algorithms and the process of protecting these rights through third party.
    i. Critical understanding of research methods in higher education and steps involved from idea generation through to publication and/or application can be incorporated.
    j. More emphasis can be given to data communication. This can be given by group presentations, peer assessments etc.
    k. Basic knowledge of social media-based profile creation e.g. LinkedIn profile that will facilitate future job prospects.
19. Assessment
    a. More emphasis on project-based assessments (instead of exams) would help students to get experience of team work and other aspects of project management.

CRITERIA	Philosophy	Statistical Machine Learning	Advanced Deep Learning	Data management and Visualization	Graph Analytics for Big Data	Big Data Integration and Processing	Semantics Web	Business Intelligence	Data Entrepreneurship in Action	Seminar 01	Seminar 02	Master Thesis
PROGRAMMING												
General knowledge on different software/algorithm, their application, strenghts and limitations.												
Knowledge and skill to undertake requirement analysis. Ability to collect and summarise the requirements. Also understand the market demand/trend in context of software/algorithm development.												
Ability to work in a team environment and understanding of the importance of communication in a multi disciplinary/institutional team environment.												
Demonstrate the knowledge and skill of project management. Particularly the differentiation and applicability of different types of project management approaches and the relevance to the project in hand.												
Demonstrate critical understanding of the software/algorithm design approaches and availability of design tools												
Demonstrate critical understanding of the software/algorithm development approaches and availability of development tools/languages.												
Critical knowledge about the importance of software/algorithm evaluation/testing and the steps of releasing.												
Critical knowledge about the importance of software/algorithm security related issues.												
Demonstrate critical knowledge on both object oriented and scripting languages.												
Ability to code in distributed cloud as well as stand-alone environement using wide range of programming languages.												
Good knowledge on privacy, security issues related to software/hardware/algorithm.												
Critical understanding of code sharing, archiving and protection through cloud-based repository platforms.												
Good knowledge about different cloud based computing platforms, e.g. AWS												
Understanding about opensource/stand-alone/cloud-based software/hardware/algorithm.												
Critical knowledge on computing performance and how to address performance related challenges.												
KNOWLEDGE MANAGEMENT												
Good understanding about big data and their impact on business and society.	60			80	95	80	75					
Critical knowledge about different steps of data management pipeline – from collection to data analytics	65	75	95		92		75					
Critical understanding about different data types, sources of data types and their strenghts and limitations.			65	70	98	98						
Critical understanding about different data collection processes and resources along with regulations associated to the processes including ethics and permission requirements.			85		94		85					
Demonstrate the knowledge on data privacy and it differs in different domains – manufacturing, medical etc.												
Critical understanding about data quality, particularly standard process of measuring data quality.												
Demonstrate the appreciation towards the importance of data cleaning and knowledge on different steps, tools used for data cleaning.							85					
Understanding about the data validation process and its importance towards data management.					75							
Good knowledge about different data warehousing techniques and technical details for implementing different types of databases.				95			94					
Updated knowledge about data warehousing for both structured and unstructured data.				85			94					
Critical knowledge about the approach to query the data as per the requirements of the users.				95		98						
Good understanding about the process of data transformation in relation to the users requirements.				94		90	85					
Provide detailed knowledge and skill on different data security related issues and how to identify different types of threats.												
Explain different types of data protection measures that can be taken to safeguard data breach.												
Understand institutional policies related to data security and privacy and tools required to implement these policies.												
Good understanding about intellectual property rights and governance laws.												
KNOWLEDGE ABSTRACTION												
Demonstrate good knowledge of algebra/stat and calculus that enables students to refresh their high school math knowledge.		80	98									
Provide detailed understanding of distribution and probability statistics.		95										
Critical understanding on intelligent agent and different types of logic representations.		82	95									
Different algorithms (structured, unstructured, adaptive) – their applications and refinement of algorithm parameters.		65	95					95				
Critical knowledge on algorithm testing, evaluation and optimisation.			94									
Good understanding of ethics and policy related to algorithm development and deployment.			85									
Up to date knowledge on state of the art algorithms implemented by different stakeholders.			95					85				
Basic understanding on how to protect intellectual property rights related to algorithms and the process of protecting these rights through third party.			65									
Provide detailed understanding of algorithms used for image analysis and object detection from including timelapse and multimodal images.												
KNOWLEDGE REPRESENTATION/COMMUNICATION												
Evaluate and apply data visualisation grammar and principle to the whole of the visualisation process and the resulting presentations.				50								
Evaluate the capabilities of different visualisation tools and programming languages, both proprietary and open-source, to support the discovery and display of critical and valuable answers hidden in small, medium and large data.				85								
Implement analysis and visualisation techniques using realistic data sources from disparate disciplines and using the most appropriate visualisation tools in order to identify the valuable questions and to develop well justified, actionable answers.				85	62							
Good understanding of processes to know the audience in terms of their requirements, domain knowledge etc.				71			80					
Differentiate between exploratory and explanatory data visualisation and or knowledge representation.							90					
RESEARCH/SOFT SKILLS												
Understanding of core aspects of philosophy and its sub discipline.	98									85		
Explore and express different ideas of philosophy and their relevance to research in higher education.	97										85	
Introduction to the main ideas of logic and methods of recognising relevant information to construct persuasive arguments.											85	
Develop analytical, argumentative problem-solving skills.	80									75		
Develop critical thinking and problem-solving skill in a collaborative manner.	75						80	70		70		
Critical understanding of research methods in higher education and steps involved from idea generation through to publication and or application.							70				75	
Critical understanding of project management and role of different team members at different levels along with the interplay.	85			84			70	70		80		
Knowledge about different administrative and operational tasks required for successful project management.	80			84				80		71		
Basic understanding of activities related to dissemination of research outcome to wider audience.				85			85					
Basic knowledge of social media-based profile creation e.g. LinkedIn profile that will facilitate future job prospects.												