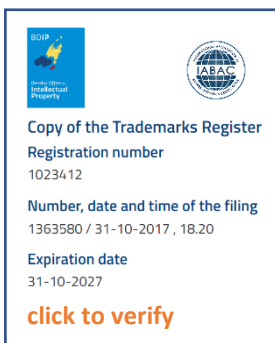




Certified Data Science Developer (CDSD)

CODE: DS2020

Syllabus and Examination



The International Association for Data Science Certification (IABAC®) is a globally recognized professional association dedicated to growing and enhancing the field of applied Data Science and Business Analytics.

IABAC™ founding principles are based on Edison Data Science Framework (EDSF), a European commission initiative, with the goal of aligning data science skills to industry requirements. IABAC was founded in the year 2017 as an association and, in the year 2018, registered as The Netherlands B.V (equivalent of English Private Limited).

KVK-nummer (Commercial Registration Number): 73234893 (click to verify)

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN_CDSD_2020_Syllabus_V1.0

CONTENTS

| | | |
|-----|--------------------------------|---|
| 1 | Introduction | 3 |
| 2 | Course Syllabus | 3 |
| 2.1 | Software Engineering Practices | 3 |
| 2.2 | Object Oriented programming | 3 |
| 2.3 | Web Development | 3 |
| 2.4 | ETL Pipelines | 3 |
| 2.5 | Natural Language Processing | 4 |
| 2.6 | Machine Learning Pipelines | 4 |
| 2.7 | Experiment Design | 4 |
| 2.8 | Deployment and A/B Testing | 4 |
| 3 | Examination | 5 |
| 3.1 | Pre-requisite qualifications | 5 |
| 3.2 | Material permitted | 5 |
| 3.3 | Exam Duration and Format | 5 |
| 3.4 | Exam mode | 5 |
| 3.5 | Pass Criteria | 5 |
| 3.6 | Results Timeline | 5 |
| 3.7 | Certificate Issuance | 5 |
| 4 | IABAC® Knowledge Areas Mapping | 6 |
| 5 | Bloom's Taxonomy Reference | 8 |

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN_CDSD_2020_Syllabus_V1.0

1 INTRODUCTION

This document is intended to provide information on Certified Data Science Developer (CDS-D-DS2020) certification for registered training providers to structure the course curriculum as per IABAC syllabus guidelines and for individuals, who are preparing for IABAC CDS-D certification exam.

2 COURSE SYLLABUS

2.1 SOFTWARE ENGINEERING PRACTICES

- Write clean, modular, and well-documented code
- Refactor code for efficiency
- Create unit tests to test programs
- Write useful programs in multiple scripts
- Track actions and results of processes with logging
- Conduct and receive code reviews

2.2 OBJECT ORIENTED PROGRAMMING

- Understand when to use object oriented programming
- Build and use classes
- Understand magic methods
- Write programs that include multiple classes, and follow good code structure
- Learn how large, modular Python packages, such as pandas and scikit-learn, use object oriented programming
- Portfolio Exercise: Build your own Python package

2.3 WEB DEVELOPMENT

- Learn about the components of a web app
- Build a web application that uses Flask, Plotly, and the Bootstrap framework
- Portfolio Exercise: Build a data dashboard using a dataset of your choice and deploy it to a web application

2.4 ETL PIPELINES

- Understand what ETL pipelines are
- Access and combine data from CSV, JSON, logs, APIs, and databases
- Standardize encodings and columns
- Normalize data and create dummy variables
- Handle outliers, missing values, and duplicated data
- Engineer new features by running calculations
- Build a SQLite database to store cleaned data

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDS-D® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN_CDS-D_2020_Syllabus_V1.0

2.5 NATURAL LANGUAGE PROCESSING

- Prepare text data for analysis with tokenization, lemmatization, and removing stop words
- Use scikit-learn to transform and vectorize text data
- Build features with bag of words and tfidf
- Extract features with tools such as named entity recognition and part of speech tagging
- Build an NLP model to perform sentiment analysis

2.6 MACHINE LEARNING PIPELINES

- Understand the advantages of using machine learning pipelines to streamline the data preparation and modelling process
- Chain data transformations and an estimator with scikit-learn Pipeline
- Use feature unions to perform steps in parallel and create more complex workflows

2.7 EXPERIMENT DESIGN

- Understand how to set up an experiment, and the ideas associated with experiments vs. observational studies
- Defining control and test conditions
- Choosing control and testing groups

2.8 DEPLOYMENT AND A/B TESTING

- How it works and its limitations
- Sources of Bias: Novelty and Recency Effects
- Multiple Comparison Techniques (FDR, Bonferroni, Tukey)
- Portfolio Exercise: Using a technical screener from Starbucks to analyse the results of an experiment and write up your findings

3 EXAMINATION

3.1 PRE-REQUISITE QUALIFICATIONS

1. Essential knowledge in
 - a. Python, R or any popular programming languages.
 - b. Machine Learning
2. Training: Though formal training is not mandatory; it is recommended to attend IABAC® registered course for Certified Data Science Developer through Registered Education Partners.

3.2 MATERIAL PERMITTED

1. The examination is an 'open book'
2. Candidates can refer to any material

3.3 EXAM DURATION AND FORMAT

1. Exam format is the project submission
2. The assessment duration is 8 hours
3. The project is graded for conceptual understanding and NLP model performance

3.4 EXAM MODE

1. Project needs to be submitted at IABAC Project Submit page, as per exam guidelines
2. Any copied work, ideas, concepts or a piece of text need to be marked with reference as per IABAC project plagiarism guidelines

3.5 PASS CRITERIA

1. The candidate needs to score assessment grade A+, A, B+, B, C+, C as a PASS Criteria
2. The candidate will be awarded grade F in case of failing to meet the pass criteria
3. The results will be declared after validation of the project as per guidelines

3.6 RESULTS TIMELINE

1. The preliminary results are usually released within **9 days** of the exam date
2. The official results are usually released within **15 days** from the exam date

3.7 CERTIFICATE ISSUANCE

1. IABAC® e-certificate will be issued through the candidate registered email
2. The e-certificate is digital verifiable at <https://www.iabac.org/verify-certificate>
3. The candidate has license to share digital certificate validation in professional networking portals such as www.linkedin.com
4. The candidate has a license to print physical copy (hardcopy) of the certificate

4 IABAC® KNOWLEDGE AREAS MAPPING

| Knowledge Area | Syllabus Details | Bloom's Index |
|---|---|---------------|
| <p>KAG1-DSDA: Data Analytics group including Machine Learning, Statistical Methods, and Business Analytics</p> | <ul style="list-style-type: none"> ● Case Study on Statistical Analysis ● Curating the Data and performing, Discrete Mathematics, Probabilistic Reasoning ● Statistical Methods, including Descriptive Statistics, Exploratory Data Analysis (EDA) and Confirmatory Data Analysis (CDA) ● Case Study & Creating Machine Learning Model ● With detailed implementation of algorithms: Artificial Intelligence, Natural Language Processing ● Knowledge Representation and Reasoning ● Data Mining and knowledge discovery ● Text analysis, Data Mining, Text Analytics including Statistical, Linguistic, and Structural Techniques to analyse Structured and Unstructured data ● Creating Predictive Forecasting Models ● Decision Analysis and Decision Support Systems ● Data Mining | 6 |
| <p>KAG2-DSENG: Data Science Engineering group including Software and Infrastructure Engineering</p> | <ul style="list-style-type: none"> ● Set Up Infrastructure and Big Data Applications ● Computer Networks for high-performance computing and Big Data Infrastructure ● Cloud Enabled Applications development ● Modelling and Simulation ● Modelling and Simulation Theory and Techniques (general and domain oriented) ● Large Scale Modelling and Simulation Systems ● Set up Big Data (Data Science) Applications Design ● Programming Languages for Big Data Analytics: R, Python, others ● Models and Languages for complex interlinked Data Presentation and Visualisation | 5 |
| <p>KAG3-DSDM: Data Management group including Data Curation, Preservation and Data Infrastructure</p> | <ul style="list-style-type: none"> ● Creating Database Models and Data Curation ● Data Modelling, Databases and Database Management Systems, Data Models and Query Languages, Database Administration ● Set up Data Management and Enterprise Data Infrastructure ● Data management, including Reference and Master Data, Data Warehousing and Business Intelligence, Data storage and Operations ● Data Archives/Storage Compliance and Certification Metadata, Linked data, Provenance | 4 |

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDS® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN_CDSD_2020_Syllabus_V1.0

| | | |
|---|--|---|
| | <ul style="list-style-type: none"> ● Data Infrastructure, Data Management and Organisation Research Data Infrastructure, Open Science, Open Data, Open Access, Data Infrastructure Compliance and Certification, Ethical Principle and Data Privacy | |
| KAG4-DSRM: Scientific and Research Methods group | <ul style="list-style-type: none"> ● Scientific/Research Methods ● Research Methodology, Paradigms and Research Cycle, Modelling and Experiment Planning ● Data Selection and Quality Evaluation ● Use Case Analysis: Research Infrastructures and Projects Research Data Management plan and Ethical Issues | 6 |
| KAG5-DSBPM: Business Process Management group | <ul style="list-style-type: none"> ● Business Process Management ● Business Processes and Operations, Project Scope and Risk Management ● Business Analysis - Organisation and Management ● Business Analysis - Planning and Monitoring ● Requirements Analysis and Design Definition ● Requirements Life Cycle Management (from inception to retirement) Solution Evaluation and Improvements Recommendation ● Business analysis and Enterprise Organisation ● Agile Data Driven Methodologies, Processes and Enterprises ● Use Case Analysis: Business and Industry | 4 |
| KAG6-DSDK: Data Science Domain Knowledge group includes domain specific knowledge | <ul style="list-style-type: none"> ● Applied Data Science use cases in Domains, HR, Retail, Fraud Analytics, Finance Trends, Health Care, Infrastructure Management | 2 |

5 BLOOM'S TAXONOMY REFERENCE

| Bloom's Learning Index | Description |
|------------------------|--|
| 1 | Remembering: Recall or retrieve previous learned information. |
| 2 | Understanding: Comprehending the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words. |
| 3 | Applying: Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the workplace. |
| 4 | Analysing: Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences. |
| 5 | Evaluating: Make judgments about the value of ideas or materials. |
| 6 | Creating: Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure. |

Copyright © IABAC 2017-2020.

IABAC®, IABAC logo®, DSF®, CDS®, CAIE®, CNLPE®, CCVE®, CDSHR®, CDSFIN®, CVAE®, CDLE®, CDSHC®, CDSMKT®, CDS®, DSCMGR®, AICE®, CDSD® and CMLE® are registered trademarks of IABAC BV, The Netherlands. Registration Number (KVK-nummer) :73234893. Reproduction of any part of this material requires written permission of IABAC BV.

All rights reserved. EN_CDSD_2020_Syllabus_V1.0