FDP INSTITUTE® by CAIA

INTRO TO THE FDP PROGRAM

April 10 - April 24, 2023

Create your FDP profile to gain access to the official FDP Candidate Study Guide

Brought to you by: CAIA ASSOCIATION®

"If, after reviewing this document, you conclude that the FDP Charter could be an essential step in your continued education and can help you gain a competitive edge in the data-driven finance industry, visit the FDP website, and create your FDP profile, and receive a copy of the complete study guide. Once you have decided to pursue the FDP Charter, register for the exam, and gain access to the sample exam and additional learning tools."
– *Hossein Kazemi*

INTRODUCTION TO THE FINANCIAL DATA PROFESSIONAL (FDP) PROGRAM

The FDP Institute® was founded by the Chartered Alternative Investment Analyst Association® to create the FDP® charter. It is the only globally recognized professional designation in financial data science, an increasingly important part of the financial services industry.

The digital revolution has disrupted the financial industry in recent years. It is critical for industry practitioners to have a working knowledge of the increasingly important roles played by big data, machine learning, and artificial intelligence in the financial industry. The FDP Institute has designed this self-study program to provide finance professionals with an efficient path to learn about financial data science's essential aspects. The FDP curriculum introduces Candidates to the central concepts of machine learning and big data, including ethical and privacy issues and their roles in various financial industry segments. Candidates will earn their FDP Charter once they pass the FDP exam and complete the online class requirements, which can be done before or after the FDP exam.

The university faculty and industry practitioners who have helped create the FDP Charter program bring years of experience in the financial services industry. Consequently, the curriculum is consistent with recent advances in data science applications to the financial industry.

Passing the FDP examination is an important accomplishment that will require significant preparation. All Candidates will need to study and become familiar with the FDP curriculum material to develop the knowledge and skills necessary to be successful on examination day.

This study guide is organized to facilitate quick learning and easy retention. Each topic is structured around learning objectives (LOs) that define the content to be tested on the exam. The learning objectives are an important way for Candidates to organize their studies as they form the basis for examination questions. All learning objectives reflect the FDP curriculum content, and all exam questions are written to address the learning objectives directly. A Candidate who can meet all learning objectives in the study guide should be well prepared for the exam. For these reasons, we believe that the FDP Institute has built a rigorous program with high standards while also maintaining an awareness of the value of Candidates' time.

Candidates for the FDP Charter must complete the FDP exam and the online requirements. Since the FDP program is designed for finance professionals, it is assumed that Candidates understand the central concepts of financial economics. Candidates are expected to have knowledge of various financial institutions and instruments' roles and characteristics and the financial models these institutions employ to value the instruments and measure their risk. These concepts are covered in CAIA®, CFA®, and FRM® exams and dedicated undergraduate or graduate courses covering financial markets, investments, and risk management.

FDP PROGRAM: ONLINE REQUIREMENTS

FDP Candidates must complete the following two components with a passing score before obtaining their FDP Charters.

- **FDP exam.**
- **Online classes covering Python or R programming.**

The FDP exam will not contain any coding questions. However, FDP Candidates must demonstrate some Python or R programming language knowledge before they obtain their FDP charter. FDP Candidates who do not have a verifiable academic background in Python or R programming can demonstrate their understanding of these languages by completing the online classes listed below. The online classes can be completed before or after a Candidate completes the FDP exam.

The FDP Institute recommends DataCamp's introductory online courses (https://www.datacamp.com) for completing the FDP Charter's requirement. The list of online classes offered by DataCamp appears on the FDP Institute's website and in this Study Guide.

The approved online classes offered by DataCamp are available as soon as a Candidate creates an account on DataCamp. Limited free access to their classes is available. DataCamp courses assume no prior knowledge of Python or R.

The Candidate Handbook, which can be found on the FDP website, describes the procedure for sending proof of successful completion of the online classes to the FDP Institute.

The following classes are recommended to complete the FDP Charter's programming knowledge requirement. These recommendations assume that a Candidate has no prior Python or R programming knowledge. If a Candidate has some knowledge of these languages, the Candidate is encouraged to take more advanced Python or R programming classes at DataCamp. If a Candidate has a verifiable academic background in Python or R, the Candidate can seek an exemption from the online classes. The approval of prior academic knowledge in Python or R programming is at the sole discretion of the FDP Institute. Please contact the FDP institute to learn more about this option.

FDP Candidates can satisfy the coding requirement of the FDP program by completing two (2) Python or two (2) R classes offered by DataCamp. DataCamp classes can be accessed through its website at https://www.datacamp.com/. Candidates are responsible for the cost of classes offered at DataCamp. Candidates are encouraged to take advantage of the limited free access offered by DataCamp to evaluate its teaching method. The classes listed below are short, and depending on the Candidate's background, each can be completed in four (4) to six (6) hours. Besides the classes mentioned below, Candidates have the option to complete any two (2) classes in either Python or R.

DataCamp: Python

1. **Introduction to Python**
   https://www.datacamp.com/courses/intro-to-python-for-data-science
2. **Intermediate Python**
   https://www.datacamp.com/courses/intermediate-python-for-data-science

DataCamp: R

1. **Introduction to R**
   https://www.datacamp.com/courses/free-introduction-to-r
2. **Intermediate R**
   https://www.datacamp.com/courses/intermediate-r

FDP EXAMINATION

The FDP examination, administered twice annually, is a four-hour computer-administered examination offered at test centers worldwide. The FDP examination consists of eighty (80) multiple choice questions weighted as 75% of the total points and two (2) to four (4) constructed response questions (multi-part essay type) weighted as 25% of the total points. The FDP exam will not contain any Python or R programming questions.

The FDP program is organized to facilitate quick learning and easy retention based on the study guide. Each topic is structured around learning objectives and keywords that define the content to be tested on the exam. The learning objectives and keywords are an important way for Candidates to organize their studies as they form the basis for examination questions. All learning objectives reflect the FDP curriculum content, and all examination questions are written to address the learning objectives directly.

For additional information about the FDP examination, please see the Candidate Handbook, which can be found on the FDP Institute website.

SAMPLE EXAM AND PRACTICE QUESTIONS

A sample exam is available for the Candidates to assist with their study efforts. This sample exam contains eighty (80) multiple choice questions and several multi-part constructed response questions. There is also a set of practice questions available to Candidates. The set of practice questions contains more questions than the number of questions in the actual exam. In addition to helping the Candidates learn the topic material, the questions can also help the Candidates get familiar with the style and conventions used. An example is a simplifying convention of using the natural logarithm to solve any problem requiring the calculation of logarithm on the exam. This convention is announced at the beginning of the sample exam and on the actual exam. This convention is also described in the Candidate Handbook.

OTHER STUDY TOOLS AND RESOURCES

In addition to this Study Guide and the Candidate Handbook, the FDP Institute website directs Candidates to the readings covered in the curriculum. The readings are detailed below by topic area and include textbooks, often used across topics, and individual articles that are usually topic-specific. Both types of readings can be purchased from Amazon or the publisher, and whenever possible, they are posted on the FDP Institute website.

Page Number References for Keywords

For Candidates' convenience, six (6) articles published by PMR Journals are provided in one collection titled "Alternative Data and Machine Learning in the Financial Industry: *A Collection of Articles from PMR Journals*." It is available at a discounted price of $99 for registered Candidates. There are two sets of page numbers in this collection: one corresponds to the

collection's table of contents. In contrast, the other corresponds to each article's page number in the original journal. The page numbers appearing next to the keywords refer to the page numbers as they appeared in the original article.

Note: *Check if your employer has a subscription to Portfolio Management Research (PMR) as this might provide free access to the six (6) PMR readings.*

THE FDP CURRICULUM: OUTLINE

Candidates for the FDP Charter will have to enroll in the self-study program created by the FDP Institute and follow its carefully designed Study Guide. To become an FDP Charterholder, Candidates must pass the FDP exam and submit their certificates of completion for the required online classes. The rest of this document discusses the FDP curriculum. Below is the outline of the curriculum:

Topics		Approximate Weight %
1.	Introduction to Data Science	5-12
2.	Linear and Logistic Regression, Support Vector Machines, Regularization, and Time Series	10-15
3.	Decision Trees, Supervised Segmentation, and Ensemble Methods	10-15
4.	Classification, Clustering, and Naïve Bayes	5-12
5.	Neural Networks and Reinforcement Learning	5-12
6.	Performance Evaluation, Back-Testing, and False Discoveries	5-12
7.	Text Mining	5-12
8.	Ethical and Privacy Issues	5-12
9.	Fintech Applications	25-40

THE FDP CURRICULUM: COMPLETE READING LIST

The following is a complete list of all curriculum readings for the April 2023 FDP exam.

Two (2) out of the three (3) books and six (6) articles from the Alternative Data and Machine Learning in the Financial Industry: A Collection of Articles from PMR Journals must be purchased. One of the books is available free of charge from the authors' website.

The Topics in Financial Data Science articles are available free of charge. Candidates may access all materials from the authors' or publishers' websites or via the FDP website. Please use the web link below to access all curriculum materials.

https://fdpinstitute.org/Curriculum-Materials

A. Books

1. Provost, F., and T. Fawcett (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media Inc., 1st Edition. Chapters 1-10. Candidates should visit the book's errata page.

2. John C. Hull (2021). Machine Learning in Business: An Introduction to the World of Data Science. Independently Published by GFS Press, 3rd Edition. Chapters 1-11.

3. James, G., D. Witten, T. Hastie, and R. Tibshirani (2021). An Introduction to Statistical Learning: With Applications in R. Springer, 2nd Edition. Chapters 1, 2 (sections 1,2), Chapter 3 (sections 1-3), Chapter 6 (sections 1-3), Chapter 8 (sections 1,2). Candidates should visit the book's errata page.

B. Alternative Data and Machine Learning in the Financial Industry: *A Collection of Articles From PMR Journals*

1. Das S., M. Donini, J. Gelman, K. Haas, M. Hardt, J. Katzman, K. Kenthapadi, P. Larroy, P. Yilmaz, and M. B. Zafar (2021). Fairness Measures for Machine Learning in Finance. The Journal of Financial Data Science, 3(4): 33-64. Reading 8.3

2. Ekster, G. and Kolm, P. N. (2021). Alternative Data in Investment Management: Usage, Challenges, and Valuation. The Journal of Financial Data Science, 3(4): 10-32. Reading 9.1

3. Åstebro, T. (2021). An Inside Peek at AI Use in Private Equity. The Journal of Financial Data Science, 3(3): 97-107. Reading 9.5

4. Li, Y., Z. Simon, and D. Turkington. (2022). Investable and Interpretable Machine Learning for Equities. The Journal of Financial Data Science, 4(1): 54-74. Reading 9.6

5. López de Prado, M. (2018). The 10 Reasons Most Machine Learning Funds Fail. The Journal of Portfolio Management, 44 (6): 120-133. Reading 9.7

6. Harvey, C. R., and Y. Liu. (2014). Evaluating Trading Strategies. The Journal of Portfolio Management, 40(5): 108-118. Reading 9.8

C. Topics in Financial Data Science

1. Das, S., and H. Kazemi (2022). Time Series: A Financial Perspective. The FDP Institute. This reading is provided by the FDP Institute free of charge. Reading 2.4
2. Colquhoun, D. (2014). An Investigation of the False Discovery Rate and the Misinterpretation of p-values. Royal Society Open Science, 1 (3): 1-16. Reading 6.3
3. Zhao, F. (2017). Natural Language Processing – Part I: Primer. S&P Global: Market Intelligence. Reading 7.3
4. Institute of International Finance (May 2019). Machine Learning Thematic Series Part II: Bias and Ethical Implications. Reading 8.2
5. OECD (2021). Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers. Reading 9.2
6. Financial Stability Board (2017). Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications. Reading 9.3
7. Zappa, D., M. Borrelli, G.P. Clemente, and N. Savelli. (2021). Text Mining in Insurance: From Unstructured Data to Meaning. Variance Journal, 14(1). Reading 9.4
8. Amler, H., L. Eckey, S. Faust, M. Kaiser, P. Sandner, and B. Schlosser. (2021). DeFi-ning DeFi: Challenges & Pathway. Reading 9.9
9. Nadini, M., L. Alessandretti, F. D. Giacinto, M. Martino, L. M. Aiello, and A. Baronchelli. (2021). Mapping the NFT Revolution: Market Trends, Trade Networks, and Visual Features. Reading 9.10

ACTION WORDS

In each learning objective that appears below, action words are used to direct Candidates' focus of study. The following table contains the list of all action words used in this study guide and their definitions.

Action Word	Meaning
Analyze	To examine methodically and in detail the constitution or structure of the information or concept covered by the LO. This is similar to offering an explanation and an interpretation. It is used chiefly to explain relationships.
Apply	To bring into action, use or employ a concept or a mathematical relationship (equation). If the LO is about an equation, the Candidate must memorize the equation (see Recognize below).
Calculate	It is similar to Apply but is related to a mathematical concept and equation. If the LO is about an equation, the Candidate must memorize the equation (see Recognize below).
Compare	Estimate, measure, or note the similarity or dissimilarity between two concepts or definitions.
Contrast	Similar to Compare. In this case, the emphasis is on the differences.
Define	A general action word. The Candidate is expected to state or describe precisely the nature, scope, or meaning of a concept. If the LO is about a mathematical equation, the Candidate is not expected to memorize the exact equation but is expected to describe its essential aspects.
Describe	Similar to Define. The Candidate should give an account in words of concepts covered by the LO. The Candidate is expected to cover all the relevant characteristics, qualities, or relationships the LO covers. If the LO is about a mathematical equation, the Candidate is not expected to memorize the exact equation but is expected to describe its essential aspects.
Discuss	It is similar to Analyze. To provide details about a key word or concept. If the LO is about an equation, the Candidate does not need to memorize the equation, but must know its uses and applications.

Action Word	Meaning
Explain	Similar to Describe. The Candidate is expected to clarify an idea, problem, or relationship by describing it in more detail or revealing relevant facts or ideas. If the LO is about a mathematical equation, the Candidate is not expected to memorize the exact equation but is expected to describe its essential aspects.
Identify	The Candidate is expected to recognize or establish as being a particular model, concept, or relationship. The LO may expect the Candidate to verify a given relationship or recognize a particular pattern. If the LO is about a mathematical equation, the Candidate is not expected to memorize the exact equation but is expected to describe its essential aspects.
Interpret	Similar to Explain. The Candidate is expected to give or provide an explanation for the observed pattern, relationship, or information. If the LO is about a mathematical equation, the Candidate is not expected to memorize the exact equation but is expected to describe its essential aspects.
List	The Candidate is expected to learn the list of related items or concepts covered by the LO. The Candidate is not expected to describe the members of the list. A separate LO may state that some or all of the members of the list must be explained.
Recognize	The Candidate is expected to identify an equation or model that has appeared in the readings. The Candidate is not expected to memorize the equation. The Candidate is expected to apply the equation or make some calculations using the equation provided on the exam.

LEARNING OBJECTIVES

Topic 1. Introduction to Data Science

Reading 1.1 Provost, F. and T. Fawcett (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media Inc., 1st Edition. Chapters 1 and 2.

Keywords

Data mining (p. 2)
Data Science (p. 2)
Data-driven decision making (p.5)
Big data (p. 8)
Classification (p. 20)
Regression (p. 21)
Similarity matching (p. 21)
Clustering (p. 21)
Co-occurrence grouping (p. 21)
Profiling (p. 22)
Link prediction (p. 22)
Data reduction (p. 22)
Causal modeling (p. 23)
Unsupervised learning (p. 24)
Supervised learning (p. 24)
Leak (p. 30)

Learning Objectives

Demonstrate proficiency in the following areas:

1.1.1 Data Analytic Thinking (Ch. 1)

For example:

A. List examples of data mining in finance, marketing, and customer relationship management.

B. Contrast data science with data mining.

C. Describe the two types of decisions that can benefit from data-driven decision making.

D. Describe the reason for the early adoption of automated decision making by finance and telecommunications industries.

E. Contrast data science with data processing.

F. Describe the usage of big data.

G. Explain why both appropriate data and data scientists are required to extract useful knowledge from data.

H. Explain why it is necessary to understand data science even if someone is not going to use data science directly.

I. List and describe the four fundamental concepts of data science.

1.1.2 Business Problems and Data Science Solutions (Ch. 2)

For example:

A. Describe when each type of data mining algorithm, such as classification, regression, similarity matching, clustering, co-occurrence grouping, profiling, link-prediction, data reduction, and causal modeling, should be used.

B. Explain the differences between regression and classification.

C. Contrast supervised learning with unsupervised learning.

D. List the algorithms that can be used for supervised and unsupervised learning.

E. Contrast data mining with the use of data mining results.

F. List and describe the steps used in Cross Industry Standard Process for Data Mining (CRISP-DM).

G. Explain the reason for having an iterative process involved in CRISP-DM.

H. Describe the characteristics of credit card and Medicare fraud.

I. List the reasons for deploying the data mining system itself rather than the models produced by a data mining system.

Reading 1.2 John C. Hull (2021). Machine Learning in Business: An Introduction to the World of Data Science. Independently Published by GFS Press, 3rd Edition. Chapter 1.

Keywords

Machine learning (p. 1)
Artificial intelligence (p. 1)
Features (p. 6)
Labels (p. 6)
Semi-supervised learning (p. 7)
Training set (p. 8)
Root-mean squared error (p. 9)
Bias-variance tradeoff (p. 15)
Numerical feature (p. 16)
Categorical feature (p. 16)
Outliers (p. 17)
Bayes' Theorem (p. 18)

Learning Objectives

Demonstrate proficiency in the areas of:

1.2.1 Introduction

For example:

A. List the advantages for the society of replacing human decision-making with machines.

B. Contrast machine learning to statistics.

C. Describe a training set, validation set, and test set.

D. Define instances.

E. Analyze the relationship between model error and model complexity.

F. Define bias and variance in the context of machine learning.

G. List the usage of the training set, validation set, and test set.

H. List and explain different data cleaning issues.

I. List the type of models that are least and most affected by outliers.

J. Calculate conditional probability using Bayes' Theorem.

Reading 1.3 James, G., D. Witten, T. Hastie, and R. Tibshirani. An Introduction to Statistical Learning: With Applications in R. Springer, 2nd Edition. Chapters 1, 2.1, and 2.2.

Keywords

Statistical learning (p. 1)
Flexible (p. 22)
Thin plate spline (p. 23)
Classification problems (p. 28)
Quantitative variables (p. 28)
Qualitative response (p. 28)
Binary response (p. 28)
Predictors (p. 29)
Mean squared error (MSE) (p. 29)
Test MSE (p. 30)
Test data (p. 30)
Training MSE (p. 30)
Degrees of freedom (p. 31)
Expected test MSE (p. 34)
Bias (p. 35)
Error rate (p.37)
Indicator variable (p. 37)
Training error (p. 37)
Test error (p. 37)
Bayes classifier (p. 37)
Conditional probability (p. 37)
Bayes decision boundary (p. 38)
Bayes error rate (p. 38)
K-nearest neighbors (p. 39)

Learning Objectives

Demonstrate proficiency in the areas of:

1.3.1 Organization and Resources of the Book a An Introduction to Statistical Learning: With Applications in R (Ch. 1)

This chapter is assigned to facilitate your studies, but no exam questions will be drawn from this chapter.

1.3.2 Statistical Learning (Ch. 2.1)

For example:

A. Explain why we estimate a function with data, including the role of input and output variables and their synonyms.

B. Explain various error terms (reducible and irreducible), the expected value of error squared, and the variance of error terms.

C. Compare and contrast parametric and non-parametric learning methods.

D. Describe the trade-offs between prediction accuracy, flexibility, and model interpretability, including the role of overfitting.

E. Explain when a supervised learning model is preferable to unsupervised or semi-supervised learning models.

F. Explain how the appropriateness of regression problems relative to classification problems may be related to whether responses are quantitative or qualitative.

1.3.3 Assessing Model Accuracy (Ch. 2.2)

For example:

A. Recognize, explain, and apply the equation for mean squared error.

B. Explain the goal of measuring the quality of fit by minimizing training and test mean square errors (MSEs) and the implications of different levels of flexibility (degrees of freedom) for both training and test MSEs.

C. Explain the purpose of cross-validation.

D. Explain the bias-variance trade-off with an MSE decomposition into three fundamental quantities.

E. Explain the salient features of a simple Bayes classifier (for two classes), including the Bayes decision boundary and Bayes error rate.

F. Calculate the Bayes error rate.

G. Explain and apply the Bayesian classifier.

H. Explain how the K-nearest neighbors (KNN) classifier is related to the Bayes classifier and how the choice of K impacts results.

I. Calculate the conditional probability of a point belonging to a particular class.

J. Analyze the relationship between the value of K and the bias-variance tradeoff for a KNN classifier.

K. Explain what happens to the decision boundary as K increases in a KNN classifier.

FDP EDITORIAL STAFF

Hossein Kazemi, Ph.D., CFA, Senior Advisor, CAIA Association
Satya Das, CAIA, CFA, Senior Advisor, Curriculum and Exams, FDP Institute
Kathryn Wilkens, Ph.D., CAIA, Consultant, FDP Institute
Kim Durand, Project & General Operations Manager, FDP Institute

Design: Gabriele Villamena, Sichtwerk, Inc.