

A: Mathematical Fundamentals and Statistics

1. Basic Linear Algebra

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

DS-014	E	2.00	KA01.01
--------	---	------	---------

Understand how to calculate the inner product of vectors, and be able to express linear expressions as inner products of vectors.

[Note] Scope of high school mathematics B (C in the new course) (added to connect with high school mathematics, since some high school students have not taken this course)

DS-015	E	2.00	KA01.01
--------	---	------	---------

Understand how to calculate matrices, matrices and vectors correctly, and express multiple linear expressions as products of matrices.

DS-016	E	1.00	KA01.01
--------	---	------	---------

Understand the definition of an inverse matrix and the ability to solve simultaneous equations in matrix notation by computing the inverse matrix.

DS-017	E	1.00	KA01.01
--------	---	------	---------

Understand the meaning of eigenvectors and eigenvalues.

2. Differential and Integral Calculus

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

DS-018	E	1.00	KA01.01
--------	---	------	---------

Understand that derivatives calculated by differentiation are expressions for determining slope.

DS-019	E	1.00	KA01.01
--------	---	------	---------

Understand how to calculate partial derivatives for functions of two or more variables.

DS-020	E	1.00	KA01.01
--------	---	------	---------

Understand the relationship between integration and area, and be able to explain that probability can be obtained by definite integration of probability density function.

IPSJ-01	T1	1.00	KA01.01
Understand the exponential function, logarithmic function and their basic properties. [Note] Range of high school mathematics II (added for connection with high school mathematics, since some high school students have not taken this course)			
IPSJ-02	T1	1.00	KA01.01
Understand the base of the natural logarithm (e). [Note] Range of high school mathematics III (added for connection with high school mathematics, since some high school students have not taken this course)			
IPSJ-03	T1	1.00	KA01.01
Understand the definition and calculation of the limit of a function. [Note] Range of high school mathematics III (added for connection with high school mathematics, since some high school students have not taken this course)			
IPSJ-04	T1	2.00	KA01.01
Understand the derivatives of elementary functions. [Note] Range of high school mathematics III (added for connection with high school mathematics, since some high school students have not taken this course)			
IPSJ-05	T1	2.00	KA01.01
Understand the basic relationship of differentiation (differentiation of composite functions, differentiation of products). [Note] Range of high school mathematics III (added for connection with high school mathematics, since some high school students have not taken this course)			
IPSJ-06	T1	1.00	KA01.01
Understand the definition of partial differentiation.			
IPSJ-07	T1	2.00	KA01.01
Understand the inverse operation of differentiation and primitive functions. [Note] Range of high school mathematics III (added for connection with high school mathematics, since some high school students have not taken this course)			
IPSJ-08	T1	2.00	KA01.01
Understand the calculation of definite integrals. [Note] Range of high school mathematics III (added for connection with high school mathematics, since some high school students have not taken this course)			

3. Mathematical Statistics

Type	Knowledge	SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
		DS-001a	T1	0.50	KA01.01
Understand the concept of the number of cases, permutations and combinations. [Note] Range of high school mathematics A (added for connection with high school mathematics, since some high school students have not taken this course)					

IPSSJ Data Science Curriculum Standard (April 2021)

DS-002	T1	0.00	KA01.01
Understand the concept of probability and be able to explain the meaning and difference between simultaneous probability and conditional probability.			
[Note] Covered by mathematical, data science, and AI model curriculum (literacy level)			
DS-003	T1	0.00	KA01.01
To be able to explain the differences in methods for calculating the mean (additive mean), median, and mode.			
[Note] Covered by mathematical, data science, and AI model curriculum (literacy level)			
DS-004a	T1	0.00	KA01.01
To be able to explain the meaning and definition of variance and standard deviation.			
[Note] Covered by mathematical, data science, and AI model curriculum (literacy level)			
DS-005	T1	0.50	KA01.01
To be able to explain the difference between the population mean, sample mean, unbiased variance and sample variance.			
DS-006	T1	0.50	KA01.01
Know the values of variance and mean of the standard normal distribution.			
DS-007	T1	0.00	KA01.01
Explain the difference between a correlation function and a causal relationship.			
[Note] Covered by mathematical, data science, and AI model curriculum (literacy level)			
DS-008	T1	0.25	KA01.01
Explain the difference between nominal, ordinal, interval, and proportional scales.			
DS-009	T1	0.00	KA01.01
Explain the denominator and numerator of common correlation coefficients (Pearson).			
[Note] Covered by mathematical, data science, and AI model curriculum (literacy level)			
DS-010	T1	1.00	KA01.01
Explain five or more typical probability distributions.			
DS-011	T1	0.50	KA01.01
Know how binomially distributed events can be approximated as the sample size increases.			
DS-012a	T1	1.00	KA01.01
Know the difference in application of Pearson's product rate correlation coefficient, Klammer's correlation coefficient, and Spearman's rank correlation coefficient.			
DS-013	T1	1.00	KA01.01
To be able to explain Bayes' theorem.			

IPSJ Data Science Curriculum Standard (April 2021)

DS-028	E	1.00	KA01.01	KA01.05
To be able to explain the difference between point estimation and interval estimation.				
DS-029	E	1.00	KA01.01	KA01.05
Explain the difference between the null hypothesis and the alternative hypothesis.				
DS-030	E	1.00	KA01.01	KA01.05
Explain the meaning of Type 1 error, Type 2 error, p-value, and significance level.				
DS-031	E	1.00	KA01.01	KA01.05
Explain the difference between a one-tailed test and a two-tailed test.				
DS-032a	E	2.00	KA01.01	KA01.05
To be able to explain the difference in testing methods when there is correspondence between data and when there is not.				
IPSJ-09	E	2.00	KA01.01	
Understand the test of the mean and the test of the difference of means (t-test with and without correspondence between groups).				
Type Skill				
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
DS-001b	T1	0.50	KA01.01	
To be able to calculate permutations and combinations using the expressions nPr and nCr .				
[Note] Range of high school mathematics A (added for connection with high school mathematics, since some high school students have not taken this course)				
DS-004b	T1	0.00	KA01.01	
To be able to calculate the variance and standard deviation of given data.				
[Note] Covered by mathematical, data science, and AI model curriculum (literacy level)				
DS-012b	T1	0.50	KA01.01	
Be able to calculate the strength of a relationship when variables are either quantitative or qualitative.				
DS-032b	E	2.00	KA01.01	KA01.05
To be able to select and apply the appropriate test method considering the existence of correspondence between the data to be estimated.				

B1: Data Mining

1. Measure of proximity

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-16-01	T1	0.50	KA01.03	KA01.04 Understands the basic properties of indicators.
ACM-16-02	T1	0.50	KA01.03	KA01.04 Lk norm: special case ? Understand Euclidean distance and Manhattan distance.
ACM-16-03	T1	0.50	KA01.03	KA01.04 Understand the use of scores and rankings, and the desirable properties of scores and ranking methods.
ACM-16-04	T1	0.50	KA01.03	KA01.04 Understand the normalization of data to make it comparable.
ACM-16-05	T2	0.50	KA01.03	KA01.04 Understand metrics related to text.
ACM-16-06	T2	0.50	KA01.03	KA01.04 Understand metrics such as correlation coefficient for a series of data.
ACM-16-07	T2	0.50	KA01.03	KA01.04 Understand similarity metrics based on relationships in graphs, such as SimRank.
ACM-16-08	T2	0.50	KA01.03	KA01.04 Understand graph-based metrics.
ACM-16-09	T2	0.50	KA01.03	KA01.04 Understand metrics for measuring similarity of time series, e.g., dynamic time stretching methods.

IPJSJ Data Science Curriculum Standard (April 2021)

Type	Skill	SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
	ACM-16-10	T1	0.50	KA01.03	KA01.04
	Explain and compare the concept of measurement methods and their relevance to different types of data (nominal, ordinal, interval, proportional).				
	ACM-16-11	T1	0.50	KA01.03	KA01.04
	Be able to choose appropriate metrics for comparison of different types of data.				
	DS-023	E	5.00	KA01.05	KA01.01
	Evaluate the accuracy of a model using the ROC curve and AUC (Area under the curve).				
	DS-024	E	5.00	KA01.05	KA01.01
	To be able to understand and evaluate the accuracy of evaluation measures such as confusion matrix (cross table of positive and negative distributions), Accuracy, Precision, Recall, and F-value.				
	DS-025	E	5.00	KA01.05	KA01.01
	Understand evaluation measures such as MSE (Mean Squared Error) and MAE (Mean Absolute Error), and be able to evaluate their accuracy.				
	DS-026	E	5.00	KA01.05	KA01.01
	Understand the mechanism of the holdout method and cross-validation method, and be able to create training data, validation data for parameter tuning, and test data.				
	DS-027	E	5.00	KA01.05	KA01.01
	Understand that for data whose structure changes over time, training data should be past data and test data should be future data.				
	DS-036	T1	5.00	KA01.05	
	Be able to create histograms with appropriate interval data settings and explain how the data varies.				
	DS-037	T1	5.00	KA01.05	
	To be able to create crosstabulation tables with appropriate axis settings and understand the bias of the data between attributes.				
	DS-038	T2	5.00	KA01.05	
	To be able to draw scatter plots of quantitative variables and understand the relationship between two variables.				
Type	Disposition	SEQ#	Priority	Assigned Time (h)	DS-BoK KA#

ACM-16-12 T1 0.50 KA01.03 KA01.04

Be able to approach the problem carefully and critically, but with a rich imagination, knowing that there are typically many approaches.

2. Data Preparation

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-17-01	T1	0.50	KA01.03 KA01.04

Understand data collection, its relevance to problem solving, the importance of expertise, and be open to expert opinion.

ACM-17-02	T1	0.50	KA01.03 KA01.04
-----------	----	------	----------------------

Have data sources, including databases, IoT, photos, videos, and online sources: Understand that there is enough data to be used for a specific purpose.

ACM-17-03	T1	0.50	KA01.03 KA01.04
-----------	----	------	----------------------

Understand ethical considerations in acquiring and using data for specific purposes, concerns about personal information when linking data, and concerns about potential bias in the data.

ACM-17-04	T1	0.50	KA01.03 KA01.04
-----------	----	------	----------------------

Data Wrangling – Understands quality considerations in handling errors and omissions in data, data cleansing, data validation, data profiling, data transformation, and combining appropriate data sets.

ACM-17-05	T1	0.50	KA01.03 KA01.04
-----------	----	------	----------------------

Understand methods for handling dataset issues such as imbalance, insufficient or unnecessary attributes, automatic or manual approaches and trade-offs between the two.

ACM-17-06	T1	0.50	KA01.03 KA01.04
-----------	----	------	----------------------

Understand the concept of features, feature extraction, feature representation, feature selection, and feature generation.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-17-07	T1	0.50	KA01.03 KA01.04

Explain the connection between the process of constructing a question and the process of acquiring data to answer the question.

ACM-17-08	T1	0.50	KA01.03 KA01.04
-----------	----	------	----------------------

Demonstrate the ability to understand specific domains that require appropriate interaction with experts.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-17-09	T1	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Use summary statistics and data visualization to formulate exploratory data analysis and inferences.

ACM-17-10	T1	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Describe possible challenges in a dataset, show their impact and how they can be solved.

ACM-17-11	T1	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Identify the various methods of feature generation and explain the advantages and implications of each.

ACM-17-12	T1	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Describe the similarities and differences between feature selection and feature generation, and show how feature generation produces features.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-17-13	T1	0.50	KA01.03	KA01.04

Be able to set an appropriate confidence level in data selection and preparation, and understand the importance of working with high quality data.

3. Information Extraction

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-18-01	E	1.75	KA01.03	KA01.04

Understand the applications where information extraction is useful.

ACM-18-02	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Understand entity and relation extraction.

ACM-18-03	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Understand rule-based information extraction methods and their applications.

ACM-18-04	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Understand statistics-based information extraction methods and their applications.

ACM-18-05	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Understand the problems that can occur in extracted data.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-18-06	E	1.75	KA01.03	KA01.04

Able to design schemas according to the requirements and data of the application.

ACM-18-07	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Able to write information extraction rules according to the application.

ACM-18-08	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Be able to apply rules and learning algorithms such as model learning and relational prediction to the information extraction task.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-18-09	E	1.75	KA01.03	KA01.04

Understand that there are various methods for extracting information from data.

4. Cluster Analysis

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-19-01	T1	0.50	KA01.03	KA01.04

Identify appropriate similarity measures for clustering.

ACM-19-02	T1	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Understand how to evaluate the quality of clustering.

ACM-19-03	T1	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Understand the k-means method, taking into account the need for iteration.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-19-04	T1	0.50	KA01.03	KA01.04						
Understand the density-based algorithm.										
ACM-19-05	T1	0.50	KA01.03	KA01.04						
Understand the applications of clustering.										
ACM-19-06	T2	0.50	KA01.03	KA01.04						
Understand Mean shift clustering.										
ACM-19-07	T2	0.50	KA01.03	KA01.04						
Understand agglomerative clustering.										
ACM-19-08	T2	0.50	KA01.03	KA01.04						
Understand the grid-based algorithm.										
ACM-19-09	T2	0.50	KA01.03	KA01.04						
Understand strategies for speeding up and parallelizing clustering algorithms.										
<table border="1"> <thead> <tr> <th>Type</th> <th>Skill</th> </tr> </thead> <tbody> <tr> <td>SEQ#</td> <td>Priority</td> <td>Assigned Time (h)</td> <td>DS-BoK KA#</td> </tr> </tbody> </table>					Type	Skill	SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
Type	Skill									
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#							
ACM-19-10	T1	0.50	KA01.03	KA01.04						
Explain the importance of feature selection for clustering.										
ACM-19-11	T1	0.50	KA01.03	KA01.04						
Provide guidelines for selecting initialization criteria for k-means algorithms.										
ACM-19-12	T2	0.50	KA01.03	KA01.04						
Compare clustering methods, highlighting their relative advantages and disadvantages.										
ACM-19-13	T2	0.50	KA01.03	KA01.04						
Indicate situations where different clustering methods should be used and situations where alternative methods are more desirable.										
ACM-19-14	T2	0.50	KA01.03	KA01.04						
Apply multiple algorithms to a test set of data and compare the results.										

ACM-19-15	T2	0.50	KA01.03	KA01.04
Present an explanation that highlights the benefits and value of clustering.				

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-19-16	T1	0.50	KA01.03	KA01.04
Create an insightful and positive perspective on the role of clustering in data science.				

ACM-19-17	T1	0.50	KA01.03	KA01.04
Understand the importance of scalable and efficient clustering algorithms that can be used in realistic scenarios.				

5. Classification and Regression

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-20-01	T1	0.50	KA01.03	KA01.04
Understand the considerations for feature selection for classification.				

ACM-20-02	T1	0.50	KA01.03	KA01.04
Understand instance-based methods such as KNN.				

ACM-20-03	T1	0.50	KA01.03	KA01.04
Understand decision tree models.				

ACM-20-04	T1	0.50	KA01.03	KA01.04
Understands probabilistic models and naive Bayes.				

ACM-20-05	T2	0.50	KA01.03	KA01.04
Understands rule-based methods.				

ACM-20-06	T2	0.50	KA01.03	KA01.04
Understands support vector machines.				

IP SJ Data Science Curriculum Standard (April 2021)

ACM-20-07 T2 0.50 KA01.03 KA01.04
Understand neural networks.

ACM-20-08 T2 0.50 KA01.03 KA01.04
Understand real-world applications of classification and regression.

ACM-20-09 T2 0.50 KA01.03 KA01.04
Understand deep learning and related software (Caffe, TensorFlow, PyTorch).

ACM-20-10 E 1.75 KA01.03 KA01.04
Understand speedup and parallelization strategies.

Type	Skill
------	-------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-20-11	T1	0.50	KA01.03 KA01.04
Explain the importance of feature selection for classification and regression.			

ACM-20-12 T1 0.50 KA01.03 KA01.04
Describe criteria such as prediction accuracy and understandability of the trained model that can contribute to the choice of which method to use.

ACM-20-13 T2 0.50 KA01.03 KA01.04
Identify the relationship between regression and classification.

ACM-20-14 T2 0.50 KA01.03 KA01.04
Recognize the key aspects that may benefit from the use of classification or regression models.

ACM-20-15 T2 0.50 KA01.03 KA01.04
Identify and implement software that can handle each method.

ACM-20-16 T2 0.50 KA01.03 KA01.04
Be able to choose a classification method, justify the reasons for choosing it, and demonstrate the ability to apply it to moderately complex cases.

DS-021 E 5.00 KA01.05 KA01.01
Be able to explain the least squares method, regression counting, and standard errors in simple regression analysis.

DS-022	E	5.00	KA01.05	KA01.01
Explain partial and standard partial regression coefficients and multiple correlations in multiple regression analysis.				

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-20-17	T1	0.50	KA01.03	KA01.04
Understand the importance of scalable and efficient classification and regression algorithms that can be used in realistic scenarios.				

ACM-20-18	T2	0.50	KA01.03	KA01.04
Explain the connection between classification and regression, and more generally with statistics and machine learning.				

6. Pattern Mining

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-21-01	T2	0.50	KA01.03	KA01.04
Understand the concept of correlated pattern mining.				

ACM-21-02	T2	0.50	KA01.03	KA01.04
Understand the considerations for computational complexity.				

ACM-21-03	T2	0.50	KA01.03	KA01.04
Understand correlation rule mining, Apriori algorithm, and frequent pattern mining algorithm.				

ACM-21-04	T2	0.50	KA01.03	KA01.04
Understand series pattern mining and GSP algorithm.				

ACM-21-05	T2	0.50	KA01.03	KA01.04
Understand efficient and concurrent pattern mining algorithms.				

ACM-21-06	T2	0.50	KA01.03	KA01.04
Understand the applications.				

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-21-07	T2	0.50	KA01.03	KA01.04
List the various areas where the Apriori algorithm can be used under normal circumstances to produce beneficial effects.				

ACM-21-08	T2	0.50	KA01.03	KA01.04
Identify an implementation of the Apriori algorithm and apply it to a meaningful application.				

ACM-21-09	T2	0.50	KA01.03	KA01.04
Compare and contrast the usefulness of pattern mining algorithms.				

DS-033	E	1.00	KA01.02	
To be able to explain the difference between classification (discriminant) models in supervised learning and grouping (clustering) in unsupervised learning.				

DS-034	E	1.00	KA01.02	
To be able to explain the difference between hierarchical cluster analysis and non-hierarchical cluster analysis.				

DS-035	E	2.00	KA01.02	
Understand how to read a dendrogram in hierarchical cluster analysis and be able to interpret it appropriately.				

DS-077	E	0.50	KA01.03	
To be able to evaluate the relationship between condition X and event Y using lift values.				

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-21-10	T2	0.50	KA01.03	KA01.04
Recognize that pattern mining is a very broad topic and has many applications.				

7. Outlier Detection

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
------	----------	-------------------	------------	--

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-22-01 T2 0.50 KA01.03 KA01.04
 Understand the definition of the concept of outliers.

ACM-22-02 T2 0.50 KA01.03 KA01.04
 General Methods – Can build a model from data and then note that some data points do not fit it.

ACM-22-03 T2 0.50 KA01.03 KA01.04
 Understand parametric methods to identify numerical outliers in one dimension, such as Z-scores.

ACM-22-04 T2 0.50 KA01.03 KA01.04
 Understand the use of probability distribution functions.

ACM-22-05 T2 0.50 KA01.03 KA01.04
 Use depth-first methods ? Identify the convex hull that can be expected for a set of points and then the inner or outer points; related graphical approaches

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-22-06	T2	0.50	KA01.03	KA01.04
Apply a wide range of outlier detection algorithms.				

ACM-22-07 T2 0.50 KA01.03 KA01.04
 Compare and contrast parametric and non-parametric methods for outlier detection.

ACM-22-08 T2 0.50 KA01.03 KA01.04
 Explain how outlier detection methods can aid in plagiarism detection, financial fraud, network intrusion detection, and other areas of application.

ACM-22-09 T2 0.50 KA01.03 KA01.04
 Use appropriate examples to illustrate the importance of outlier detection.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-22-10	T2	0.50	KA01.03	KA01.04
Develops a critical and broad perspective on outlier analysis and detection.				

8. Time Series Data

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
------	----------	-------------------	------------	--

ACM-23-01	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

The nature of time series data, including a comparison with sequential temporal data.

ACM-23-02	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Data Transformation – Understand denoising and normalization of time series data.

ACM-23-03	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Understand stationary time series and non-stationary time series.

ACM-23-04	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Understand the transformation of time series data to discrete series data.

ACM-23-05	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Time series data forecasting ? Understand the prediction of future values based on past values.

ACM-23-06	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Time series motifs ? Understand patterns that appear frequently in time series data.

ACM-23-07	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Understand time series clustering and classification.

ACM-23-08	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Outlier Detection in Time Series Data – Understand outliers as points and shapes.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
------	----------	-------------------	------------	--

ACM-23-09	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Be able to provide multiple situations where meaningful time series data exists and show the importance of mining the data.

ACM-23-10	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Demonstrate with examples when it is desirable to convert time series data into series data.

ACM-23-11	E	1.75	KA01.03	KA01.04
-----------	---	------	---------	---------

Explain the techniques used in clustering and classification of time series data.

DS-066	T2	5.00	KA01.05	
--------	----	------	---------	--

To be able to explain what time series data is and its basic handling. (Time series graphs, periodicity, moving average, etc.)

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-23-12	E	1.75	KA01.03	KA01.04

Data mining of time series data has been found to be very important in several important applications.

9. Web Data Mining

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-24-01	T2	0.50	KA01.03	KA01.04

Understand the process of web scraping and web spidering, and web crawling as it relates to web access.

ACM-24-02	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Understand the ethical guidelines associated with accessing web data.

ACM-24-03	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Understand the structure and functionality of software libraries for accessing web data.

ACM-24-04	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Understand knowledge discovery approaches to web data, such as community discovery and link prediction.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
------	----------	-------------------	------------	--

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-24-05	T2	0.50	KA01.03	KA01.04
Compare and contrast the capabilities and ease of use of two different libraries for accessing the Web.				

ACM-24-06	T2	0.50	KA01.03	KA01.04
Demonstrate how to use software to scrape accurate data from publicly available websites.				

ACM-24-07	T2	0.50	KA01.03	KA01.04
Given a set of constraints, develop software to retrieve data from the Web.				

ACM-24-08	T2	0.50	KA01.03	KA01.04
Develop efficient algorithms for discovering knowledge from the web.				

Type	Disposition
------	-------------

SEQ#	Priority	Assigned Time (h)	DS-BoK	KA#
ACM-24-09	T2	0.50	KA01.03	KA01.04
Facilitate and encourage access to quality data, taking into account the ethical framework.				

B2: Artificial Intelligence

1. General

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-59-01	T1	0.25	KA01.02
Understand the history of AI.			

ACM-59-02	T1	0.25	KA01.02
Understand the reality of AI (what it is, what it can do) and the gaps in perception.			

ACM-59-03	T1	0.25	KA01.02
Understand the subfields of AI: knowledge representation, logic and probabilistic reasoning, planning, recognition, natural language processing, learning, robotics (physical and virtual).			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-59-04	T1	0.25	KA01.02
Be able to explain the main branches of AI in order to recognize useful concepts and methods when they are needed in data science.			

ACM-59-05	T2	0.50	KA01.02
Explain clearly what AI is, how it collects and uses data to make it work, and how it collects and generates data that can be used by data scientists.			

ACM-59-06	T2	0.50	KA01.02
Explain quantitatively how (physical and virtual) robots, agents, and multi-agent systems collect and use data to embed AI and make it work.			

ACM-59-07	T2	0.50	KA01.02
Explain how the data collected or generated by AI may be useful in data science applications.			

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-59-08	T1	0.25	KA01.02
-----------	----	------	---------

Understand that AI is not a new field, but rather one with a long and rich history.

ACM-59-09	T2	0.50	KA01.02
-----------	----	------	---------

Know the main concerns of AI and the kinds of problems they address, in order to know where to find approaches when needed, and to avoid rediscovering existing methods.

2. Knowledge Representation and Reasoning (Logic-based Models)

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-60-01	T2	0.50	KA01.02

Understand predicate logic and its use cases.

ACM-60-02	T2	0.50	KA01.02
-----------	----	------	---------

Automated reasoning: Understand forward and backward reasoning.

ACM-60-03	T2	0.50	KA01.02
-----------	----	------	---------

Understand reasoning integrated in large systems (e.g., Watson).

ACM-60-04	T2	0.50	KA01.02
-----------	----	------	---------

Understands ontologies, knowledge graphs (e.g., protege, ConceptNet, YAGO, UMLS).

ACM-60-05	E	0.50	KA01.02
-----------	---	------	---------

Automated reasoning: understands derivations, theorem proving.

ACM-60-06	E	0.50	KA01.02
-----------	---	------	---------

Understand the language for automated reasoning.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-60-07	T2	0.50	KA01.02

Able to express statements of natural language in predicate logic.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-60-08	T2	0.50	KA01.02
Able to express statements of predicate logic in natural language.			
ACM-60-09	T2	0.50	KA01.02
To be able to describe the use cases and limitations of predicate logic.			
ACM-60-10	T2	0.50	KA01.02
Give examples of algorithms and systems that perform efficient automated reasoning.			
ACM-60-11	T2	0.50	KA01.02
Describe logic-based automated reasoning with forward and backward reasoning, for example.			
ACM-60-12	T2	0.50	KA01.02
Give examples of how reasoning is integrated into large data-driven systems.			
ACM-60-13	E	0.50	KA01.02
Describe a particular method of automated theorem proving.			
ACM-60-14	E	0.50	KA01.02
Describe what an ontology is, the situations in which it can be used (e.g., question answering), and how it can be used (e.g., to assist in disambiguation), giving examples of existing techniques.			
ACM-60-15	E	0.50	KA01.02
Describe how an ontology is constructed.			
ACM-60-16	E	0.50	KA01.02
Implement reasoning problems of moderate size.			
Type Disposition			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-60-17	T2	0.50	KA01.02
Have a good understanding of the benefits and limitations of logic-based knowledge representation.			
ACM-60-18	T2	0.50	KA01.02
Be aware of the rich history of formal logic and logic-based algorithms for use in specific applications.			

3. Knowledge Representation and Reasoning (Probability-based Models)

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-61-01	T1	0.25	KA01.02
-----------	----	------	---------

Understand basic concepts: random variables, axioms of probability, independence, conditional probability, and marginal probability. Also refer to probability, which is fundamental knowledge in the field of data science (not in the field of computer science).

ACM-61-02	T1	0.25	KA01.02
-----------	----	------	---------

Understand causal models.

ACM-61-03	T2	0.50	KA01.02
-----------	----	------	---------

Understand Bayesian networks.

ACM-61-04	T2	0.50	KA01.02
-----------	----	------	---------

Understand Markov decision processes.

ACM-61-05	E	0.50	KA01.02
-----------	---	------	---------

Understand reinforcement learning.

ACM-61-06	E	0.50	KA01.02
-----------	---	------	---------

Understand stochastic logic models (e.g., Markov logic networks).

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-61-07	T1	0.25	KA01.02
-----------	----	------	---------

Justify the need for probabilistic inference.

ACM-61-08	T1	0.25	KA01.02
-----------	----	------	---------

Define basic concepts such as random variables and independence.

ACM-61-09	T1	0.25	KA01.02
-----------	----	------	---------

Be able to state the axioms of probability.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-61-10	T1	0.25	KA01.02
Be able to model simple systems and answer questions using the basic concepts and axioms above.			
ACM-61-11	T1	0.25	KA01.02
Describe what a causal model is and how it is used.			
ACM-61-12	T2	0.50	KA01.02
Describe what a Bayesian network is, giving examples of small and medium sizes.			
ACM-61-13	T2	0.50	KA01.02
Show situations in which Bayesian networks can be useful (e.g., medical diagnosis problems).			
ACM-61-14	T2	0.50	KA01.02
Demonstrate how Bayesian networks can be used for inference, understand that accurate inference is not feasible in most cases, and give examples of more efficient approaches to inference (e.g., belief propagation).			
ACM-61-15	T2	0.50	KA01.02
Be able to identify independence relations exhibited by Bayesian networks.			
ACM-61-16	T2	0.50	KA01.02
Describe what a Markov decision process is, giving examples of small and medium sizes.			
ACM-61-17	T2	0.50	KA01.02
Show situations in which Markov decision processes can be useful (e.g., optimization problems, control problems).			
ACM-61-18	T2	0.50	KA01.02
Show how Markov decision processes can be used for inference.			
ACM-61-19	E	0.50	KA01.02
To be able to construct a Bayesian network for small/medium scale problems.			
ACM-61-20	E	0.50	KA01.02
To be able to apply learning algorithms to construct Bayesian networks for small- and medium-scale problems.			
ACM-61-21	E	0.50	KA01.02
Describe how the parameters of a Markov decision process are learned, and give examples of algorithms that learn these parameters.			

ACM-61-22	E	0.50	KA01.02
-----------	---	------	---------

Be able to apply reinforcement learning algorithms to appropriate problems.

ACM-61-23	E	0.50	KA01.02
-----------	---	------	---------

Give examples of stochastic logic models, such as Markov logic networks, identifying useful applications.

ACM-61-24	E	0.50	KA01.02
-----------	---	------	---------

Construct probabilistic logic models for small- and medium-scale problems.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-61-25	T1	0.25	KA01.02

Understand the merits and limitations of probability-based knowledge representation and methods for reasoning on it.

4. Planning and Search Strategies

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-62-01	T2	0.50	KA01.02

Understand the representation of the state space of possible solutions to a problem.

ACM-62-02	T2	0.50	KA01.02
-----------	----	------	---------

Understand breadth-first search and depth-first search in state space (without prior information).

ACM-62-03	T2	0.50	KA01.02
-----------	----	------	---------

Understand heuristic search (with prior information) in state space (e.g., A* search).

ACM-62-04	E	0.50	KA01.02
-----------	---	------	---------

Understand probabilistic search algorithms (e.g., genetic algorithms, annealing methods).

ACM-62-05	E	0.50	KA01.02
-----------	---	------	---------

Understand constraint satisfaction problems and related methods.

IP SJ Data Science Curriculum Standard (April 2021)

Type	Skill			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-62-06	T2	0.50	KA01.02	Explain how a solution to a problem can be regarded as a state in a possible solution space.
ACM-62-07	T2	0.50	KA01.02	Be able to model a given problem as a search in a multidimensional state space.
ACM-62-08	T2	0.50	KA01.02	Explain how breadth- and depth-first search can be used to explore a solution space modeled as a graph.
ACM-62-09	T2	0.50	KA01.02	Explain how heuristics can be used to (potentially) increase the speed of graph/state space exploration.
ACM-62-10	E	0.50	KA01.02	A search without prior information can be applied to find a solution to a problem modeled as a state space (where the graph representing the state is constructed as the search is performed rather than given as input).
ACM-62-11	E	0.50	KA01.02	Heuristics can be designed for small-scale problems.
ACM-62-12	E	0.50	KA01.02	It is possible to apply prior information search to small/medium problems.
ACM-62-13	E	0.50	KA01.02	To be able to apply stochastic search to small/medium scale problems.
ACM-62-14	E	0.50	KA01.02	Explain how stochastic search algorithms address the problem of searching a large space (e.g., avoiding locally optimal solutions), and how stochastic search algorithms address the problem of locally searching a promising solution space.
ACM-62-15	E	0.50	KA01.02	Explain how solutions to problems involve constraints on certain variables and relationships between them: Describe methods for revealing these constraints.
ACM-62-16	E	0.50	KA01.02	Be able to implement search algorithms.

ACM-62-17	E	0.50	KA01.02
Model a small-scale problem as a constraint satisfaction problem.			

ACM-62-18	E	0.50	KA01.02
Apply constraint satisfaction algorithms to small- and medium-scale problems.			

Type	Disposition
------	-------------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-62-19	T2	0.50	KA01.02

Understand that there can be multiple acceptable solutions in the state space and multiple ways to find them. Different solutions and problem solving approaches should be used depending on external conditions such as optimality requirements and time const

ACM-62-20	T2	0.50	KA01.02
-----------	----	------	---------

Have a good understanding of the relationship between algorithms, heuristics, and the optimality of solutions to problems.

B3: Machine Learning

1. General

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-63-01	T1	0.25	KA01.02
-----------	----	------	---------

Understand the history of machine learning.

ACM-63-02	T1	0.25	KA01.02
-----------	----	------	---------

Understand the major tasks of machine learning, including supervised learning, unsupervised learning, reinforcement learning, and deep learning.

ACM-63-03	T1	0.25	KA01.02
-----------	----	------	---------

Understand the differences between symbolic and numerical learning, and between statistical and structural/syntactic approaches.

ACM-63-04	T1	0.25	KA01.02
-----------	----	------	---------

Understand learning algorithms as a principle-based optimization approach.

ACM-63-05	T1	0.25	KA01.02
-----------	----	------	---------

Understand that "doing machine learning" is a process and a method of data mining.

ACM-63-06	T1	0.25	KA01.02
-----------	----	------	---------

Understand the importance of robust evaluation.

ACM-63-07	T1	0.25	KA01.02
-----------	----	------	---------

Understand the difficulties in machine learning, including data quality and the need for regularization.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-63-08	T1	0.25	KA01.02
-----------	----	------	---------

Compare the goals, inputs, and outputs of supervised learning, unsupervised learning, reinforcement learning, and deep learning.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-63-09	T1	0.25	KA01.02
Know that different kinds of data-driven problems can be solved by different approaches, and be able to link them appropriately.			
ACM-63-10	T1	0.25	KA01.02
Roughly explain that machine learning models and algorithms are techniques based on mathematical and statistical foundations.			
ACM-63-11	T1	0.25	KA01.02
Be able to follow the process of "doing machine learning" as a data mining technique: understand the problem that the customer wants to solve, collect the data relevant to solving the problem, convert the raw data into features, select an appropriate mach			
ACM-63-12	T1	0.25	KA01.02
We can discuss the trade-off between fitting to training data and generalization to new data, and how model complexity, number of data, and number of features affect this trade-off. We can relate this to the role of hyperparameters and their settings.			
ACM-63-13	T1	0.25	KA01.02
Have a good understanding of the trade-offs between performance, interpretability, and scalability. Recognize that different optimization functions and techniques may produce different tradeoffs.			
ACM-63-14	T2	0.50	KA01.02
Can follow the derivation of simple optimization functions and learning methods from the most basic principles (e.g., follow decision tree principles from information theory, logistic regression from maximum likelihood or stochastic gradient descent, prin			
ACM-63-15	T2	0.50	KA01.02
Analyze the performance of a model using bootstrapping and statistical significance tests.			
ACM-63-16	T2	0.50	KA01.02
Understand how to efficiently migrate a model to a product and how to choose the tools that will support the migration from the beginning.			
ACM-63-17	T2	0.50	KA01.02
Understand which tools to use depending on the size of the data: for big data it is essential to choose machine learning tools that can run in parallel, otherwise those that take more time in the learning process are acceptable.			
ACM-63-18	T2	0.50	KA01.02
They are aware of the availability of state-of-the-art machine learning tools.			
ACM-63-19	E	0.50	KA01.02
In particular, we can describe automatic machine learning (meta-learning), including how it automates the machine learning pipeline, including data preprocessing, model selection, model structure exploration, and hyperparameter tuning.			

IPJSJ Data Science Curriculum Standard (April 2021)

DS-123 E 1.00 KA01.02
 Know the names of at least three analytical methods for machine learning, and be able to give an overview of the methods.

DS-124 E 1.00 KA01.02
 Have used machine learning models under instruction, and understand what problems they can solve.

DS-125 T1 0.50 KA01.02
 Understand the difference between “supervised learning” and “unsupervised learning”.

DS-126 T1 0.50 KA01.02
 Explain what overlearning is and the problems it can cause.

DS-127 E 0.50 KA01.02
 To be able to explain what a curse of dimensionality is and the problems it causes.

DS-128 E 0.50 KA01.02
 Explain the need for annotation in supervised learning.

DS-129 E 0.50 KA01.02
 Understand the risk that the output of the model will behave in a discriminatory manner when the observed data contains bias, or when the learned predictive model recognizes minority data as noise.

DS-130 E 0.50 KA01.02
 Understand the difference between global explanations (e.g., the contribution of each variable per model) and local explanations (e.g., the contribution of each variable per record to be predicted) in machine learning.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-63-20	T1	0.25	KA01.02
-----------	----	------	---------

Machine learning is gaining popularity these days, but we fully understand that it is not a recent invention. Look for existing methods before assuming that new ones are needed.

ACM-63-21	T1	0.25	KA01.02
-----------	----	------	---------

Fully understand that machine learning is not a set of ad hoc tricks and should be used responsibly.

ACM-63-22	T1	0.25	KA01.02
-----------	----	------	---------

Understand that “doing machine learning” in general is not a simple process of applying a machine learning program to a conveniently formatted dataset, but rather a process of moving towards a customer goal.

ACM-63-23	T1	0.25	KA01.02
-----------	----	------	---------

Fully understand that there are many aspects of comparing trained models, ranging from empirical error minimization to model size and complexity to human interpretability.

ACM-63-24	T1	0.25	KA01.02
-----------	----	------	---------

Understand their responsibility to state results fairly and to compare them honestly, considering all aspects of model comparison, including quality, efficiency, and interpretability.

2. Supervised Learning

Type	Knowledge
------	-----------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-64-01	T1	0.25	KA01.02
-----------	----	------	---------

Understand the main tasks of supervised learning: regression and classification.

ACM-64-02	T1	0.25	KA01.02
-----------	----	------	---------

Understand the use cases of regression and classification.

ACM-64-03	T1	0.25	KA01.02
-----------	----	------	---------

Understand the key concerns and tradeoffs in machine learning, including the relationship between model complexity and generalizability, understand the bias-variance tradeoff, and understand Occam's Razor, which is the motivation for using simple models.

ACM-64-04	T1	0.25	KA01.02
-----------	----	------	---------

Understand the need to partition training, test, and validation data. Understand the definitions of training error and generalization error.

ACM-64-05	T1	0.25	KA01.02
-----------	----	------	---------

Understand common metrics for classification tasks (e.g., percent correct, sensitivity, specificity, goodness of fit, recall, F1 score, area under the ROC curve, Riglet) and common metrics for regression tasks (e.g., root mean square error, mean absolute

ACM-64-06	T1	0.25	KA01.02
-----------	----	------	---------

Understand the need for validation data, the process of cross-validation and its goals: tuning hyperparameters and evaluating model performance.

ACM-64-07	T1	0.25	KA01.02
-----------	----	------	---------

Understand the criteria for evaluating the quality of training, test, and validation data, such as number of data and class hierarchy.

ACM-64-08	T1	0.25	KA01.02
-----------	----	------	---------

Understand at least one classification/regression algorithm that is linear or nonlinear (e.g., linear regression, linear classification, logistic regression, nearest neighbor, naive Bayes, decision tree learning).

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-64-09 T1 0.25 KA01.02
 Understand general extensions to basic algorithms involving polynomial features and ensembles (e.g., bagging models, boosting models, random forests).

ACM-64-10 T2 0.50 KA01.02
 Understand approaches for determining whether a model has a large bias or a large variance (e.g., performance on training data vs. performance on test data, learning curves).

ACM-64-11 T2 0.50 KA01.02
 Understand at least two reasons for increasing or decreasing the number of features, respectively, and the tradeoffs involved.

ACM-64-12 T2 0.50 KA01.02
 Understand how supervised learning classifiers can be adapted to multiclass classification problems and how two-class classification models can be extended to multiclass classification problems.

ACM-64-13 T2 0.50 KA01.02
 Understand how macroscopic/microscopic metrics can be used to describe performance.

ACM-64-14 T2 0.50 KA01.02
 Understand one or more advanced supervised learning algorithms (e.g., kernel SVMs, neural networks).

ACM-64-15 E 0.50 KA01.02
 Be able to derive supervised learning algorithms from basic principles.

Type	Skill
------	-------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-64-16 T1 0.25 KA01.02
 Be able to express the performance of classification models using confusion matrices.

ACM-64-17 T1 0.25 KA01.02
 Compare the strengths and weaknesses of evaluation metrics for classification/regression tasks.

ACM-64-18 T1 0.25 KA01.02
 Compare the trade-offs between two or more practical classification/regression algorithms.

ACM-64-19 T1 0.25 KA01.02
 Apply two or more classification/regression algorithms to small and medium datasets.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-64-20 T1 0.25 KA01.02
 Compare training error and generalization error in terms of information available from the trained model.

ACM-64-21 T1 0.25 KA01.02
 Compare the performance of algorithms using various metrics.

ACM-64-22 T1 0.25 KA01.02
 Able to apply two or more augmented algorithms to small/medium/large datasets (e.g., ensemble algorithms), and compare the performance of the algorithms using various metrics.

ACM-64-23 T1 0.25 KA01.02
 Justify when extended algorithms, such as polynomial features and ensembles, are appropriate based on the problems they can address.

ACM-64-24 T2 0.50 KA01.02
 Be able to apply more than one classification/regression algorithm to a large dataset.

ACM-64-25 T2 0.50 KA01.02
 One or more extension algorithms can be applied to large datasets.

ACM-64-26 T2 0.50 KA01.02
 Apply bias/variance reduction methods.

ACM-64-27 T2 0.50 KA01.02
 Apply feature addition/selection to problems of medium/large size.

ACM-64-28 T2 0.50 KA01.02
 Apply advanced supervised learning algorithms (e.g., kernel SVMs, neural networks).

ACM-64-29 E 0.50 KA01.02
 Derive simple optimization functions and learning algorithms from basic principles (e.g., maximum likelihood and stochastic gradient descent to derive logistic regression). These techniques can be extended to similar models.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-64-30	T1	0.25	KA01.02

Understand how important algorithm selection and metrics are to the quality of the trained model. They know that these choices affect important stakeholders (to whom they provide the developed models) and that they should be made with them.

ACM-64-31 T1 0.25 KA01.02

They fully understand the importance of applying principled model evaluation methods that they are confident in.

3. Unsupervised Learning

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-65-01	T1	0.25	KA01.02
-----------	----	------	---------

Understand the main tasks of unsupervised learning, including clustering and dimensionality reduction.

ACM-65-02	T1	0.25	KA01.02
-----------	----	------	---------

Understand use cases for both tasks (e.g., data exploration/summarization/visualization, feature selection, data compression, data denoising, prototype learning, recommendation systems, topic modeling).

ACM-65-03	T1	0.25	KA01.02
-----------	----	------	---------

Understand one or more clustering algorithms (e.g., k-means, hierarchical clustering).

ACM-65-04	T1	0.25	KA01.02
-----------	----	------	---------

Understand the tradeoffs between connectivity-based clustering and centroid-based clustering.

ACM-65-05	T1	0.25	KA01.02
-----------	----	------	---------

Understand one or more dimensionality reduction algorithms (e.g., principal component analysis).

ACM-65-06	T1	0.25	KA01.02
-----------	----	------	---------

Understand the similarities and differences in feature transformation, feature selection, and feature projection.

ACM-65-07	T2	0.50	KA01.02
-----------	----	------	---------

Understand one or more advanced clustering algorithms (e.g., density-based algorithms such as Gaussian mixture models).

ACM-65-08	T2	0.50	KA01.02
-----------	----	------	---------

Understand one or more advanced dimensionality reduction algorithms (e.g., independent component analysis, non-negative matrix factorization).

ACM-65-09	E	0.50	KA01.02
-----------	---	------	---------

Understand one or more mathematical methods for efficient implementation of algorithms (e.g., matrix factorization, singular value decomposition, eigenvalue decomposition for principal component analysis).

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-65-10 E 0.50 KA01.02
 Understand one or more developmental algorithms (e.g., spectral clustering, kernel k-means, kernel principal component analysis, latent Dirichlet allocation).

ACM-65-11 E 0.50 KA01.02
 Understand the connection between principal component analysis and self-encoders, and the generalization to nonlinear dimensionality reduction.

ACM-65-12 E 0.50 KA01.02
 Able to derive unsupervised learning algorithms from basic principles.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-65-13 T1 0.25 KA01.02
 Be able to apply one or more clustering/dimensionality reduction algorithms to small/medium/large datasets.

ACM-65-14 T1 0.25 KA01.02
 Describe the performance of unsupervised learning algorithms using various metrics (e.g., visualization, comparison with correct data (if possible), calculation of metrics such as cluster density, and indirect metrics of usefulness in other applications).

ACM-65-15 T1 0.25 KA01.02
 Describe and apply methods for describing how to choose hyperparameters, such as the number of clusters for k-means or the number of elements for principal component analysis.

ACM-65-16 T2 0.50 KA01.02
 Compare the trade-offs of two or more clustering algorithms.

ACM-65-17 T2 0.50 KA01.02
 Compare the trade-offs of two or more dimensionality reduction algorithms.

ACM-65-18 E 0.50 KA01.02
 Apply advanced unsupervised learning algorithms.

ACM-65-19 E 0.50 KA01.02
 Derive simple optimization functions and learning algorithms from basic principles (e.g., derive principal component analysis using variance minimization and eigenvalues). These techniques can be extended to similar models.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-65-20 T1 0.25 KA01.02

Understand how important algorithm selection and evaluation metrics are to the quality of the trained model. They know that these choices affect important stakeholders (to whom they provide the developed models) and that they should be made with them.

ACM-65-21 T1 0.25 KA01.02

They fully understand the importance of applying a principled model evaluation method that they are confident in.

ACM-65-22 T2 0.50 KA01.02

Understand that unsupervised learning provides techniques for exploring, understanding, summarizing, and visualizing data.

ACM-65-23 T2 0.50 KA01.02

Understand that unsupervised learning can be a useful preprocessing step to improve the quality and efficiency of supervised learning algorithms.

4. Mixed Learning

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-66-01 E 0.50 KA01.02

Understand examples of problems and domains in machine learning where data structures and correlations between data points can be used effectively in learning models (e.g., time series prediction, serial prediction, recommendation systems).

ACM-66-02 E 0.50 KA01.02

Understand how the assumptions of time dependency and information shared by each data point can be used effectively in learning.

ACM-66-03 E 0.50 KA01.02

Understand the drawbacks of using a supervised/unsupervised learning approach instead of a mixed approach (e.g., model interpretability and performance issues).

ACM-66-04 T2 0.50 KA01.02

Understands one or more standard approaches to the problem (e.g., using hidden Markov models for series prediction, using enhanced filtering for recommendation systems).

ACM-66-05 T2 0.50 KA01.02

Understand the need to partition training and test data in this context.

ACM-66-06 T2 0.50 KA01.02

Understand common evaluation metrics for selected tasks (e.g., recall rate, fit rate, F1 score in recommendation systems).

ACM-66-07 T2 0.50 KA01.02

Understand the criteria for evaluating the quality of training, validation, and test data for a selected problem.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-66-08	E	0.50	KA01.02
-----------	---	------	---------

Be able to map a problem to a learning framework for it, i.e., map data to inputs and outputs, deliberate about setting hyperparameters, and run an appropriate learning algorithm.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-66-09	E	0.50	KA01.02
-----------	---	------	---------

We recognize that the difficulties that exist in machine learning models can generally be more pronounced in these situations (e.g., time heterogeneity, data sparsity).

5. Deep Learning

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-67-01	T2	0.50	KA01.02
-----------	----	------	---------

Understand how multilayer neural networks (including non-deep networks) learn and encode input features into higher-order features.

ACM-67-02	T2	0.50	KA01.02
-----------	----	------	---------

Understand the architecture, objectives, and properties of common deep learning approaches such as forward propagation networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and LSTMs.

ACM-67-03	T2	0.50	KA01.02
-----------	----	------	---------

Understand the practical challenges of common deep learning approaches (e.g., choice of deep learning architecture, availability of sufficient data, possibility of overlearning, length of training time, interpretability).

ACM-67-04	T2	0.50	KA01.02
-----------	----	------	---------

Understand examples of regularization techniques in deep learning architectures, such as early termination, parameter sharing, and dropout.

ACM-67-05	T2	0.50	KA01.02
-----------	----	------	---------

Understand examples of methods to solve other challenges in deep learning, such as GPU-powered tools and distributed systems.

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-67-06	T2	0.50	KA01.02						
Be able to choose the right tool to handle the size of the data: in particular, to handle the demands of big data using deep learning tools that can run in parallel.									
ACM-67-07	T2	0.50	KA01.02						
Is aware of the availability of state-of-the-art deep learning tools.									
ACM-67-08	T2	0.50	KA01.02						
Understand one or more common learning algorithms that appear in the context of deep learning (e.g., how back propagation is used in forward propagation networks, how back propagation is used to learn higher-order features in convolutional networks, how b									
ACM-67-09	T2	0.50	KA01.02						
Understand what convolutional operations are and why they are useful (e.g., detecting perpendicular edges in an image).									
ACM-67-10	T2	0.50	KA01.02						
Understand pooling: understand examples of pooling functions such as max pooling and practical examples.									
ACM-67-11	T2	0.50	KA01.02						
Understand the challenge of long/short term data dependency in recurrent neural networks: Understand one or more solutions such as LSTM.									
ACM-67-12	E	0.50	KA01.02						
Understand deep generative models, such as adversarial generative networks (GANs), and the applications in which they are used.									
ACM-67-13	E	0.50	KA01.02						
Understand the practical challenges of these approaches (e.g., convergence, mode collapse).									
ACM-67-14	E	0.50	KA01.02						
Understand approaches to addressing and mitigating the above effects.									
<table border="1"> <thead> <tr> <th>Type</th> <th>Skill</th> </tr> </thead> <tbody> <tr> <td>SEQ#</td> <td>Priority</td> </tr> <tr> <td>Assigned Time (h)</td> <td>DS-BoK KA#</td> </tr> </tbody> </table>				Type	Skill	SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
Type	Skill								
SEQ#	Priority								
Assigned Time (h)	DS-BoK KA#								
ACM-67-15	T2	0.50	KA01.02						
Be able to determine the most appropriate type of deep learning approach for a given dataset and task.									
ACM-67-16	T2	0.50	KA01.02						
Be able to apply pre-trained models to a dataset using the deep learning toolset (e.g. PyTorch, Tensorflow).									

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-67-17 T2 0.50 KA01.02
 Can train models on datasets, including network settings, using the deep learning toolkit (e.g. PyTorch, Tensorflow).

ACM-67-18 E 0.50 KA01.02
 Be able to apply a generative approach (from scratch) for a specific goal using the deep learning toolset.

ACM-67-19 E 0.50 KA01.02
 Be able to configure the toolkit to work for a given system architecture.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-67-20	T1	0.25	KA01.02
Understand the potential negative consequences of using machine learning models that are difficult or impossible to interpret and explain.			

ACM-67-21	T1	0.25	KA01.02
Fully understand that there are a number of problems that can overwhelm the capabilities of deep learning.			

ACM-67-22	T1	0.25	KA01.02
Understands and is fully aware of the social and political concerns around deep faking.			

C: Modeling and Simulation

1. Sampling

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
DS-039	E	3.00	KA01.06

To be able to explain what a sample error is.

DS-040	E	4.50	KA01.06
--------	---	------	---------

To be able to outline the design of experiments.

2. Data Visualization

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
DS-041	E	4.50	KA01.06

To be able to outline the range of objectives in visualization (from simply supporting work in the field to dynamically displaying the relationships between elements in big data).

DS-042	T1	1.00	KA01.06
--------	----	------	---------

To be able to identify appropriate candidates for the vertical and horizontal axes in scatter plots.

DS-043	T1	1.00	KA01.06
--------	----	------	---------

To be able to select appropriate candidates for stratification (comparison axes), for example, when selecting attributes in a stacked vertical bar chart.

DS-044	T2	3.00	KA01.06
--------	----	------	---------

To be able to reduce the amount of data to an appropriate level by sampling and ensemble averaging.

DS-045	T2	1.38	KA01.06
--------	----	------	---------

Determine the appropriate information density (e.g., data ink ratio).

DS-046	T2	3.00	KA01.06
--------	----	------	---------

Understand the basics of axis expression to avoid unnecessary exaggeration (e.g., the reference point of the Y axis in a column chart should start at "0" and the axis should not be cut off).

IPJSJ Data Science Curriculum Standard (April 2021)

DS-047 T2 2.50 KA01.06
 Understand the effects of emphases and clearly inappropriate emphases (e.g., position and size expressions are more effective than color expressions for measurement data).

DS-048 T2 3.00 KA01.06
 Be able to use BI tools, spreadsheets, and other tools to create charts and graphs for one- to three-dimensional comparisons, depending on the purpose (comparison, composition, distribution, change, etc.).

DS-049 T2 2.00 KA01.06
 To be able to visualize changes in charts and graphs in a straightforward manner using animation (e.g., to show how a demographic histogram changes over time).

DS-050 T1 1.00 KA01.06
 To be able to visualize data as a sign that conveys the meaning of the data to people other than the data analysis department.

DS-051 T2 2.00 KA01.06
 To be able to explain the concept of Voronoi diagrams and how to use them.

DS-052 T1 2.00 KA01.06
 To be able to visualize multivariate comparisons by extending charts in one to three dimensions appropriately (e.g., parallel coordinates, scatterplot matrices, table lenses, heat maps).

DS-053 E 1.50 KA01.06
 Understand the importance of visualizing and viewing data in order to understand its properties.

DS-054 E 3.00 KA01.06
 Be able to select an appropriate representation method to find outliers.

DS-055 E 3.00 KA01.06
 Be able to identify the basic perspectives in data visualization (e.g., finding singularities, differences, trends, and relationships).

3. Analysis Process

Type	Skill
------	-------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
DS-056	E	3.50	KA01.06

To be able to select necessary data, analysis methods, visualization, etc., if the scope and contents of study are clearly defined.

4. Understanding and Verifying Data

Type	Skill			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
DS-057	E	3.00	KA01.06	Understand the message of numbers and graphs when they come in contact with statistical information in news articles, etc.
DS-058	T1	0.80	KA01.06	Be able to check single graphs for aggregation errors.
DS-059	T1	2.00	KA01.06	Be able to correctly verify data items and the quantity and quality of data under instructions, and explain the results.
DS-060	E	3.00	KA01.06	Understand the importance of considering the background of the data and not relying on it.
DS-061	E	3.00	KA01.06	Understand the importance of setting up the starting points for aggregation and comparison targets in order to correctly highlight facts from data.
DS-062	E	1.50	KA01.06	To be able to explain the triggers, timing, and frequency of occurrence of data that are usually used in business, and to understand basic statistics.
DS-063	E	3.00	KA01.06	To be able to aggregate data in accordance with the purpose of what the data is being aggregated for and what kind of knowledge is being obtained.
DS-064	E	3.00	KA01.06	To be able to extract direct implications from analyses and figures. (e.g., variability, significance, distribution trend, specificity, relevance, inflection point, high/low relevance)
DS-065	E	3.00	KA01.06	To be able to interpret the numerical values of distribution results objectively without being influenced by assumptions.

5. Language Processing

Type	Skill			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	

DS-071 E 1.50 KA01.06

To be able to perform typical cleaning processes for text data (e.g., lowercasing, number substitution, half-width conversion, symbol removal, stemming, etc.) appropriately according to the task.

DS-072 E 1.50 KA01.06

To be able to explain the concepts of morphological analysis and entailment analysis.

6. Image and Video Processing

Type	Skill
------	-------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

DS-073 E 1.50 KA01.06

Know the mechanism of digital representation of images and typical image formats.

DS-074 T1 2.00 KA01.06

To be able to perform appropriate color conversion and simple filter processing on images according to the purpose.

DS-075 T1 2.00 KA01.06

Be able to perform typical cleaning processes (resizing, battening, standardization, etc.) on image data as appropriate to the task.

DS-076 T2 1.50 KA01.06

Understand the mechanism of digital representation of video and typical video formats, and be able to use existing methods for extracting images from video.

D1: Fundamental Computer Science

1. Fundamental Computer Architecture

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-09-01	T1	0.50	KA02.01
Understand the concept of the “power barrier”.			

ACM-09-02	T1	0.50	KA02.01
Understand bits, bytes, and words.			

ACM-09-03	T1	0.50	KA02.01
Understand the representation of numerical data.			

ACM-09-04	T1	1.00	KA02.01
Understand CPUs and GPUs.			

ACM-09-05	T2	0.50	KA02.01
Understand the representation of non-numeric data.			

ACM-09-06	T2	0.50	KA02.01
Understand multi-core and multi-processing.			

ACM-09-07	T2	0.50	KA02.01
Understand the basic structure of Neumann architecture.			

ACM-09-08	T2	0.50	KA02.01
Understand parallel architectures (e.g., SIMD, MIMD).			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-09-09	T1	0.50	KA02.01
-----------	----	------	---------

Be able to explain the meaning of the “power barrier” from the perspective of further improving processor performance and promoting the use of parallel processing.

ACM-09-10	T1	1.00	KA02.01
-----------	----	------	---------

Explain the impact of fixed-length numerical expressions on accuracy.

ACM-09-11	T1	2.00	KA02.01
-----------	----	------	---------

Explain the role of the CPU (in comparison to the specialized purpose of the GPU).

ACM-09-12	T2	1.00	KA02.01
-----------	----	------	---------

Explain the internal representation of non-numeric data such as characters, strings, and images.

ACM-09-13	T2	0.50	KA02.01
-----------	----	------	---------

Explain the difference between multi-core and multi-processor systems.

ACM-09-14	T2	1.00	KA02.01
-----------	----	------	---------

Explain the structure of the classical von Neumann architecture and its major functional units.

ACM-09-15	T2	0.50	KA02.01
-----------	----	------	---------

Explain the concept of parallel processing beyond the classical von Neumann model.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-09-16	T1	1.00	KA02.01

Understand the advantages and limitations of data representation and processor speed in modern computing devices.

2. Fundamentals of Storage System

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-10-01	T1	1.00	KA02.01

Understand storage systems and technologies.

ACM-10-02	T1	1.00	KA02.01
-----------	----	------	---------

Understand registers, caches, and RAM.

ACM-10-03	T1	1.00	KA02.01
-----------	----	------	---------

Understand virtual memory.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-10-04	T1	1.00	KA02.01
-----------	----	------	---------

Explain the major memory technologies (e.g., SRAM, DRAM, flash, magnetic disk) and their relative cost and performance.

ACM-10-05	T1	1.00	KA02.01
-----------	----	------	---------

Explain how effective latency is reduced by the use of a storage hierarchy.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-10-06	T1	0.50	KA02.01
-----------	----	------	---------

Understand the tradeoffs between expensive, fast storage devices and inexpensive, slow storage devices.

3. Fundamentals of Operating System

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-11-01	T1	1.00	KA02.01
-----------	----	------	---------

Understand the role and purpose of operating systems (OS).

ACM-11-02	T1	0.50	KA02.01
-----------	----	------	---------

Understand the types of security threats and how to prevent them.

ACM-11-03	T2	0.50	KA02.01
-----------	----	------	---------

Understand OS with network functions, client-server OS, and distributed OS.

ACM-11-04 T2 0.50 KA02.01
 Understand reliability and availability.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-11-05	T1	1.00	KA02.01

Explain the purpose and functions of modern operating systems.

ACM-11-06 T1 1.00 KA02.01
 Identify potential threats to operating systems (e.g., software vulnerabilities, authentication issues, malware, etc.) and the types of security features designed to prevent them.

ACM-11-07 T2 1.00 KA02.01
 Describe networked, client-server, and distributed operating systems and how they differ from single-user operating systems.

ACM-11-08 T2 0.50 KA02.01
 Explain the importance of reliability and availability, and the fault-tolerance methods used to guarantee both.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-11-09	T1	0.50	KA02.01

Understand the important role of the OS in providing an interface between humans and system resources, and between system resources.

ACM-11-10 T1 0.50 KA02.01
 Be able to behave appropriately to avoid system attacks on the OS.

4. File System

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-12-01	T1	0.50	KA02.01

Understand files (data, metadata, manipulation, organization).

ACM-12-02	T1	0.50	KA02.01
Understand directories (content and structure).			

ACM-12-03	T1	0.50	KA02.01
Understand file protection.			

ACM-12-04	T2	0.50	KA02.01
Understand files (ordered files, unordered files).			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-12-05	T1	1.00	KA02.01
Compare and contrast the various approaches to file organization, recognizing their advantages and disadvantages.			

ACM-12-06	T1	0.50	KA02.01
Describe the levels of file protection and the mechanisms for setting them.			

ACM-12-07	T2	0.50	KA02.01
Compare and explain the differences between ordered and unordered files.			

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-12-08	T1	1.00	KA02.01
Understand the importance of good file organization and the importance of protecting files from inappropriate access.			

5. Network

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-13-01	T1	1.00	KA02.01
Understand the components of a network (host, router, switch, ISP, access point, firewall).			

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-13-02	T1	0.50	KA02.01																																						
Local Area Network (LAN): Topology of LAN (e.g., bus type, ring type)																																									
ACM-13-03	T1	1.00	KA02.01																																						
Understand the structure of the Internet (e.g., Internet Service Providers (ISPs), content providers).																																									
ACM-13-04	T2	0.50	KA02.01																																						
Understand circuit switched networks and packet switched networks.																																									
ACM-13-05	T2	0.50	KA02.01																																						
Understand the hierarchical network structure.																																									
ACM-13-06	T2	1.00	KA02.01																																						
Understand naming and name resolution methods (DNS, IP addresses, Uniform Resource Identifier (URI), etc.).																																									
ACM-13-07	T2	1.00	KA02.01																																						
Understand the basic protocols (TCP, IP).																																									
ACM-13-08	T2	0.50	KA02.01																																						
Understand HTTP/HTTPS as an application layer protocol.																																									
<table border="1"> <thead> <tr> <th>Type</th> <th>Skill</th> </tr> </thead> <tbody> <tr> <th>SEQ#</th> <th>Priority</th> <th>Assigned Time (h)</th> <th>DS-BoK KA#</th> </tr> <tr> <td>ACM-13-09</td> <td>T1</td> <td>1.00</td> <td>KA02.01</td> </tr> <tr> <td colspan="4">Name the major components of a computer network.</td> </tr> <tr> <td>ACM-13-10</td> <td>T1</td> <td>0.50</td> <td>KA02.01</td> </tr> <tr> <td colspan="4">Know that a LAN can be configured in various topologies.</td> </tr> <tr> <td>ACM-13-11</td> <td>T1</td> <td>1.00</td> <td>KA02.01</td> </tr> <tr> <td colspan="4">Explain the structure of the (high-level) Internet.</td> </tr> <tr> <td>ACM-13-12</td> <td>T2</td> <td>0.50</td> <td>KA02.01</td> </tr> <tr> <td colspan="4">Explain the difference between circuit switching and packet switching.</td> </tr> </tbody> </table>				Type	Skill	SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	ACM-13-09	T1	1.00	KA02.01	Name the major components of a computer network.				ACM-13-10	T1	0.50	KA02.01	Know that a LAN can be configured in various topologies.				ACM-13-11	T1	1.00	KA02.01	Explain the structure of the (high-level) Internet.				ACM-13-12	T2	0.50	KA02.01	Explain the difference between circuit switching and packet switching.			
Type	Skill																																								
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#																																						
ACM-13-09	T1	1.00	KA02.01																																						
Name the major components of a computer network.																																									
ACM-13-10	T1	0.50	KA02.01																																						
Know that a LAN can be configured in various topologies.																																									
ACM-13-11	T1	1.00	KA02.01																																						
Explain the structure of the (high-level) Internet.																																									
ACM-13-12	T2	0.50	KA02.01																																						
Explain the difference between circuit switching and packet switching.																																									
ACM-13-09	T1	1.00	KA02.01																																						
Name the major components of a computer network.																																									
ACM-13-10	T1	0.50	KA02.01																																						
Know that a LAN can be configured in various topologies.																																									
ACM-13-11	T1	1.00	KA02.01																																						
Explain the structure of the (high-level) Internet.																																									
ACM-13-12	T2	0.50	KA02.01																																						
Explain the difference between circuit switching and packet switching.																																									

ACM-13-13 T2 1.00 KA02.01
 Explain the hierarchical structure of a typical network architecture.

ACM-13-14 T2 0.50 KA02.01
 Identify similarities and differences between names and addresses in networks.

ACM-13-15 T2 1.00 KA02.01
 Explain how basic protocols such as TCP and IP work.

ACM-13-16 T2 1.00 KA02.01
 Explain how application layer protocols such as HTTPS work.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-13-17	T1	1.00	KA02.01

Understand the complexities of transmitting information over a network and how to mitigate problems that may occur during transmission.

6. Web and Web Programming

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-14-01	T1	0.50	KA02.01

Understand the relationship between the Internet and the World Wide Web.

ACM-14-02 T1 0.50 KA02.01
 Understand web application vulnerabilities and security attacks (e.g., SQL injection, distributed denial of service (DoS) attacks).

ACM-14-03 T2 0.50 KA02.01
 Understand the detection and countermeasures against security attacks.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-14-04 T1 1.00 KA02.01
 Understand the relationship between the Internet and the World Wide Web.

ACM-14-05 T1 2.00 KA02.01
 Design and implement a simple browser-based application.

ACM-14-06 T1 0.50 KA02.01
 Explain common web application vulnerabilities and security attacks.

ACM-14-07 T1 1.50 KA02.01
 Use web programming languages (e.g., HTML5, JavaScript, PHP, CSS).

ACM-14-08 T2 0.50 KA02.01
 Recognize and apply methods to protect against security attacks.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-14-09	T1	0.50	KA02.01
Recognize the potential risks of creating and using web applications in the most secure way possible.			

7. Compiler and Interpreter

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-15-01	T1	0.50	KA02.01
Understand programs that take (other) programs as input (interpreters, compilers, type checkers, document generators).			
ACM-15-02	T1	1.00	KA02.01
Understand interpretation and execution by interpreters, compilation to native code, and compilation to portable intermediate representations.			
ACM-15-03	T1	1.00	KA02.01
Understand syntax and parsing, semantics and evaluation.			

ACM-15-04 T1 0.50 KA02.01
 Understand examples of languages that can be classified as interpreted or compiled.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-15-05	T1	1.00	KA02.01
-----------	----	------	---------

To be able to explain how a program that processes another program handles the program as input data.

ACM-15-06 T1 0.50 KA02.01
 Explain the advantages and disadvantages of languages and code for interpreted and compiled languages.

ACM-15-07 T1 0.50 KA02.01
 Distinguish between syntax, parsing, semantics, and evaluation.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-15-08	T1	0.50	KA02.01
-----------	----	------	---------

Understand the speed trade-offs between interpreted and compiled code.

ACM-15-09 T1 0.50 KA02.01
 Understand the tradeoff between compiling to native code and compiling to a portable intermediate representation.

ACM-15-10 T1 0.50 KA02.01
 Understand the usefulness of interpreters in code development.

8. Algorithmic Thinking and Problem Solving

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-35-01	T1	0.25	KA04.01
-----------	----	------	---------

Definition of Algorithm

ACM-35-02 T1 0.25 KA04.01
 The importance of algorithms in the problem solving process.

ACM-35-03 T1 0.25 KA04.01
 At least one formal method for solving a problem.

ACM-35-04 T1 0.25 KA04.01
 Basic object-oriented design concepts and principles (abstraction, encapsulation and information hiding, separation of behavior and implementation)

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-35-05 T1 0.25 KA04.01
 The use of some formal notation other than code, such as flowcharts or pseudocode, to describe the solution to the problem.

ACM-35-06 T1 0.25 KA04.01
 Describe the problem solution using a formal notation other than code, such as a flowchart or pseudocode. Illustrate the flow of data, such as input, transformation, and output, using some formal means of problem solution, such as a data flow diagram.

ACM-35-07 T1 0.25 KA04.01
 Identify the inputs and outputs, such as data, hyperparameters, and user responses, that are essential to implementing a program that solves the problem.

ACM-35-08 T1 0.25 KA04.01
 Identify the data components and behaviors of multiple abstract data types. (See PDA Data Structures)

ACM-35-09 T2 0.25 KA04.01
 Use at least one formal method to solve a problem.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-35-10 T1 0.25 KA04.01
 Explain exactly what an algorithm is and what a program is. Algorithms are not the same as programs.

ACM-35-11 T1 0.25 KA04.01
 Understand accurately that there is a principled way to divide a large problem into implementable solutions and to express those solutions in some form.

9. Algorithm

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
------	----------	-------------------	------------	--

ACM-38-01	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand simple numerical algorithms such as computing the mean of a list of numbers and finding the minimum, maximum, and maximum frequency values.

ACM-38-02	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand sorting and searching.

ACM-38-03	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand linear search and binary search.

ACM-38-04	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand $O(n^2)$ sorting algorithms (insertion sort, etc.) and $O(n \log n)$ sorting algorithms (merge sort).

ACM-38-05	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand random choice algorithms for search and sort algorithms.

ACM-38-06	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand the potential usefulness of hash-based search and sort algorithms with respect to efficiency.

ACM-38-07	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand the properties of graphs: connectivity, mediation, centrality, etc.

ACM-38-08	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand graph algorithms.

ACM-38-09	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand basic algorithmic strategies, such as greedy and divide-and-conquer methods.

ACM-38-10	T1	0.50	KA01.02	KA01.03
-----------	----	------	---------	---------

Understand algorithms for solving systems of linear equations.

IP SJ Data Science Curriculum Standard (April 2021)

ACM-38-11 T2 0.75 KA01.02 KA01.03
Understand algorithms for combinatorial optimization problems.

ACM-38-12 T2 0.75 KA01.02 KA01.03
Understand heuristic optimization methods.

ACM-38-13 E 1.25 KA01.02 KA01.03
Understand hashing and hash functions.

ACM-39-01 T1 0.50 KA01.02 KA01.03
Understand the definitions of temporal and spatial computational complexity.

ACM-39-02 T1 0.50 KA01.02 KA01.03
Understand the difference between the best/average/worst case behavior of algorithms.

ACM-39-03 T1 0.50 KA01.02 KA01.03
Understand the trade-off between time and space computation.

ACM-39-04 T1 0.50 KA01.02 KA01.03
Understand the classification of algorithms.

ACM-39-05 T1 0.50 KA01.02 KA01.03
Deterministic/nondeterministic algorithms

ACM-39-06 T1 0.50 KA01.02 KA01.03
Time/space hierarchy

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK	KA#
ACM-38-14	T1	0.50	KA01.02	KA01.03
Apply simple numerical algorithms (e.g., calculate the mean, find the minimum, etc.).				

ACM-38-15 T1 0.50 KA01.02 KA01.03
Apply search and sort algorithms.

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-38-16	T1	0.50	KA01.02	KA01.03
Contrast the trade-offs between various array-based search and sorting algorithms.				
ACM-38-17	T1	0.50	KA01.02	KA01.03
Perform graph and tree traversal using the general framework of depth-first and breadth-first search.				
ACM-38-18	T1	0.50	KA01.02	KA01.03
Identify shortest paths in graphs and trees using efficient algorithms such as the greedy algorithm.				
ACM-38-19	T1	0.50	KA01.02	KA01.03
Apply linear equation solvers to appropriate problems.				
ACM-38-20	T2	0.75	KA01.02	KA01.03
Identify maximum and minimum flows in graphs and trees using efficient algorithms.				
ACM-38-21	T2	0.75	KA01.02	KA01.03
Use general algorithms for combinatorial optimization problems (e.g., branch-and-bound methods).				
ACM-38-22	T2	0.75	KA01.02	KA01.03
Apply heuristic optimization methods (particle swarm optimization, genetic algorithms, evolutionary computation).				
ACM-38-23	T2	0.75	KA01.02	KA01.03
Implement algorithms based on dynamic programming for appropriate problems.				
ACM-38-24	E	1.25	KA01.02	KA01.03
Implement and apply search and sort algorithms to distributed systems and data.				
ACM-38-25	E	1.25	KA01.02	KA01.03
Implement and compare hash functions.				
ACM-38-26	E	1.25	KA01.02	KA01.03
Implement algorithms for traversing graphs, finding shortest paths, and finding flows.				
ACM-38-27	E	1.25	KA01.02	KA01.03
Analyze random selection algorithms.				

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-39-07 T1 0.50 KA01.02 KA01.03
 Roughly compare the efficiency of algorithms (e.g., number of operations).

ACM-39-08 T1 0.50 KA01.02 KA01.03
 Be able to run algorithms on inputs of various sizes and compare their performance.

ACM-39-09 T1 0.50 KA01.02 KA01.03
 Provide examples that show how the choice of implementation and algorithm can affect the amount of time and space required for execution.

ACM-39-10 T1 0.50 KA01.02 KA01.03
 Explain how algorithms are related/connected to problem representation and data structures.

ACM-39-11 T2 0.75 KA01.02 KA01.03
 Formally apply a taxonomy to understand algorithms.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-38-28	T1	0.50	KA01.02 KA01.03

Be aware that there are often many different types of algorithms that can successfully deal with a given problem.

ACM-38-29 T1 0.50 KA01.02 KA01.03
 Recognize that the choice of algorithm can have a significant impact on execution efficiency.

ACM-38-30 T1 0.50 KA01.02 KA01.03
 Fully understands the impact of execution efficiency on stakeholders (customers, consumers, maintenance and operations personnel).

ACM-39-12 T1 0.50 KA01.02 KA01.03
 Understand that there can be trade-offs between temporal and spatial computational complexity, and be fully aware of the impact of these trade-offs on customers and users of the software.

10. Data Structure

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-37-01 T1 0.33 KA02.05

The purpose and use of basic data structures and abstract data types (lists, arrays, stacks, queues, strings, sets, records/structures, maps, hash tables).

ACM-37-02 T1 0.33 KA02.05

Basic matrix representation structures (sparse/dense, row, column): matrix representation types, advantages/disadvantages of basic matrix operations based on representation types

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-37-05	T1	1.00	KA02.05
-----------	----	------	---------

Appropriate selection of basic data structures in programming.

ACM-37-06 T1 1.00 KA02.05

Appropriate use of the standard data type library of a given programming language.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-37-11	T1	0.33	KA02.05
-----------	----	------	---------

Is thorough in implementation and data structure selection, and understands its impact on usage, efficiency (time and space), and readability.

11. Programming

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-36-01	T1	0.33	KA02.05
-----------	----	------	---------

Basic concepts: Variables and primitive data types

ACM-36-02	T1	0.33	KA02.05
-----------	----	------	---------

Basic Concepts: Expressions and Assignment Statements

ACM-36-03	T1	0.33	KA02.05
-----------	----	------	---------

Basic Concepts: Conditional Statements and Repeated Statements

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-36-04	T1	0.33	KA02.05
Basic Concepts: Recursive Functions			
ACM-36-05	T1	0.33	KA02.05
Basic Concepts: Functions and Parameter Passing			
ACM-36-06	T1	0.33	KA02.05
Basic Concepts: Simple Input and Output, Including Files and Static Data Sources			
ACM-36-07	T1	0.33	KA02.05
Basic Concepts: Exception Handling			
ACM-36-08	T1	0.33	KA02.05
Core Practices: Documentation			
ACM-36-09	T1	0.33	KA02.05
Core Practices: Testing			
ACM-36-10	T1	0.33	KA02.05
Core Practice: Version Control			
ACM-36-11	T1	0.33	KA02.05
Splitting a Program into Smaller Parts			
ACM-36-12	T1	0.33	KA02.05
Types of errors (syntax, logic, runtime), causes of errors, and how to deal with errors			
ACM-36-13	T1	0.33	KA02.05
How to query and parse data sources			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-36-18	T1	1.50	KA02.05

Be able to write programs that include the basic concepts and core practices described above.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-36-19	T1	1.00	KA02.05
Infer the execution of a code segment and give a clear overview of the computation.			
ACM-36-20	T1	1.50	KA02.05
Apply decomposition techniques to break down a program into smaller routines, etc.			
ACM-36-21	T1	1.00	KA02.05
Use appropriate techniques (database queries, API calls, regular expressions, etc.) to manipulate data in a given data source (database, spreadsheet, document, XML, etc.).			
ACM-36-22	T1	1.00	KA02.05
Be able to develop programs using recursion and iteration.			
ACM-36-23	T1	1.00	KA02.05
Use consistent documentation and programming style standards that contribute to the readability and maintainability of software.			
ACM-36-24	T1	1.00	KA02.05
Apply techniques for testing and debugging programs.			
DS-111	T1	4.00	KA02.06
Be able to programmatically implement data processing (extraction, processing, analysis, etc.) for handling small-scale structured data (CSV, RDB, etc.) based on design documents.			
DS-112	E	4.00	KA02.06
Design and implement programs using APIs to pass data in standard formats such as JSON and XML.			
DS-113	E	8.00	KA02.06
Be able to analyze data and create reports using interactive development environments such as Jupyter Notebook and Rstudio.			
DS-114	E	8.00	KA03.02
Know, write, and execute SQL syntax (DML, DDL, use of JOINS, aggregate functions and GROUP BY, vertical and horizontal conversion using CASE statements, use of subqueries and EXISTS, etc.).			
Type Disposition			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-36-35	T1	0.33	KA02.05
Demonstrate a strong commitment to using software engineering concepts and design principles in programming practice.			

ACM-36-36	T1	0.33	KA02.05
Be prepared to go beyond what is directly taught. Understand that programming constructs and methods are general and useful in many contexts.			

ACM-36-37	T1	0.33	KA02.05
Have the attitude to discover ways to go beyond simple methods. The data scientist should not be limited to reworking existing solutions.			

12. Numerical Computation

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-40-01	T1	0.33	KA02.05
Random Number Generator			

ACM-40-02	T1	0.33	KA02.05
Simulation of probability distributions			

ACM-40-03	T1	0.33	KA02.05
Limitations of bit-based numerical representations and their impact on error accumulation (overflow, underflow, roundoff, truncation) of results			

ACM-40-04	T1	0.33	KA02.05
Impact of numerical representation on computational complexity.			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-40-07	T1	0.33	KA02.05
Explain how numerical algorithms and processes affect the execution of simulations, data sampling, and data generation.			

ACM-40-08	T1	0.33	KA02.05
Describe appropriate numerical algorithms for performing data analysis, recognizing their limitations and numerical constraints.			

ACM-40-09	T1	0.33	KA02.05
Effectively use random number generators and simulated probability distributions to achieve reproducibility in data analysis using nondeterministic algorithms.			

ACM-40-10 T1 0.33 KA02.05

Use random number generators and simulated probability distributions to achieve reproducibility in data analysis using nondeterministic algorithms that satisfy appropriate statistical and numerical constraints.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-40-12	T1	0.33	KA02.05
-----------	----	------	---------

Sensitive to the advantages and limitations of (pseudo-)random number generation.

ACM-40-13 T1 0.33 KA02.05

Sensitive to the limitations of numerical algorithms.

D2: Big Data System

1. Problems of Scale

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-26-01	T1	0.50	KA02.01
-----------	----	------	---------

Understand the need for measurement in the context of big data, such as size, capacity, and timing.

ACM-26-02	T1	0.50	KA02.01
-----------	----	------	---------

Understand the concept of problem size.

ACM-26-03	T1	0.50	KA02.01
-----------	----	------	---------

Understand the impact of rapid growth in computational processing.

ACM-26-04	T1	0.50	KA02.01
-----------	----	------	---------

Understand the impact of rapid data growth on storage systems.

ACM-26-05	T1	0.50	KA02.01
-----------	----	------	---------

Understand the need to focus on simplicity.

ACM-26-06	T1	0.50	KA02.01
-----------	----	------	---------

Understand problem solving approaches to deal with increasing number of agents/processes.

ACM-26-07	T1	0.50	KA02.01
-----------	----	------	---------

Understand approaches to solving scale problems while addressing scalability.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-26-08	T1	0.50	KA02.01
-----------	----	------	---------

Explain the reasons for big data applications leading to increased complexity and the nature of that complexity.

ACM-26-09	T1	0.50	KA02.01
Explain why it is important to emphasize simplicity while not being overly simplistic.			

ACM-26-10	T1	0.50	KA02.01
Describe common steps to reduce complexity.			

ACM-26-11	T1	1.00	KA02.01
Evaluate the data scale and speed of an application based on its description.			

ACM-26-12	T2	0.50	KA02.01
Perform computational tasks at multiple levels of scale.			

Type	Disposition
------	-------------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-26-13	T1	0.50	KA02.01
Respond appropriately to difficulties caused by scale.			

ACM-26-14	T1	0.50	KA02.01
Behave appropriately to solve scale problems.			

2. Big Data Computing Architectures

Type	Knowledge
------	-----------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-27-01	E	0.50	KA02.02
Mechanisms to support efficient and fast I/O.			

ACM-27-02	E	1.00	KA02.02
Concepts and requirements for data-oriented high-performance computing			

ACM-27-03	E	2.00	KA02.02
Consideration and understanding of memory and caches, including cache coherence.			

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-27-04	E	2.00	KA02.02
Understanding the advantages and limitations of various parallel computers, such as multicore, grid computing, GPUs, shared memory, distributed memory, shared memory computing, vector processing, etc.			

ACM-27-05	E	0.50	KA02.02
Classification of parallel computers by Flynn			

ACM-27-06	E	1.00	KA02.02
Understanding of implementation schemes to support parallel processing			

ACM-27-07	E	1.00	KA02.02
Hierarchical structure of parallel storage			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-27-08	E	1.00	KA02.02
Understanding of Various Approaches to Realize High Speed I/O			

ACM-27-09	E	1.00	KA02.02
A (qualitative) description of the challenges in achieving high speed I/O			

ACM-27-10	E	2.00	KA02.02
Comparison and contrast of various parallel computer architectures.			

ACM-27-11	E	2.00	KA02.02
A (qualitative) explanation of which parallel computer architecture is appropriate for an application.			

ACM-27-12	E	2.00	KA02.02
Selecting the appropriate system architecture for a given computational model or framework based on computational patterns and data characteristics			

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-27-13	E	0.50	KA02.02
Knowledge of hardware trends that support data science applications			

3. Parallel Computing Frameworks

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-28-01	E	0.50	KA02.01
-----------	---	------	---------

Purpose and definition of parallel computing models

ACM-28-02	E	1.00	KA02.01
-----------	---	------	---------

Classification of models

ACM-28-03	E	2.00	KA02.01
-----------	---	------	---------

Distributed systems

ACM-28-04	E	2.00	KA02.01
-----------	---	------	---------

Grid search (parameter tuning)

ACM-28-05	E	2.00	KA02.01
-----------	---	------	---------

Iterative processes for communication and coordination (in data processing)

ACM-28-06	E	1.50	KA02.01
-----------	---	------	---------

Understanding the challenges of task-oriented and data-parallel work partitioning

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-28-07	E	1.00	KA02.01
-----------	---	------	---------

Modeling of parallel computing systems

ACM-28-08	E	2.00	KA02.01
-----------	---	------	---------

Evaluating efficient and effective parallel computing

ACM-28-09	E	2.00	KA02.01
-----------	---	------	---------

Methods for designing and developing parallel systems for processing large data sets

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-28-10	E	1.00	KA02.01
Familiarity with the evaluation and design of complex (potentially) systems			

4. Distributed Data Storage

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-29-01	T2	0.50	KA02.02
Methods for storing large amounts of data in storage consisting of a wide variety of devices			

ACM-29-02	T2	0.50	KA02.02
Hierarchical storage			

ACM-29-03	T2	1.00	KA02.02
Understanding of clean, consistent, and typical data			

ACM-29-04	T2	1.00	KA02.02
Data maintenance and protection			

ACM-29-05	T2	0.50	KA02.02
Data acquisition and extraction challenges			

ACM-29-06	T2	1.00	KA02.02
Understanding the strengths and limitations of techniques using hashing, filtering, and sampling			

ACM-29-07	T2	1.00	KA02.02
Data backup			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-29-08 T2 1.00 KA02.02
 Understanding the role of storage hierarchies in Big Data

ACM-29-09 T2 1.00 KA02.02
 An overview of the benefits of redundancy in Big Data

ACM-29-10 T2 1.00 KA02.02
 Explain how excessive redundancy in Big Data can compromise system efficiency.

ACM-29-11 T2 1.00 KA02.02
 Describe methods for protecting and managing data to ensure it is useful and up-to-date for Big Data applications

ACM-29-12 E 2.00 KA02.02
 Develop a distributed data storage system that can be easily scaled.

ACM-29-13 E 2.00 KA02.02
 Design storage systems that take into account techniques such as backup, migration, and compression in data-oriented systems to ensure scalability, usability, efficiency, and security.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-29-14	T2	1.00	KA02.02

Creation of beneficial arrangements in the design of storage mechanisms to support big data applications.

5. Parallel Programming

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-30-01	T2	0.50	KA02.03

Concurrency control and parallel processing and distributed systems

ACM-30-02 T2 0.50 KA02.03
 Limitations of Parallel Processing due to Overhead

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-30-03 T2 1.00 KA02.03
 Understanding different approaches to concurrency control and parallel processing.

ACM-30-04 T2 1.00 KA02.03
 Understanding of parallel algorithms with respect to methods adapted to hardware architectures, including load balancing.

ACM-30-05 T2 0.50 KA02.03
 Understanding of typical parallel programming paradigms such as MapReduce.

ACM-30-06 T2 0.50 KA02.03
 Understanding the complexity of parallel/parallel algorithms

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-30-07 T2 0.50 KA02.03
 Explain the limitations of concurrency/parallelism with respect to scale

ACM-30-08 T2 0.50 KA02.03
 Understanding of parallel processing overhead for specific algorithms

ACM-30-09 T2 1.00 KA02.03
 Understanding of implementation schemes for data oriented parallel programs

ACM-30-10 T2 2.00 KA02.03
 Development and operation of data-oriented parallel computing systems for data processing and data scale

ACM-30-11 T2 2.00 KA02.03
 Development and optimization of data-oriented parallel programs

ACM-30-12 T2 2.00 KA02.03
 Design, implementation and tuning of algorithms based on parallel programming paradigms

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-30-13	T2	0.50	KA02.03
Recognition of excessive parallel processing overhead in specific cases			

ACM-30-14	T2	0.50	KA02.03
Understanding and deepening the understanding of parallel systems in conjunction with case studies			

6. Techniques for Big Data Applications

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-31-01	T2	0.50	KA02.05
Necessity of technologies to support big data processing			

ACM-31-02	T2	0.50	KA02.05
Hash processing (hashing)			

ACM-31-03	T2	0.50	KA02.05
Sampling, Filtering			

ACM-31-04	T2	0.50	KA02.05
Data characteristics and overview			

ACM-31-05	T2	0.50	KA02.05
Limitations of hashing, sampling, and filtering			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-31-06	T2	0.50	KA02.05
Illustrate the role of hashing in big data			

ACM-31-07	T2	0.50	KA02.05
Explain the criteria for using sampling and filtering			

ACM-31-08	T2	1.00	KA02.05
Practice sample selection in applications using big data by following the specified guidelines.			

ACM-31-09	T2	0.50	KA02.05
Discuss various filtering methods based on specific use cases.			

ACM-31-10	T2	2.00	KA02.05
Design data structures and analyze their performance based on available storage capacity and expected accuracy.			

Type	Disposition
------	-------------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-31-11	T2	0.50	KA02.05
Awareness of pitfalls such as bias in sampling and filtering			

7. Cloud Computing

Type	Knowledge
------	-----------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-32-01	T2	0.50	KA02.03
Understand cloud computing and its benefits			
ACM-32-02	T2	0.50	KA02.03
Data center architecture			
ACM-32-03	T2	0.50	KA02.03
Challenges related to cloud computing			
ACM-32-04	T2	0.50	KA02.03
Understand the different methods that support cloud computing			
ACM-32-05	T2	1.00	KA02.03
Distributed File Systems			

IPSJ Data Science Curriculum Standard (April 2021)

ACM-32-06 T2 0.50 KA02.03
Cloud services to support big data applications

ACM-32-07 T2 1.00 KA02.03
Virtualization technologies

ACM-32-08 T2 1.00 KA02.03
Security issues related to cloud computing, cloud storage, virtual machines and other cloud technologies

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-32-09 T2 1.00 KA02.03
Understanding the overview of tasks to be performed in a cloud system

ACM-32-10 T2 2.00 KA02.03
Data center design

ACM-32-11 T2 2.00 KA02.03
Understanding of the various cloud services offered to support big data applications

ACM-32-12 T2 2.00 KA02.03
Selection and application of cloud services to support specific big data applications

ACM-32-13 T2 2.00 KA02.03
Design a security strategy for the cloud

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-32-14 T2 1.00 KA02.03
Responsible behavioral attitude towards the use of cloud services

ACM-32-15 T2 1.00 KA02.03
Understanding the different cloud services currently available such as Amazon, Google, Microsoft, etc.

8. Complexity Theory

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-33-01	E	1.00	KA02.01
-----------	---	------	---------

Issues related to efficiency and computational complexity of algorithms

ACM-33-02	E	1.00	KA02.01
-----------	---	------	---------

Understanding of computational complexity in terms of its importance in terms of big data and its use in terms of concurrency control/parallel processing

ACM-33-03	E	0.50	KA02.01
-----------	---	------	---------

Limitations on the concept of complexity

ACM-33-04	E	1.00	KA02.01
-----------	---	------	---------

Evaluation of complexity in commonly used algorithms including concurrency control/parallel processing

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-33-05	E	1.00	KA02.01
-----------	---	------	---------

Explanation of why mathematical analysis alone is not sufficient for efficiency considerations.

ACM-33-06	E	2.00	KA02.01
-----------	---	------	---------

Identify problems in terms of computer resource limitation, time limitation, data size, etc., and analyze whether the problems can be solved or approximately solved within the given conditions in terms of complexity.

ACM-33-07	E	2.00	KA02.01
-----------	---	------	---------

Provide specific examples of how to evaluate the efficiency of algorithms used in big data processing.

ACM-33-08	E	2.00	KA02.01
-----------	---	------	---------

Select appropriate algorithms for specific applications involving big data and explain the scaling challenges.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-33-09 E 0.50 KA02.01
 Foster a knowledge of complexity and a proactive attitude toward it

ACM-33-10 E 0.50 KA02.01
 Awareness of the limitations of complexity

9. Software Support for Big Data Applications

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-34-01	T2	1.00	KA02.05
-----------	----	------	---------

Understand the need for a programming environment to support big data applications.

ACM-34-02	T2	1.00	KA02.05
-----------	----	------	---------

Concepts of auto-scaling and serverless computing

ACM-34-03	T2	1.00	KA02.05
-----------	----	------	---------

Consideration of the potential for more sophisticated web services to support data transfer, analysis, and machine learning in the context of Big Data

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-34-04	T2	2.00	KA02.05
-----------	----	------	---------

Compare and contrast the use of auto-scaling and serverless computing.

ACM-34-05	T2	2.00	KA02.05
-----------	----	------	---------

Understand the relationship between auto-scaling and load balancing.

ACM-34-06	T2	1.00	KA02.05
-----------	----	------	---------

Outline the importance of buffer size in streaming applications.

ACM-34-07	T2	2.00	KA02.05
-----------	----	------	---------

Understand web services that facilitate face recognition and video streaming applications.

IP SJ Data Science Curriculum Standard (April 2021)

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-34-08	T2	0.50	KA02.05
-----------	----	------	---------

Recommend a reflective approach to the use of web services, taking into account possible biases, or lack thereof.

ACM-34-09	T2	0.50	KA02.05
-----------	----	------	---------

Deepening the understanding of Big Data applications

ACM-34-10	T2	0.50	KA02.05
-----------	----	------	---------

In the context of big data applications, we recommend a focus on appropriate simplicity.

E1: Security and Privacy

1. Fundamentals of Security

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-08-01	T1	0.27	KA03.03

Have a strong sense of responsibility for handling the role of data cleaning in data use. or between governments.

ACM-08-02	T1	0.27	KA03.03
-----------	----	------	---------

Laws and responsibilities regarding transnational privacy and data security.

ACM-08-03	T1	0.27	KA03.03
-----------	----	------	---------

An understanding of the extent to which internationally engaged organizations should consider differences in privacy laws, regulations, and standards across the jurisdictions in which they operate.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-08-04	T1	0.27	KA03.03

An explanation of how law and technology can help organizations protect their information systems from cyber attacks in the context of international, national, and local jurisdictional structures.

ACM-08-05	T1	0.27	KA03.03
-----------	----	------	---------

Explain the requirements of the General Data Protection Regulation (GDPR) and the privacy agreements between nations, such as between the United States and the United Kingdom, that allow for the transfer of personal data.

ACM-08-06	T1	0.27	KA03.03
-----------	----	------	---------

Explain how Section 5 of the US Federal Trade Commission, state data security laws, state data breach notification laws, Health Insurance Portability and Accountability Act (HIPAA), Gramm-Leach-Bliley Act (GLBA), and US-CERT information sharing, the Cyber

DS-115	T1	1.00	KA02.04
--------	----	------	---------

Explain the three elements of security (confidentiality, availability, and integrity) using concrete examples.

DS-116	T1	0.50	KA02.04
--------	----	------	---------

Always be aware of the types of serious risks (loss, leakage, service outage, etc.) caused by malware, etc.

IPSSJ Data Science Curriculum Standard (April 2021)

DS-117	E	1.50	KA02.04
Be able to set the access level for each user to OS, network, application, and data according to the procedure.			
DS-118	T1	2.00	KA02.04
Understand that unencrypted data can be easily misused by unauthorized persons, and be able to encrypt and decrypt data using software according to the degree of confidentiality.			
DS-119	E	0.50	KA02.04
Understand that digital signatures and public key infrastructure (PKI) are necessary to prove that a document is not spoofed or forged.			
DS-120	E	2.00	KA02.04
Be able to use hash functions to detect data tampering.			

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-08-07	T1	0.27	KA03.03
Act ethically in data governance policies and actions.			
ACM-08-08	T1	0.27	KA03.03
Be comfortable with the harm of data loss resulting from security and privacy failures.			
ACM-08-09	T1	0.27	KA03.03
Maintain the highest ethical standards regarding legal and social responsibility for data.			

2. Privacy Protection

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-43-01	T1	0.33	KA02.04
Sensitive data exposed by social engineering and social media.			
ACM-43-02	T1	0.33	KA02.04
Trade-off between privacy rights and the need for transparency in information dissemination.			

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-43-03 T1 0.33 KA02.04
 Ethical responsibilities regarding the disclosure, transmission, and sharing of information obtained using analytical tools.

ACM-44-01 T1 0.33 KA02.04
 The importance of encrypting data before transmitting it over any channel.

ACM-44-02 T1 0.33 KA02.04
 The tradeoff in computational time between using encrypted and unencrypted data for statistical analysis.

ACM-45-01 T1 0.33 KA02.04
 Concepts and techniques for achieving authentication, authorization, access control, and data privacy.

ACM-45-02 T1 0.33 KA02.04
 Hierarchical defense systems to maximize confidentiality, integrity, and availability.

ACM-46-01 T1 0.33 KA02.04
 Importance of security protocols to enable secure communication over insecure channels.

ACM-46-02 T1 0.33 KA02.04
 Importance of privacy protocols that allow individual communication over secure transmission channels.

ACM-46-03 T1 0.33 KA02.04
 Internet/communication protocols that can guarantee confidential communication between applications and servers.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-43-06 T1 0.33 KA02.04
 Demonstrate an understanding of the data sensitivities that arise during data entry.

ACM-43-07 T1 0.33 KA02.04
 Scenarios in which data cleaning must be considered prior to information processing.

ACM-43-08 T1 0.33 KA02.04
 Understand techniques for ensuring data privacy when processing raw data, such as data scoping and protection techniques.

ACM-44-06	T1	0.33	KA02.04
Identify cryptographic tools and mechanisms to reduce the risk of data leakage while considering computational cost.			
ACM-44-07	T1	1.00	KA02.04
Train a variety of audiences, including individuals, organizations, and government agencies, on data encryption processes that impact privacy requirements.			
ACM-44-08	T1	0.33	KA02.04
Illustrate the use of cryptographic techniques to provide privacy, such as message authentication codes, digital signatures, authenticated encryption, and hash trees.			
ACM-44-09	T1	0.33	KA02.04
Understand the tradeoffs between the cost of processing simple text data and encrypted data.			
ACM-45-06	T1	0.33	KA02.04
Explain how data privacy in a system can affect the security of the system.			
ACM-45-07	T1	1.00	KA02.04
Discuss the trade-off between data transparency and data privacy.			

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-43-11	T1	0.33	KA02.04
An understanding of the ethical aspects of providing data to any organization (entity) that may lead to a violation of data privacy.			
ACM-43-12	T1	0.33	KA02.04
Understands that proper data handling raises both public and private issues, and handles data accurately and ethically through computers and transmission channels.			
ACM-44-11	T1	0.33	KA02.04
Be sensitive to the need for different methods of encryption.			
ACM-45-09	T1	0.33	KA02.04
Act cautiously to protect the information in a given computer system.			

3. Data Security

IP SJ Data Science Curriculum Standard (April 2021)

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-47-01	T1	0.33	KA02.04
Qualitative Indicators			

ACM-47-02	T1	0.33	KA02.04
Importance of security for data assets			

ACM-47-03	T1	0.33	KA02.04
Different types of security objectives needed			

ACM-47-04	T1	0.33	KA02.04
Data sources and assets			

ACM-47-05	T1	0.33	KA02.04
Access control and access management to data assets			

ACM-50-01	T1	0.33	KA02.04
Insight into data transactions over the network in data-driven applications.			

ACM-50-02	T1	0.33	KA02.04
Network and Web Protocols			

ACM-50-03	T1	0.33	KA02.04
Available and (or) operational security modules in communication protocols			

ACM-50-04	T1	0.33	KA02.04
Operations (storage, retrieval, remote computation) on data networks and the Web			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-47-09	T1	0.33	KA02.04
Understand the flow of data in an application.			

ACM-47-10	T1	1.00	KA02.04
Derive the key security objectives to be achieved.			
ACM-47-11	T1	1.00	KA02.04
Explain which data assets should be protected and why.			
ACM-50-05	T1	1.00	KA02.04
Analyze and adjust communication protocols to achieve security.			
ACM-50-06	T1	0.33	KA02.04
Explain the properties and operating principles of network and Web protocols.			
ACM-50-07	T1	0.33	KA02.04
Understand how to perform data communication between various entities on the network and the Web.			

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-50-08	T1	0.33	KA02.04
Commitment to security measures in network/web protocols.			

4. Data Integrity

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-51-01	T1	0.33	KA02.04
Concept of logical consistency			
ACM-51-02	T1	0.33	KA02.04
Types of consistency constraints in database systems			
ACM-51-03	T1	0.33	KA02.04
Entity Consistency, Referential Consistency, Domain Consistency, User Defined Consistency			

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-52-01	T1	0.33	KA02.04																														
Concepts of Physical Consistency																																	
ACM-52-02	T1	0.33	KA02.04																														
Physical and hardware methods to ensure data consistency, such as RAID, redundant hardware, uninterruptible power supplies, error correcting memory, and serial clusters.																																	
ACM-53-01	T1	0.33	KA02.04																														
Common threats to data consistency including human error, software error, transmission error, malware, insider threats, cyber attacks, and compromised hardware.																																	
ACM-53-02	T1	0.33	KA02.04																														
Data and information tampering																																	
ACM-53-03	T1	0.33	KA02.04																														
Data provenance assurance																																	
ACM-54-01	T1	0.33	KA02.04																														
The role of hashing algorithms in maintaining consistency.																																	
ACM-54-02	T1	0.33	KA02.04																														
The role of message authentication codes (MACs) and their derivatives.																																	
ACM-54-03	T1	0.33	KA02.04																														
Cyclic Redundancy Check (CRC) and checksum for consistency.																																	
ACM-54-04	T1	0.33	KA02.04																														
Digital signature mechanisms (RSA and ECDSA)																																	
<table border="1"> <thead> <tr> <th>Type</th> <th>Skill</th> <th>Priority</th> <th>Assigned Time (h)</th> <th>DS-BoK KA#</th> </tr> </thead> <tbody> <tr> <td>SEQ#</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ACM-51-04</td> <td></td> <td>T1</td> <td>0.33</td> <td>KA02.04</td> </tr> <tr> <td colspan="5">Be able to explain the concept of logical consistency.</td> </tr> <tr> <td>ACM-52-03</td> <td></td> <td>T1</td> <td>0.33</td> <td>KA02.04</td> </tr> <tr> <td colspan="5">Explain the concept of physical consistency.</td> </tr> </tbody> </table>				Type	Skill	Priority	Assigned Time (h)	DS-BoK KA#	SEQ#					ACM-51-04		T1	0.33	KA02.04	Be able to explain the concept of logical consistency.					ACM-52-03		T1	0.33	KA02.04	Explain the concept of physical consistency.				
Type	Skill	Priority	Assigned Time (h)	DS-BoK KA#																													
SEQ#																																	
ACM-51-04		T1	0.33	KA02.04																													
Be able to explain the concept of logical consistency.																																	
ACM-52-03		T1	0.33	KA02.04																													
Explain the concept of physical consistency.																																	

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-52-04 T1 0.33 KA02.04
 Explain the physical and hardware methods to ensure physical consistency.

ACM-53-04 T1 0.33 KA02.04
 List the common types of security threats that affect data integrity.

ACM-53-05 T1 0.33 KA02.04
 Describe the potential vulnerabilities of different hash functions, such as SHA-1 and MD5.

ACM-54-05 T1 0.33 KA02.04
 Explain how hash algorithms and MAC methods can be used to establish data consistency.

ACM-54-06 T1 0.33 KA02.04
 Explain digital signature schemes and their necessity in terms of consistency.

ACM-54-07 T1 0.33 KA02.04
 Compare and contrast different consistency preservation techniques in terms of performance and security.

ACM-54-08 T1 0.33 KA02.04
 Understand how consistency models can be used to ensure the origin and validity of data in domains with multiple data owners.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-51-05	T1	0.33	KA02.04
-----------	----	------	---------

Be able to explain logical consistency with confidence.

ACM-52-05 T1 0.33 KA02.04
 Be confident in using hardware methods to address physical consistency.

ACM-53-06 T1 0.33 KA02.04
 Confidently describe common security threats.

ACM-54-09 T1 0.33 KA02.04
 Thoroughly understand how to address data integrity using various methods and techniques.

5. Analysis for Security

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-56-01	T1	0.33	KA02.04
-----------	----	------	---------

Statistical methods for exploratory data analysis of security data, including descriptive statistics, summary plots, outlier detection, point estimation, hypothesis testing, test statistics, linear regression, and generalized linear regression.

ACM-56-02	T1	0.33	KA02.04
-----------	----	------	---------

Computer vision-based methods for static or dynamic (security) threat classification and malware detection, including malware display techniques, transfer learning, and hierarchical ensemble neural networks (HeNet).

ACM-57-01	T1	0.33	KA02.04
-----------	----	------	---------

Basic concepts on adversarial machine learning, types of attacks on machine learning, and frameworks for defending machine learning from attacks.

ACM-57-02	T1	0.33	KA02.04
-----------	----	------	---------

Adversarial machine learning techniques such as fast gradient codes, iterative fast gradients, universal paradoxical perturbations, etc.

ACM-57-03	T1	0.33	KA02.04
-----------	----	------	---------

Defensive techniques such as adversarial training etc. to protect machine learning models

ACM-57-04	T1	0.67	KA02.04
-----------	----	------	---------

1. Machine learning techniques with explanatory power for security applications. Explanations include local explanations on a sample-by-sample basis and global explanations that consider the entire dataset. 2. Know how to apply model-independent explanat

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-56-03	T1	1.00	KA02.04
-----------	----	------	---------

Security applications can be transformed into problems that can leverage machine learning.

ACM-56-04	T1	1.00	KA02.04
-----------	----	------	---------

Design malware detection solutions with static or dynamic (security) detection mechanisms using malware display techniques, transfer learning, hierarchical ensemble neural networks (HeNet), etc.

ACM-56-05	T1	0.33	KA02.04
-----------	----	------	---------

Explain the results of machine learning models in security applications to people with different backgrounds.

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-57-05	T1	1.00	KA02.04
Recognize that there are blind spots where detection does not occur and evaluate the resilience of machine learning.			
ACM-57-06	T1	1.00	KA02.04
Improve the resilience of machine learning by introducing adversarial training.			
ACM-57-07	T1	0.33	KA02.04
Investigate machine learning algorithms and explain the effectiveness of the models to security experts.			
ACM-57-08	T1	1.00	KA02.04
Define machine learning metrics and communicate with various stakeholders to address accountability and vulnerability issues.			
ACM-57-09	T1	0.33	KA02.04
Explain why resilience and vulnerability are important for machine learning used in security and privacy applications.			
ACM-57-10	T1	1.00	KA02.04
Apply AI methods with explanatory power such as LIME, LEMNA, and TCAV to security applications.			
ACM-57-11	T1	1.00	KA02.04
Apply AI methods with explanatory power, such as LIME, to machine learning models built for security applications, and be able to select models based on their reliability scores.			
Type Disposition			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-56-06	T1	0.33	KA02.04
Understand the various perspectives of computer vision, natural language processing, classical data analysis, etc. for approaches to threat detection and malware intelligence and their recognition.			
ACM-57-12	T1	0.67	KA02.04
Evaluate machine learning not only for accuracy of classification, false positive rate, and precision, but also for robustness and vulnerability.			

E2: Human-Computer Interaction

1. General

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-68-01	T1	0.33	KA02.05
-----------	----	------	---------

Contexts for considering human-computer interfaces: data visualization, web pages, multimedia materials, educational materials, and general computing environments considering navigation

ACM-68-02	T1	0.33	KA02.05
-----------	----	------	---------

Theories, models, principles, guidelines, and standards applicable to the design and implementation of user interfaces

ACM-68-03	T1	0.33	KA02.05
-----------	----	------	---------

Different measures for interfaces: effectiveness and attractiveness

ACM-68-04	T1	0.33	KA02.05
-----------	----	------	---------

Color and multimedia, ergonomics and the use of web services

ACM-68-05	T1	0.33	KA02.05
-----------	----	------	---------

Cognitive models that influence interaction

ACM-68-06	T1	0.33	KA02.05
-----------	----	------	---------

Scope, advantages and disadvantages of augmented reality (AR)

ACM-68-07	T1	0.33	KA02.05
-----------	----	------	---------

Software support to aid data understanding through analysis and visualization

ACM-68-08	T1	0.33	KA02.05
-----------	----	------	---------

Accessibility considerations for different groups of users, including those with special needs

ACM-69-01	T1	0.33	KA02.05
-----------	----	------	---------

The role of visualization in data science

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-69-02	T1	0.33	KA02.05
Visualization with diagrams (including historical and contemporary examples)			
ACM-69-03	T1	0.33	KA02.05
Characteristics of effective visualization			
ACM-69-04	T1	0.33	KA02.05
Suitability of different techniques for different data and users			
ACM-69-05	T1	0.33	KA02.05
Dashboards and interactive visualization			
ACM-69-06	T1	0.33	KA02.05
Software to support visualization			
ACM-69-07	T1	0.33	KA02.05
Visualization-based reasoning			
ACM-69-08	T1	0.33	KA02.05
Preparation for visualization – scaling, role of colors			
ACM-69-09	T1	0.33	KA02.05
Various graphs: tables, scatter plots, pie charts, histograms, graphs, data maps (including representations using pixels, glyphs, graphs, maps)			
Type Skill			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-68-09	T1	1.00	KA02.05
Can justify the adoption of a user-centric approach to analyzing and displaying data.			
ACM-68-10	T1	1.00	KA02.05
Critique how considerations such as attention, perception, cognition, speech, and behavior affect the usability of an interface in different situations.			
ACM-68-11	T1	1.00	KA02.05
Explain how formal documents (theories, models, guidelines, etc.) affect the analysis and visualization of data.			

IPSJ Data Science Curriculum Standard (April 2021)

ACM-68-12	T1	1.00	KA02.05
Explain the desired impact of the interface on users with disabilities and different age groups (including children).			
ACM-68-13	T1	1.00	KA02.05
Outline the ways in which bias can be recognized in an interface.			
ACM-68-14	T1	1.00	KA02.05
Outline the range of software that can be employed to assist in the analysis and presentation of data.			
ACM-68-15	T1	1.00	KA02.05
Explain the added value and possibilities that AR can bring to an interface.			
ACM-69-10	T1	0.50	KA02.05
Explain some well-known examples of commonly used visualizations.			
ACM-69-11	T1	0.50	KA02.05
Identify the various roles that visualization plays in data science.			
ACM-69-12	T1	1.00	KA02.05
Given a set of data that needs to be used for a specific purpose, implement an effective visualization for it.			
ACM-69-13	T1	0.50	KA02.05
Explain the role of visualization in data classification and identify methods to facilitate it.			
ACM-69-14	T1	1.00	KA02.05
Be able to implement a variety of visualizations for different data formats and software.			
Type Disposition			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-68-16	T1	0.33	KA02.05
Actively and proactively recognize the important role of interfaces in influencing all aspects of usability.			
ACM-69-15	T1	0.33	KA02.05
Appropriately recognize the role of visualization.			

F: Software Engineering

1. Software Design and Development

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-41-01	T1	3.00	KA02.06
Understand standards for coding and design.			

ACM-41-02	T1	1.50	KA02.06
Understand the integration of information management and database systems.			

ACM-41-03	T1	1.50	KA02.06
Understand the software life cycle.			

ACM-41-04	T1	1.50	KA02.06
Understand the data lifecycle.			

ACM-41-05	T2	1.50	KA02.06
Understand project management methodologies.			

ACM-41-06	E	3.00	KA02.06
Understand integration with embedded, process control, and communication systems.			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-41-07	T1	1.50	KA02.06
Describe coding standards.			

ACM-41-08	T1	1.50	KA02.06
Describe design standards for projects.			

IP SJ Data Science Curriculum Standard (April 2021)

ACM-41-09 T1 1.50 KA02.06
Describe how to integrate and interoperate with information management/database systems.

ACM-41-10 T1 1.50 KA02.06
Describe the scope and types of different testing methodologies and needs across multiple disciplines.

ACM-41-11 T1 3.00 KA02.06
Independently implement a small software project that meets the design specifications.

ACM-41-12 T1 6.00 KA02.06
Work in a team to complete a software project that meets the design specifications.

ACM-41-13 T1 3.00 KA02.06
Able to follow given design, documentation, and implementation standards.

ACM-41-14 T1 3.00 KA02.06
Execute the basic software lifecycle in a simple program.

ACM-41-15 T1 3.00 KA02.06
Perform the basic data (science) lifecycle using simple data processing software.

ACM-41-16 T1 1.50 KA02.06
Integrate and interact with information management systems and database systems.

ACM-41-17 T2 2.00 KA02.06
Implement a given project management methodology.

ACM-41-18 T2 1.50 KA02.06
Work in teams to plan and design software projects that meet stakeholder specifications.

ACM-41-19 T2 1.50 KA02.06
Lead a team to completion of a project while meeting stakeholder requirements.

ACM-41-20 T2 1.50 KA02.06
Implement the data science lifecycle to make data-driven decisions at appropriate stages of the software lifecycle.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-41-21 E 6.00 KA02.06

Integrate and interoperate with embedded systems, process control systems, and communication systems.

DS-099 T1 1.00 KA03.02

Able to create small Excel datasets using some data extraction methods from databases.

DS-100 E 2.00 KA03.02

Be able to organize the requirements for an analysis system that collects and utilizes open data.

IPJSJ-13 E 3.00 KA02.06

Practice agile development.

[Note] Ref. SWEBOK 2021

IPJSJ-14 E 3.00 KA02.06

DevOps practices.

[Note] Ref. SWEBOK 2021

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-41-22 T1 2.00 KA02.06

Recognize the value of teams built on respect, diversity, and collaboration, and behave in a collaborative and ethical manner.

ACM-41-23 T1 1.50 KA02.06

Believes in the significance of adhering to project coding and design standards.

ACM-41-24 T1 1.50 KA02.06

Collaborative, flexible, and a good listener, with the ability to present and negotiate ideas.

ACM-41-25 T1 1.50 KA02.06

You have a life-cycle mindset and a strong commitment to working with data and software projects.

ACM-41-26 T1 1.50 KA02.06

Awareness of the benefits of test-driven development.

ACM-41-27 T2 1.50 KA02.06

Adheres to the principles of respect, listening, and responsibility, and leads projects to completion in an ethical manner.

ACM-41-28

T2

1.50

KA02.06

Is committed to adhering to project coding and design standards, and promotes and encourages team members to do so as well.

2. Software Testing

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-42-01	T1	3.00	KA02.06

Understand the paradigm and need for various types of testing (unit testing, dynamic testing, integration testing, interface testing, regression testing, system testing, security testing).

ACM-42-02	T2	3.00	KA02.06
-----------	----	------	---------

Understand potential security issues in programs (e.g., buffer overflow, race conditions, improper initialization including privilege selection, failure to inspect input data, assumptions about success or correctness, failure to verify assumptions).

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-42-03	T1	2.00	KA02.06

Define and explain the scope and types of different testing paradigms and needs in various disciplines.

ACM-42-04	T1	2.00	KA02.06
-----------	----	------	---------

To be able to design basic tests (unit tests, dynamic tests, integration tests).

ACM-42-05	T2	1.50	KA02.06
-----------	----	------	---------

Extract representative data from big data datasets in order to test algorithms on small datasets before running programs on large datasets.

ACM-42-06	T2	2.00	KA02.06
-----------	----	------	---------

Write test specifications for interface testing, user testing, regression testing, system testing, and security testing.

ACM-42-08	T2	3.00	KA02.06
-----------	----	------	---------

Evaluate the results of program execution using statistical significance tests.

ACM-42-09	T2	1.50	KA02.06
-----------	----	------	---------

Explain the types of risks that are considered in software systems.

ACM-42-10 T2 1.50 KA02.06
 Explain secure and defensive coding practices.

ACM-42-11 E 6.00 KA02.06
 Design, develop, and execute tests in multiple application areas.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-42-13	T1	1.50	KA02.06

Have a strong commitment to driving basic software and data projects from a test-driven perspective (unit testing, execution testing, integration testing).

ACM-42-14 T2 1.50 KA02.06

A strong commitment to approach software and data project development from a test-driven perspective (especially security testing, interface testing, user testing, regression testing, and system testing).

ACM-42-15 E 3.00 KA02.06

A strong commitment to approaching software and data projects holistically from the multifaceted perspective of test-driven development.

3. Collaboration with Related Fields

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
IPSJ-10	E	4.50	KA02.06

Ability to develop software including AI design patterns.

[Note] Ref. SWEBOK 2021

IPSJ-11 E 4.50 KA02.06

Software development including IoT design patterns.

[Note] Ref. SWEBOK 2021

IPSJ-12 E 3.00 KA02.06

Value creation and value proposition.

[Note] Ref. SWEBOK 2021

G1: Data Collection, Management and Governance

1. Fundamentals of Database

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

IPSJ-15	T1	1.00	KA03.02
---------	----	------	---------

Understand the role and purpose of databases.

IPSJ-16	T1	1.00	KA03.02
---------	----	------	---------

Understand and explain the models and characteristics of relational and non-relational databases.

IPSJ-17	T1	1.00	KA03.02
---------	----	------	---------

Explain persistence and data independence in database systems.

IPSJ-18	T2	1.00	KA03.02
---------	----	------	---------

Explain the purpose and function of modern database systems.

IPSJ-19	T2	1.00	KA03.02
---------	----	------	---------

Describe the availability, reliability, and recovery of database systems.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

DS-104	T1	0.50	KA03.02
--------	----	------	---------

To be able to determine whether the data being handled is structured data (customer data, product data, inventory data, etc.) or unstructured data (miscellaneous text, audio, images, video, etc.)

DS-105	T1	1.00	KA03.02
--------	----	------	---------

Understand relationships between tables by reading ER diagrams.

DS-106	E	2.00	KA03.02
--------	---	------	---------

To be able to normalize tables using normalization methods (first normalization to third normalization).

2. Data Collection

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-01-01	T1	0.27	KA03.03
-----------	----	------	---------

Data sources

ACM-01-02	T1	0.27	KA03.03
-----------	----	------	---------

pull-based, push-based approach.

ACM-01-03	T1	0.27	KA03.03
-----------	----	------	---------

Collection of a variety of data based on the characteristics of the data to be collected

ACM-01-04	T1	0.27	KA03.03
-----------	----	------	---------

Techniques to facilitate data collection

ACM-01-05	T1	0.27	KA03.03
-----------	----	------	---------

Data discretization techniques

ACM-01-06	T1	0.27	KA03.03
-----------	----	------	---------

Security and privacy standards and best practice skills

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-01-07	T1	0.27	KA03.03
-----------	----	------	---------

Selection of data sources to leverage data

ACM-01-08	T1	0.27	KA03.03
-----------	----	------	---------

Design techniques for data collection based on characteristics of the data source and intended use of the data

ACM-01-09	T1	0.27	KA03.03
-----------	----	------	---------

Subsequent procedures, including data discretization, transfer, and storage

ACM-01-10	T2	0.27	KA03.03
Design of acceleration and parallelization strategies for data acquisition based on data utilization			
ACM-02	T2	6.00	KA03.03
Able to extract information from document files, web pages, and multimedia data. See Knowledge of Information Extraction in Data Mining (ACM-16).			
DS-101	E	1.00	KA03.02
Outline the functions (SDK, API, etc.) provided by the target platform.			
DS-102	E	1.00	KA03.02
To be able to use web crawlers and scripting tools to collect static content on websites as data for analysis.			
DS-103	T1	2.00	KA03.02
To be able to implement functions for storing data at the collection destination using the functions provided by the target platform (HTTP, FTP, SSH, etc.).			

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-01-11	T1	0.27	KA03.03
Demonstrate business acumen in the ability to evaluate tradeoffs between accuracy and efficiency in data collection.			

3. Data Manipulation

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
DS-078	T1	0.50	KA03.02 KA02.05
Be able to convert name scale variables into dummy variables.			
DS-079	T1	0.50	KA03.02 KA02.05
Know what standardization is, and be able to standardize appropriately.			
DS-080	T1	0.50	KA03.02 KA02.05
Know what outliers, abnormal values, and missing values are, and be able to detect, remove, and deal with variables appropriately under instructions.			

IPJSJ Data Science Curriculum Standard (April 2021)

DS-081	E	0.50	KA03.02	KA02.05
Able to design and execute categorization and binning of bins as necessary, based on analysis requirements and distribution of each variable.				
DS-082	T1	0.50	KA03.02	KA02.05
Able to filter data of hundreds of thousands of records by specifying conditions. (e.g., extract data that matches or does not match a specific value, extract data in a specific range, extract substrings, etc.)				
DS-083	E	0.50	KA03.02	KA02.05
To be able to use regular expressions to extract data that matches the conditions. (e.g. judging whether an email address meets the format)				
DS-084	T1	0.50	KA03.02	KA02.05
To be able to sort by specific columns among records for data with hundreds of thousands of records, and to be able to sort by columns for data with thousands of records.				
DS-085	T1	0.50	KA03.02	KA02.05
Able to perform inner join, outer join, and self-join based on a single condition for data of several hundred thousand records, and to perform UNION processing.				
DS-086	T1	0.50	KA03.02	KA02.05
For data with hundreds of thousands of records, be able to remove records with null values, unexpected or out-of-range data, or convert them to the specified values.				
DS-087	T1	0.50	KA03.02	KA02.05
For data with hundreds of thousands of records, be able to express one value with another specified value, such as converting against a specified list or converting from prefecture to geocode.				
DS-088	T1	0.50	KA03.02	KA02.05
Extract data randomly or at regular intervals from hundreds of thousands of records.				
DS-089	T1	0.50	KA03.02	KA02.05
Calculate the sum, maximum value, minimum value, and number of records for hundreds of thousands of records.				
DS-090	T1	0.50	KA03.02	KA02.05
To be able to perform arithmetic operations on data with hundreds of thousands of records, and to convert numerical data into other data types such as date and time data.				

4. Data Sharing

Type	Skill			
SEQ#	Priority	Assigned Time (h)	DS-BoK	KA#

IPJSJ Data Science Curriculum Standard (April 2021)

DS-091 T1 0.25 KA03.02 KA02.05
 To be able to convert and export the results of processing and analysis into a specified format such as CSV, XML, JSON, or Excel.

DS-092 T1 0.50 KA03.02 KA02.05
 To be able to insert records of processing and analysis results according to the table specifications of the destination DB.

DS-093 E 1.00 KA03.02 KA02.05
 Obtain necessary data using Web API (REST) and Web services (SOAP) for data acquisition.

DS-094 T1 0.50 KA03.02 KA02.05
 Download necessary data files from FTP servers, file sharing servers, etc., and import them into spreadsheet software such as Excel for use.

DS-095 E 1.00 KA03.02 KA02.05
 To be able to publish a new report by using the report editing function of BI tools.

DS-096 E 1.00 KA03.02 KA02.05
 To be able to extract necessary data and create graphs by using the free search function of BI tools.

5. Data Accumulation

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-03-01	T2	1.20	KA03.03

Data representation: numeric, text, image, data accuracy.

ACM-03-02	T2	1.20	KA03.03
-----------	----	------	---------

Text data processing: word clusters, word counts, ITF-IDF, n-grams, lexical analysis, parsing, semantic analysis, stop word filtering, stemming, basic applications

ACM-03-03	T2	1.20	KA03.03
-----------	----	------	---------

Image processing, data representation: integers, features, image operators, multi-dimensional matrices of video operators, object recognition, high-dimensional feature extraction

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-03-04 T2 1.20 KA03.03

Be able to write programs to perform basic operations on data of various types. To compute basic statistics, extract n-grams, and perform operations on images.

DS-107 E 1.00 KA03.02

Be able to connect to DWH appliances (Oracle Exadata, IBM Integrated Analytics System, Teradata, etc.) and extract composite data from multiple tables.

DS-108 E 2.00 KA03.02

Understand the basic structure and configuration of Hadoop and Spark distributed technologies.

DS-109 E 1.00 KA03.02

Able to access NoSQL data stores (Hbase, Cassandra, Mongo DB, CouchDB, Redis, Amazon, DynamoDB, Cloudant, Azure Cosmos DB, etc.) via API and register new data.

DS-110 E 1.00 KA03.02

Be able to connect to and store data in cloud storage services (Amazon S3, Google Cloud Storge, IBM Cloud Object Storage, etc.).

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-03-05	T2	1.20	KA03.03

Be accurate in selecting data types for encoding information.

6. System Operation

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
DS-097	E	2.00	KA03.01

To be able to construct and operate a system of 1 to 10 servers with instructions.

DS-098 E 2.00 KA03.01

Be able to perform routine operations such as backup and archive creation for databases with hundreds of thousands of records.

7. Data Management

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

IP SJ Data Science Curriculum Standard (April 2021)

ACM-04-01	T1	0.27	KA03.03
Concepts and usage scenarios for government databases, data warehouses, and mediator-based information integration.			
ACM-04-02	T1	0.27	KA03.03
Schema Mapping Concepts and Approaches			
ACM-04-03	T1	0.27	KA03.03
Data mapping concepts and approaches			
ACM-04-04	T1	0.27	KA03.03
Data semantic transformation concepts and approaches			
ACM-04-05	T1	0.27	KA03.03
Methods for integrating data across domains			
ACM-05-01	T1	0.30	KA03.03
The role of compression and reduction in data processing.			
ACM-05-02	T1	0.30	KA03.03
Different approaches to data sampling			
ACM-05-03	T1	0.30	KA03.03
Data filtering techniques			
ACM-05-04	T1	0.30	KA03.03
Data compression techniques			
ACM-06-01	T1	0.25	KA03.03
Data Transformation Pipeline			
ACM-06-02	T1	0.25	KA03.03
Simple functional transformation methods and their applications			
ACM-06-03	T1	0.25	KA03.03
Data standardization and its applications			

IPSJ Data Science Curriculum Standard (April 2021)

ACM-06-04 T1 0.25 KA03.03
Data normalization and its applications

ACM-06-05 T1 0.25 KA03.03
Data Encoding Approaches and Their Applications

ACM-06-06 T1 0.25 KA03.03
Data smoothing approaches and their applications

ACM-07-01 T1 0.30 KA03.03
Aspects of Data Quality

ACM-07-02 T1 0.30 KA03.03
Approaches to improve data quality

ACM-07-03 T1 0.30 KA03.03
Data cleaning algorithms, including entity resolution, truth discovery, and rule-based data.

ACM-07-04 T1 0.30 KA03.03
Various forms of data quality rules: functional dependencies (FD), conditional functional dependencies (CFD), conditional inclusion dependencies (CIND), and matching rules. (CIND: conditional inclusion dependencies), and matching dependencies (MD: matchin

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-04-06	T1	0.27	KA03.03
Choosing a scheme for data integration. Choose a data integration scheme, i.e., traditional data integration or cross-domain data integration.			

ACM-04-07	T1	0.27	KA03.03
Select the data integration architecture according to the application functionality.			

ACM-04-08	T1	0.27	KA03.03
Select or develop suitable algorithms for schema mapping, data mapping, and semantic transformation of data.			

ACM-04-09	T1	0.27	KA03.03
Develop appropriate algorithms for integrating data across domains.			

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-05-05 T1 0.30 KA03.03
 Be able to investigate whether data reduction and compression is necessary.

ACM-05-06 T1 0.30 KA03.03
 To perform data sampling and filtering.

ACM-05-07 T2 0.30 KA03.03
 Analyze the characteristics of data sampling.

ACM-05-08 T2 0.30 KA03.03
 Select data compression techniques based on requirements for computation, communication, and storage.

ACM-05-09 T2 0.30 KA03.03
 Develop approaches to data compression that are appropriate to the query.

ACM-06-07 T1 0.25 KA03.03
 Evaluating the dimension and range of data versus the dimension and range required by the application.

ACM-06-08 T1 0.25 KA03.03
 To determine the process of data transformation.

ACM-06-09 T1 0.25 KA03.03
 Selecting the appropriate data algorithm for a particular task.

ACM-06-10 T1 0.25 KA03.03
 Evaluate the effectiveness of the data transformation.

ACM-07-05 T1 0.30 KA03.03
 Assessing data quality.

ACM-07-06 T1 0.30 KA03.03
 Create data cleaning rules according to the requirements of the application and the semantics of the data.

ACM-07-07 T1 0.30 KA03.03
 Develop data cleaning pipelines according to data quality requirements.

ACM-07-08 T1 0.30 KA03.03

Develop algorithms for efficient and effective data cleaning.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-04-10	T1	0.27	KA03.03
-----------	----	------	---------

Be keenly aware of the challenges posed by heterogeneous data sources.

ACM-04-11	T1	0.27	KA03.03
-----------	----	------	---------

Be aware of the role of AI in data integration.

ACM-05-10	T1	0.30	KA03.03
-----------	----	------	---------

Attention is paid to a detailed evaluation of the trade-offs between effectiveness and efficiency of data computation.

ACM-06-11	T1	0.25	KA03.03
-----------	----	------	---------

Be sensitive to the importance of data transformation for the use of the data.

ACM-06-12	T1	0.25	KA03.03
-----------	----	------	---------

Be sensitive to the relationship between data transformation and data quality.

ACM-07-09	T1	0.30	KA03.03
-----------	----	------	---------

Be sensitive to harm caused by data quality issues.

ACM-07-10	T1	0.30	KA03.03
-----------	----	------	---------

Have a strong sense of responsibility for handling the role of data cleaning in data use.

8. Information Retrieval

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-25-01	T2	0.50	KA01.03 KA01.04
-----------	----	------	----------------------

Understand the techniques used to measure the efficiency of the search process.

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-25-02	T2	0.50	KA01.03	KA01.04
Understand the use of various techniques for storing and organizing data, i.e., the use of encoding functions, to ensure that information is extracted efficiently.				
ACM-25-03	T2	0.50	KA01.03	KA01.04
Understand the concept of search strategy, i.e., the role of broadening and narrowing the search.				
ACM-25-04	T2	0.50	KA01.03	KA01.04
Understand keyword selection in the search process, i.e., the use of Boolean operators.				
ACM-25-05	T2	0.50	KA01.03	KA01.04
Understand the retrieval of ordinal data.				
ACM-25-06	T2	0.50	KA01.03	KA01.04
Understand techniques for searching text-based materials.				
ACM-25-07	T2	0.50	KA01.03	KA01.04
Document set retrieval: Understand strategies for listing the names of selected items.				
ACM-25-08	T2	0.50	KA01.03	KA01.04
Feature identification and feature extraction for non-text based data: Understand the search strategies used for photos, audio, and video.				
ACM-25-09	T2	0.50	KA01.03	KA01.04
Understand the role of hashing, indexing and filtering.				
ACM-25-10	T2	0.50	KA01.03	KA01.04
Understand approaches to searching text-based materials.				
ACM-25-11	T2	0.50	KA01.03	KA01.04
Understand techniques for creating and searching relational databases.				
ACM-25-12	T2	0.50	KA01.03	KA01.04
Understand the various relational, non-relational, and other database formats.				
ACM-25-13	T2	0.50	KA01.03	KA01.04
Understand web-based information retrieval, the web viewed as a graph of interconnected nodes, relevance metrics from graph theory, PageRank and relevance metrics to facilitate web-based retrieval.				

IPSJ Data Science Curriculum Standard (April 2021)

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-25-14	T2	0.50	KA01.03	KA01.04

Be able to devise a search strategy for a given information retrieval task.

ACM-25-15	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Explain the ethical concerns that may be associated with the information retrieval process.

ACM-25-16	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Identify situations where it is appropriate to use parallelization techniques to speed up the search.

ACM-25-17	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Outline the key elements of an effective strategy to underpin web-based search.

ACM-25-18	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Identify software that can be used for information retrieval tasks related to images, sound recordings, and video.

ACM-25-19	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Use SQL to create and use relational database structures.

ACM-25-20	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

Explain the role that information retrieval can play in the operation of a digital library.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-25-21	T2	0.50	KA01.03	KA01.04

Recognize and appreciate the importance of various considerations that underpin an efficient and effective approach to information retrieval.

G2: Professionalism

1. Continuing Professional Development

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-73-01	T1	0.38	KA04.01	KA05.01
Be able to demonstrate the meaning of competencies and capabilities.				

ACM-73-02	T1	0.38	KA04.01	KA05.01
Acquisition of expertise/proficiency or development of competency; role of journals, conferences, courses, webinars				

ACM-73-03	T1	0.38	KA04.01	KA05.01
Changes in technology and their impact on competency				

ACM-73-04	T1	0.38	KA04.01	KA05.01
CPD and the role of professional associations (profession-specific associations) in professional activities				

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-73-05	T1	0.38	KA04.01	KA05.01
Explaining the importance of maintaining competency for professional data scientists.				

ACM-73-06	T1	0.38	KA04.01	KA05.01
To explain how professionals typically develop and become proficient in competencies, and to illustrate the backward advantages.				

ACM-73-07	T1	0.38	KA04.01	KA05.01
Assert the importance of the role of professional associations in career development.				

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
------	----------	-------------------	------------	--

ACM-73-08 T1 0.38 KA04.01 KA05.01

The field of data science is a rapidly changing field. I am proactive and passionate about recognizing that it is essential to know how to stay current.

2. Communication

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-74-01	T1	0.30	KA04.01 KA05.01
Various forms of communication – electronic, written and spoken – and how to use them effectively.			

ACM-74-02	T1	0.30	KA04.01 KA05.01
Technical literature relevant to data science			

ACM-74-03	T1	0.30	KA04.01 KA05.01
Relevant audiences for data scientist-related communications? Small groups, large groups, professionals and non-professionals, young people, senior managers, machines? And the elements of effective communication with each			

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-74-04	T1	0.30	KA04.01 KA05.01
Assessing the state of the technical literature related to data science.			

ACM-74-05	T1	0.30	KA04.01 KA05.01
To create a literature to guide the technical development of colleagues.			

ACM-74-06	T1	0.30	KA04.01 KA05.01
Create presentations for a variety of audiences interested in data science.			

ACM-74-07	T1	0.30	KA04.01 KA05.01
Designing status reports for senior management to illustrate key initiatives arising from data science investigations, incorporating common issues around change management where appropriate.			

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

ACM-74-08	T1	0.30	KA04.01	KA05.01
-----------	----	------	---------	---------

Understand how to respond to, adjust to, and effectively use the latest information about relevant technologies, and keep abreast of new development opportunities.

ACM-74-09	T1	0.30	KA04.01	KA05.01
-----------	----	------	---------	---------

Actively and spontaneously assess the importance of new learning and new experiences.

ACM-74-10	T1	0.30	KA04.01	KA05.01
-----------	----	------	---------	---------

Has an accurate and focused understanding of the strengths and weaknesses of knowledge.

3. Teamwork

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-75-01	T1	0.33	KA04.01	KA05.01

Team Member Selection. The need to complement the abilities and skills of team members.

ACM-75-02	T1	0.33	KA04.01	KA05.01
-----------	----	------	---------	---------

Team dynamics, team discipline.

ACM-75-03	T1	0.33	KA04.01	KA05.01
-----------	----	------	---------	---------

Elements of effective team operation.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-75-04	T1	0.33	KA04.01	KA05.01

Describing the steps for dealing with conflicts among team members.

ACM-75-05	T1	0.33	KA04.01	KA05.01
-----------	----	------	---------	---------

Summarize the considerations that should be made in selecting a team to conduct a particular data science study.

ACM-75-06	T1	0.33	KA04.01	KA05.01
-----------	----	------	---------	---------

Recognize the qualities required of the leader of a team conducting a data science research study.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-75-07	T1	0.33	KA04.01	KA05.01
Cooperate respectfully, and act sensitively and appropriately in the organization and operation of the team.				

ACM-75-08	T1	0.33	KA04.01	KA05.01
When working with others, put aside unimportant differences, cooperate, and act appropriately.				

4. Economic Considerations

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-76-01	T2	0.46	KA04.01	KA05.01
Cost and Value of High Quality Datasets and Their Maintenance				

ACM-76-02	T2	0.46	KA04.01	KA05.01
Justifying the cost of data science activities				

ACM-76-03	T2	0.46	KA04.01	KA05.01
Estimating project costs				

ACM-76-04	T2	0.46	KA04.01	KA05.01
Facilitating data science				

ACM-76-05	T2	0.46	KA04.01	KA05.01
Automation resulting from data science activities				

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-76-06	T2	0.46	KA04.01	KA05.01
Evaluate the value of the dataset to the organization, taking into account the need to maintain it.				

ACM-76-07	T2	0.46	KA04.01	KA05.01
Discussing examples of data that the organization should collect on a regular basis and designing a data collection process that identifies the attributes the data should have and determines the format of the data set to be handled, taking quality into ac				

ACM-76-08	T2	0.46	KA04.01	KA05.01
Evaluate the cost (as a resource) of collecting high quality data for a specific purpose.				

ACM-76-09	T2	0.46	KA04.01	KA05.01
To justify and measure the cost of placing data science activities within an organization.				

ACM-76-10	T2	0.46	KA04.01	KA05.01
To infer the value to an organization of undertaking a survey or research project.				

ACM-76-11	T2	0.46	KA04.01	KA05.01
To monitor the resources required to conduct research within the organization and compare them to those required to outsource the activity.				

ACM-76-12	T2	0.46	KA04.01	KA05.01
Evaluate the costs associated with automating an activity.				

Type	Disposition
------	-------------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-76-13	T2	0.46	KA04.01 KA05.01
Respect the costs associated with data science activities and act appropriately.			

5. Privacy and Confidentiality

Type	Knowledge
------	-----------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-77-01	T1	0.25	KA04.01 KA05.01
Freedom of Information			

ACM-77-02	T1	0.25	KA04.01 KA05.01
Data protection regulations such as the EU General Data Protection Regulation (GDPR) – see [5].			

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-77-03	T1	0.25	KA04.01	KA05.01
Privacy laws				

ACM-77-04	T1	0.25	KA04.01	KA05.01
How to protect data confidentiality				

ACM-77-05	T1	0.25	KA04.01	KA05.01
Risks to privacy and confidentiality				

ACM-77-06	T1	0.25	KA04.01	KA05.01
International aspects				

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-77-07	T1	0.25	KA04.01	KA05.01
Describe the technical mechanisms for protecting data confidentiality.				

ACM-77-08	T1	0.25	KA04.01	KA05.01
To compare the privacy laws of different countries and explain the challenges that arise from these differences.				

ACM-77-09	T1	0.25	KA04.01	KA05.01
To recognize the privacy and confidentiality challenges arising from the use of video, voice and facial recognition software.				

ACM-77-10	T1	0.25	KA04.01	KA05.01
To be aware of the context in which each privacy law should be applied, taking into account international standards.				

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-77-11	T1	0.25	KA04.01	KA05.01
Be sensitive to incorporating and maintaining aspects of privacy and confidentiality so that data science activities can be trusted.				

Type Contextual Issue

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
------	----------	-------------------	------------	--

ACM-77-12	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

Legal frameworks related to privacy and security vary from country to country.

6. Ethical Considerations

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-78-01	T1	0.25	KA04.01	KA05.01

Ethical issues in maintaining competencies and capabilities

ACM-78-02	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

Confidentiality issues related to data and its use.

ACM-78-03	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

EU General Data Protection Regulation (GDPR) – see [5].

ACM-78-04	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

The need for data and data samples that are properly representative of a situation

ACM-78-05	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

Bias in algorithms and data, and awareness of possible types of bias. Mechanisms to identify and avoid bias.

ACM-78-06	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

Transparency and accountability of algorithms.

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-78-07	T1	0.25	KA04.01	KA05.01

Indication of some situations in which the data scientist may be overstepping his or her bounds, and characterization of measures to mitigate these situations.

ACM-78-08	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

To display techniques for ensuring that datasets and algorithms do not contain bias.

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-78-09 T1 0.25 KA04.01 KA05.01
 To contemplate the benefits of joining a data science professional network.

DS-121 T1 0.50 KA03.01
 Have ethics appropriate for a person handling data. (Do not falsify, alter, or plagiarize data, etc.)

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-78-10	T1	0.25	KA04.01	KA05.01
Address the deeper ethical issues surrounding the collection and use of data.				

ACM-78-11 T1 0.25 KA04.01 KA05.01
 Address issues of prejudice and preconceptions and work proactively to resolve them.

ACM-78-12 T1 0.25 KA04.01 KA05.01
 Be independent and self-motivated in the advancement of data science.

7. Legal Considerations

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-79-01	T2	0.60	KA04.01	KA05.01
Computer Crimes Related to Data Science				

ACM-79-02 T2 0.60 KA04.01 KA05.01
 Cyber security

ACM-79-03 T2 0.60 KA04.01 KA05.01
 Crime prevention

ACM-79-04 T2 0.60 KA04.01 KA05.01
 Mechanisms to detect criminal behavior. Importance of various approaches

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-79-05 T2 0.60 KA04.01 KA05.01
 Recovery mechanisms. Maintaining 100% uptime

ACM-79-06 T2 0.60 KA04.01 KA05.01
 Laws against Computer Crime

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-79-07	T2	0.60	KA04.01 KA05.01
Evaluate various mechanisms for detecting stated forms of criminal activity.			

ACM-79-08 T2 0.60 KA04.01 KA05.01
 Be able to justify the desirability of multiple and diverse approaches to countering the threat.

DS-122 T1 1.00 KA03.01
 Understand the laws and regulations related to personal information (e.g., Personal Information Protection Act, EU General Data Protection Regulation (GDPR)) and an overview of anonymously processed information, and be able to explain the points to be pro

DS-145 E 0.50 KA05.01
 To be able to explain the difference between a contract and an associate contract.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-79-09	T2	0.60	KA04.01 KA05.01
Demonstrate kindness and compassion while adopting a responsible and ethical attitude when encountering behavior that appears to be criminal.			

Type Contextual Issue

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-79-10	T2	0.60	KA04.01 KA05.01
The legal structure varies from country to country.			

8. Intellectual Property

Type Knowledge

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-80-01	E	0.27	KA04.01	KA05.01
Patents, copyrights, trademarks, trade secrets, moral rights, and trademarks				

ACM-80-02	E	0.27	KA04.01	KA05.01
What data science-related intellectual property can and cannot be protected, and what types of protection are available.				

ACM-80-03	E	0.27	KA04.01	KA05.01
Types of data science-related intellectual property that are legally protectable and types that are not, and the types of protection available				

ACM-80-04	E	0.27	KA04.01	KA05.01
IP-related rules, ownership of IP, the territoriality of IP rights, including the impact of international agreements (e.g., European Directive on the Protection of Trade Secrets), and IP term limits				

ACM-80-05	E	0.27	KA04.01	KA05.01
Types of IP rights that are automatically granted and those that require registration, and an overview of the process of obtaining protected IP rights.				

ACM-80-06	E	0.27	KA04.01	KA05.01
The possibility of infringing the rights of others. Proper use of protected intellectual property.				

Type Skill

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-80-07	E	0.27	KA04.01	KA05.01
Describe the types of intellectual property relevant to a data scientist.				

ACM-80-08	E	0.27	KA04.01	KA05.01
Differentiate between patents, copyrights, designs, and trademarks, and describe their use in data science.				

ACM-80-09	E	0.27	KA04.01	KA05.01
Explain the role of trade secrets in the context of data science.				

ACM-80-10	E	0.27	KA04.01	KA05.01
To explain the process of registering intellectual property rights.				

IPJSJ Data Science Curriculum Standard (April 2021)

ACM-80-11 E 0.27 KA04.01 KA05.01
 To explain the challenges associated with intellectual property ownership and moral rights.

ACM-80-12 E 0.27 KA04.01 KA05.01
 To evaluate the risks involved in using protected intellectual property and how to overcome those risks in a justifiable manner.

Type **Disposition**

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-80-13	E	0.27	KA04.01 KA05.01

Be perceptive and sensitive to the existence and importance of intellectual property, and to the responsibilities and opportunities presented by intellectual property.

Type **Contextual Issue**

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-80-14	E	0.27	KA04.01 KA05.01

Thoroughness and adaptability in dealing with the ethical and legal frameworks surrounding intellectual property varies from country to country. Get advice from a patent attorney.

9. On Automation

Type **Knowledge**

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
ACM-81-01	E	0.42	KA04.01 KA05.01

Automation, its benefits and reasons

ACM-81-02 E 0.42 KA04.01 KA05.01
 Concerns specific to automation in critical situations

ACM-81-03 E 0.42 KA04.01 KA05.01
 Transparency and accountability in algorithms

Type **Skill**

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

IPSSJ Data Science Curriculum Standard (April 2021)

ACM-81-04 E 0.42 KA04.01 KA05.01
 Explaining to a non-technical audience the extent to which decisions are automated in certain situations.

ACM-81-05 E 0.42 KA04.01 KA05.01
 To be able to analyze the impact of requirements on design on computerized decision making.

ACM-81-06 E 0.42 KA04.01 KA05.01
 To explain the benefits of automation in various situations.

ACM-81-07 E 0.42 KA04.01 KA05.01
 Identify the steps necessary to ensure that a decision-making system can be audited.

Type Disposition

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
ACM-81-08	E	0.42	KA04.01	KA05.01
Be perceptive and sensitive to the challenges associated with automation and its impact on employment.				

ACM-81-09 E 0.42 KA04.01 KA05.01
 Take a respectful and ethical approach to the issue of automation.

G3: Business Fundamentals

1. Logical Thinking

Type	Skill
------	-------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

DS-135	T1	0.67	KA04.01
--------	----	------	---------

Be able to recognize duplication of data and events.

DS-136	T1	0.67	KA04.01
--------	----	------	---------

Able to correctly verbalize the implications of analysis results in the case of commonly observed phenomena.

DS-137	E	1.00	KA04.01
--------	---	------	---------

Understand the general structure of a paper. (Introduction 竊誕 approach 竊池 results 竊壇 discussion, introduction 竊知 ain argument 竊團 onclusion, etc.)

DS-138	E	2.00	KA04.01
--------	---	------	---------

Be able to theoretically summarize a document of one figure or several figures. (Issue background, approach, results of discussion, implications, next steps)

DS-139	T1	1.00	KA04.01
--------	----	------	---------

To be able to understand the other party's argument promptly when the other party points out the lack of argument or logical breakdown in the report.

2. Project Management

Type	Skill
------	-------

SEQ#	Priority	Assigned Time (h)	DS-BoK KA#
------	----------	-------------------	------------

DS-140	E	0.67	KA04.02
--------	---	------	---------

To be able to explain the difference between waterwall development and agile development.

DS-141	T1	0.67	KA04.02
--------	----	------	---------

To be able to follow instructions and follow a schedule, and to be able to complete one's own work as requested by the team leader.

DS-142	T1	0.67	KA04.02
--------	----	------	---------

Be able to quickly and appropriately report any delays or failures in the tasks for which they are responsible.

3. Business Implementation

Type	Skill			
SEQ#	Priority	Assigned Time (h)	DS-BoK KA#	
DS-067	E	2.00	KA05.02	Understand the key variables (KPIs) of the business for which you are responsible, in addition to the general revenue equation.
DS-068	E	2.00	KA05.02	Explain the market size, major players, dominant business models, challenges and opportunities in the business domain for which they are responsible.
DS-069	T1	2.00	KA05.02	Understand the basic framework of the issues in the business domain for which they are primarily responsible. (Organize procurement activities in terms of 5 forces, CRM issues in terms of RFM, etc.)
DS-070	T1	1.00	KA05.02	Given a hypothesis or a known problem, be able to identify and secure access to the necessary data.
DS-131	T1	0.67	KA05.01	Recognize the importance of logic and data in business, and be able to act based on an analytical and data-driven mindset.
DS-132	T1	0.67	KA05.01	Understand that Analyzing data without setting a purpose or goal is not meaningful.
DS-133	T1	0.67	KA05.01	Understand the importance of verbalizing issues and hypotheses.
DS-134	T1	0.67	KA05.01	Understand the importance of having access to primary information, such as by visiting the field and conducting interviews.
DS-143	T1	0.67	KA05.01	Understand the importance of looking at data with a hypothesis from a business perspective, and the possibility that results that differ from the hypothesis may still be important insights.
DS-144	E	0.67	KA05.01	Understand the importance of looking at data with a hypothesis, and that even if the results are different from the hypothesis, they may be important findings.

IPSSJ Data Science Curriculum Standard (April 2021)

DS-146

T1

0.50

KA05.01

Understand the importance of monitoring the results and the degree of improvement.