

データサイエンス・カリキュラム標準案 (専門教育レベル)¹

一般社団法人 情報処理学会

2021年2月9日

1. データサイエンス・カリキュラム標準の策定趣旨

データサイエンス(DS)分野の資格制度やカリキュラム標準に関する様々な取り組みが世界中で進んでいる。こうした中で、情報処理学会が日本における情報処理の専門家集団としての立場からデータサイエンス教育の体系化に貢献することは社会的な重要性が高い。

情報処理学会は、30年近くに渡って J17 等、大学レベルの情報教育に関するカリキュラム標準を策定した実績があり、初中等および大学における情報分野の教育活動に様々な取り組みを推進してきた。また、米国 ACM, IEEE-CS, IFIP 等、情報分野における海外の主要学会とも連携している。

J17 カリキュラム標準は文部科学省委託事業として取り組んだが、2017年度時点では DS 分野のカリキュラム標準を策定するには至らなかった。

しかし、2019年度末の時点において、DS 分野のカリキュラム標準(専門教育レベル)の検討が米国および欧州でも進んでおり、国際的通用性を確保するための標準化が視野に入りつつある。さらに、データサイエンティスト資格と DS 標準カリキュラムの連携に向けた取り組みも情報処理学会内で進捗しつつあり、日本における情報分野の DS 標準カリキュラム(専門教育レベル)を策定するための基盤が構築されつつある。

こうした状況を踏まえ、情報処理学会は、情報学の立場から見た DS カリキュラム標準(専門教育レベル)の策定を推進している。本カリキュラム標準は、数理・データサイエンス教育強化拠点コンソーシアムが策定を進めている、「数理・データサイエンス・AI モデルカリキュラム(リテラシーレベルおよび応用基礎レベル)」とは異なり、データサイエンスを専門とする大学・学部レベルの専門教育プログラムが参照することを意図している。

2. データサイエンス・カリキュラム標準の策定方針

A) 国際的通用性を確保する

- ACM Data Science カリキュラム(Final Draft Report. <http://dstf.acm.org/>)を参照する。
- 欧州 EDISON Data Science Framework より、DS-BoK および DS Model Curriculum を参照する(<https://edison-project.eu/edison/edison-data-science-framework-edsf/>)。EDISON DS-BoK は、統計、AI、データ工学、ソフトウェア工学、コンピュータサイエンス、ビジネス基礎等を含んでおり、ACM および DS スキルチェックリストの範囲をカバーしている。各教育項目には DS-BoK の KA (Knowledge Area) 番号を付す。
- ACM Data Science カリキュラムおよび DS-BoK の全体構成は、本文書の末尾に示す。

¹ 本資料(pdf版)には目次の代わりにしおりを設定しています。

- B) 情報処理学会のデータサイエンティスト資格(策定中)との連携を図る
- データサイエンティスト協会・DS スキルチェックリストから★レベルの項目(必須項目および選択項目)を参照する(https://www.datascientist.or.jp/common/docs/PR_skillcheck_ver3.00.pdf)。同スキルチェックリストも、本文書の末尾に示す。
- C) データサイエンスを専門とする学生に求められる知識・スキルの具体化を図る。
- EDISON DS-BoK は、求められる知識・スキルの具体性が低いため、ACM Data Science カリキュラムおよび DS スキルチェックリスト等を活用して具体化を図る。
 - ACM Data Science カリキュラムを参照する教育項目には、「ACM」で始まる通し番号を付す。
 - DS スキルチェックリストを参照する教育項目には、「DS」で始まる通し番号を付す。
 - それ以外の教育項目には、「IPSJ」で始まる通し番号を付す。
- D) 60 単位程度(標準的な授業時間:約 675 時間)の規模とする(表 1 参照)。
- 各教育項目に対して、T1(必須項目・100%履修が求められる)、T2(準必須項目・80%程度の履修が求められる)、E(選択項目・50%程度の履修が求められる)の優先度を付す。
 - 各教育項目に対して時間数を割り当てる。各領域に時間数を割り当てる際には EDISON DS Model Curriculum の時間配分を参照する。
 - 有効教育時間(各教育項目について、割り当て時間数と履修率を乗じた値の合計値)が 675 時間程度(60 単位×11.25 時間/単位)になるようにカリキュラムを設計する。
- E) 数理・データサイエンス・AI モデルカリキュラム(リテラシーレベル)の学修を前提とする。
- 同カリキュラムに含まれる教育項目には、原則として時間を割り当てない。

3. 情報処理学会・データサイエンス教育委員会の構成

本カリキュラム標準は、情報処理学会データサイエンス教育委員会によって作成された。

- 加藤 浩(放送大学・委員長)
- 掛下 哲郎(佐賀大学・副委員長)
- 中谷 多哉子(放送大学・幹事)
- 石井 一夫(久留米大学)
- 石川 佳治(名古屋大学)
- 上松 恵理子(武蔵野学院大学)
- 内田 暁(東京大学)
- 奥村 晴彦(元三重大学)
- 佐伯 諭(株式会社電通・データサイエンティスト協会)
- 高橋 克己(NTT セキュアプラットフォーム研究所)
- 高橋 尚子(國學院大學)
- 高橋 範光(株式会社チェンジ・データサイエンティスト協会)
- 中野 美由紀(津田塾大学)
- 松尾 豊(東京大学)
- 松原 仁(東京大学)
- 村田 剛志(東京工業大学)
- 北川 源四郎(統計数理研究所・オブザーバ)

中分類	割り当て時間数(h)				有効教育 時間数(h)
	T1	T2	E	総計	
A:基礎数学と数理統計学	18.25		19.00	37.25	27.75
B1:データマイニング	28.00	46.50	78.00	152.50	104.20
B2:人工知能	3.25	17.00	12.50	32.75	23.10
B3:機械学習	13.75	19.50	13.50	46.75	36.10
C:モデリングとシミュレーション	11.80	18.38	47.00	77.18	50.00
D1:コンピュータサイエンス基礎	81.16	22.50	26.25	129.91	112.28
D2:ビッグデータ・システム	7.00	59.50	47.00	113.50	78.10
E1:セキュリティとプライバシー	41.93	0.00	4.00	45.93	43.93
E2:ヒューマン・コンピュータ・インタラクション	16.83	0.00	0.00	16.83	16.83
F:ソフトウェア工学	53.50	24.50	38.00	116.00	92.10
G1:データ収集, 管理, ガバナンス	24.01	13.17	17.00	54.18	43.04
G2:プロフェッショナリズム	16.14	12.01	7.99	36.14	29.74
G3:ビジネス基礎	10.50		8.33	18.83	14.67
総計	326.12	233.05	318.57	877.74	671.85

表 1: 情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)の全体構成

A: 基礎数学と数理統計学

1. 線形代数学基礎

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
DS-014	ベクトルの内積に関する計算方法を理解し、一次式をベクトルの内積で表現できる。 【備考】高校数学B(新課程ではC)の範囲(未履修の高校生もいるため、高校数学との接続を考慮して追加)	E	2.00	KA01.01
DS-015	行列同士、および行列とベクトルの計算方法を正しく理解し、複数の一次式を行列の積で表現できる。 【備考】新課程の高校数学Cの範囲(未履修の高校生もいるため、高校数学との接続を考慮して追加)	E	2.00	KA01.01
DS-016	逆行列の定義、および逆行列を求めることにより行列表記された連立方程式を解くことができることを理解している。	E	1.00	KA01.01
DS-017	固有ベクトルおよび固有値の意味を理解している。	E	1.00	KA01.01

2. 微分・積分学

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
DS-018	微分により計算する導関数が傾きを求めるための式であることを理解している。	E	1.00	KA01.01
DS-019	2変数以上の関数における偏微分の計算方法を理解している。	E	1.00	KA01.01
DS-020	積分と面積の関係を理解し、確率密度関数を定積分することで確率が得られることを説明できる。	E	1.00	KA01.01
IPSJ-01	指数関数、対数関数とその基本性質を理解している。 【備考】高校数学Ⅱの範囲(未履修の高校生もいるため、高校数学との接続を考慮して追加)	T1	1.00	KA01.01

IPSJ-02	T1	1.00	KA01.01
自然対数の底(e)を理解している。 【備考】高校数学Ⅲの範囲(未履修の高校生もいるため, 高校数学との接続を考慮して追加)			
IPSJ-03	T1	1.00	KA01.01
関数の極限の定義と計算を理解している。 【備考】高校数学Ⅲの範囲(未履修の高校生もいるため, 高校数学との接続を考慮して追加)			
IPSJ-04	T1	2.00	KA01.01
初等関数の導関数を理解している。 【備考】高校数学Ⅲの範囲(未履修の高校生もいるため, 高校数学との接続を考慮して追加)			
IPSJ-05	T1	2.00	KA01.01
微分法の基本的な関係(合成関数の微分法, 積の微分法)を理解している。 【備考】高校数学Ⅲの範囲(未履修の高校生もいるため, 高校数学との接続を考慮して追加)			
IPSJ-06	T1	1.00	KA01.01
偏微分の定義を理解している。 【備考】高校数学Ⅲの範囲(未履修の高校生もいるため, 高校数学との接続を考慮して追加)			
IPSJ-07	T1	2.00	KA01.01
微分の逆演算と原始関数を理解している。 【備考】高校数学Ⅲの範囲(未履修の高校生もいるため, 高校数学との接続を考慮して追加)			
IPSJ-08	T1	2.00	KA01.01
定積分の計算を理解している。 【備考】高校数学Ⅲの範囲(未履修の高校生もいるため, 高校数学との接続を考慮して追加)			

3. 数理統計学

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
DS-001a	場合の数, 順列・組み合わせの概念を理解している。 【備考】高校数学Aの範囲(未履修の高校生も存在)	T1	0.50	KA01.01
DS-002	確率の概念を理解し, 同時確率と条件付き確率の意味や違いを説明できる。 【備考】数理・データサイエンス・AIモデルカリキュラム(リテラシーレベル)によりカバー	T1	0.00	KA01.01
DS-003	平均(相加平均), 中央値, 最頻値の算出方法の違いを説明できる。 【備考】数理・データサイエンス・AIモデルカリキュラム(リテラシーレベル)によりカバー	T1	0.00	KA01.01
DS-004a	分散と標準偏差の意味と定義を説明できる。 【備考】数理・データサイエンス・AIモデルカリキュラム(リテラシーレベル)によりカバー	T1	0.00	KA01.01

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

DS-005	T1	0.50	KA01.01	
母(集団)平均と標本平均, 不偏分散と標本分散の違いを説明できる.				
DS-006	T1	0.50	KA01.01	
標準正規分布の分散と平均の値を知っている.				
DS-007	T1	0.00	KA01.01	
相関関数と因果関係の違いを説明できる. 【備考】数理・データサイエンス・AIモデルカリキュラム(リテラシーレベル)によりカバー				
DS-008	T1	0.25	KA01.01	
名義尺度, 順序尺度, 間隔尺度, 比例尺度の違いを説明できる.				
DS-009	T1	0.00	KA01.01	
一般的な相関係数(ピアソン)の分母と分子を説明できる. 【備考】数理・データサイエンス・AIモデルカリキュラム(リテラシーレベル)によりカバー				
DS-010	T1	1.00	KA01.01	
5つ以上の代表的な確率分布を説明できる.				
DS-011	T1	0.50	KA01.01	
二項分布の事象もサンプル数が増えていくとどのような分布に近似されるかを知っている.				
DS-012a	T1	1.00	KA01.01	
ピアソンの積率相関係数とクラメールの連関係数とスピアマンの順位相関係数の適用場面の違いを知っている.				
DS-013	T1	1.00	KA01.01	
ベイズの定理を説明できる.				
DS-028	E	1.00	KA01.01	KA01.05
点推定と区間推定の違いを説明できる.				
DS-029	E	1.00	KA01.01	KA01.05
帰無仮説と対立仮説の違いを説明できる.				
DS-030	E	1.00	KA01.01	KA01.05
第1種の過誤, 第2種の過誤, p値, 有意水準の意味を説明できる.				

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

DS-031 E 1.00 KA01.01 KA01.05
片側検定と両面検定の違いを説明できる.

DS-032a E 2.00 KA01.01 KA01.05
データ間に対応のある場合と無い場合の検定手法の違いを説明できる.

IPSJ-09 E 2.00 KA01.01
平均値の検定と平均値の差の検定(群間の対応あり, なしを含むt検定)を理解している.

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

DS-001b	T1	0.50	KA01.01
---------	----	------	---------

順列や組合せを式 nPr , nCr を用いて計算できる.

【備考】高校数学Aの範囲(未履修の高校生も存在)

DS-004b	T1	0.00	KA01.01
---------	----	------	---------

与えられたデータにおける分散と標準偏差が計算できる.

【備考】数理・データサイエンス・AIモデルカリキュラム(リテラシーレベル)によりカバー

DS-012b	T1	0.50	KA01.01
---------	----	------	---------

変数が量的, 質的どちらの場合でも関係の強さを算出できる.

DS-032b	E	2.00	KA01.01 KA01.05
---------	---	------	-----------------

推定する対象となるデータの対応の有無を考慮した上で適切な検定手法を選択し, 適用できる.

B1: データマイニング

1. 近接度の尺度

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-16-01	指標の基本的性質を理解している。	T1	0.50	KA01.03 KA01.04
ACM-16-02	Lkノルム: 特殊事例 - ユークリッド距離, マンハッタン距離を理解している。	T1	0.50	KA01.03 KA01.04
ACM-16-03	スコアと順位付けの使用, スコアとランキング手法の望ましい性質を理解している。	T1	0.50	KA01.03 KA01.04
ACM-16-04	比較可能にするためのデータの正規化を理解している。	T1	0.50	KA01.03 KA01.04
ACM-16-05	テキストに関するメトリクスを理解している。	T2	0.50	KA01.03 KA01.04
ACM-16-06	データの系列に対する相関係数などのメトリクスを理解している。	T2	0.50	KA01.03 KA01.04
ACM-16-07	SimRankなどのグラフの中の関係性に基づく類似度のメトリクスを理解している。	T2	0.50	KA01.03 KA01.04
ACM-16-08	グラフベースのメトリクスを理解している。	T2	0.50	KA01.03 KA01.04
ACM-16-09	時系列の類似度を測るメトリクス, 例えば動的時間伸縮法を理解している。	T2	0.50	KA01.03 KA01.04

種別	スキル	優先度	割り当て時間数(h)	DS-BoK KA番号
----	-----	-----	------------	-------------

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-16-10 T1 0.50 KA01.03 KA01.04
測定手法の概念や異なる種類のデータ(名義, 順序, 間隔, 比例)に対する関連性についても説明・比較できる.

ACM-16-11 T1 0.50 KA01.03 KA01.04
種々のデータの比較のために適切なメトリクスを選べる.

DS-023 E 5.00 KA01.05 KA01.01
ROC曲線, AUC(Area under the curve), を用いてモデルの精度を評価できる.

DS-024 E 5.00 KA01.05 KA01.01
混同行列(正誤分布のクロス表), Accuracy, Precision, Recall, F値をいった評価尺度を理解し, 精度を評価できる.

DS-025 E 5.00 KA01.05 KA01.01
MSE(Mean Square Error), MAE(Mean Absolute Error)といった評価尺度を理解し, 精度を評価できる.

DS-026 E 5.00 KA01.05 KA01.01
ホールドアウト法, 交差検証(クロスバリデーション)法の仕組みを理解し, 学習データ, パラメータチューニング用の検証データ, テストデータを作成できる.

DS-027 E 5.00 KA01.05 KA01.01
時間の経過とともに構造が変わっていくデータの場合は, 学習データを過去のデータとし, テストデータを未来のデータとすることを理解している.

DS-036 T1 5.00 KA01.05
適切な区間データ設定でヒストグラムを作成し, データのバラつき方を説明できる.

DS-037 T1 5.00 KA01.05
適切な軸設定でクロス集計表を作成し, 属性間のデータの偏りを把握できる.

DS-038 T2 5.00 KA01.05
量的変数の散布図を描き, 2変数の関係性を把握できる.

種別 態度

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-16-12 T1 0.50 KA01.03 KA01.04
典型的には多くのアプローチが存在すると分かりながら, 注意深く批判的にしかし豊かな想像力のあるアプローチができる.

2. データの準備

種別 知識		優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
ACM-17-01	T1	0.50	KA01.03 KA01.04	データ収集, それを持つ問題解決への関連性, 専門知識の重要性を理解し, 専門家の意見に対してオープンである.
ACM-17-02	T1	0.50	KA01.03 KA01.04	データベース, IoT, 写真, 動画, オンラインの情報源を含むデータ源があること: 特定の目的に利用するデータが十分なことを理解している.
ACM-17-03	T1	0.50	KA01.03 KA01.04	特定の目的に際してのデータ取得と使用における倫理的な考慮, データを紐づける際の個人情報に対する懸念, データに含まれる潜在的なバイアスに対する懸念を理解している.
ACM-17-04	T1	0.50	KA01.03 KA01.04	データラングリング(前処理・加工)ー データに含まれるエラー・抜け漏れの取り扱い, データクレンジング, データ妥当性の検証, データプロファイリング, データの変換, 適切なデータセットの結合: 質への考慮を理解している.
ACM-17-05	T1	0.50	KA01.03 KA01.04	不均衡・不十分・不要な属性などのデータセットの課題を扱う手法, 自動または手作業によるアプローチと両者のトレードオフについて理解している.
ACM-17-06	T1	0.50	KA01.03 KA01.04	特徴量の概念, 特徴抽出, 特徴表現, 特徴選択, 特徴生成を理解している.
種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
ACM-17-07	T1	0.50	KA01.03 KA01.04	問いの組み立て過程と問いに答えるためのデータ取得過程の繋がりについて説明できる.
ACM-17-08	T1	0.50	KA01.03 KA01.04	専門家と適切にやり取りする必要がある特定の領域を理解する能力を示すことができる.
ACM-17-09	T1	0.50	KA01.03 KA01.04	探索的データ解析や推論を成すにあたって, 要約統計量やデータの可視化を使用できる.
ACM-17-10	T1	0.50	KA01.03 KA01.04	データセットにおいて起こりうる課題を記述し, それらがもたらす影響とどう解決しうるかを示すことができる.
ACM-17-11	T1	0.50	KA01.03 KA01.04	特徴量生成の種々の手法を同定でき, それぞれの利点と含意を説明できる.

ACM-17-12	T1	0.50	KA01.03	KA01.04
特徴量選択と特徴量生成の類似点と相違点を記述することができ、特徴量生成がどのように特徴量を生み出すのかを示すことができる。				

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
------	-----	------------	-------------	--

ACM-17-13	T1	0.50	KA01.03	KA01.04
データ選択と準備において、適切な信頼度を設定でき、また、質の高いデータを扱うことの重要性を理解している。				

3. 情報抽出

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
------	-----	------------	-------------	--

ACM-18-01	E	1.75	KA01.03	KA01.04
情報抽出が役に立つ応用先を理解している。				

ACM-18-02	E	1.75	KA01.03	KA01.04
エンティティと関係抽出を理解している。				

ACM-18-03	E	1.75	KA01.03	KA01.04
ルールベース情報抽出手法とその応用先を理解している。				

ACM-18-04	E	1.75	KA01.03	KA01.04
統計ベース情報抽出手法とその応用先を理解している。				

ACM-18-05	E	1.75	KA01.03	KA01.04
抽出されたデータにおいて起こりうる問題を理解している。				

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
------	-----	------------	-------------	--

ACM-18-06	E	1.75	KA01.03	KA01.04
応用先の要件とデータに応じて、スキーマを設計できる。				

ACM-18-07	E	1.75	KA01.03	KA01.04
応用先に応じた情報抽出ルールが書ける。				

ACM-18-08 E 1.75 KA01.03 KA01.04
 情報抽出タスクのためにルールやモデル学習や関係予測などの学習アルゴリズムを適用できる。

種別	態度	通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
		ACM-18-09	E	1.75	KA01.03	KA01.04

データからの情報抽出において種々の手法があることを理解している。

4. クラスタ分析

種別	知識	通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
		ACM-19-01	T1	0.50	KA01.03	KA01.04
		クラスタリングのための適切な類似度指標を同定できる。				
		ACM-19-02	T1	0.50	KA01.03	KA01.04
		クラスタリングの質の評価を理解している。				
		ACM-19-03	T1	0.50	KA01.03	KA01.04
		k-平均法について反復することの必要性を考慮しながら理解している。				
		ACM-19-04	T1	0.50	KA01.03	KA01.04
		Density-basedアルゴリズムを理解している。				
		ACM-19-05	T1	0.50	KA01.03	KA01.04
		クラスタリングの応用先を理解している。				
		ACM-19-06	T2	0.50	KA01.03	KA01.04
		Mean shiftクラスタリングを理解している。				
		ACM-19-07	T2	0.50	KA01.03	KA01.04
		凝集型クラスタリングを理解している。				
		ACM-19-08	T2	0.50	KA01.03	KA01.04
		Grid-basedアルゴリズムを理解している。				

ACM-19-09 T2 0.50 KA01.03 KA01.04
 クラスタリングアルゴリズムの高速化と並列化の戦略を理解している。

種別 スキル				
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-19-10	T1	0.50	KA01.03	KA01.04
クラスタリングのための特徴選択の重要性を説明できる。				
ACM-19-11	T1	0.50	KA01.03	KA01.04
k-平均アルゴリズムの初期化規準選択の指針を提供できる。				
ACM-19-12	T2	0.50	KA01.03	KA01.04
相対的な利点と欠点を強調しながら、クラスタリング手法を比較できる。				
ACM-19-13	T2	0.50	KA01.03	KA01.04
種々のクラスタリング法が使われるべき状況や、代替手法がより望ましい状況を示すことができる。				
ACM-19-14	T2	0.50	KA01.03	KA01.04
複数のアルゴリズムをデータのテストセットに適用して結果を比較できる。				
ACM-19-15	T2	0.50	KA01.03	KA01.04
クラスタリングの効能と価値を強調する説明が提示できる。				

種別 態度				
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-19-16	T1	0.50	KA01.03	KA01.04
データサイエンスにおけるクラスタリングの役割について、見識に基づく肯定的な視点を創造することができる。				
ACM-19-17	T1	0.50	KA01.03	KA01.04
現実的なシナリオで用いられる大規模化可能で効率的なクラスタリングアルゴリズムの重要性を理解している。				

5. 分類・回帰

種別 知識				
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-20-01 T1 0.50 KA01.03 KA01.04
分類のための特徴量選択に関して考慮すべき事柄を理解している.

ACM-20-02 T1 0.50 KA01.03 KA01.04
KNNなどのインスタンスベース手法を理解している.

ACM-20-03 T1 0.50 KA01.03 KA01.04
決定木モデルを理解している.

ACM-20-04 T1 0.50 KA01.03 KA01.04
確率モデル, ナイーブベイズを理解している.

ACM-20-05 T2 0.50 KA01.03 KA01.04
ルールベース手法を理解している.

ACM-20-06 T2 0.50 KA01.03 KA01.04
サポートベクターマシンを理解している.

ACM-20-07 T2 0.50 KA01.03 KA01.04
ニューラルネットワークを理解している.

ACM-20-08 T2 0.50 KA01.03 KA01.04
分類と回帰の実社会適用を理解している.

ACM-20-09 T2 0.50 KA01.03 KA01.04
深層学習とそれに関連するソフトウェア(Caffe, TensorFlow, PyTorch)を理解している.

ACM-20-10 E 1.75 KA01.03 KA01.04
高速化と並列化戦略を理解している.

種別 スキル

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-20-11 T1 0.50 KA01.03 KA01.04
分類と回帰のための特徴量選択の重要性を説明できる.

ACM-20-12 T1 0.50 KA01.03 KA01.04
どちらの手法を使うべきか選択に資しうる, 予測精度・学習済みモデルの理解可能性などの基準を記述することができる.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-20-13 T2 0.50 KA01.03 KA01.04
回帰と分類の関連について同定できる。

ACM-20-14 T2 0.50 KA01.03 KA01.04
分類モデルまたは回帰モデルの利用から利益が得られるかもしれない重要な局面を認識できる。

ACM-20-15 T2 0.50 KA01.03 KA01.04
各手法を扱えるソフトウェアを特定でき、実際に扱える。

ACM-20-16 T2 0.50 KA01.03 KA01.04
分類手法を選び、それを選んだ理由を正当化でき、そして程々に複雑な事例に対しても適用する能力を示せる。

DS-021 E 5.00 KA01.05 KA01.01
単回帰分析について最小二乗法、回帰計数、標準誤差の説明ができる。

DS-022 E 5.00 KA01.05 KA01.01
重回帰分析において偏回帰係数と標準偏回帰係数、重相関関係について説明できる。

種別 態度

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-20-17 T1 0.50 KA01.03 KA01.04
現実的なシナリオで用いられる大規模化可能で効率的な分類・回帰アルゴリズムの重要性を理解している。

ACM-20-18 T2 0.50 KA01.03 KA01.04
分類と回帰、より一般に統計や機械学習とのつながりについて説明できる。

6. パターンマイニング

種別 知識

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-21-01 T2 0.50 KA01.03 KA01.04
相関パターンマイニングの概念を理解している。

ACM-21-02 T2 0.50 KA01.03 KA01.04
計算量に関して考慮すべき事項を理解している。

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-21-03 T2 0.50 KA01.03 KA01.04
 相関規則マイニング, Aprioriアルゴリズム, 頻出パターンマイニングアルゴリズムを理解している.

ACM-21-04 T2 0.50 KA01.03 KA01.04
 系列パターンマイニング, GSPアルゴリズムを理解している.

ACM-21-05 T2 0.50 KA01.03 KA01.04
 効率的で並行実行できるパターンマイニングアルゴリズムを理解している.

ACM-21-06 T2 0.50 KA01.03 KA01.04
 応用先を理解している.

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-21-07 T2 0.50 KA01.03 KA01.04
 有益な効果を生むために通常の下でAprioriアルゴリズムが使われる様々な領域を列挙することができる.

ACM-21-08 T2 0.50 KA01.03 KA01.04
 Aprioriアルゴリズムの実装を同定でき, それを意味のある適用先に適用できる.

ACM-21-09 T2 0.50 KA01.03 KA01.04
 パターンマイニングアルゴリズムの有用性を比較対照できる.

DS-033 E 1.00 KA01.02
 教師あり学習の分類(判別)モデルと教師なし学習のグループ化(クラスタリング)の違いを説明できる.

DS-034 E 1.00 KA01.02
 階層クラスタ分析と非階層クラスタ分析の違いを説明できる.

DS-035 E 2.00 KA01.02
 階層クラスタ分析において, デンドログラムの見方を理解し, 適切に解釈できる.

DS-077 E 0.50 KA01.03
 条件Xと事象Yの関係性をリフト値を用いて評価できる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-21-10 T2 0.50 KA01.03 KA01.04
 パターンマイニングがとて広いトピックであり, 多くの適用先があることを認識している.

7. 外れ値検出

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-22-01	T2	0.50	KA01.03 KA01.04

外れ値という概念の定義を理解している.

ACM-22-02	T2	0.50	KA01.03 KA01.04
-----------	----	------	-----------------

一般的手法 - データからモデルを構築した上で, あるデータ点はそれにフィットしないことに注意を払うことができる.

ACM-22-03	T2	0.50	KA01.03 KA01.04
-----------	----	------	-----------------

Zスコアなど, 1次元において数値的な外れ値を特定するパラメトリック手法を理解している.

ACM-22-04	T2	0.50	KA01.03 KA01.04
-----------	----	------	-----------------

確率分布関数の用途を理解している.

ACM-22-05	T2	0.50	KA01.03 KA01.04
-----------	----	------	-----------------

深さ優先手法の利用 - 点の集合に対して期待する凸包を特定した上で, 内側の点または外側の点を; 関連するグラフィカルアプローチ

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-22-06	T2	0.50	KA01.03 KA01.04

幅広い外れ値検出のアルゴリズムを適用できる.

ACM-22-07	T2	0.50	KA01.03 KA01.04
-----------	----	------	-----------------

外れ値検出におけるパラメトリック手法, ノンパラメトリック手法を比較対照できる.

ACM-22-08	T2	0.50	KA01.03 KA01.04
-----------	----	------	-----------------

外れ値検出手法が剽窃検知, 金融詐欺, ネットワーク侵入検知や他の適用分野を支援する方法を説明できる.

ACM-22-09	T2	0.50	KA01.03 KA01.04
-----------	----	------	-----------------

適切な例を使って外れ値検出が大切であることを説明できる.

種別	態度	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-22-10	T2	0.50	KA01.03 KA01.04
外れ値解析と外れ値検出に関する批判的で幅広い視点を涵養している。					

8. 時系列データ

種別	知識	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-23-01	E	1.75	KA01.03 KA01.04
Time seriesデータの性質について. Sequential temporal dataとの比較も含む。					
		ACM-23-02	E	1.75	KA01.03 KA01.04
データ変換 - 時系列データのノイズ除去, 正規化を理解している。					
		ACM-23-03	E	1.75	KA01.03 KA01.04
定常時系列と非定常時系列を理解している。					
		ACM-23-04	E	1.75	KA01.03 KA01.04
時系列データの離散系列データへの変換を理解している。					
		ACM-23-05	E	1.75	KA01.03 KA01.04
時系列データ予測 - 過去の値に基づいて未来の値を予測することを理解している。					
		ACM-23-06	E	1.75	KA01.03 KA01.04
時系列モチーフ - 時系列データにおいて頻繁に現れるパターンを理解している。					
		ACM-23-07	E	1.75	KA01.03 KA01.04
時系列クラスタリングと分類を理解している。					
		ACM-23-08	E	1.75	KA01.03 KA01.04
時系列データにおける外れ値検出 - 点および形状としての外れ値を理解している。					
種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-23-09 E 1.75 KA01.03 KA01.04
 意味のある時系列データが存在する状況を複数提供でき、そのデータをマイニングすることの重要性を示せる。

ACM-23-10 E 1.75 KA01.03 KA01.04
 時系列データを系列データに変換することが望ましいのがどういう時なのか、例を使って示せる。

ACM-23-11 E 1.75 KA01.03 KA01.04
 時系列データのクラスタリングと分類で使われるテクニックについて説明できる。

DS-066 T2 5.00 KA01.05
 時系列データとは何か、その基礎的な扱いについて説明できる。(時系列グラフ, 周期性, 移動平均など)

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-23-12	E	1.75	KA01.03 KA01.04
-----------	---	------	-----------------

いくつかの重要な適用先において、時系列データのデータマイニングが非常に重要であることが分かっている。

9. Webデータマイニング

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-24-01	T2	0.50	KA01.03 KA01.04
-----------	----	------	-----------------

ウェブスクレイピングとウェブスパイダリングの過程, ウェブアクセスと関連づけられたウェブクロールリングを理解している。

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

ウェブデータへのアクセスに関連する倫理指針を理解している。

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

ウェブデータにアクセスするためのソフトウェアライブラリの構造と機能性を理解している。

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

コミュニティ発見やリンク予測などの, ウェブデータに対する知識発見アプローチを理解している。

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-24-05 T2 0.50 KA01.03 KA01.04
ウェブアクセスのための2つの異なるライブラリについて, 利用するのに必要な能力や利用の容易さについて比較対照できる.

ACM-24-06 T2 0.50 KA01.03 KA01.04
公的に利用可能なウェブサイトから正確なデータをスクレイプするソフトウェアの利用方法を示せる.

ACM-24-07 T2 0.50 KA01.03 KA01.04
所与の制約に応じて, ウェブからデータを取得するソフトウェアを開発できる.

ACM-24-08 T2 0.50 KA01.03 KA01.04
ウェブから知識を発見するための効率的アルゴリズムを開発できる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
------	-----	------------	--------	------

ACM-24-09	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

倫理的な枠組みを考慮に入れつつ, 質の高いデータへのアクセスを促進・助長することができる.

10. 情報検索

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
------	-----	------------	--------	------

ACM-25-01	T2	0.50	KA01.03	KA01.04
-----------	----	------	---------	---------

検索プロセスの効率性を測定するために使うテクニックを理解している.

ACM-25-02 T2 0.50 KA01.03 KA01.04
情報が効率的に抽出されるために, データを保存し組織化するための様々な手法, すなわち符号化機能の利用について理解している.

ACM-25-03 T2 0.50 KA01.03 KA01.04
検索戦略の概念, すなわち検索を広げることと狭めることの役割を理解している.

ACM-25-04 T2 0.50 KA01.03 KA01.04
検索過程におけるキーワード選択, すなわちブール演算子の利用方法を理解している.

ACM-25-05 T2 0.50 KA01.03 KA01.04
順序データの検索を理解している.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-25-06	T2	0.50	KA01.03	KA01.04	テキストベースの資料を検索するためのテクニックを理解している。
ACM-25-07	T2	0.50	KA01.03	KA01.04	ドキュメント集合の検索:選択された項目の名前をリスト化する戦略を理解している。
ACM-25-08	T2	0.50	KA01.03	KA01.04	非テキストベースデータの特徴量同定と特徴量抽出:写真, 音声, 動画に用いられる検索戦略を理解している。
ACM-25-09	T2	0.50	KA01.03	KA01.04	ハッシュ化, インデキシングとフィルタリングの役割を理解している。
ACM-25-10	T2	0.50	KA01.03	KA01.04	テキストベースの資料を検索するアプローチを理解している。
ACM-25-11	T2	0.50	KA01.03	KA01.04	リレーショナルデータベースを作成し検索するテクニックを理解している。
ACM-25-12	T2	0.50	KA01.03	KA01.04	種々のリレーショナルデータベースや, 非リレーショナルデータベースや, それ以外のデータベースの形式を理解している。
ACM-25-13	T2	0.50	KA01.03	KA01.04	ウェブベースの情報検索, 相互に結合されたノードのグラフとして捉えられたウェブ, グラフ理論から得られる関連指標, ウェブベースの検索を促進するPageRankと関連指標を理解している。
種別 スキル					
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号		
ACM-25-14	T2	0.50	KA01.03	KA01.04	所与の情報検索タスクに対して検索戦略を工夫できる。
ACM-25-15	T2	0.50	KA01.03	KA01.04	情報検索過程に関連するかもしれない倫理的懸念について説明できる。
ACM-25-16	T2	0.50	KA01.03	KA01.04	検索をスピードアップするための並列化技術の利用が適切な状況を特定できる。
ACM-25-17	T2	0.50	KA01.03	KA01.04	ウェブベースの検索を下支えする効果的な戦略の主要要素の概要を説明できる。

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-25-18 T2 0.50 KA01.03 KA01.04
 画像, 録音, 動画に関連する情報検索タスクにおいて使えるソフトウェアを特定できる.

ACM-25-19 T2 0.50 KA01.03 KA01.04
 SQLを使って, リレーショナルデータベース構造を作成し利用できる.

ACM-25-20 T2 0.50 KA01.03 KA01.04
 デジタル図書館の運用において, 情報検索が果たし得る役割を説明できる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
------	-----	------------	--------	------

ACM-25-21 T2 0.50 KA01.03 KA01.04

情報検索のための効率的で効果的なアプローチを下支えする種々の考慮の重要性を認識し, その大切さを理解している.

B2: 人工知能

1. 一般

種別	知識	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
----	----	------	-----	------------	-------------

ACM-59-01	T1	0.25	KA01.02
-----------	----	------	---------

AIの歴史を理解している。

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

AIの現実(AIとは何か, AIは何ができるか)と認識の隔たりを理解している。

ACM-59-03	T1	0.25	KA01.02
-----------	----	------	---------

AIのサブフィールドを理解している: 知識表現, 論理・確率推論, プランニング, 認識, 自然言語処理, 学習, ロボティクス(物理・仮想)

種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
----	-----	------	-----	------------	-------------

ACM-59-04	T1	0.25	KA01.02
-----------	----	------	---------

データサイエンスにおいて必要となったときに, 有用な概念や手法を認知するためにAIの主な分枝を説明することができる。

ACM-59-05	T2	0.50	KA01.02
-----------	----	------	---------

AIとは何か, AIはAIが動作するようにデータを収集し利用するとともに, データサイエンティストが利用できるデータを収集し生成する, ということを明確に説明できる。

ACM-59-06	T2	0.50	KA01.02
-----------	----	------	---------

(物理的な, あるいは仮想的な)ロボットやエージェント, マルチエージェントシステムがAIを埋め込み, 動くようにするためにデータをどのように収集し利用するかを定量的に説明することができる。

ACM-59-07	T2	0.50	KA01.02
-----------	----	------	---------

AIによって収集あるいは生成されたデータはデータサイエンスの応用先に有用な可能性があることを説明できる。

種別	態度	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
----	----	------	-----	------------	-------------

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

AIは新しい分野というわけではなく, むしろ長く豊かな歴史を持った分野であることを十分に理解している。

ACM-59-09	T2	0.50	KA01.02
必要なときにどこでアプローチを見つけるべきか知っておくために、既存の手法の再発見を避けるために、AIの主な関心事およびそれらが取り組んでいる問題の種類を知っている。			

2. 知識表現と推論(論理ベースモデル)

種別	知識			
通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
ACM-60-01	T2	0.50	KA01.02	
述語論理とその使用事例を理解している。				
ACM-60-02	T2	0.50	KA01.02	
自動推論: 前向き推論, 後ろ向き推論を理解している。				
ACM-60-03	T2	0.50	KA01.02	
大規模なシステムに統合された推論を理解している(例: Watson)。				
ACM-60-04	T2	0.50	KA01.02	
オントロジー, 知識グラフ(例: protege, ConceptNet, YAGO, UMLS)を理解している。				
ACM-60-05	E	0.50	KA01.02	
自動推論: 導出, 定理証明を理解している。				
ACM-60-06	E	0.50	KA01.02	
自動推論のための言語を理解している。				
種別	スキル			
通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
ACM-60-07	T2	0.50	KA01.02	
自然言語の言明を述語論理で表現することができる。				
ACM-60-08	T2	0.50	KA01.02	
述語論理の言明を自然言語で表現することができる。				
ACM-60-09	T2	0.50	KA01.02	
述語論理の使用事例と限界を述べることができる。				

ACM-60-10	T2	0.50	KA01.02
効率的な自動推論を行うアルゴリズムおよびシステムの例を挙げることができる。			
ACM-60-11	T2	0.50	KA01.02
例えば前向き推論・後ろ向き推論による論理ベースの自動推論を記述することができる。			
ACM-60-12	T2	0.50	KA01.02
推論がデータ駆動の大規模なシステムに統合されている事例を挙げることができる。			
ACM-60-13	E	0.50	KA01.02
自動定理証明の特定の手法を記述することができる。			
ACM-60-14	E	0.50	KA01.02
既存の技術の例を挙げながら、オントロジーとは何か、それが使える状況(例えば質問応答)、それがどのように使われるか(例えば曖昧さの解消を補助するために)を記述することができる。			
ACM-60-15	E	0.50	KA01.02
オントロジーがどのように構築されるか記述できる。			
ACM-60-16	E	0.50	KA01.02
中程度のサイズの推論問題を実装できる。			

種別	態度	優先度	割り当て時間数(h)	DS-BoK KA番号
----	----	-----	------------	-------------

ACM-60-17	T2	0.50	KA01.02
論理ベースの知識表現のメリットや限界を十分に理解している。			
ACM-60-18	T2	0.50	KA01.02
特定の応用先に利用するために、形式的な論理や論理ベースのアルゴリズムの豊かな歴史を意識している。			

3. 知識表現と推論(確率ベースモデル)

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
----	----	-----	------------	-------------

ACM-61-01	T1	0.25	KA01.02
基本的な概念: 確率変数, 確率の公理, 独立性, 条件付き確率, 周辺確率を理解している。(コンピュータサイエンスの分野ではなくデータサイエンスの分野において基本的な知識である確率も参照すること。			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-61-02	T1	0.25	KA01.02
因果モデルを理解している.			
ACM-61-03	T2	0.50	KA01.02
ベイジアンネットワークを理解している.			
ACM-61-04	T2	0.50	KA01.02
マルコフ決定過程を理解している.			
ACM-61-05	E	0.50	KA01.02
強化学習を理解している.			
ACM-61-06	E	0.50	KA01.02
確率論理モデルを理解している(例:マルコフ論理ネットワーク).			
種別	スキル		
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-61-07	T1	0.25	KA01.02
確率的な推論の必要性を正当化できる.			
ACM-61-08	T1	0.25	KA01.02
確率変数, 独立性などの基本的な概念を定義できる.			
ACM-61-09	T1	0.25	KA01.02
確率の公理を述べることができる.			
ACM-61-10	T1	0.25	KA01.02
上の基本的な概念と公理を使って単純なシステムをモデル化し, 問題に答えることができる.			
ACM-61-11	T1	0.25	KA01.02
因果モデルとは何か, それはどのように使われるかを記述することができる.			
ACM-61-12	T2	0.50	KA01.02
小規模/中規模のサイズの例を挙げながらベイジアンネットワークとは何か述べることができる.			
ACM-61-13	T2	0.50	KA01.02
ベイジアンネットワークが有用になり得る状況を示すことができる(例:医療診断の問題).			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-61-14	T2	0.50	KA01.02	ベイジアンネットワークがどのようにして推論に利用されるのかを示すことができる, たいていの場合で正確な推論は実行不可能であることを理解している, より効率的な推論のアプローチの例を述べる(例: 信念の伝播).
ACM-61-15	T2	0.50	KA01.02	ベイジアンネットワークが示す独立性の関係を同定することができる.
ACM-61-16	T2	0.50	KA01.02	小規模/中規模なサイズの例を挙げながら, マルコフ決定過程とは何かを述べる(例: 小規模/中規模なサイズの例を挙げながら, マルコフ決定過程とは何かを述べる).
ACM-61-17	T2	0.50	KA01.02	マルコフ決定過程が有用になり得る状況を示すことができる(例: 最適化問題, 制御問題).
ACM-61-18	T2	0.50	KA01.02	マルコフ決定過程がどのようにして推論に利用されるのかを示すことができる.
ACM-61-19	E	0.50	KA01.02	小規模/中規模な問題に対してベイジアンネットワークを構成することができる.
ACM-61-20	E	0.50	KA01.02	小規模/中規模な問題に対してベイジアンネットワークを構成する学習アルゴリズムを適用することができる.
ACM-61-21	E	0.50	KA01.02	マルコフ決定過程のパラメータがどのように学習されるのか述べる(例: そのパラメータを学習するアルゴリズムの例を挙げる).
ACM-61-22	E	0.50	KA01.02	適切な問題に対して強化学習アルゴリズムを適用することができる.
ACM-61-23	E	0.50	KA01.02	有用な応用先を同定しながら, マルコフ論理ネットワークなどの確率論理モデルの例を挙げる(例: 有用な応用先を同定しながら, マルコフ論理ネットワークなどの確率論理モデルの例を挙げる).
ACM-61-24	E	0.50	KA01.02	小規模/中規模な問題に対して確率論理モデルを構成することができる.
種別	態度			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	

ACM-61-25 T1 0.25 KA01.02
 確率ベースの知識表現およびその上で推論を行う手法のメリットや限界を十分に理解している。

4. プランニング・探索戦略

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-62-01	T2	0.50	KA01.02
-----------	----	------	---------

問題に対して考えられる解の状態空間の表現を理解している。

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

状態空間における(事前情報のない)幅優先探索, 深さ優先探索を理解している。

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

状態空間における(事前情報のある)ヒューリスティック探索を理解している(例:A*探索)。

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

確率的な探索アルゴリズムを理解している(例: 遺伝的アルゴリズム, 焼きなまし法)。

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

制約充足問題とそれに関する手法を理解している。

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-62-06	T2	0.50	KA01.02
-----------	----	------	---------

問題に対する解が考えられる解空間における状態であるとのようにみなすことができるかを説明できる。

ACM-62-07	T2	0.50	KA01.02
-----------	----	------	---------

与えられた問題に対して, それを多次元の状態空間における探索としてモデル化することができる。

ACM-62-08	T2	0.50	KA01.02
-----------	----	------	---------

幅優先探索と深さ優先探索がグラフとしてモデル化された解空間を探索するのにどのように使われるかを説明できる。

ACM-62-09	T2	0.50	KA01.02
-----------	----	------	---------

ヒューリスティクスがグラフ/状態空間の探索の速度を(潜在的に)上げるのにどのように使われるかを説明できる。

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-62-10 E 0.50 KA01.02
 事前情報なしの探索を状態空間としてモデル化された問題に対する解を見つけるために適用することができる(ここで状態を表現するグラフは入力として与えられるというよりは探索が実行される中で構築されていく).

ACM-62-11 E 0.50 KA01.02
 小規模な問題に対してヒューリスティクスを設計することができる.

ACM-62-12 E 0.50 KA01.02
 小規模/中規模な問題に対して事前情報ありの探索を適用することができる.

ACM-62-13 E 0.50 KA01.02
 小規模/中規模な問題に対して確率的な探索を適用することができる.

ACM-62-14 E 0.50 KA01.02
 確率的な探索アルゴリズムはどのようにして空間を広く探索するという問題に対処しているのか説明できる(例: 局所最適解を避ける), 確率的な探索アルゴリズムはどのようにして有望な解空間の局所的な探索に対処しているのかを説明できる.

ACM-62-15 E 0.50 KA01.02
 問題に対する解がある変数の制約およびそれらの間の関係性にどのように関与するのか説明できる: これらの制約を明らかにする手法を記述できる.

ACM-62-16 E 0.50 KA01.02
 探索アルゴリズムを実装することができる.

ACM-62-17 E 0.50 KA01.02
 小規模な問題を制約充足問題としてモデル化することができる.

ACM-62-18 E 0.50 KA01.02
 小規模/中規模な問題に対して制約充足アルゴリズムを適用することができる.

種別 態度

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-62-19 T2 0.50 KA01.02
 状態空間には複数の許容解が存在し得るとともに, それらを見つける方法も複数存在し得るということをも十分に理解している. 種々の解法および問題解決のアプローチは, 最適性の要請や時間制約などの外部の条件によって使われるべきである.

ACM-62-20 T2 0.50 KA01.02
 アルゴリズム, ヒューリスティクス, そして問題に対する解の最適性の関係を十分に理解している.

B3: 機械学習

1. 一般

種別	知識	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-63-01	T1	0.25	KA01.02
		機械学習の歴史を理解している。			
		ACM-63-02	T1	0.25	KA01.02
		教師あり学習, 教師なし学習, 強化学習, 深層学習を含む機械学習の主要なタスクを理解している。			
		ACM-63-03	T1	0.25	KA01.02
		記号的な学習と数値的な学習, および統計的なアプローチと構造的・統語的なアプローチの違いを理解している。			
		ACM-63-04	T1	0.25	KA01.02
		原理に基づいた最適化のアプローチとしての学習アルゴリズムを理解している。			
		ACM-63-05	T1	0.25	KA01.02
		「機械学習を行うこと」はデータマイニングの一手法であり, 一つのプロセスであると理解している。			
		ACM-63-06	T1	0.25	KA01.02
		頑健な評価の重要性を理解している。			
		ACM-63-07	T1	0.25	KA01.02
		データの質, 正則化の必要性を含む機械学習における困難を理解している。			

種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-63-08	T1	0.25	KA01.02
		教師あり学習, 教師なし学習, 強化学習, 深層学習の目標, 入出力を比較することができる。			
		ACM-63-09	T1	0.25	KA01.02
		異なる種類のデータ駆動の問題はそれぞれ異なるアプローチによって解くことができることを知っており, それらを適切に結びつけることができる。			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-63-10	T1	0.25	KA01.02	機械学習のモデル・アルゴリズムは数学的・統計的な基盤に基づく技術であることをおおまかに説明できる。
ACM-63-11	T1	0.25	KA01.02	データマイニングの手法として「機械学習を行うこと」の過程をたどることができる:顧客が解きたいと思っている問題を理解し, その問題を解くのに関係するデータを収集し, 生のデータを特徴量に変換し, 適切な機械学習の手法を選択し, チューニングを行い, (しばしばベースラインに対する)性能を評価し, 結果や洞察を述べる。
ACM-63-12	T1	0.25	KA01.02	訓練データに対するフィッティングと新しいデータに対する汎化のトレードオフ, およびモデルの複雑度, データ数, 特徴量数がどのようにこのトレードオフに影響するかについて議論することができる。これをハイパーパラメータの役割, 設定と関連づけることができる。
ACM-63-13	T1	0.25	KA01.02	性能, 解釈可能性, 大規模化可能性のトレードオフを十分に理解している。異なる最適化の関数や技術は異なるトレードオフを生み出すことがあることを認識している。
ACM-63-14	T2	0.50	KA01.02	単純な最適化関数や学習方法の導出を最も基本的な原理から追うことができる(例:決定木の原理を情報理論から追う, ロジスティック回帰を最尤法や確率的勾配降下法から追う, 主成分分析を分散最小化や固有値から追う)。
ACM-63-15	T2	0.50	KA01.02	ブートストラップ法や統計学的有意差検定を用いてモデルの性能を分析することができる。
ACM-63-16	T2	0.50	KA01.02	どのように効率的にモデルを製品に移行するか, 初期から移行をサポートするツールをどのように選ぶのかを理解している。
ACM-63-17	T2	0.50	KA01.02	データサイズによってどのツールを使うべきか理解している:ビッグデータに対しては並列実行できる機械学習ツールを選ぶことは必須であるし, そうでなければ学習過程により多くの時間を要するものも許容される。
ACM-63-18	T2	0.50	KA01.02	最先端の機械学習ツールが利用可能であることに気づいている。
ACM-63-19	E	0.50	KA01.02	特にデータの前処理, モデル選択, モデル構造探索, ハイパーパラメータチューニングを含む機械学習のパイプラインをどのように自動化しているかを含め, 自動機械学習(メタ学習)について記述することができる。
DS-123	E	1.00	KA01.02	機械学習にあたる解析手法の名称を3つ以上知っており, 手法の概要を説明できる。

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

DS-124	E	1.00	KA01.02	指示を受けて機械学習のモデルを使用したことがあり, どのような問題を解決することができる理解している.
DS-125	T1	0.50	KA01.02	「教師あり学習」「教師なし学習」の違いを理解している.
DS-126	T1	0.50	KA01.02	過学習とは何か, それをもたらす問題について説明できる.
DS-127	E	0.50	KA01.02	次元の呪いとは何か, その問題について説明できる.
DS-128	E	0.50	KA01.02	教師あり学習におけるアノテーションの必要性を説明できる.
DS-129	E	0.50	KA01.02	観測されたデータにバイアスが含まれる場合や, 学習した予測モデルが少数派のデータをノイズと認識してしまった場合などに, モデルの出力が差別的な振る舞いをしてしまうリスクを理解している.
DS-130	E	0.50	KA01.02	機械学習における大域的(global)な説明(モデル単位の各変数の寄与度など)と局所的(local)な説明(予測するレコード単位の各変数の寄与度など)の違いを理解している.
種別	態度			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-63-20	T1	0.25	KA01.02	機械学習は最近人気を博しているが, 最近の発明ではないことを十分理解している. 新しい発明が必要だと考える前に, 既存の手法を探ることができる.
ACM-63-21	T1	0.25	KA01.02	機械学習は場当たりのトリックの集合ではなく, 責任を持って使用するべきだということを十分に理解している.
ACM-63-22	T1	0.25	KA01.02	一般に「機械学習を行う」というのは都合よくフォーマットされたデータセットに機械学習プログラムを適用するといった単純なプロセスではなく, 顧客の目標へと向かうプロセスであるということを理解している.
ACM-63-23	T1	0.25	KA01.02	学習済みのモデルを比較する側面として, 経験誤差最小化からモデルのサイズ・複雑度, 人間の解釈可能性まで様々あることを十分に理解している.

ACM-63-24 T1 0.25 KA01.02

質, 効率, 解釈可能性などモデル比較のためのすべての側面を考えた上で結果を公正に述べ, 率直に比較する責任を理解している.

2. 教師あり学習

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-64-01		T1	0.25	KA01.02
教師あり学習の主要なタスクを理解している: 回帰と分類				
ACM-64-02		T1	0.25	KA01.02
回帰と分類の使用例を理解している.				
ACM-64-03		T1	0.25	KA01.02
モデルの複雑度と汎化性の関係を含む機械学習における重要な懸念事項やトレードオフを理解している, バイアス・バリエーションのトレードオフを理解している, 単純なモデルを使うモチベーションであるところのオッカムの剃刀を理解している.				
ACM-64-04		T1	0.25	KA01.02
訓練データ, テストデータ, 検証データを分割する必要性を理解している. 訓練誤差と汎化誤差の定義を理解している.				
ACM-64-05		T1	0.25	KA01.02
分類タスクにおける一般的な評価指標(例: 正解率, 感度, 特異性, 適合率, 再現率, F1スコア, ROC曲線下面積, リグレット)および回帰タスクにおける一般的な評価指標(例: 二乗平均平方根誤差, 平均絶対誤差, 決定係数)を理解している.				
ACM-64-06		T1	0.25	KA01.02
検証データの必要性, 交差検証のプロセスとその目標: ハイパーパラメータのチューニングおよびモデル性能の評価を理解している.				
ACM-64-07		T1	0.25	KA01.02
データ数やクラス階層のような, 訓練データ・テストデータ・検証データの質を評価する基準を理解している.				
ACM-64-08		T1	0.25	KA01.02
分類/回帰アルゴリズムで, 線形/非線形なものをそれぞれ1つ以上理解している(例: 線形回帰, 線形分類, ロジスティック回帰, 近傍法, ナイーブベイズ, 決定木学習).				
ACM-64-09		T1	0.25	KA01.02
多項式の特徴量やアンサンブルを含む基本的なアルゴリズムへの一般的な拡張を理解している(例: バギングモデル, ブースティングモデル, ランダムフォレスト).				

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-64-10	T2	0.50	KA01.02	あるモデルについてバイアスが大きいのかバリエーションが大きいのか判断するためのアプローチを理解している(例: 訓練データに対する性能とテストデータに対する性能の比較, 学習曲線).
ACM-64-11	T2	0.50	KA01.02	特徴量を増やしたり減らしたりする理由をそれぞれ2つ以上理解しており, そのトレードオフも理解している.
ACM-64-12	T2	0.50	KA01.02	教師あり学習の分類器がどのように多クラス分類問題に適応するのか, 2クラス分類モデルがどのように多クラス分類問題に拡張することができるかを理解している.
ACM-64-13	T2	0.50	KA01.02	巨視的な/微視的な指標を用いてどのように性能を表現するか理解している.
ACM-64-14	T2	0.50	KA01.02	1つ以上の発展的な教師あり学習のアルゴリズムを理解している(例: カーネルSVM, ニューラルネットワーク).
ACM-64-15	E	0.50	KA01.02	基本的な原理から教師あり学習のアルゴリズムを導出することができる.

種別	スキル			
通し番号	優先度	割り当て時間数(h)		DS-BoK KA番号
ACM-64-16	T1	0.25		KA01.02
				混同行列を用いて分類モデルの性能を表現できる.
ACM-64-17	T1	0.25		KA01.02
				分類/回帰タスクの評価指標の長所/短所を比較することができる.
ACM-64-18	T1	0.25		KA01.02
				2つ以上の実用的な分類/回帰アルゴリズムのトレードオフを比較することができる.
ACM-64-19	T1	0.25		KA01.02
				小規模/中規模なデータセットに対して2つ以上の分類/回帰アルゴリズムを適用することができる.
ACM-64-20	T1	0.25		KA01.02
				学習済みモデルから得られる情報という観点で訓練誤差と汎化誤差を比較することができる.
ACM-64-21	T1	0.25		KA01.02
				様々な指標を用いてアルゴリズムの性能を比較することができる.

ACM-64-22 T1 0.25 KA01.02
 小規模/中規模/大規模なデータセットに対して2つ以上の拡張アルゴリズムを適用することができる(例:アンサンブルアルゴリズム), 様々な指標を用いてアルゴリズムの性能を比較することができる.

ACM-64-23 T1 0.25 KA01.02
 多項式の特徴量やアンサンブルなどの拡張アルゴリズムが適切であるのはどんなときか, それぞれが対処可能な問題に基づいて正当化することができる.

ACM-64-24 T2 0.50 KA01.02
 大規模なデータセットに対して2つ以上の分類/回帰アルゴリズムを適用することができる.

ACM-64-25 T2 0.50 KA01.02
 大規模なデータセットに対して1つ以上の拡張アルゴリズムを適用することができる.

ACM-64-26 T2 0.50 KA01.02
 バイアス/バリエーションを軽減する手法を適用することができる.

ACM-64-27 T2 0.50 KA01.02
 中規模/大規模なサイズの問題に対して特徴量の追加/選択を適用することができる.

ACM-64-28 T2 0.50 KA01.02
 発展的な教師あり学習のアルゴリズムを適用することができる(例:カーネルSVM, ニューラルネットワーク).

ACM-64-29 E 0.50 KA01.02
 単純な最適化関数および学習アルゴリズムを基本的な原理から導出することができる(例:最尤法や確率的勾配降下法を用いてロジスティック回帰を導出する). これらの技術を類似のモデルに拡張することができる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-64-30	T1	0.25	KA01.02
アルゴリズムの選択と評価指標は学習されたモデルの質にとってどれくらい重要であるか十分理解している. これらの選択は重要な利害関係者(開発したモデルを提供する相手)に影響を及ぼすこと, 重要な利害関係者とともによすべきであるということを知っている.			

ACM-64-31	T1	0.25	KA01.02
自分たちが自信を持っている, 原理に基づいたモデルの評価方法を適用することの重要性を十分理解している.			

3. 教師なし学習

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-65-01 T1 0.25 KA01.02
クラスタリングや次元削減などを含む教師なし学習の主要なタスクを理解している。

ACM-65-02 T1 0.25 KA01.02
両方のタスクの使用例を理解している(例: データ探索/要約/可視化, 特徴量選択, データ圧縮, データのノイズ除去, プロトタイプ学習, 推薦システム, トピックモデリング)。

ACM-65-03 T1 0.25 KA01.02
1つ以上のクラスタリングアルゴリズムを理解している(例: k平均法, 階層クラスタリング)。

ACM-65-04 T1 0.25 KA01.02
connectivity-basedクラスタリングとcentroid-basedクラスタリングのトレードオフを理解している。

ACM-65-05 T1 0.25 KA01.02
1つ以上の次元削減アルゴリズムを理解している(例: 主成分分析)。

ACM-65-06 T1 0.25 KA01.02
特徴量変換, 特徴量選択, 特徴量の射影の類似点, 相違点を理解している。

ACM-65-07 T2 0.50 KA01.02
1つ以上の発展的なクラスタリングアルゴリズムを理解している(例: ガウス混合モデルのような密度ベースのアルゴリズム)。

ACM-65-08 T2 0.50 KA01.02
1つ以上の発展的な次元削減アルゴリズムを理解している(例: 独立成分分析, 非負値行列因子分解)。

ACM-65-09 E 0.50 KA01.02
アルゴリズムを効率的に実装するための数学的な手法を1つ以上理解している(例: 行列分解, 特異値分解, 主成分分析のための固有値分解)。

ACM-65-10 E 0.50 KA01.02
1つ以上の発展的なアルゴリズムを理解している(例: スペクトラルクラスタリング, カーネルk平均法, カーネル主成分分析, 潜在ディリクレ配分法)。

ACM-65-11 E 0.50 KA01.02
主成分分析と自己符号化器とのつながりを理解している, 非線形次元削減への一般化を理解している。

ACM-65-12 E 0.50 KA01.02
基本的な原理から教師なし学習のアルゴリズムを導出することができる。

種別 スキル

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-65-13 T1 0.25 KA01.02
 小規模/中規模/大規模なデータセットに対して1つ以上クラスタリング/次元削減アルゴリズムを適用することができる。

ACM-65-14 T1 0.25 KA01.02
 様々な指標を用いて教師なし学習アルゴリズムの性能を表現することができる(例:可視化,(可能ならば)正解データとの比較, クラスタ密度のような指標の計算, 他の応用先での有用性という間接的な指標)。

ACM-65-15 T1 0.25 KA01.02
 k平均法のクラスタの数や主成分分析の要素数などのハイパーパラメータをどのように選択するかを記述する手法を説明でき, かつ適用できる。

ACM-65-16 T2 0.50 KA01.02
 2つ以上のクラスタリングアルゴリズムのトレードオフを比較することができる。

ACM-65-17 T2 0.50 KA01.02
 2つ以上の次元削減アルゴリズムのトレードオフを比較することができる。

ACM-65-18 E 0.50 KA01.02
 発展的な教師なし学習のアルゴリズムを適用することができる。

ACM-65-19 E 0.50 KA01.02
 単純な最適化関数および学習アルゴリズムを基本的な原理から導出することができる(例:分散最小化や固有値を用いて主成分分析を導出する)。これらの技術を類似のモデルに拡張することができる。

種別 態度

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-65-20 T1 0.25 KA01.02
 アルゴリズムの選択と評価指標は学習されたモデルの質にとってどれくらい重要であるか十分理解している。これらの選択は重要な利害関係者(開発したモデルを提供する相手)に影響を及ぼすこと, 重要な利害関係者とともになすべきであるということを知っている。

ACM-65-21 T1 0.25 KA01.02
 自分たちが自信を持っている, 原理に基づいたモデルの評価方法を適用することの重要性を十分理解している。

ACM-65-22 T2 0.50 KA01.02
 教師なし学習はデータ探索・理解・要約・可視化に役立つ技術を提供することを十分理解している。

ACM-65-23 T2 0.50 KA01.02
 教師なし学習は教師あり学習アルゴリズムの質や効率を向上させる有用な前処理段階となり得ることを十分理解している。

4. 混合手法

種別 知識		優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
ACM-66-01	E	0.50	KA01.02	データ構造やデータ点どうしの相関関係が学習モデルで有効活用されるような機械学習の問題や領域の例を理解している(例:時系列予測, 系列予測, 推薦システム).
ACM-66-02	E	0.50	KA01.02	時間依存性や各データ点に共有される情報の仮定が学習においてどのように有効活用され得るのか理解している.
ACM-66-03	E	0.50	KA01.02	混合アプローチではなく教師あり学習/教師なし学習アプローチを利用することの欠点を理解している(例:モデルの解釈可能性や性能などの問題).
ACM-66-04	T2	0.50	KA01.02	問題に対して標準的なアプローチを1つ以上理解している(例:系列予測には隠れマルコフモデルを用いる, 推薦システムには強調フィルタリングを用いる).
ACM-66-05	T2	0.50	KA01.02	この文脈で訓練データとテストデータを分割する必要性を理解している.
ACM-66-06	T2	0.50	KA01.02	選択されたタスクに対する一般的な評価指標を理解している(例:再現率, 適合率, 推薦システムにおけるF1スコア).
ACM-66-07	T2	0.50	KA01.02	選択された問題に対する訓練データ・検証データ・テストデータの質を評価する基準を理解している.
種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
ACM-66-08	E	0.50	KA01.02	問題をそれに対する学習の枠組みに対応づけることができる, すなわちデータを入出力に対応させ, ハイパーパラメータの設定について熟慮し, 適切な学習アルゴリズムを実行することができる.
種別 態度		優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
ACM-66-09	E	0.50	KA01.02	機械学習モデルに存在する困難は一般にこれらの状況ではより顕著になり得ることを認識している(例:時間の不均一性, データの過疎性).

5. 深層学習

種別	知識	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-67-01	T2	0.50	KA01.02
	(深層ではないネットワークも含めて)多層ニューラルネットワークはどのように学習し, 入力の特徴量を高次の特徴量に符号化しているかを理解している.				
		ACM-67-02	T2	0.50	KA01.02
	順伝播型ネットワーク, 畳み込みニューラルネットワーク(CNN), 再帰型ニューラルネットワーク(RNN), LSTMなどの一般的な深層学習のアーキテクチャ, 目的, 特性を理解している.				
		ACM-67-03	T2	0.50	KA01.02
	一般的な深層学習アプローチの実践的な課題を理解している(例: 深層学習のアーキテクチャの選択, 十分なデータの確保, 過学習の可能性, 学習時間の長さ, 解釈可能性).				
		ACM-67-04	T2	0.50	KA01.02
	早期終了, パラメータ共有, ドロップアウトなどの深層学習アーキテクチャにおける正則化の手法の例を理解している.				
		ACM-67-05	T2	0.50	KA01.02
	GPUを搭載したツールや分散システムなどの深層学習における他の課題を解決する手法の例を理解している.				
		ACM-67-06	T2	0.50	KA01.02
	データサイズに対応できる適切なツールを選択することができる: 特にビッグデータの要求を並列実行できる深層学習ツールを用いて処理する.				
		ACM-67-07	T2	0.50	KA01.02
	最先端の深層学習ツールが利用可能であることに気づいている.				
		ACM-67-08	T2	0.50	KA01.02
	深層学習の文脈で現れる, よく用いられる学習アルゴリズムを1つ以上理解している(例: 順伝播型ネットワークにおいて逆伝播がどのように利用されるか, 畳み込みネットワークにおいて高次の特徴量を学習するのに逆伝播がどのように利用されるか, 再帰型ネットワークにおいて時間ステップを通して逆伝播がどのように利用されるか).				
		ACM-67-09	T2	0.50	KA01.02
	畳み込み操作とは何か, なぜそれが有用なのかを理解している(例: 画像に含まれる垂直なエッジを検出する).				
		ACM-67-10	T2	0.50	KA01.02
	プーリングを理解している: maxプーリングなどのプーリング関数の例と実用例を理解している.				

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-67-11 T2 0.50 KA01.02
 再帰型ニューラルネットワークにおける長期/短期のデータ依存性の課題を理解している:LSTM
 など1つ以上の解決策を理解している.

ACM-67-12 E 0.50 KA01.02
 敵対的生成ネットワーク(GAN)などの深層生成モデルおよびそれらが利用されている応用先を理
 解している.

ACM-67-13 E 0.50 KA01.02
 これらのアプローチの実践的な課題を理解している(例:収束, モード崩壊).

ACM-67-14 E 0.50 KA01.02
 上の影響に対処したりそれを軽減したりするアプローチを理解している.

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-67-15 T2 0.50 KA01.02
 与えられたデータセットおよびタスクに対して, 最も適切な深層学習アプローチの種類を決定する
 ことができる.

ACM-67-16 T2 0.50 KA01.02
 深層学習の道具一式を用いて学習モデル済みデータをデータセットに適用することができる(例:
 PyTorch, Tensorflow).

ACM-67-17 T2 0.50 KA01.02
 深層学習の道具一式を用いて, ネットワークの設定も含めてデータセットに対してモデルを学習さ
 せることができる(例:PyTorch, Tensorflow).

ACM-67-18 E 0.50 KA01.02
 深層学習の道具一式を用いて特定の目標のために生成的アプローチを(1から)適用することがで
 きる.

ACM-67-19 E 0.50 KA01.02
 与えられたシステムのアーキテクチャに対してツールキットが動作するように設定を行うことがで
 きる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-67-20 T1 0.25 KA01.02
 解釈したり説明したりするのが困難または不可能な機械学習モデルを用いることの潜在的で負の
 影響を理解している.

ACM-67-21 T1 0.25 KA01.02
 深層学習の能力が必要以上になってしまうような問題も多数存在することを十分理解している.

ディープフェイク周辺の社会的・政治的懸念について理解し,十分に認識している.

C: モデリングとシミュレーション

1. サンプルング

種別	スキル	優先度	割り当て時間数(h)	DS-BoK KA番号
----	-----	-----	------------	-------------

DS-039 E 3.00 KA01.06
標本誤差とは何かを説明できる.

DS-040 E 4.50 KA01.06
実験計画法の概要を説明できる.

2. データ可視化

種別	スキル	優先度	割り当て時間数(h)	DS-BoK KA番号
----	-----	-----	------------	-------------

DS-041 E 4.50 KA01.06
可視化における目的の広がりについて概略を説明できる(単に現場の作業支援する場合から、ビッグデータ中の要素間の関連性をダイナミックに表示する場合など).

DS-042 T1 1.00 KA01.06
散布図などの軸だしにおいて、縦軸・横軸の候補を適切に洗い出せる.

DS-043 T1 1.00 KA01.06
積み上げ縦棒グラフでの属性の選択など、適切な層化(比較軸)の候補を出せる.

DS-044 T2 3.00 KA01.06
サンプルングやアンサンプル平均によって適量にデータ量を減らすことができる.

DS-045 T2 1.38 KA01.06
適切な情報濃度を判断できる(データインク比など).

DS-046 T2 3.00 KA01.06
不必要な誇張をしないための軸表現の基礎を理解できている(コラムチャートのY軸の基準点は「0」からを原則とし軸を切らないなど).

DS-047 T2 2.50 KA01.06
 強調表現がもたらす効果と、明らかに不適切な強調表現を理解している(計量データに対しては位置やサイズ表現が色表現よりも効果的など)。

DS-048 T2 3.00 KA01.06
 1～3次元の比較において目的(比較, 構成, 分布, 変化など)に応じ, BIツール, スプレッドシートなどを用いて図表化できる。

DS-049 T2 2.00 KA01.06
 端的に図表の変化をアニメーションで可視化できる(人口動態のヒストグラムが経年変化する様子を表現するなど)。

DS-050 T1 1.00 KA01.06
 データ解析部門以外の方に, データの意味を伝えるサインとしての可視化ができる。

DS-051 T2 2.00 KA01.06
 ボロノイ図の概念と活用方法を説明できる。

DS-052 T1 2.00 KA01.06
 1～3次元の図表を拡張した多変量の比較を適切に可視化できる(平行座標, 散布図行列, テーブルレンズ, ヒートマップなど)。

DS-053 E 1.50 KA01.06
 データの性質を理解するために, データを可視化し眺めて考えることの重要性を理解している。

DS-054 E 3.00 KA01.06
 外れ値を見出すための適切な表現手法を選択できる。

DS-055 E 3.00 KA01.06
 データの可視化における基本的な視点を挙げるができる(特異点, 相違性, 傾向性, 関連性を見出すなど)。

3. 分析プロセス

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

DS-056	E	3.50	KA01.06
--------	---	------	---------

スコープ, 検討範囲・内容が明快に設定されていれば, 必要なデータ, 分析手法, 可視化などを適切に選択できる。

4. データの理解・検証

種別	スキル	優先度	割り当て時間数(h)	DS-BoK KA番号
DS-057		E	3.00	KA01.06
ニュース記事などで統計情報に接したときに、数字やグラフの持つメッセージを理解できる。				
DS-058		T1	0.80	KA01.06
単独のグラフに対して、集計ミスなどがないかチェックできる。				
DS-059		T1	2.00	KA01.06
データ項目やデータの量・質について、指示のもと正しく検証し、結果を説明できる。				
DS-060		E	3.00	KA01.06
データが生み出された背景を考え、鵜呑みにはしないことの重要性を理解している。				
DS-061		E	3.00	KA01.06
データから事実を正しく浮き彫りにするために、集計の切り口や比較対象の設定が重要であることを理解している。				
DS-062		E	1.50	KA01.06
普段業務で扱っているデータの発生トリガー・タイミング・頻度などを説明でき、また基本統計量を把握している。				
DS-063		E	3.00	KA01.06
何のために集計しているか、どのような知見を得たいのか、目的に即して集計できる。				
DS-064		E	3.00	KA01.06
分析・図表から直接的な意味合いを抽出できる。(バラツキ, 有意性, 分布傾向, 特異性, 関連性, 変曲点, 関連度の高低など)				
DS-065		E	3.00	KA01.06
想定に影響されず、分布結果の数値を客観的に解釈できる。				

5. 言語処理

種別	スキル	優先度	割り当て時間数(h)	DS-BoK KA番号
DS-071		E	1.50	KA01.06
テキストデータに対する代表的なクリーニング処理(小文字化, 数値置換, 半角変換, 記号除去, ステミングなど)をタスクに応じて適切に実施できる。				

DS-072 E 1.50 KA01.06
 形態素解析や係り受け解析の概念を説明できる.

6. 画像・動画処理

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

DS-073	E	1.50	KA01.06
--------	---	------	---------

画像のデジタル表現の仕組みと代表的な画像フォーマットを知っている.

DS-074	T1	2.00	KA01.06
--------	----	------	---------

画像に対して、目的に応じた適切な色変換や簡単なフィルタ処理などを行うことができる.

DS-075	T1	2.00	KA01.06
--------	----	------	---------

画像データに対する代表的なクリーニング処理(リサイズ, バタニング, 標準化など)をタスクに応じて適切に実施できる.

DS-076	T2	1.50	KA01.06
--------	----	------	---------

動画のデジタル表現の仕組みと代表的な動画フォーマットを理解しており、動画から画像を抽出する既存方法を使うことができる.

D1: コンピュータサイエンス基礎

1. 基本的なコンピュータアーキテクチャ

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-09-01	「電力の壁」の概念を理解している。	T1	0.50	KA02.01
ACM-09-02	ビット, バイト, ワードを理解している。	T1	0.50	KA02.01
ACM-09-03	数値データの表現を理解している。	T1	0.50	KA02.01
ACM-09-04	CPUとGPUを理解している。	T1	1.00	KA02.01
ACM-09-05	非数値データの表現を理解している。	T2	0.50	KA02.01
ACM-09-06	マルチコアとマルチプロセッシングを理解している。	T2	0.50	KA02.01
ACM-09-07	ノイマン型アーキテクチャの基本構成を理解している。	T2	0.50	KA02.01
ACM-09-08	並列アーキテクチャ(例: SIMD, MIMD)を理解している。	T2	0.50	KA02.01
種別	スキル	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-09-09	プロセッサの性能をさらに向上させ, 並列処理の活用を推進する観点から, 「電力の壁」の意味を説明できる。	T1	0.50	KA02.01

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-09-10	T1	1.00	KA02.01
固定長の数値表現の精度への影響を説明できる。			
ACM-09-11	T1	2.00	KA02.01
(GPUの特化された目的と比較対照して)CPUの役割を説明できる。			
ACM-09-12	T2	1.00	KA02.01
文字, 文字列, 画像などの非数値データの内部表現を説明できる。			
ACM-09-13	T2	0.50	KA02.01
マルチコアシステムとマルチプロセッサシステムの違いを説明できる。			
ACM-09-14	T2	1.00	KA02.01
古典的なフォンノイマンアーキテクチャの構成とその主要な機能ユニットを説明できる。			
ACM-09-15	T2	0.50	KA02.01
古典的なフォンノイマンモデルを超えた並列処理の概念を説明できる。			

種別	態度	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-09-16	T1	1.00	KA02.01	
最新のコンピューティングデバイスにおけるデータ表現とプロセッサ速度の利点と制限を理解している。				

2. ストレージシステムの基礎

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-10-01	T1	1.00	KA02.01	
記憶システムとその技術を理解している。				
ACM-10-02	T1	1.00	KA02.01	
レジスタ, キャッシュ, RAMを理解している。				
ACM-10-03	T1	1.00	KA02.01	
仮想記憶を理解している。				

種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-10-04	T1	1.00	KA02.01	

主要な記憶技術(例:SRAM, DRAM, フラッシュ, 磁気ディスク)と, それらの相対的なコストと性能を説明できる.

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

記憶階層の使用により, 実効的な遅延がどのように減少するかを説明できる.

種別 態度		優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-10-06	T1	0.50	KA02.01	

高価で高速な記憶装置と安価で低速な記憶装置のトレードオフを理解している.

3. オペレーティングシステムの基礎

種別 知識		優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-11-01	T1	1.00	KA02.01	

オペレーティングシステム(OS)の役割と目的を理解している.

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

セキュリティの脅威の種類と対策の方法を理解している.

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

ネットワーク機能を持つOS, クライアントサーバーOS, 分散OSを理解している.

ACM-11-04	T2	0.50	KA02.01
-----------	----	------	---------

信頼性と可用性について理解している.

種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-11-05	T1	1.00	KA02.01	

最新のOSの目的と機能を説明できる.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-11-06	T1	1.00	KA02.01
OSに対する潜在的な脅威(例:ソフトウェアの脆弱性, 認証の問題, マルウェアなど)と, それらを防ぐように設計されたセキュリティ機能の種類を挙げることができる.			
ACM-11-07	T2	1.00	KA02.01
ネットワーク機能を持つOS, クライアントサーバーOS, 分散OSについて, また, それらが単一ユーザーのOSとどう異なるかを説明できる.			
ACM-11-08	T2	0.50	KA02.01
信頼性と可用性の重要性について, また, 両者を保証するためのフォールトトレランスの手法を説明できる.			

種別	態度			
通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
ACM-11-09	T1	0.50	KA02.01	
人間とシステムリソース間, およびシステムリソース間のインターフェイスを提供する上でのOSの重要な役割を理解している.				
ACM-11-10	T1	0.50	KA02.01	
OSへのシステム攻撃を避けるために適切に振る舞える.				

4. ファイルシステム

種別	知識			
通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
ACM-12-01	T1	0.50	KA02.01	
ファイル(データ, メタデータ, 操作, 編成)について理解している.				
ACM-12-02	T1	0.50	KA02.01	
ディレクトリ(コンテンツと構造)について理解している.				
ACM-12-03	T1	0.50	KA02.01	
ファイル保護について理解している.				
ACM-12-04	T2	0.50	KA02.01	
ファイル(順編成ファイル, 順編成でないファイル)について理解している.				

種別	スキル			
通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-12-05	T1	1.00	KA02.01
-----------	----	------	---------

ファイル編成のさまざまなアプローチについて、それらの長所と短所を認識しながら比較対照できる。

ACM-12-06	T1	0.50	KA02.01
-----------	----	------	---------

ファイル保護のレベルとそれらを設定するための仕組みを説明できる。

ACM-12-07	T2	0.50	KA02.01
-----------	----	------	---------

順編成ファイル、順編成でないファイルを比較し違いを説明できる。

種別	態度
-----------	-----------

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-12-08	T1	1.00	KA02.01
-----------	----	------	---------

優れたファイル編成の重要性と、不適切なアクセスからファイルを保護することの重要性を理解している。

5. ネットワーク

種別	知識
-----------	-----------

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-13-01	T1	1.00	KA02.01
-----------	----	------	---------

ネットワークの構成要素(ホスト、ルータ、スイッチ、ISP、アクセスポイント、ファイアウォール)を理解している。

ACM-13-02	T1	0.50	KA02.01
-----------	----	------	---------

ローカルエリアネットワーク(LAN):LANのトポロジー(例:バス型, リング型)

ACM-13-03	T1	1.00	KA02.01
-----------	----	------	---------

インターネットの構成(インターネットサービスプロバイダ(ISP), コンテンツプロバイダなど)を理解している。

ACM-13-04	T2	0.50	KA02.01
-----------	----	------	---------

回線交換ネットワークとパケット交換ネットワークを理解している。

ACM-13-05	T2	0.50	KA02.01
-----------	----	------	---------

階層型のネットワーク構成を理解している。

ACM-13-06	T2	1.00	KA02.01
-----------	----	------	---------

名前付けと名前解決の方法(DNS, IPアドレス, URI (Uniform Resource Identifier)など)を理解している。

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-13-07 T2 1.00 KA02.01
 基本的なプロトコル(TCP, IP)を理解している.

ACM-13-08 T2 0.50 KA02.01
 アプリケーション層のプロトコルとしてのHTTP/HTTPSを理解している.

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-13-09 T1 1.00 KA02.01
 コンピュータネットワークの主要な構成要素を挙げるができる.

ACM-13-10 T1 0.50 KA02.01
 LANがさまざまなトポロジーで構成できることを知っている.

ACM-13-11 T1 1.00 KA02.01
 (高レベルの)インターネットの構成を説明できる.

ACM-13-12 T2 0.50 KA02.01
 回線交換とパケット交換の違いを説明できる.

ACM-13-13 T2 1.00 KA02.01
 典型的なネットワークアーキテクチャの階層的な構成を説明できる.

ACM-13-14 T2 0.50 KA02.01
 ネットワークにおける名前とアドレスの類似点, 相違点を挙げるができる.

ACM-13-15 T2 1.00 KA02.01
 TCPやIPなどの基本的なプロトコルがどう機能するかを説明できる.

ACM-13-16 T2 1.00 KA02.01
 HTTPSのようなアプリケーション層のプロトコルがどう機能するかを説明できる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-13-17 T1 1.00 KA02.01
 ネットワークを介して情報を送信することの複雑さ, および送信中に発生する可能性のある問題を軽減するための仕組みを理解している.

6. WebとWebプログラミング

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-14-01	T1	0.50	KA02.01

インターネットとWorld Wide Webの関係を理解している。

ACM-14-02	T1	0.50	KA02.01
-----------	----	------	---------

Webアプリケーションの脆弱性とセキュリティ攻撃(例:SQLインジェクション, 分散型DoS(Denial of Service)攻撃)について理解している。

ACM-14-03	T2	0.50	KA02.01
-----------	----	------	---------

セキュリティ攻撃の検出と対策を理解している。

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-14-04	T1	1.00	KA02.01

インターネットとWorld Wide Webの関係を説明できる。

ACM-14-05	T1	2.00	KA02.01
-----------	----	------	---------

簡単なブラウザベースのアプリケーションの設計と実装が行える。

ACM-14-06	T1	0.50	KA02.01
-----------	----	------	---------

一般的なWebアプリケーションの脆弱性とセキュリティ攻撃を説明できる。

ACM-14-07	T1	1.50	KA02.01
-----------	----	------	---------

Webプログラミング言語(例:HTML5, JavaScript, PHP, CSS)を使用できる。

ACM-14-08	T2	0.50	KA02.01
-----------	----	------	---------

セキュリティ攻撃から保護する方法を認識し適用できる。

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-14-09	T1	0.50	KA02.01

Webアプリケーションの作成と使用をできるかぎり安全に行うための、潜在的なリスクを認識している。

7. コンパイラとインタプリタ

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

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

(他の)プログラムを入力として受け取るプログラム(インタプリタ, コンパイラ, タイプチェッカ, 文書ジェネレータ)を理解している.

ACM-15-02	T1	1.00	KA02.01
-----------	----	------	---------

インタプリタによる解釈実行, ネイティブコードへのコンパイル, ポータブルな中間表現へのコンパイルについて理解している.

ACM-15-03	T1	1.00	KA02.01
-----------	----	------	---------

構文と構文解析, セマンティクスと評価について理解している.

ACM-15-04	T1	0.50	KA02.01
-----------	----	------	---------

インタプリタ型とコンパイル型に分類される言語の例について理解している.

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

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

他のプログラムを処理するプログラムが, 入力データとしてのプログラムをどのように扱うかを説明できる.

ACM-15-06	T1	0.50	KA02.01
-----------	----	------	---------

インタプリタ型言語とコンパイル型言語について, 言語やコードの利点・欠点を説明できる.

ACM-15-07	T1	0.50	KA02.01
-----------	----	------	---------

構文と構文解析とセマンティクスと評価の区別を区別できる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

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

インタプリタ型コードとコンパイル型コードの速度のトレードオフを正しく理解している.

ACM-15-09	T1	0.50	KA02.01
-----------	----	------	---------

ネイティブコードへのコンパイルとポータブルな中間表現へのコンパイルの柔軟性のトレードオフを理解している.

ACM-15-10 T1 0.50 KA02.01
 コード開発におけるインタプリタの有用性を理解している。

8. アルゴリズム思考と問題解決

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-35-01	T1	0.25	KA04.01
アルゴリズムの定義			

ACM-35-02	T1	0.25	KA04.01
問題解決プロセスにおけるアルゴリズムの重要性			

ACM-35-03	T1	0.25	KA04.01
問題を解決するための少なくとも一つの正式な手法			

ACM-35-04	T1	0.25	KA04.01
基本的なオブジェクト指向設計の概念と原理(抽象化, カプセル化と情報隠蔽, 振る舞いと実装の分離)			

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-35-05	T1	0.25	KA04.01
フローチャートや擬似コードといったコード以外の形式的な表記を用いて問題解決策を記述すること。			

ACM-35-06	T1	0.25	KA04.01
データフロー図などの何らかの形式的手段による問題解決策によって, 入力, 変換, 出力といったデータの流れを図示すること。			

ACM-35-07	T1	0.25	KA04.01
問題を解決するプログラムを実装するために不可欠な, データ, ハイパーパラメーター, ユーザーの応答などの入力と出力を識別すること。			

ACM-35-08	T1	0.25	KA04.01
複数の抽象データ型のデータコンポーネントと振る舞いを識別すること。(PDAデータ構造を参照)			

ACM-35-09	T2	0.25	KA04.01
問題を解決するために, 少なくとも1つの形式手法を使用すること。			

種別	態度	優先度	割り当て時間数(h)	DS-BoK KA番号
----	----	-----	------------	-------------

ACM-35-10	T1	0.25	KA04.01
-----------	----	------	---------

アルゴリズムとプログラムの正確に説明する. アルゴリズムはプログラムとは違う.

ACM-35-11	T1	0.25	KA04.01
-----------	----	------	---------

大きな問題を実装可能な解決策に分割し, それらの解決策を何らかの形式で表現するための原理的なやり方があることを正確に理解している.

9. アルゴリズム

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
----	----	-----	------------	-------------

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

数字のリストの平均値の計算や最小値・最大値・最頻値の探索などの単純な数値アルゴリズムを理解している.

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

ソートと探索を理解している.

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

線形探索, 二分探索を理解している.

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

$O(n^2)$ のソートアルゴリズム(挿入ソートなど), $O(n\log n)$ のソートアルゴリズム(マージソート)を理解している.

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

探索・ソートアルゴリズムに関する乱択アルゴリズムを理解している.

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

ハッシュベースの探索・ソートアルゴリズムの効率に関する潜在的な有用性を理解している.

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

グラフの性質: 連結性, 媒介性, 中心性などを理解している.

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

グラフアルゴリズムを理解している.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-38-09 T1 0.50 KA01.02 KA01.03
貪欲法, 分割統治法などの基本的なアルゴリズム戦略を理解している.

ACM-38-10 T1 0.50 KA01.02 KA01.03
線形方程式系を解くアルゴリズムを理解している.

ACM-38-11 T2 0.75 KA01.02 KA01.03
組み合わせ最適化問題のアルゴリズムを理解している.

ACM-38-12 T2 0.75 KA01.02 KA01.03
ヒューリスティックな最適化手法を理解している.

ACM-38-13 E 1.25 KA01.02 KA01.03
ハッシュ化とハッシュ関数を理解している.

ACM-39-01 T1 0.50 KA01.02 KA01.03
時間計算量・空間計算量の定義を理解している.

ACM-39-02 T1 0.50 KA01.02 KA01.03
アルゴリズムの最良の/平均的な/最悪の場合の振る舞いの違いを理解している.

ACM-39-03 T1 0.50 KA01.02 KA01.03
時間計算量と空間計算量のトレードオフを理解している.

ACM-39-04 T1 0.50 KA01.02 KA01.03
アルゴリズムの分類を理解している.

ACM-39-05 T1 0.50 KA01.02 KA01.03
決定的/非決定的アルゴリズム

ACM-39-06 T1 0.50 KA01.02 KA01.03
時間/空間階層

種別 スキル

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-38-14 T1 0.50 KA01.02 KA01.03
単純な数値アルゴリズムを適用できる(例: 平均値を算出する, 最小値を得るなど).

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-38-15	T1	0.50	KA01.02	KA01.03	探索とソートのアルゴリズムを適用できる.
ACM-38-16	T1	0.50	KA01.02	KA01.03	種々の配列ベース検索とソートアルゴリズムのトレードオフについて対照できる.
ACM-38-17	T1	0.50	KA01.02	KA01.03	深さ優先探索・幅優先探索の一般的な枠組みを用いてグラフ・木の走査を実行することができる.
ACM-38-18	T1	0.50	KA01.02	KA01.03	貪欲アルゴリズムなどの効率的なアルゴリズムを用いてグラフ・木における最短経路を同定することができる.
ACM-38-19	T1	0.50	KA01.02	KA01.03	適切な問題に対して線型方程式系ソルバーを適用することができる.
ACM-38-20	T2	0.75	KA01.02	KA01.03	効率的なアルゴリズムを用いてグラフ・木の最大/最小フローを同定することができる.
ACM-38-21	T2	0.75	KA01.02	KA01.03	組み合わせ最適化問題に対して一般的なアルゴリズムを利用することができる(例:分枝限定
ACM-38-22	T2	0.75	KA01.02	KA01.03	ヒューリスティックな最適化手法(粒子群最適化, 遺伝的アルゴリズム, 進化的計算)を適用できる.
ACM-38-23	T2	0.75	KA01.02	KA01.03	適切な問題に対して動的計画法に基づくアルゴリズムを実装できる.
ACM-38-24	E	1.25	KA01.02	KA01.03	分散システム・データに対して探索・ソートアルゴリズムを実装・適用することができる.
ACM-38-25	E	1.25	KA01.02	KA01.03	ハッシュ関数を実装・比較することができる.
ACM-38-26	E	1.25	KA01.02	KA01.03	グラフの走査・最短経路の探索・フローを求めるアルゴリズムを実装できる.

ACM-38-27	E	1.25	KA01.02	KA01.03	乱択アルゴリズムを解析できる。
ACM-39-07	T1	0.50	KA01.02	KA01.03	アルゴリズムの効率について大雑把な比較をすることができる(例: 演算回数)。
ACM-39-08	T1	0.50	KA01.02	KA01.03	様々なサイズの入力に対してアルゴリズムを実行し, その性能を比較することができる。
ACM-39-09	T1	0.50	KA01.02	KA01.03	実装やアルゴリズムの選択が実行時の時間・空間計算量に影響をもたらすことを示すような例を提示することができる。
ACM-39-10	T1	0.50	KA01.02	KA01.03	問題の表現およびデータ構造とアルゴリズムがどのように関連しているか/結びついているかを説明することができる。
ACM-39-11	T2	0.75	KA01.02	KA01.03	アルゴリズムを理解するために分類法を形式的に適用することができる。

種別	態度	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-38-28	T1	0.50	KA01.02	KA01.03	ある問題にうまく対処できるアルゴリズムは往々にして様々な種類のものがあることを意識している。
ACM-38-29	T1	0.50	KA01.02	KA01.03	アルゴリズムの選択が実行効率に重大な影響をもたらすことを認知している。
ACM-38-30	T1	0.50	KA01.02	KA01.03	実行効率がステークホルダー(顧客, 消費者, 保守運用担当者)にもたらす影響を十分に理解している。
ACM-39-12	T1	0.50	KA01.02	KA01.03	時間計算量・空間計算量の間にはトレードオフが存在し得ることを理解し, このトレードオフがソフトウェアの顧客やユーザーに影響することを十分に認識している。

10. データ構造

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-38-28	T1	0.50	KA01.02	KA01.03	ある問題にうまく対処できるアルゴリズムは往々にして様々な種類のものがあることを意識している。
ACM-38-29	T1	0.50	KA01.02	KA01.03	アルゴリズムの選択が実行効率に重大な影響をもたらすことを認知している。
ACM-38-30	T1	0.50	KA01.02	KA01.03	実行効率がステークホルダー(顧客, 消費者, 保守運用担当者)にもたらす影響を十分に理解している。
ACM-39-12	T1	0.50	KA01.02	KA01.03	時間計算量・空間計算量の間にはトレードオフが存在し得ることを理解し, このトレードオフがソフトウェアの顧客やユーザーに影響することを十分に認識している。

ACM-37-01 T1 0.33 KA02.05
 基本的なデータ構造と抽象データ型(リスト, 配列, スタック, キュー, 文字列, セット, レコード/構造体, マップ, ハッシュテーブル)の目的と使い方

ACM-37-02 T1 0.33 KA02.05
 基本的な行列表現構造(疎/密, 行, 列):行列表現型, 表現型に基づく基本的な行列操作の長所/短所

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

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

プログラミングにおける基本的なデータ構造を適切に選択できる.

ACM-37-06 T1 1.00 KA02.05
 与えられたプログラミング言語の標準的なデータ型ライブラリを適切に使うことができる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

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

実装やデータ構造の選択が使用量, 効率(時間と空間), 可読性に影響を与えることを認識している.

11. プログラミング

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

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

基本概念:変数とプリミティブデータ型

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

基本概念:式と代入文

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

基本概念:条件文と繰り返し文

ACM-36-04	T1	0.33	KA02.05
-----------	----	------	---------

基本概念:再帰関数

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-36-05 T1 0.33 KA02.05
 基本概念:関数とパラメタ渡し

ACM-36-06 T1 0.33 KA02.05
 基本概念:ファイルおよび静的データソースを含む単純な入出力

ACM-36-07 T1 0.33 KA02.05
 基本概念:例外処理

ACM-36-08 T1 0.33 KA02.05
 コアプラクティス:ドキュメンテーション

ACM-36-09 T1 0.33 KA02.05
 コアプラクティス:テスト

ACM-36-10 T1 0.33 KA02.05
 コアプラクティス:バージョン管理

ACM-36-11 T1 0.33 KA02.05
 より細かい部品へのプログラムの分割

ACM-36-12 T1 0.33 KA02.05
 エラーの種類(構文, ロジック, 実行時), エラーが起きる原因, エラーへの対処方法

ACM-36-13 T1 0.33 KA02.05
 データソースへの問い合わせおよび解析方法

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-36-18	T1	1.50	KA02.05
-----------	----	------	---------

上述した基本概念とコアプラクティスを含むプログラムの読み, 書き, デバッグができる.

ACM-36-19	T1	1.00	KA02.05
-----------	----	------	---------

コードセグメントの実行を追跡し, 計算の概要を明確にする.

ACM-36-20	T1	1.50	KA02.05
-----------	----	------	---------

分解の技術を適用して, プログラムをより小さなルーチン等に分解する.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-36-21	T1	1.00	KA02.05
適切な手法(データベース問合せ, API呼び出し, 正規表現等)を利用し, 指定されたデータソース(データベース, スプレッドシート, 文書, XML等)のデータを操作する.			
ACM-36-22	T1	1.00	KA02.05
再帰および反復を使用したプログラム開発ができる.			
ACM-36-23	T1	1.00	KA02.05
ソフトウェアの可読性と保守性に貢献する一貫したドキュメントとプログラミングスタイルの標準を利用できる.			
ACM-36-24	T1	1.00	KA02.05
プログラムのテストとデバッグのための手法を適用できる.			
DS-111	T1	4.00	KA02.06
小規模な構造化データ(CSV, RDBなど)を扱うデータ処理(抽出・加工・分析など)を, 設計書に基づき, プログラム実装できる.			
DS-112	E	4.00	KA02.06
JSON, XMLなど標準的なフォーマットのデータを受け渡すために, APIを使用したプログラムを設計・実装できる.			
DS-113	E	8.00	KA02.06
Jupyter NotebookやRstudioなどの対話型の開発環境を用いて, データの分析やレポートの作成ができる.			
DS-114	E	8.00	KA03.02
SQLの構文を一通り知っていて, 記述・実行できる(DML, DDLの理解, 各JOINの使い分け, 集計関数とGROUP BY, CASE文を使用した縦横変換, 副問い合わせやEXISTSの活用など).			

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-36-35	T1	0.33	KA02.05
プログラミングの実践におけるソフトウェア工学の概念と設計原理の重要性を理解している.			
ACM-36-36	T1	0.33	KA02.05
直接教えられたことを超えたことができるようになる. プログラミングの構成と方法は一般的であり, 多くの文脈で有用であることを理解する.			
ACM-36-37	T1	0.33	KA02.05
単純な方法を超える方法を発見する態度を持つ. データサイエンティストは, 既存のソリューションを手直しかつに縛られるべきではない.			

12. 数値計算

種別 知識			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-40-01 乱数発生器	T1	0.33	KA02.05
ACM-40-02 確率分布のシミュレーション	T1	0.33	KA02.05
ACM-40-03 ビットを用いた数値表現の限界と結果の誤差蓄積(オーバーフロー, アンダーフロー, ラウンドオフ, 切り捨て)への影響	T1	0.33	KA02.05
ACM-40-04 数値表現の計算複雑性への影響	T1	0.33	KA02.05
種別 スキル			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-40-07 数値計算アルゴリズムとプロセスがシミュレーションの実行, データサンプリング, データ生成にどのような影響を与えるかを説明できる.	T1	0.33	KA02.05
ACM-40-08 データ解析を実行するための適切な数値計算アルゴリズムを, その限界と数値的な制約を認識した上で説明できる.	T1	0.33	KA02.05
ACM-40-09 非決定性アルゴリズムを用いたデータ解析の再現性を実現するために, 乱数発生器とシミュレートされた確率分布を効果的に使用できる.	T1	0.33	KA02.05
ACM-40-10 乱数発生器と模擬確率分布を使用して, 適切な統計的・数値的制約を満たす非決定性アルゴリズムを用いたデータ分析に再現性を持たせる.	T1	0.33	KA02.05
種別 態度			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-40-12 (擬似)乱数生成の利点と限界を理解している.	T1	0.33	KA02.05
ACM-40-13 数値計算アルゴリズムの限界を理解している.	T1	0.33	KA02.05

D2:ビッグデータ・システム

1. 規模に伴う課題

種別	知識			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-26-01	T1	0.50	KA02.01	サイズ, キャパシティ, タイミングなど, ビッグデータのコンテキストにおける測定の必要性を理解している.
ACM-26-02	T1	0.50	KA02.01	問題のサイズの概念を理解している.
ACM-26-03	T1	0.50	KA02.01	計算処理が急速に増大することの影響を理解している.
ACM-26-04	T1	0.50	KA02.01	データの急速な増大に対する記憶システムへの影響について理解している.
ACM-26-05	T1	0.50	KA02.01	単純さを重視する必要性を理解している.
ACM-26-06	T1	0.50	KA02.01	増加するエージェント/プロセス数に対処するための問題解決のアプローチについて理解している.
ACM-26-07	T1	0.50	KA02.01	スケーラビリティに対応しつつスケールの問題を解決するためのアプローチについて理解している.
種別	スキル			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-26-08	T1	0.50	KA02.01	複雑さの増大につながるビッグデータアプリケーションの理由と, その複雑さの性質を説明できる.
ACM-26-09	T1	0.50	KA02.01	過度の単純さではないが, 単純さを強調することの重要性の理由を説明できる.

ACM-26-10 T1 0.50 KA02.01
 複雑さを軽減するための一般的な手順を説明できる。

ACM-26-11 T1 1.00 KA02.01
 アプリケーションのデータスケールと速度を, その記述をもとに評価できる。

ACM-26-12 T2 0.50 KA02.01
 複数のスケールのレベルで計算タスクを実行できる。

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-26-13	T1	0.50	KA02.01

スケールにより生じる困難に適切に対応する。

ACM-26-14 T1 0.50 KA02.01
 スケールの問題を解決するため適切に振る舞う。

2. ビッグデータを扱うコンピュータアーキテクチャ

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-27-01	E	0.50	KA02.02

効率よく, かつ, 高速な入出力を支援する機構

ACM-27-02 E 1.00 KA02.02
 データ指向の高性能計算に関する概念と要求

ACM-27-03 E 2.00 KA02.02
 キャッシュコヒーレンスを含むメモリおよびキャッシュへの検討と理解

ACM-27-04 E 2.00 KA02.02
 マルチコア, グリッドコンピューティング, GPU, 共有メモリ, 分散メモリ, メモリ共有型計算, ベクター処理等様々な並列計算機の長所と限界の理解

ACM-27-05 E 0.50 KA02.02
 フリンによる並列計算機分類

ACM-27-06	E	1.00	KA02.02
並列処理を支援する実装方式への理解			
ACM-27-07	E	1.00	KA02.02
並列ストレージの階層構造			

種別		スキル	
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-27-08	E	1.00	KA02.02
高速な入出力実現のための様々なアプローチの理解			
ACM-27-09	E	1.00	KA02.02
高速な入出力実現のための課題の(定性的な)説明			
ACM-27-10	E	2.00	KA02.02
様々な並列計算機アーキテクチャの比較, 対照			
ACM-27-11	E	2.00	KA02.02
アプリケーションにとって, いかなる並列計算機アーキテクチャが適切かの(定性的な)説明			
ACM-27-12	E	2.00	KA02.02
計算パターンおよびデータの特徴から得られる計算モデル, フレームワークにとって, 適切なシステムアーキテクチャの選択			

種別		態度	
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-27-13	E	0.50	KA02.02
データサイエンスアプリケーションを支援するハードウェア動向に関する造詣			

3. 並列計算のフレームワーク

種別		知識	
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-28-01	E	0.50	KA02.01
並列計算モデルの目的および定義			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-28-02	E	1.00	KA02.01
モデルの分類			
ACM-28-03	E	2.00	KA02.01
分散システム			
ACM-28-04	E	2.00	KA02.01
グリッドサーチ(パラメタチューニング)			
ACM-28-05	E	2.00	KA02.01
(データ処理における)通信および調整のための繰り返し処理			
ACM-28-06	E	1.50	KA02.01
タスク指向およびデータパラレルによる作業の分割における課題の理解			

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-28-07	E	1.00	KA02.01
並列計算システムのモデル化			
ACM-28-08	E	2.00	KA02.01
効率的かつ効果的な並列計算の評価			
ACM-28-09	E	2.00	KA02.01
大規模データを処理する並列システムの設計と開発の手法			

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-28-10	E	1.00	KA02.01
複雑(になりうる)システムの評価およびデザインへの造詣			

4. 分散データストレージ

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-29-01 T2 0.50 KA02.02
多種多様なデバイスからなるストレージにおける大容量データの格納手法

ACM-29-02 T2 0.50 KA02.02
階層ストレージ

ACM-29-03 T2 1.00 KA02.02
クリーンかつコンシステントかつ典型的なデータへの理解

ACM-29-04 T2 1.00 KA02.02
データの保守, 保護

ACM-29-05 T2 0.50 KA02.02
データの取得, 抽出の課題

ACM-29-06 T2 1.00 KA02.02
ハッシュ, フィルタリング, サンプリングを用いた技術における長所と限界への理解

ACM-29-07 T2 1.00 KA02.02
データのバックアップ

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-29-08 T2 1.00 KA02.02
ビッグデータに関するストレージ階層構造の役割の理解

ACM-29-09 T2 1.00 KA02.02
ビッグデータにおける冗長性の利点の概説

ACM-29-10 T2 1.00 KA02.02
ビッグデータにおける過剰な冗長性がシステムの効率を損なうことの説明

ACM-29-11 T2 1.00 KA02.02
ビッグデータアプリケーションのために, 有用かつ最新の状態を確約するデータの保護と管理に関する手法の説明

ACM-29-12 E 2.00 KA02.02
スケールリングが容易に可能な分散データストレージシステムの開発

ACM-29-13	E	2.00	KA02.02
スケーラビリティ, ユーザビリティ, 効率およびセキュリティを担保したデータ指向システムにおけるバックアップ, マイグレーション, 圧縮などの技術を考慮したストレージシステムの設計			

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-29-14	T2	1.00	KA02.02
ビッグデータアプリケーションを支援するストレージ機構の設計において有益な配置の創生			

5. 並列プログラミング

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-30-01	T2	0.50	KA02.03
並行制御および並列処理および分散システム			

ACM-30-02	T2	0.50	KA02.03
オーバーヘッドによる並列処理の限界			

ACM-30-03	T2	1.00	KA02.03
並行制御および並列処理に関する異なる手法の理解			

ACM-30-04	T2	1.00	KA02.03
負荷平衡等を含むハードウェアアーキテクチャに適合した手法に関する並列アルゴリズムの理解			

ACM-30-05	T2	0.50	KA02.03
MapReduce等の典型的な並列プログラミングパラダイムの理解			

ACM-30-06	T2	0.50	KA02.03
並列/並行アルゴリズムの複雑さの理解			

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-30-07	T2	0.50	KA02.03
スケールに関する並行制御/並列処理の限界について説明			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-30-08	T2	0.50	KA02.03
特定のアルゴリズムにおける並列処理オーバーヘッドの理解			
ACM-30-09	T2	1.00	KA02.03
データ指向並列プログラムの実装方式に関する理解			
ACM-30-10	T2	2.00	KA02.03
データ処理およびデータスケールに対応したデータ指向並列計算システムの開発・運用			
ACM-30-11	T2	2.00	KA02.03
データ指向並列プログラムの開発と最適化			
ACM-30-12	T2	2.00	KA02.03
並列プログラミングパラダイムに基づいたアルゴリズムの設計, 実装と調整			

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-30-13	T2	0.50	KA02.03
特定の事例で過剰に生じる並列処理オーバーヘッドの認識			
ACM-30-14	T2	0.50	KA02.03
事例に併せた並列システムに関する理解とその深化			

6. ビッグデータアプリケーション開発技術

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-31-01	T2	0.50	KA02.05
ビッグデータ処理を支援する技術の必要性			
ACM-31-02	T2	0.50	KA02.05
ハッシュ処理(ハッシング)			
ACM-31-03	T2	0.50	KA02.05
サンプリング, フィルタリング			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-31-04	T2	0.50	KA02.05
データの特性と概要			

ACM-31-05	T2	0.50	KA02.05
ハッシング, サンプルング, フィルタリングの限界			

種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
----	-----	------	-----	------------	-------------

ACM-31-06	T2	0.50	KA02.05
ビッグデータにおけるハッシングの役割の例示			

ACM-31-07	T2	0.50	KA02.05
サンプルングとフィルタリングを用いる際の基準について説明できる			

ACM-31-08	T2	1.00	KA02.05
ビッグデータを利用するアプリケーションにおいて, 指定されたガイドラインを遵守して, 標本選択の実践			

ACM-31-09	T2	0.50	KA02.05
具体的な利用事例に基づいた, 様々なフィルタリング手法について考察			

ACM-31-10	T2	2.00	KA02.05
利用可能な記憶容量と期待される精度に基づき, データ構造を設計し, その性能を解析			

種別	態度	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
----	----	------	-----	------------	-------------

ACM-31-11	T2	0.50	KA02.05
サンプルングおよびフィルタリング実行時のバイアスなどの陥穽への気づき			

7. クラウドコンピューティング

種別	知識	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
----	----	------	-----	------------	-------------

ACM-32-01	T2	0.50	KA02.03
クラウドコンピューティングとその利点に関する理解			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-32-02	T2	0.50	KA02.03
データセンタのアーキテクチャ			
ACM-32-03	T2	0.50	KA02.03
クラウドコンピューティングに関する課題			
ACM-32-04	T2	0.50	KA02.03
クラウドコンピューティングを支援する異なる手法の理解			
ACM-32-05	T2	1.00	KA02.03
分散ファイルシステム			
ACM-32-06	T2	0.50	KA02.03
ビッグデータアプリケーションを支援するクラウドサービス			
ACM-32-07	T2	1.00	KA02.03
仮想化技術			
ACM-32-08	T2	1.00	KA02.03
クラウドコンピューティング, クラウドストレージ, 仮想マシン等のクラウド技術に関するセキュリティ課題			
種別	スキル		
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-32-09	T2	1.00	KA02.03
クラウドシステムで実行するタスクの概要の理解			
ACM-32-10	T2	2.00	KA02.03
データセンターの設計			
ACM-32-11	T2	2.00	KA02.03
ビッグデータアプリケーション支援のために提供されている様々なクラウドサービスの理解			
ACM-32-12	T2	2.00	KA02.03
特定のビッグデータアプリケーションを支援するためのクラウドサービスの選定と適用			
ACM-32-13	T2	2.00	KA02.03
クラウドに関するセキュリティ戦略の設計			

種別	態度			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	

ACM-32-14	T2	1.00	KA02.03	
クラウドサービスの利用に関する責任ある行動姿勢				

ACM-32-15	T2	1.00	KA02.03	
アマゾン, グーグル, マイクロソフト等現在利用可能な異なるクラウドサービスの理解				

8. 複雑さの理論

種別	知識			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	

ACM-33-01	E	1.00	KA02.01	
アルゴリズムの効率性, 計算量に関する課題				

ACM-33-02	E	1.00	KA02.01	
ビッグデータという観点からの重要性, 並行制御/並列処理の観点からの利用として, 計算量に関する理解				

ACM-33-03	E	0.50	KA02.01	
複雑さの概念に関する限界				

ACM-33-04	E	1.00	KA02.01	
並行制御/並列処理も含む一般的に利用されているアルゴリズムにおける複雑さの評価				

種別	スキル			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	

ACM-33-05	E	1.00	KA02.01	
効率性を考慮する際, 数学的解析だけでは不十分な理由の説明				

ACM-33-06	E	2.00	KA02.01	
計算機資源の制限, 時間の制限およびデータサイズ等から課題を洗い出し, 複雑さの観点から与えられた条件内で, その課題が解決可能, あるいは, おおよそ可能かどうかを解析				

ACM-33-07	E	2.00	KA02.01	
ビッグデータ処理に用いられるアルゴリズムの効率性をいかに評価するかについて具体的に例示				

ACM-33-08	E	2.00	KA02.01
ビッグデータを含む特定のアプリケーションに適切なアルゴリズムを選択し、スケーリングの課題 に関し説明			

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-33-09	E	0.50	KA02.01
複雑性に関する造詣と、積極的に取り組む態度を醸成			

ACM-33-10	E	0.50	KA02.01
複雑さによる限界についての認識			

9. ビッグデータアプリケーションのためのソフトウェア支援

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-34-01	T2	1.00	KA02.05
ビッグデータアプリケーションを支援するためのプログラミング環境の必要性とその理解			

ACM-34-02	T2	1.00	KA02.05
自動スケーリングおよびサーバーレスコンピューティングの概念			

ACM-34-03	T2	1.00	KA02.05
ビッグデータという観点からデータの転送, 解析, 機械学習への支援のための, より精緻なウェブ サービスの可能性について考察			

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-34-04	T2	2.00	KA02.05
自動スケーリングおよびサーバーレスコンピューティングの利用についての比較, 対照			

ACM-34-05	T2	2.00	KA02.05
自動スケーリングと負荷バランスの関係性についての理解			

ACM-34-06	T2	1.00	KA02.05
ストリーミングアプリケーションにおけるバッファサイズの重要性について概説			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-34-07	T2	2.00	KA02.05
顔認識およびビデオストリーミングのアプリケーションを容易にするウェブサービスの理解			

種別	態度		
----	----	--	--

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-34-08	T2	0.50	KA02.05
考えられうる偏り, その他の欠如なども考慮したウェブサービス利用に関する省察的(リフレクティブ)アプローチの推奨			

ACM-34-09	T2	0.50	KA02.05
ビッグデータアプリケーションに関する理解の深化			

ACM-34-10	T2	0.50	KA02.05
ビッグデータアプリケーションという観点から, 適切な単純さについて注視を推奨			

E1: セキュリティとプライバシー

1. セキュリティ基礎

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-08-01	T1	0.27	KA03.03
-----------	----	------	---------

個人の間, 組織間, あるいは政府間のプライバシー要求の関係

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

国境を越えたプライバシーとデータセキュリティに関する法律と責任

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

国際的に関与している組織が, 事業を行う管轄区域全体でのプライバシーに関する法, 規制, および標準の差異をどのくらい考慮すべきかに関する理解.

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-08-04	T1	0.27	KA03.03
-----------	----	------	---------

国際的, 国内的, 地域的に存在している司法の構造という文脈で, 組織がサイバー攻撃から情報システムを保護するために, 法と技術がいかに関わるかを説明できること.

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

一般データ保護規則(GDPR)の要求と, 個人データの転送を可能にする米国と英国間などの国家間のプライバシー保護に関する合意を説明できること.

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

米国連邦取引委員会のセクション5, 各州のデータセキュリティ法, 各州のデータ侵害通知法, 医療保険の相互運用性説明法(HIPAA), グラムリーチブライリー法(GLBA), およびUS-CERTによる情報共有, 2015年のサイバーセキュリティ法, その他の法律が, データセキュリティにどのような影響を与えるかを説明できること.

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

セキュリティの3要素(機密性, 可用性, 完全性)について具体的な事例を用いて説明できる.

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

マルウェアなどによる深刻なリスクの種類(消失, 漏洩, サービスの停止など)を常に意識している.

DS-117	E	1.50	KA02.04
--------	---	------	---------

OS, ネットワーク, アプリケーション, データに対するユーザーごとのアクセスレベルを手順に従い設定できる.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

DS-118 T1 2.00 KA02.04
暗号化されていないデータは、不正取得された際に容易に不正利用される恐れがあることを理解し、データの機密度合いに応じてソフトウェアを使用した暗号化と復号ができる。

DS-119 E 0.50 KA02.04
なりすましや偽造された文書でないことを証明するために電子署名と公開鍵認証基盤(PKI: public key infrastructure)が必要であることを理解している。

DS-120 E 2.00 KA02.04
ハッシュ関数を用いて、データの改ざんを検出できる。

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-08-07 T1 0.27 KA03.03
データガバナンスのポリシーとアクションにおいて倫理的に行動する。

ACM-08-08 T1 0.27 KA03.03
セキュリティ、およびプライバシーの障害によって生ずるデータ損失の害について間違いの無い態度をとる。

ACM-08-09 T1 0.27 KA03.03
データの法的小よび社会的責任に関して最高の倫理基準を維持している。

2. プライバシー保護

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-43-01 T1 0.33 KA02.04
ソーシャルエンジニアリングやソーシャルメディアを利用することで暴かれる機微データ

ACM-43-02 T1 0.33 KA02.04
プライバシー権と情報発信による透明性の必要性のトレードオフ

ACM-43-03 T1 0.33 KA02.04
分析ツールを用いて得た情報の開示、送信、共有に関する倫理的責任

ACM-44-01 T1 0.33 KA02.04
どのようなチャンネルでも送信する前にデータを暗号化することの重要性

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-44-02 T1 0.33 KA02.04
統計解析のために暗号化されたデータと暗号化されていないデータを使用する場合の計算時間のトレードオフ

ACM-45-01 T1 0.33 KA02.04
認証, 権限, アクセス制御, データプライバシーを実現するための概念と技術

ACM-45-02 T1 0.33 KA02.04
機密性, 完全性, 可用性を最大化する階層化防御システム

ACM-46-01 T1 0.33 KA02.04
安全でないチャンネルで安全な通信を可能にするセキュリティプロトコルの重要性

ACM-46-02 T1 0.33 KA02.04
安全な伝送路上における個別通信を可能とするプライバシープロトコルの重要性

ACM-46-03 T1 0.33 KA02.04
アプリケーションとサーバー間の秘密通信を保証できるインターネット/通信プロトコル

種別 スキル

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-43-06 T1 0.33 KA02.04
データ入力時に生じるデータの機微に関する理解

ACM-43-07 T1 0.33 KA02.04
情報処理を行う前にデータクリーニングを検討しなくてはならない場合についての認識

ACM-43-08 T1 0.33 KA02.04
データの範囲指定や保護技術など, 生データを処理する際にデータプライバシーを確保するための技術の理解

ACM-44-06 T1 0.33 KA02.04
計算コストを考慮しながら, データ漏洩のリスクを低減するための暗号化ツールやメカニズムを特定できる.

ACM-44-07 T1 1.00 KA02.04
プライバシー要件に影響を与えるデータ暗号化プロセスについて, 個人, 組織, 政府機関などのさまざまな対象者を訓練できる.

ACM-44-08 T1 0.33 KA02.04
メッセージ認証符号, デジタル署名, 認証された暗号化, ハッシュツリーなど, プライバシーを提供するための暗号技術の使用について例示できる.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-44-09 T1 0.33 KA02.04
 単純なテキストデータおよび暗号化されたデータの処理コストのトレードオフについて理解している。

ACM-45-06 T1 0.33 KA02.04
 システムのデータプライバシーがシステムのセキュリティにどのような影響を与えるかについて説明できる。

ACM-45-07 T1 1.00 KA02.04
 データの透明性とデータプライバシーのトレードオフについて議論できる。

種別 態度

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-43-11 T1 0.33 KA02.04
 いかなる組織(実体)に対してもデータの提供はデータプライバシーの侵害につながる可能性があることへの理解

ACM-43-12 T1 0.33 KA02.04
 コンピュータシステムや伝送路を通して不適切なデータの扱いにより、公的にも私的にも社会で起こりうる課題についての理解

ACM-44-11 T1 0.33 KA02.04
 暗号化のために異なる手法が必要なことへの理解

ACM-45-09 T1 0.33 KA02.04
 与えられたコンピュータシステム内の情報を保護している。

3. データセキュリティ

種別 知識

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-47-01 T1 0.33 KA02.04
 定性的指標

ACM-47-02 T1 0.33 KA02.04
 データ資産に対するセキュリティの重要性

ACM-47-03 T1 0.33 KA02.04
 必要とされる様々な種類のセキュリティ目的

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-47-04	T1	0.33	KA02.04
データソースと資産			
ACM-47-05	T1	0.33	KA02.04
データ資産へのアクセス制御とアクセス管理			
ACM-50-01	T1	0.33	KA02.04
データ駆動型アプリケーションにおけるネットワーク上でのデータランザクションに関する洞察			
ACM-50-02	T1	0.33	KA02.04
ネットワークとWebのプロトコル			
ACM-50-03	T1	0.33	KA02.04
通信プロトコルにおける利用可能かつ(または)稼働可能なセキュリティモジュール			
ACM-50-04	T1	0.33	KA02.04
データネットワーク及びWeb上での操作(格納, 検索, リモート計算)			

種別	スキル			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-47-09	T1	0.33	KA02.04	
アプリケーションにおけるデータの流れを理解できる.				
ACM-47-10	T1	1.00	KA02.04	
達成すべき重要なセキュリティ目標を導き出せる.				
ACM-47-11	T1	1.00	KA02.04	
どのデータ資産を保護対象とすべきか, 理由を説明できる.				
ACM-50-05	T1	1.00	KA02.04	
セキュリティを実現するために, 通信プロトコルを解析, 調整できる.				
ACM-50-06	T1	0.33	KA02.04	
ネットワークとWebのプロトコルの特性と動作原理を説明できる.				
ACM-50-07	T1	0.33	KA02.04	
ネットワークおよびWeb上の様々な実体間でデータ通信を行う方法を理解している.				

種別	態度			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-50-08	T1	0.33	KA02.04	ネットワーク/ウェブプロトコルを実行する上でのセキュリティ対策の重要性を理解している。

4. データー一貫性

種別	知識			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-51-01	T1	0.33	KA02.04	論理的な一貫性の概念
ACM-51-02	T1	0.33	KA02.04	データベースシステムにおける一貫性制約の種類
ACM-51-03	T1	0.33	KA02.04	エンティティの一貫性, 参照一貫性, ドメインの一貫性, ユーザー定義の一貫性
ACM-52-01	T1	0.33	KA02.04	物理的な一貫性の概念
ACM-52-02	T1	0.33	KA02.04	RAID, 冗長ハードウェア, 無停電電源装置, エラー訂正メモリ, シビアクラスタなど, データの一貫性を確保するための物理的, ハードウェア的な方法
ACM-53-01	T1	0.33	KA02.04	ヒューマンエラー, ソフトウェアエラー, 伝送エラー, マルウェア, 内部者の脅威, サイバー攻撃, および侵害されたハードウェアを含む一般的なデータ一貫性に対する脅威
ACM-53-02	T1	0.33	KA02.04	データおよび情報の改ざん
ACM-53-03	T1	0.33	KA02.04	データの出所保証
ACM-54-01	T1	0.33	KA02.04	一貫性を保持するハッシュアルゴリズムの役割

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-54-02	T1	0.33	KA02.04
メッセージ認証符号(MACs)とその派生符号の役割			
ACM-54-03	T1	0.33	KA02.04
一貫性を実現するための巡回冗長検査(CRC)とチェックサム			
ACM-54-04	T1	0.33	KA02.04
電子署名の仕組み(RSAおよびECDSA)			
種別 スキル			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-51-04	T1	0.33	KA02.04
論理的ー貫性の概念を説明できる.			
ACM-52-03	T1	0.33	KA02.04
物理的ー貫性の概念を説明できる.			
ACM-52-04	T1	0.33	KA02.04
物理的ー貫性を確保するための物理的およびハードウェアによる方法を説明できる.			
ACM-53-04	T1	0.33	KA02.04
データの完全性に影響を与えるセキュリティ上の脅威の一般的なタイプを列挙できる.			
ACM-53-05	T1	0.33	KA02.04
SHA-1 や MD5 などの異なるハッシュ関数を持つ潜在的な脆弱性を説明できる.			
ACM-54-05	T1	0.33	KA02.04
データー貫性を確立するために、ハッシュアルゴリズムおよびMAC手法をどのように用いるか説明できる.			
ACM-54-06	T1	0.33	KA02.04
一貫性保持の観点から、デジタル署名スキームおよびその必要性について説明できる.			
ACM-54-07	T1	0.33	KA02.04
性能およびセキュリティの観点から、異なる一貫性保持技術の比較や対照ができる.			
ACM-54-08	T1	0.33	KA02.04
複数のデータ所有者が存在するドメインにおいて、データの出所および妥当性を確保するために、一貫性モデルをどのように利用するかを理解している.			

種別	態度			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-51-05	T1	0.33	KA02.04	論理的・一貫性について自信をもって説明できる。
ACM-52-05	T1	0.33	KA02.04	ハードウェア手法による物理的一貫性への対応に自信を持つ。
ACM-53-06	T1	0.33	KA02.04	一般的なセキュリティ脅威について自信をもって説明できる。
ACM-54-09	T1	0.33	KA02.04	様々な方法や技術を用いたデータの整合性への対処に自信を持つ。

5. セキュリティ分析

種別	知識			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-56-01	T1	0.33	KA02.04	記述統計, 要約プロット, 外れ値検出, 点推定, 仮説検定, 検定統計, 線形回帰, 一般化線形回帰を含む, セキュリティデータの探索的データ分析のための統計手法
ACM-56-02	T1	0.33	KA02.04	マルウェア表示技術, 転移学習, 階層アンサンブルニューラルネットワーク(HeNet)等の静的あるいは動的な(セキュリティ)脅威の分類とマルウェア検知のためのコンピュータビジョンを利用した手法
ACM-57-01	T1	0.33	KA02.04	敵対的機械学習, 機械学習に対する攻撃の種類, 攻撃からの機械学習の防御フレームワークに関する基本的概念
ACM-57-02	T1	0.33	KA02.04	高速勾配符号, 反復高速勾配, 普遍的逆説的摂動などの敵対的機械学習技術
ACM-57-03	T1	0.33	KA02.04	機械学習モデルを守るための敵対的トレーニング等のような防御技術

ACM-57-04	T1	0.67	KA02.04	1. セキュリティアプリケーションのための説明能力を持つ機械学習手法. 説明には, サンプル単位で説明する局所的な説明と, データセット全体を考慮した大域的な説明が含まれる. 2. 自然画像を用いたモデルに依存しない説明を, コンピュータビジョンを用いたマルウェア検知システムに適用する方法を知る.
種別 スキル				
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-56-03	T1	1.00	KA02.04	機械学習を活用し, セキュリティアプリケーションのために利用できる.
ACM-56-04	T1	1.00	KA02.04	マルウェア表示技術, 転移学習, 階層アンサンブルニューラルネットワーク(HeNet)等を利用して静的あるいは動的な(セキュリティ)検知機構を持ったマルウェア検知ソリューションを設計できる.
ACM-56-05	T1	0.33	KA02.04	異なるバックグラウンドを持つ人に対して, セキュリティアプリケーションにおいて機械学習モデルが出した結果を説明できる.
ACM-57-05	T1	1.00	KA02.04	検知が行われない盲点があることを認識し, 機械学習のレジリエンス(弾力性, 許容性)を評価できる.
ACM-57-06	T1	1.00	KA02.04	敵対的トレーニングを導入することで, 機械学習のレジリエンスを改良できる.
ACM-57-07	T1	0.33	KA02.04	機械学習アルゴリズムにおける説明可能性を調べ, セキュリティの専門家にその機械学習モデルを選んだ理由を説明できる.
ACM-57-08	T1	1.00	KA02.04	説明可能性と脆弱性の問題に対処するために機械学習の評価基準を定義し, 様々な利害関係者とコミュニケーションできる.
ACM-57-09	T1	0.33	KA02.04	セキュリティおよびプライバシーアプリケーションで用いられる機械学習では, レジリエンスと脆弱性が重要な理由を説明できる.
ACM-57-10	T1	1.00	KA02.04	LIME, LEMNA, TCAVなどの説明能力を持つAI手法をセキュリティアプリケーションに適用できる.
ACM-57-11	T1	1.00	KA02.04	LIME等の説明能力を持つAI手法をセキュリティアプリケーション用に構築された機械学習モデルに適用し, 信頼性の高いスコアに基づいてモデルを選定できる.

種別	態度	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-56-06	T1	0.33	KA02.04
		脅威検知, マルウェアインテリジェンスへのアプローチおよびそれらを認識するために, コンピュータビジョン, 自然言語処理, 古典的データ解析等の様々な視点から理解している.			
		ACM-57-12	T1	0.67	KA02.04
		分類の正確性, 誤検知率, 精度, 再現性だけでなく, 頑健性と脆弱性についても意識して機械学習を評価する.			

E2: ヒューマン・コンピュータ・インタラクション

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-68-01	通し番号	T1	0.33	KA02.05
ヒューマン・コンピュータ・インターフェースを検討するためのコンテキスト: データの可視化, ウェブページ, マルチメディア資料, 教材, ナビゲーションを考慮した一般的なコンピューティング環境				
ACM-68-02		T1	0.33	KA02.05
ユーザーインターフェースの設計や実装に適用できる理論, モデル, 原理, ガイドライン, 標準				
ACM-68-03		T1	0.33	KA02.05
インターフェイスに対する異なる尺度: 有効性と魅力				
ACM-68-04		T1	0.33	KA02.05
カラーとマルチメディア, 人間工学とウェブサービスの利用				
ACM-68-05		T1	0.33	KA02.05
インタラクションに影響を与える認知モデル				
ACM-68-06		T1	0.33	KA02.05
拡張現実(AR)のスコープ, 利点, 欠点				
ACM-68-07		T1	0.33	KA02.05
分析や可視化を通じてデータ理解を支援するためのソフトウェアサポート				
ACM-68-08		T1	0.33	KA02.05
特別なニーズを持つユーザーを含む, さまざまなグループのユーザーに対するアクセシビリティへの配慮				
ACM-69-01		T1	0.33	KA02.05
データサイエンスにおける可視化の役割				
ACM-69-02		T1	0.33	KA02.05
図解を用いた可視化(歴史的な例や現代的な例を含む)				

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-69-03 T1 0.33 KA02.05
効果的な可視化の特徴

ACM-69-04 T1 0.33 KA02.05
異なるデータやユーザーに対する異なる技術の適合性

ACM-69-05 T1 0.33 KA02.05
ダッシュボードとインタラクティブな可視化

ACM-69-06 T1 0.33 KA02.05
可視化を支援するソフトウェア

ACM-69-07 T1 0.33 KA02.05
可視化に基づく推論

ACM-69-08 T1 0.33 KA02.05
可視化の準備 - スケーリング, 色の役割

ACM-69-09 T1 0.33 KA02.05
様々なグラフ: 表, 散布図, 円グラフ, ヒストグラム, グラフ, データマップ(ピクセル, グリフ, グラフ, マップを用いた表現を含む)

種別 スキル

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-68-09 T1 1.00 KA02.05
データの分析や表示を行うために, ユーザー中心アプローチの採用を正当化できる.

ACM-68-10 T1 1.00 KA02.05
注意, 知覚, 認識, 発話, 動きに対する考慮がインターフェイスのユーザビリティにどのように影響するかを説明し, 分析と表示におけるそれらの役割を示すことができる.

ACM-68-11 T1 1.00 KA02.05
公式文書(理論, モデル, ガイドライン等)がデータの分析や可視化に与える影響を説明できる.

ACM-68-12 T1 1.00 KA02.05
障害のあるユーザーや年齢の異なるグループ(子供を含む)がインターフェイスに与える望ましい影響を説明できる.

ACM-68-13 T1 1.00 KA02.05
インターフェイスにおいてバイアスが認識される方法の概要を説明できる.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-68-14	T1	1.00	KA02.05
データの分析および表示を支援するために採用できるソフトウェアの範囲を概説できる.			
ACM-68-15	T1	1.00	KA02.05
ARを活用することでインターフェイスにもたらされる付加価値と可能性について説明できる.			
ACM-69-10	T1	0.50	KA02.05
一般的に使用されている可視化の有名な例を説明できる.			
ACM-69-11	T1	0.50	KA02.05
データサイエンスにおいて可視化が果たす様々な役割を明らかにできる.			
ACM-69-12	T1	1.00	KA02.05
特定の目的のために使用する必要がある一連のデータがある場合, そのための効果的な可視化を実装できる.			
ACM-69-13	T1	0.50	KA02.05
データ分類における可視化の役割を説明し, それを促進する手法を特定できる.			
ACM-69-14	T1	1.00	KA02.05
複数のソフトウェアや可視化手法を用いて, 複数のデータの可視化を実現できる.			
種別	態度		
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-68-16	T1	0.33	KA02.05
ユーザビリティのあらゆる側面に影響を与えるインターフェイスの重要な役割を積極的に認識している.			
ACM-69-15	T1	0.33	KA02.05
見える化の役割の認識を適切に認識している.			

F:ソフトウェア工学

1. ソフトウェアの設計と開発

種別 知識				
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-41-01	T1	3.00	KA02.06	コーディングと設計に関する標準を理解している.
ACM-41-02	T1	1.50	KA02.06	情報管理とデータベースシステムの統合を理解している.
ACM-41-03	T1	1.50	KA02.06	ソフトウェアライフサイクルを理解している.
ACM-41-04	T1	1.50	KA02.06	データ・ライフサイクルを理解している.
ACM-41-05	T2	1.50	KA02.06	プロジェクト・マネジメントの方法論を理解している.
ACM-41-06	E	3.00	KA02.06	組み込み, プロセス制御および通信システムとの統合を理解している.
種別 スキル				
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-41-07	T1	1.50	KA02.06	コーディング標準を説明できる.
ACM-41-08	T1	1.50	KA02.06	プロジェクトにおける設計標準を説明できる.
ACM-41-09	T1	1.50	KA02.06	情報管理/データベースシステムとの統合や相互連携の方法を記述できる.

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-41-10	T1	3.00	KA02.06	複数の分野において異なるテスト方法論およびニーズの範囲と種類を定義し、説明できる。
ACM-41-11	T1	3.00	KA02.06	設計仕様を満たす小規模なソフトウェアプロジェクトを、独力で完成させることができる。
ACM-41-12	T1	6.00	KA02.06	設計仕様を満たすソフトウェアプロジェクトを、チームで完成させることができる。
ACM-41-13	T1	3.00	KA02.06	与えられた設計、文書化、および実装の基準を遵守できる。
ACM-41-14	T1	3.00	KA02.06	単純なプログラムで基本的なソフトウェアのライフサイクルを実行できる。
ACM-41-15	T1	3.00	KA02.06	単純なデータ処理ソフトウェアを用いて基本的なデータ(サイエンス)ライフサイクルを実行できる。
ACM-41-16	T1	1.50	KA02.06	情報管理システムやデータベースシステムとの統合または相互作用が行える。
ACM-41-17	T2	1.50	KA02.06	与えられたプロジェクト管理の方法論を遵守できる。
ACM-41-18	T2	1.50	KA02.06	ステークホルダーの仕様を満たすソフトウェアプロジェクトの計画と設計をチームで実施できる。
ACM-41-19	T2	1.50	KA02.06	チームのリーダーとして、ステークホルダーの要件を満たしつつ、プロジェクトを完成まで導くことができる。
ACM-41-20	T2	1.50	KA02.06	ソフトウェアのライフサイクルの適切な段階でデータに基づく意思決定を行うために、データサイエンスのライフサイクルを実現できる。
ACM-41-21	E	6.00	KA02.06	組込みシステム、プロセス制御システム、通信システムとの統合または相互連携が行える。

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

DS-099	T1	1.00	KA03.02
データベースから何らかのデータ抽出方法を活用し、小規模なExcelのデータセットを作成できる。			
DS-100	E	2.00	KA03.02
オープンデータを収集して活用する分析システムの要件を整理できる。			
IPSJ-13	E	3.00	KA02.06
アジャイル開発を実践できる。 【備考】SWEBOK 2021準拠項目			
IPSJ-14	E	3.00	KA02.06
DevOpsを実践できる。 【備考】SWEBOK 2021準拠項目			

種別	態度	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-41-22	T1	1.50	KA02.06	尊敬, 多様性, 協調の上に築かれたチームの価値を認識している。
ACM-41-23	T1	1.50	KA02.06	プロジェクトのコーディングと設計基準を遵守することの意義を理解している。
ACM-41-24	T1	1.50	KA02.06	協調性と柔軟性があり, 聞き上手で, アイデアを提示する能力と交渉力がある。
ACM-41-25	T1	1.50	KA02.06	ライフサイクルを踏まえたマインドセットを持ち, データやソフトウェアプロジェクトに取り組むことに強いコミットメントを持つ。
ACM-41-26	T1	1.50	KA02.06	テスト駆動型開発の利点を認識している。
ACM-41-27	T2	1.50	KA02.06	尊敬, 傾聴, 責任などの原則に従って, プロジェクトを完成まで導いている。
ACM-41-28	T2	1.50	KA02.06	プロジェクトのコーディングおよび設計基準の遵守を促進し, 奨励している。

2. ソフトウェアテスト

種別 知識		優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-42-01	T1	3.00	KA02.06	様々なテスト(ユニットテスト, 動的テスト, 統合テスト, インターフェーステスト, 回帰テスト, システムテスト, セキュリティテスト)のパラダイムと必要性を理解している.
ACM-42-02	T2	3.00	KA02.06	プログラムにおける潜在的なセキュリティ問題(バッファ等のオーバーフロー, 競合条件, 権限の選択を含む不適切な初期化, 入力データの検査漏れ, 成功しないし正しさに関する思いこみ, 前提条件の検証漏れ)
種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-42-03	T1	3.00	KA02.06	様々な分野の異なるテストパラダイムと必要性の範囲と種類を定義し, 説明できる.
ACM-42-04	T1	3.00	KA02.06	基本的なテスト(ユニットテスト, 動的テスト, 統合テスト)の設計と実施が行える.
ACM-42-05	T2	1.50	KA02.06	大規模データを扱うプログラムを実行する前に, 小規模データでアルゴリズムをテストするために, ビッグデータのデータセットから代表的なデータを抽出できる.
ACM-42-06	T2	2.00	KA02.06	インターフェーステスト, ユーザーテスト, 回帰テスト, システムテスト, セキュリティテストを行うためのテスト仕様書を作成できる.
ACM-42-08	T2	3.00	KA02.06	統計的有意性の検定を用いてプログラムの実行結果を評価できる.
ACM-42-09	T2	1.50	KA02.06	ソフトウェアシステムで考えられるリスクの種類を説明できる.
ACM-42-10	T2	1.50	KA02.06	安全なコーディングと防御的なコーディングの実践について説明できる.
ACM-42-11	E	6.00	KA02.06	複数分野のテストの設計, 開発, 実行が行える.
種別 態度		優先度	割り当て時間数(h)	DS-BoK KA番号

ACM-42-13 T1 1.50 KA02.06
 基本的なソフトウェアおよびデータプロジェクトの推進に対して、テスト駆動(ユニットテスト, 実行テスト, 統合テスト)の観点から取り組むことに強いコミットメントを持つ。

ACM-42-14 T2 1.50 KA02.06
 テスト駆動(特にセキュリティテスト, インターフェーステスト, ユーザーテスト, 回帰テスト, システムテスト)の観点からソフトウェア開発およびデータプロジェクト開発に取り組むことに強いコミットメントを持つ。

ACM-42-15 E 3.00 KA02.06
 テスト駆動型開発における多面的な観点を踏まえて、ソフトウェアおよびデータプロジェクトに全体的に取り組むことに強いコミットメントを持つ。

3. 関連分野との連携

種別	スキル	優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
IPSJ-10		E	4.50	KA02.06
AIデザインパターンを含むソフトウェア開発が行える。 【備考】SWEBOK 2021準拠項目				
IPSJ-11		E	4.50	KA02.06
IoTデザインパターンを含むソフトウェア開発が行える。 【備考】SWEBOK 2021準拠項目				
IPSJ-12		E	3.00	KA02.06
価値創造および価値提案が行える。 【備考】SWEBOK 2021準拠項目				

G1: データ収集, 管理, ガバナンス

1. データ構造

種別	スキル			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
DS-104	T1	0.50	KA03.02	
扱うデータが, 構造化データ(顧客データ, 商品データ, 在庫データなど)か非構造化データ(雑多なテキスト, 音声, 画像, 動画など)なのかを判断できる.				
DS-105	T1	1.00	KA03.02	
ER図を読んでテーブル間のリレーションシップを理解できる.				
DS-106	E	2.00	KA03.02	
正規化手法(第一正規化~第三正規化)を用いてテーブルを正規化できる.				

2. データ収集

種別	知識			
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-01-01	T1	0.27	KA03.03	
データのソース				
ACM-01-02	T1	0.27	KA03.03	
pull-based, push-based アプローチ				
ACM-01-03	T1	0.27	KA03.03	
収集するデータの特徴に基づいた多様なデータの収集				
ACM-01-04	T1	0.27	KA03.03	
データ収集を促進するための技術				
ACM-01-05	T1	0.27	KA03.03	
データの離散化手法				

ACM-01-06 T1 0.27 KA03.03
セキュリティとプライバシーの基準とベストプラクティススキル

種別 スキル

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-01-07 T1 0.27 KA03.03
データを活用するためのデータソースの選択

ACM-01-08 T1 0.27 KA03.03
データソースとデータの活用目的の特徴に基づいたデータ収集のための設計技術

ACM-01-09 T1 0.27 KA03.03
データの離散化, 転送, 保存などを含む, その後の手続き

ACM-01-10 T2 0.27 KA03.03
データ活用に基づいてデータを取得するための加速化および並列化戦略の設計

ACM-02 T2 6.00 KA03.03
文書ファイル, Webページ, マルチメディアデータから情報を抽出できる. データマイニング(ACM-16)の情報抽出に関する知識を参照のこと.

DS-101 E 1.00 KA03.02
対象プラットフォームが提供する機能(SDKやAPIなど)の概要を説明できる.

DS-102 E 1.00 KA03.02
Webクローラー・スクイピングツールを用いてWebサイト上の静的コンテンツを分析用データとして収集できる.

DS-103 T1 2.00 KA03.02
対象プラットフォームに用意された機能(HTTP, FTP, SSHなど)を用い, データを収集先に格納するための機能を実装できる.

種別 態度

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-01-11 T1 0.27 KA03.03
データ収集における精度と効率のトレードオフを評価する能力においてビジネス的な洞察力を示す.

3. データ加工

種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
		DS-078	T1	0.50	KA03.02	KA02.05
		名義尺度の変数をダミー変数に変換できる.				
		DS-079	T1	0.50	KA03.02	KA02.05
		標準化とは何かを知っていて、適切に標準化が行える.				
		DS-080	T1	0.50	KA03.02	KA02.05
		外れ値・異常値・欠損値とは何かそれぞれ知っていて、指示のもと適切に検出と除去・変数などの対応ができる.				
		DS-081	E	0.50	KA03.02	KA02.05
		分析要件や各変数の分布などをふまえて、必要に応じてビン化のカテゴリサイズやビン化を設計・実行できる.				
		DS-082	T1	0.50	KA03.02	KA02.05
		数十万レコードのデータに対して、条件を指定してフィルタリングできる。(特定値に合致する・もしくは合致しないデータの抽出, 特定範囲のデータの抽出, 部分文字列の抽出など)				
		DS-083	E	0.50	KA03.02	KA02.05
		正規表現を活用して条件に合致するデータを抽出できる。(メールアドレスの書式を満たしているか判定するなど)				
		DS-084	T1	0.50	KA03.02	KA02.05
		数十万レコードのデータに対して、レコード間で特定カラムでのソートができ、数千レコードのデータに対して、カラム間でソートできる.				
		DS-085	T1	0.50	KA03.02	KA02.05
		数十万レコードのデータに対して、単一条件による内部結合, 外部結合, 自己結合ができ、UNION処理ができる.				
		DS-086	T1	0.50	KA03.02	KA02.05
		数十万レコードのデータに対して、NULL値や想定外・範囲外のデータを持つレコードを取り除く、または規定値に変換できる.				
		DS-087	T1	0.50	KA03.02	KA02.05
		数十万レコードのデータに対して、規程されたリストと照合して変換する、都道府県からジオコードに変換するなど、ある値を規定の別の値で表現できる.				
		DS-088	T1	0.50	KA03.02	KA02.05
		数十万レコードのデータに対して、ランダムまたは一定間隔にデータを抽出できる.				
		DS-089	T1	0.50	KA03.02	KA02.05
		数十万レコードのデータに対して、合計や最大値, 最小値, レコード数を算出できる.				

DS-090 T1 0.50 KA03.02 KA02.05
 数十万レコードのデータに対する四則演算ができ、数値データを日時データに変換するなど別のデータ型に変換できる。

4. データ共有

種別	スキル	優先度	割り当て時間数(h)	DS-BoK KA番号
DS-091	T1	T1	0.25	KA03.02 KA02.05
加工・分析処理結果をCSV, XML, JSON, Excelなどの指定フォーマット形式に変換してエクスポートできる。				
DS-092	T1	T1	0.50	KA03.02 KA02.05
加工・分析処理結果を、接続先DBのテーブル仕様に合わせてレコード挿入できる。				
DS-093	E	E	1.00	KA03.02 KA02.05
データ取得用のWeb API(REST)やWebサービス(SOAP)などを用いて、必要なデータを取得できる。				
DS-094	T1	T1	0.50	KA03.02 KA02.05
FTPサーバー、ファイル共有サーバーなどから必要なデータファイルをダウンロードして、Excelなどの表計算ソフトに取り込み活用できる。				
DS-095	E	E	1.00	KA03.02 KA02.05
BIツールのレポート編集機能を用いて新規レポートを公開できる。				
DS-096	E	E	1.00	KA03.02 KA02.05
BIツールの自由検索機能を活用し、必要なデータを抽出して、グラフを作成できる。				

5. データ蓄積

種別	知識	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-03-01	T2	T2	1.20	KA03.03
データの表現: 数値, テキスト, イメージ, データの精度				
ACM-03-02	T2	T2	1.20	KA03.03
テキストデータ処理: 単語の集まり, 語数, ITF-IDF, n-gram, 語句解析, 構文解析, 意味解析, ストップワードフィルタリング, ステミング, 基本的なアプリケーション				

ACM-03-03 T2 1.20 KA03.03
 画像処理, データ表現: 整数, 特徴, イメージオペレータ, ビデオオペレータの多次元行列, 物体認識, 高次元特徴抽出

種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-03-04	T2		1.20	KA03.03
様々な型のデータに対する基本的な操作を実行するためのプログラムを書けること. 基本統計量の計算, n-gramの抽出, イメージに対する操作の実行を行うこと.				
DS-107	E		1.00	KA03.02
DWHアプライアンス(Oracle Exadata, IBM Integrated Analytics System, Teradataなど)に接続し, 複数テーブルを複合したデータを抽出できる.				
DS-108	E		2.00	KA03.02
Hadoop・Sparkの分散技術の基本的な仕組みと構成を理解している.				
DS-109	E		1.00	KA03.02
NoSQLデータストア(Hbase, Cassandra, Mongo DB, CouchDB, Redis, Amazon, DynamoDB, Cloudant, Azure Cosmos DBなど)にAPIを介してアクセスし, 新規データを登録できる.				
DS-110	E		1.00	KA03.02
クラウド上のストレージサービス(Amazon S3, Google Cloud Starge, IBM Cloud Object Storageなど)に接続しデータを格納できる.				

種別 態度		優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-03-05	T2		1.20	KA03.03
情報をエンコードするための, データ型の選択が正確である.				

6. システム運用

種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号
DS-097	E		2.00	KA03.01
サーバー1~10台規模のシステム構築, システム運用を指示書があれば実行できる.				
DS-098	E		2.00	KA03.01
数十万レコードを持つデータベースのバックアップ・アーカイブ作成など定常運用ができる.				

7. データ管理

種別	知識	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-04-01	T1	0.27	KA03.03
		政府のデータベース, データウェアハウス, メディエーターベースの情報統合の概念と利用シナリオ			
		ACM-04-02	T1	0.27	KA03.03
		スキーママッピングの概念とアプローチ			
		ACM-04-03	T1	0.27	KA03.03
		データマッピングの概念とアプローチ			
		ACM-04-04	T1	0.27	KA03.03
		データの意味変換の概念とアプローチ			
		ACM-04-05	T1	0.27	KA03.03
		領域を横断するデータを統合するための手法			
		ACM-05-01	T1	0.30	KA03.03
		データ処理における圧縮と削減の役割			
		ACM-05-02	T1	0.30	KA03.03
		さまざまなデータサンプリングのアプローチ			
		ACM-05-03	T1	0.30	KA03.03
		データフィルタリングの技術			
		ACM-05-04	T1	0.30	KA03.03
		データ圧縮技術			
		ACM-06-01	T1	0.25	KA03.03
		データ変換パイプライン			
		ACM-06-02	T1	0.25	KA03.03
		単純な関数変換の方法とその応用			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-06-03	T1	0.25	KA03.03
データの標準化とその応用			
ACM-06-04	T1	0.25	KA03.03
データの正規化とその応用			
ACM-06-05	T1	0.25	KA03.03
データエンコーディングアプローチとその応用			
ACM-06-06	T1	0.25	KA03.03
データのスムーズなアプローチとその応用			
ACM-07-01	T1	0.30	KA03.03
データ品質の側面			
ACM-07-02	T1	0.30	KA03.03
データ品質を改善するためのアプローチ			
ACM-07-03	T1	0.30	KA03.03
名寄せ(entity resolution), 正解データの発見(truth discovery), ルールベースのデータを含むデータクレンジングアルゴリズム			
ACM-07-04	T1	0.30	KA03.03
データ品質ルール of のさまざまな形式: 機能依存(FD: functional dependencies), 条件付き機能依存関係(CFD: conditional functional dependencies), 条件付き包含依存関係(CIND: conditional inclusion dependencies), および一貫性依存関係(MD: matching dependencies)			
種別	スキル		
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
ACM-04-06	T1	0.27	KA03.03
データ統合のスキームを選択すること. すなわち, 従来のデータ統合か, 領域を横断するデータを統合するかを選択すること.			
ACM-04-07	T1	0.27	KA03.03
アプリケーションの機能に応じて, データ統合のアーキテクチャを選択すること.			
ACM-04-08	T1	0.27	KA03.03
スキーママッピング, データマッピング, およびデータの意味変換に適したアルゴリズムを選択または開発すること.			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-04-09	T1	0.27	KA03.03
領域を横断するデータを統合するための適切なアルゴリズムを開発すること.			
ACM-05-05	T1	0.30	KA03.03
データの削減と圧縮が必要か否かを調査できること.			
ACM-05-06	T1	0.30	KA03.03
データのサンプリングとフィルタリングを実行すること.			
ACM-05-07	T2	0.30	KA03.03
データサンプリングの特徴を分析すること.			
ACM-05-08	T2	0.30	KA03.03
計算, 通信, 格納に関する要求に基づいたデータ圧縮技術のを選択すること.			
ACM-05-09	T2	0.30	KA03.03
問い合わせに適したデータ圧縮のアプローチを開発すること.			
ACM-06-07	T1	0.25	KA03.03
データの次元と範囲と, アプリケーションで要求されている次元と範囲とを比較して評価すること.			
ACM-06-08	T1	0.25	KA03.03
データ変換のプロセスを決定すること.			
ACM-06-09	T1	0.25	KA03.03
特定のタスクに適したデータアルゴリズムを選択すること.			
ACM-06-10	T1	0.25	KA03.03
データ変換の有効性を評価すること.			
ACM-07-05	T1	0.30	KA03.03
データ品質を評価すること.			
ACM-07-06	T1	0.30	KA03.03
アプリケーションとデータの意味論に対する要求に従って, データクリーニングのルールを作成すること.			

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-07-07 T1 0.30 KA03.03
データ品質要求に従ってデータクリーニングパイプラインを開発すること.

ACM-07-08 T1 0.30 KA03.03
効率的かつ効果的なデータクリーニングのためのアルゴリズムを開発すること.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
------	-----	------------	-------------

ACM-04-10 T1 0.27 KA03.03
異種のデータソースがもたらす課題について鋭敏である.

ACM-04-11 T1 0.27 KA03.03
データ統合におけるAIの役割について鋭敏である.

ACM-05-10 T1 0.30 KA03.03
データ計算の有効性と効率性のトレードオフを詳細に評価することに注意を払っている.

ACM-06-11 T1 0.25 KA03.03
データを使用するためのデータ変換の重要性に対して鋭敏である.

ACM-06-12 T1 0.25 KA03.03
データ変換とデータ品質の関係に鋭敏である.

ACM-07-09 T1 0.30 KA03.03
データの品質問題による害に対して鋭敏である.

ACM-07-10 T1 0.30 KA03.03
データ使用におけるデータクリーニングの役割を取り扱うための強い責任感を持っている.

G2: プロフェッショナリズム

1. 継続研鑽

種別 知識		優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
ACM-73-01		T1	0.38	KA04.01 KA05.01
能力(コンピテンシー)の意味と能力を示すことができること.				
ACM-73-02		T1	0.38	KA04.01 KA05.01
専門知識の習得/習熟またはコンピテンシーを伸ばす; ジャーナル, カンファレンス, コース, ウェビナーの役割				
ACM-73-03		T1	0.38	KA04.01 KA05.01
技術の変更とそのコンピテンシーへの影響				
ACM-73-04		T1	0.38	KA04.01 KA05.01
CPDと専門的活動における職能団体(職業別の団体)の役割				
種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
ACM-73-05		T1	0.38	KA04.01 KA05.01
プロのデータサイエンティストにとってのコンピテンシー維持の重要性を説明すること.				
ACM-73-06		T1	0.38	KA04.01 KA05.01
プロが一般的にどのようにしてコンピテンシーを伸ばし, 熟達するか説明し, 後方の利点を説明すること.				
ACM-73-07		T1	0.38	KA04.01 KA05.01
キャリア開発における職能団体の役割の重要性を主張すること.				
種別 態度		優先度	割り当て時間数(h)	DS-BoK KA番号
通し番号				
ACM-73-08		T1	0.38	KA04.01 KA05.01
データサイエンスの分野は急速に変化する分野である. 最新情報を入手する方法を知ることが必須であると認識することに積極的, かつ情熱的に取り組んでいる.				

2. コミュニケーション

種別 知識		優先度	割り当て時間数(h)	DS-BoK KA番号	
通し番号					
ACM-74-01	T1	0.30	KA04.01	KA05.01	様々なコミュニケーション形式—電子的な書面や話し方—その有効な使い方
ACM-74-02	T1	0.30	KA04.01	KA05.01	データサイエンスに関連する技術文献
ACM-74-03	T1	0.30	KA04.01	KA05.01	データサイエンティストが関連するコミュニケーションに関連する対象者(オーディエンス)—小規模グループ, 大規模グループ, 専門家と非専門家, 若者, 上級管理者, 機械—そしてそれぞれとの有効なコミュニケーションの要素

種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号	
通し番号					
ACM-74-04	T1	0.30	KA04.01	KA05.01	データサイエンスに関連する技術専門文献の様相を評価すること.
ACM-74-05	T1	0.30	KA04.01	KA05.01	同僚の技術開発をガイドするための文献を作成すること.
ACM-74-06	T1	0.30	KA04.01	KA05.01	データサイエンスに興味のある様々なオーディエンスを対象にしたプレゼンテーションを作成すること.
ACM-74-07	T1	0.30	KA04.01	KA05.01	必要に応じて変革管理にまつわる一般的な課題を取り入れながら, データサイエンスの調査から生じる重要な取り組みを説明するために, 上級管理者向けの状況報告を設計すること.

種別 態度		優先度	割り当て時間数(h)	DS-BoK KA番号	
通し番号					
ACM-74-08	T1	0.30	KA04.01	KA05.01	関係する技術の最新情報に対応して調整し, それを効果的に活用する方法を理解し, 新しい開発の機会を見逃さない.
ACM-74-09	T1	0.30	KA04.01	KA05.01	積極的かつ自発的に, 新しい学びと新しい体験の重要性を判断する.
ACM-74-10	T1	0.30	KA04.01	KA05.01	知識に関して, その強みと弱みの理解が正確であり, 重視している.

3. チームワーク

種別 知識		優先度	割り当て時間数(h)	DS-BoK KA番号	
通し番号					
ACM-75-01		T1	0.33	KA04.01	KA05.01
チームメンバー選び.メンバーの能力やスキルを補う必要性					
ACM-75-02		T1	0.33	KA04.01	KA05.01
チームダイナミックス, チームの規律					
ACM-75-03		T1	0.33	KA04.01	KA05.01
有効なチーム運用の要素					

種別 スキル		優先度	割り当て時間数(h)	DS-BoK KA番号	
通し番号					
ACM-75-04		T1	0.33	KA04.01	KA05.01
チームメンバーの間の対立に対応するステップを説明すること.					
ACM-75-05		T1	0.33	KA04.01	KA05.01
特定のデータサイエンス調査を行うチームを選ぶ上で考慮すべき点を要約すること.					
ACM-75-06		T1	0.33	KA04.01	KA05.01
データサイエンス研究の調査を行うチームのリーダーに求められる資質を認識すること.					

種別 態度		優先度	割り当て時間数(h)	DS-BoK KA番号	
通し番号					
ACM-75-07		T1	0.33	KA04.01	KA05.01
敬意を払って協力し, チームの編成と運用に関して敏感, かつ適切に行動する.					
ACM-75-08		T1	0.33	KA04.01	KA05.01
他の人と活動するときは, 重要でない相違点は脇に置いて, 協力し, 適切に行動する.					

4. 経済的課題

種別 知識		優先度	割り当て時間数(h)	DS-BoK KA番号	
通し番号					

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-76-01 T2 0.46 KA04.01 KA05.01
質の高いデータセットとその維持のコストと価値

ACM-76-02 T2 0.46 KA04.01 KA05.01
データサイエンス活動のコストの正当化

ACM-76-03 T2 0.46 KA04.01 KA05.01
プロジェクト費用の見積もり

ACM-76-04 T2 0.46 KA04.01 KA05.01
データサイエンスの促進

ACM-76-05 T2 0.46 KA04.01 KA05.01
データサイエンス活動から生じる自動化(オートメーション)

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
------	-----	------------	--------	------

ACM-76-06 T2 0.46 KA04.01 KA05.01
維持の必要性も考慮しながら、組織にとってのデータセットの価値を評価すること。

ACM-76-07 T2 0.46 KA04.01 KA05.01
組織が定期的に収集すべきデータの事例を議論し、品質を考慮しながら、データが持つべき属性を特定し、取り扱うデータ集合の形式を決めるデータ収集プロセスを設計すること。

ACM-76-08 T2 0.46 KA04.01 KA05.01
特定の目的のために質の高いデータを収集する(いわゆる資源としての)コストを評価すること。

ACM-76-09 T2 0.46 KA04.01 KA05.01
組織の中にデータサイエンス活動を置くことを正当化し、コストを測ること。

ACM-76-10 T2 0.46 KA04.01 KA05.01
組織にとっての、ある調査・研究プロジェクトに取り組む価値を推論すること。

ACM-76-11 T2 0.46 KA04.01 KA05.01
組織内で調査を行うために必要なリソースを監視し、その活動を外部委託する場合と比較すること。

ACM-76-12 T2 0.46 KA04.01 KA05.01
ある活動の自動化にまつわるコストを評価すること。

種別	態度	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-76-13	T2	0.46	KA04.01 KA05.01
データサイエンス活動にまつわるコストを尊重し, 適切に行動する.					

5. プライバシーと守秘義務

種別	知識	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-77-01	T1	0.25	KA04.01 KA05.01
情報の自由					

		ACM-77-02	T1	0.25	KA04.01 KA05.01
EU一般データ保護規則(GDPR)など, データ保護規則 [5]参照					

		ACM-77-03	T1	0.25	KA04.01 KA05.01
プライバシーに関する法律					

		ACM-77-04	T1	0.25	KA04.01 KA05.01
データの守秘義務を守る方法					

		ACM-77-05	T1	0.25	KA04.01 KA05.01
プライバシーや守秘義務に対するリスク					

		ACM-77-06	T1	0.25	KA04.01 KA05.01
国際的な側面					

種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		ACM-77-07	T1	0.25	KA04.01 KA05.01
データの守秘義務を守るための技術的メカニズムを説明すること.					

		ACM-77-08	T1	0.25	KA04.01 KA05.01
各国のプライバシーに関する法律を比較し, その違いから生まれる課題を説明すること.					

ACM-77-09 T1 0.25 KA04.01 KA05.01
 動画, 声, 顔認識ソフトの使用から生じるプライバシーや守秘義務の課題を認識すること.

ACM-77-10 T1 0.25 KA04.01 KA05.01
 国際基準を考慮しながら, 各プライバシー関連法律をどのような文脈で適応すべきか意識すること.

種別 態度

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-77-11 T1 0.25 KA04.01 KA05.01

データサイエンス活動が信用されるよう, プライバシーや守秘義務の側面を取り入れ, 維持することに敏感に対応する.

6. 倫理的課題

種別 知識

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

ACM-78-01 T1 0.25 KA04.01 KA05.01

能力(コンピテンシー)や能力の維持における倫理的な課題

ACM-78-02 T1 0.25 KA04.01 KA05.01

データおよびその使用にまつわる守秘義務問題

ACM-78-03 T1 0.25 KA04.01 KA05.01

EU一般データ保護規則(GDPR) —[5]参照

ACM-78-04 T1 0.25 KA04.01 KA05.01

ある状況を正しく代表するデータおよびデータサンプルの必要性

ACM-78-05 T1 0.25 KA04.01 KA05.01

アルゴリズムやデータにおけるバイアスと, ありうるバイアスの種類の意識. バイアスを確認し, 避けるための仕組み

ACM-78-06 T1 0.25 KA04.01 KA05.01

アルゴリズムの透明性と説明責任

種別 スキル

通し番号 優先度 割り当て時間数(h) DS-BoK KA番号

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-78-07 T1 0.25 KA04.01 KA05.01
データサイエンティストが自分の能力の範囲を超えかねないいくつかの状況を表示し、それらの状況を緩和するための対策を特性すること。

ACM-78-08 T1 0.25 KA04.01 KA05.01
データセットやアルゴリズムにバイアスが含まれていないことを確認するためのテクニックを表示すること。

ACM-78-09 T1 0.25 KA04.01 KA05.01
データサイエンスのプロフェッショナルネットワークに入会するメリットを熟考すること。

DS-121 T1 0.50 KA03.01
データを取り扱う人間として相応しい倫理を身に着けている。(データのねつ造, 改ざん, 盗用を行わないなど)

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
------	-----	------------	--------	------

ACM-78-10	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

データ収集・使用にまつわる深い倫理的課題に対応する。

ACM-78-11	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

偏見, 先入観の問題に対処し, それらを解消するために積極的に取り組む。

ACM-78-12	T1	0.25	KA04.01	KA05.01
-----------	----	------	---------	---------

データサイエンスの進歩に対して, 自主的かつ自発的である。

7. 法的課題

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK	KA番号
------	-----	------------	--------	------

ACM-79-01	T2	0.60	KA04.01	KA05.01
-----------	----	------	---------	---------

データサイエンスに関連したコンピューター犯罪

ACM-79-02	T2	0.60	KA04.01	KA05.01
-----------	----	------	---------	---------

サイバーセキュリティ

ACM-79-03	T2	0.60	KA04.01	KA05.01
-----------	----	------	---------	---------

犯罪防止

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-79-04	T2	0.60	KA04.01	KA05.01
犯罪行為を検出するメカニズム.多様なアプローチの重要性				
ACM-79-05	T2	0.60	KA04.01	KA05.01
リカバリメカニズム.100%の稼働率の維持				
ACM-79-06	T2	0.60	KA04.01	KA05.01
対コンピュータ犯罪の法律				

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-79-07	T2	0.60	KA04.01	KA05.01
犯罪行為の記載フォームを検出するための様々なメカニズムを評価すること.				
ACM-79-08	T2	0.60	KA04.01	KA05.01
脅威に対抗するために, 複数の多様なアプローチの望ましさを正当化できること.				

DS-122 T1 1.00 KA03.01
 個人情報に関する法令(個人情報保護法, EU一般データ保護規則:GDPRなど)や, 匿名加工情報の概要を理解し, 守るべきポイントを説明できる.

DS-145 E 0.50 KA05.01
 請負契約と準委任契約の違いを説明できる.

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-79-09	T2	0.60	KA04.01	KA05.01
犯罪と思われる行為に出会ったときに, 責任をもった倫理的な態度を取りながら親切で思いやりのある姿勢を示している.				

8. 知的財産

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-80-01	E	0.27	KA04.01	KA05.01
特許, 著作権, 商標, 営業秘密, 著作者人格権および商標				

情報処理学会 データサイエンス・カリキュラム標準(専門教育レベル)

ACM-80-02	E	0.27	KA04.01	KA05.01	データサイエンス関連知的財産で守られるものと守られないもの, およびその保護の種類
ACM-80-03	E	0.27	KA04.01	KA05.01	法的保護が可能なデータサイエンス関連知的財産の種類と法的保護が不可能な種類, およびその保護の種類
ACM-80-04	E	0.27	KA04.01	KA05.01	知的財産関連規則, 知的財産の保有, 国際協定(例 営業秘密の保護に関する欧州指令)の影響を含めた知的財産権の属地主義, および知的財産の期間限定
ACM-80-05	E	0.27	KA04.01	KA05.01	自動的に与えられる知的財産権の種類と登録必要な種類と, 保護された知的財産権を獲得する過程の概要
ACM-80-06	E	0.27	KA04.01	KA05.01	他人の権利を侵害する可能性.守られている知的財産の正しい使用法
種別 スキル					
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号		
ACM-80-07	E	0.27	KA04.01	KA05.01	データサイエンティストに関連する知的財産の種類を説明すること.
ACM-80-08	E	0.27	KA04.01	KA05.01	特許, 著作権, デザイン, 商標の違いを示し, それらのデータサイエンスにおける使用を表すこと.
ACM-80-09	E	0.27	KA04.01	KA05.01	データサイエンスの観点から見た営業秘密の役割を説明すること.
ACM-80-10	E	0.27	KA04.01	KA05.01	知的財産権を登録する過程を説明すること.
ACM-80-11	E	0.27	KA04.01	KA05.01	知的財産権保有と著作者人格権にまつわる課題を説明すること.
ACM-80-12	E	0.27	KA04.01	KA05.01	守られている知的財産を使用する上でのリスクと, そのリスクの正当な乗り越え方を評価すること.
種別 態度					
通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号		

ACM-80-13	E	0.27	KA04.01	KA05.01
知的財産の存在と重要性, および知的財産によってもたらされる責任と機会に対して鋭敏で敏感に対応する.				

9. 自動化

種別 知識

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-81-01	E	0.42	KA04.01	KA05.01
自動化(オートメーション), そのメリットと理由				
ACM-81-02	E	0.42	KA04.01	KA05.01
クリティカルな状況における自動化特有の懸念				
ACM-81-03	E	0.42	KA04.01	KA05.01
アルゴリズムにおける透明性と説明責任				

種別 スキル

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-81-04	E	0.42	KA04.01	KA05.01
非技術者のオーディエンスを対象に, ある特定の場面で意思決定がどの程度自動化されているか説明すること.				
ACM-81-05	E	0.42	KA04.01	KA05.01
設計に対する要求事項が, コンピュータによる意思決定に与える影響を分析できること.				
ACM-81-06	E	0.42	KA04.01	KA05.01
様々な場面における自動化のメリットを説明すること.				
ACM-81-07	E	0.42	KA04.01	KA05.01
意思決定システムの監査が可能であることを確認するために必要なステップを特定すること.				

種別 態度

通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号	
ACM-81-08	E	0.42	KA04.01	KA05.01
自動化と, 雇用への影響にまつわる課題に対して鋭敏で敏感に対応する.				

ACM-81-09

E

0.42

KA04.01

KA05.01

自動化の問題に対して敬意に満ちた倫理的なアプローチをする.

G3: ビジネス基礎

1. 論理的思考

種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		DS-135	T1	0.67	KA04.01
		データや事象の重複に気づくことができる。			
		DS-136	T1	0.67	KA04.01
		通常見受けられる現象の場合において、分析結果の意味合いを正しく言語化できる。			
		DS-137	E	1.00	KA04.01
		一般的な論文構成について理解している。(序論→アプローチ→検討結果→考察や、序論→本論→結論 など)			
		DS-138	E	2.00	KA04.01
		1つの図表～枚数程度のドキュメントを理論立ててまとめることができる。(課題背景、アプローチ、検討結果、意味合い、ネクストステップ)			
		DS-139	T1	1.00	KA04.01
		報告に対する論拠不足や論理破壊を指摘された際に、相手の主張をすみやかに理解できる。			

2. プロジェクト管理

種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		DS-140	E	0.67	KA04.02
		ウォーターフォール開発とアジャイル開発の違いを説明できる。			
		DS-141	T1	0.67	KA04.02
		指示に従ってスケジュールを守り、チームリーダーに頼まれた自分の仕事を完遂できる。			
		DS-142	T1	0.67	KA04.02
		担当するタスクの遅延や障害などを発見した場合、迅速かつ適切に報告ができる。			

3. ビジネス実装

種別	スキル	通し番号	優先度	割り当て時間数(h)	DS-BoK KA番号
		DS-067	E	2.00	KA05.02
		一般的な収益方程式に加え, 自らが担当する業務の主要な変数(KPI)を理解している.			
		DS-068	E	2.00	KA05.02
		担当する事業領域において, 市場規模, 主要なプレイヤー, 支配的なビジネスモデル, 課題と機会について説明できる.			
		DS-069	T1	2.00	KA05.02
		主に担当する事業領域であれば, 取り扱う課題領域に対して基本的な課題の枠組みが理解できる.(調達活動の5フォースでの整理, CRM課題のRFMでの整理など)			
		DS-070	T1	1.00	KA05.02
		仮説や既知の問題が与えられた中で, 必要なデータにあたりをつけ, アクセスを確保できる.			
		DS-131	T1	0.67	KA05.01
		ビジネスにおける論理とデータの重要性を認識し, 分析的でデータドリブンな考え方にに基づき行動できる.			
		DS-132	T1	0.67	KA05.01
		「目的やゴールの設定がないままデータを分析しても, 意味合いが出ない」ことを理解している.			
		DS-133	T1	0.67	KA05.01
		課題や仮説を言語化することの重要性を理解している.			
		DS-134	T1	0.67	KA05.01
		現場に出向いてヒアリングするなど, 一次情報に接することの重要性を理解している.			
		DS-143	T1	0.67	KA05.01
		ビジネス観点で仮説を持ってデータをみることの重要性和, 仮に仮説と異なる結果となった場合にも, それが重要な知見である可能性を理解している.			
		DS-144	E	0.67	KA05.01
		分析結果を元に, 起きている事象の背景や意味合い(真実)を見ぬくことができる.			
		DS-146	T1	0.50	KA05.01
		結果, 改善の度合いをモニタリングする重要性を理解している.			

通し番号	大分類	中分類	小分類
ACM-01	Data Acquirement, Management and Governance (DG)	Data Acquisition	
ACM-02		Information Extraction	
ACM-03		Working with Various Types of Data	
ACM-04		Data Integration	
ACM-05		Data Reduction and Compression	
ACM-06		Data Transformation	
ACM-07		Data Cleaning	
ACM-08		Data Privacy and Security	
ACM-09	Computing and Computer Fundamentals (CCF)	Basic Computer Architecture	
ACM-10		Storage System Fundamentals	
ACM-11		Operating System Basics	
ACM-12		File Systems	
ACM-13		Networks	
ACM-14		The Web and Web Programming	
ACM-15		Compilers and Interpreters	
ACM-16	Data Mining (DM)	Proximity Measurement	
ACM-17		Data Preparation	
ACM-18		Information Extraction	
ACM-19		Cluster Analysis	
ACM-20		Classification and Regression	
ACM-21		Pattern Mining	
ACM-22		Outlier Detection	
ACM-23		Time Series Data	
ACM-24		Mining Web Data	
ACM-25		Information Retrieval	
ACM-26	Big Data Systems (BDS)	Problems of Scale	
ACM-27		Big Data Computing Architectures	
ACM-28		Parallel Computing Frameworks	
ACM-29		Distributed Data Storage	
ACM-30		Parallel Programming	
ACM-31		Techniques used in Big Data Applications	
ACM-32		Cloud Computing	
ACM-33		Complexity Theory	
ACM-34		Software Support for Big Data Applications	
ACM-35	Programming, Data Structures, and Algorithms (PDA)	Algorithmic Thinking & Problem Solving	
ACM-36		Programming	
ACM-37		Data Structures	
ACM-38		Algorithms	
ACM-39		Basic Complexity Analysis	
ACM-40		Numerical Computing	
ACM-41	Software Development and Maintenance (SDM)	Software Design and Development	
ACM-42		Software Testing	
ACM-43	Data Privacy, Security, Integrity, and Analysis for Security (DPSIA)	Data Privacy	Social Responsibility
ACM-44			Cryptography
ACM-45			Information Systems
ACM-46			Communication Protocols
ACM-47		Data Security	Data quality and handling for security
ACM-48			Classification of cryptographic tools
ACM-49			Security and performance trade-off
ACM-50			Network and web protocols
ACM-51		Data Integrity	Logical integrity
ACM-52			Physical integrity
ACM-53			Security threats affecting data integrity
ACM-54			Methods to ensure data integrity

通し番号	大分類	中分類	小分類
ACM-55		Analysis for Security	Data corruption and data validation
ACM-56			Machine learning (ML) algorithms and statistical methods for security
ACM-57			Machine learning (ML) robustness and explainability
ACM-58			Categories of security applications
ACM-59	Artificial Intelligence (AI)	General	
ACM-60		Knowledge Representation and Reasoning (Logic-based Models)	
ACM-61		Knowledge Representation and Reasoning (Probability-based Models)	
ACM-62		Planning and Search Strategies	
ACM-63	Machine Learning (ML)	General	
ACM-64		Supervised Learning	
ACM-65		Unsupervised Learning	
ACM-66		Applications that Require Mixed Methods	
ACM-67		Deep Learning	
ACM-68	Analysis and Presentation (AP)	Fundamental considerations	
ACM-69		Visualization	
ACM-70		User-centered design	
ACM-71		Interaction design	
ACM-72		Interface design and development	
ACM-73	Professionalism (PR)	Continuing Professional Development	
ACM-74		Communication	
ACM-75		Teamwork	
ACM-76		Economic Considerations	
ACM-77		Privacy and Confidentiality	
ACM-78		Ethical Considerations	
ACM-79		Legal Considerations	
ACM-80		Intellectual Property	
ACM-81		On Automation	

DSスキルチェックリスト★レベル

通し番号	大分類	中分類	小分類	種別
DS-001	基礎数学	統計数理基礎	順列や組合せを式 nPr , nCr を用いて計算できる.	必須
DS-002	基礎数学	統計数理基礎	条件付き確率の意味を説明できる.	必須
DS-003	基礎数学	統計数理基礎	平均(相加平均), 中央値, 最頻値の算出方法の違いを説明できる.	必須
DS-004	基礎数学	統計数理基礎	与えられたデータにおける分散と標準偏差が計算できる.	必須
DS-005	基礎数学	統計数理基礎	母(集団)平均と標本平均, 不偏分散と標本分散がそれぞれ異なることを説明できる.	必須
DS-006	基礎数学	統計数理基礎	標準正規分布の分散と平均の値を知っている.	必須
DS-007	基礎数学	統計数理基礎	相関関数と因果関係の違いを説明できる.	必須
DS-008	基礎数学	統計数理基礎	名義尺度, 順序尺度, 間隔尺度, 比例尺度の違いを説明できる.	必須
DS-009	基礎数学	統計数理基礎	一般的な相関係数(ピアソン)の分母と分子を説明できる.	必須
DS-010	基礎数学	統計数理基礎	5つ以上の代表的な確率分布を説明できる.	必須
DS-011	基礎数学	統計数理基礎	二項分布の事象もサンプル数が増えていくとどのような分布に近似されるかを知っている.	必須
DS-012	基礎数学	統計数理基礎	変数が量的, 質的どちらの場合でも関係の強さを算出できる.	必須
DS-013	基礎数学	統計数理基礎	ベイズの定理を説明できる.	必須
DS-014	基礎数学	線形代数基礎	ベクトルの内積に関する計算方法を理解し線形式をベクトルの内積で表現できる.	
DS-015	基礎数学	線形代数基礎	行列同士, および行列とベクトルの計算方法を正しく理解し, 複数の線形式を行列の積で表現できる.	
DS-016	基礎数学	線形代数基礎	逆行列の定義, および逆行列を求めることにより行列表記された連立方程式を解くことができることを理解している.	
DS-017	基礎数学	線形代数基礎	固有ベクトルおよび固有値の意味を理解している	
DS-018	基礎数学	微分・積分基礎	微分により計算する導関数が傾きを求めるための式であることを理解している	
DS-019	基礎数学	微分・積分基礎	2変数以上の関数における偏微分の計算方法を理解している	
DS-020	基礎数学	微分・積分基礎	積分と面積の関係を理解し, 確率密度関数を定積分することで確率が得られることを説明できる	
DS-021	予測	回帰/分類	単回帰分析について最小二乗法, 回帰計数, 標準誤差の説明ができる.	必須
DS-022	予測	回帰/分類	重回帰分析において偏回帰係数と標準偏回帰係数, 重相関関係について説明できる.	
DS-023	予測	評価	ROC曲線, AUC(Area under the curve), を用いてモデルの精度を評価できる.	
DS-024	予測	評価	混同行列(正誤分布のクロス表), Accuracy, Precision, Recall, F値をいった評価尺度を理解し, 精度を評価できる.	
DS-025	予測	評価	MSE(Mean Square Error), MAE (Mean Absolute Error) といった評価尺度を理解し, 精度を評価できる.	必須
DS-026	予測	評価	ホールドアウト法, 交差検証(クロスバリデーション)法の仕組みを理解し, 学習データ, パラメータチューニング用の検証データ, テストデータを作成できる.	
DS-027	予測	評価	時間の経過とともに構造が変わっていくデータの場合は, 学習データを過去のデータとし, テストデータを未来のデータとすることを理解している.	
DS-028	検定/判断	検定/判断	点推定と区間推定の違いを説明できる.	
DS-029	検定/判断	検定/判断	帰無仮説と対立仮説の違いを説明できる.	
DS-030	検定/判断	検定/判断	第1種の過誤, 第2種の過誤, p値, 有意水準の意味を説明できる.	
DS-031	検定/判断	検定/判断	片側検定と両面検定の違いを説明できる.	
DS-032	検定/判断	検定/判断	推定する対象となるデータの対応の有無を考慮した上で適切な検定手法を選択し, 適用できる.	
DS-033	グルーピング	グルーピング	教師あり学習の分類(判別)モデルと教師なし学習のグループ化(クラスタリング)の違いを説明できる.	
DS-034	グルーピング	グルーピング	階層クラスター分析と非階層クラスター分析の違いを説明できる.	
DS-035	グルーピング	グルーピング	階層クラスター分析において, デンドログラムの見方を理解し, 適切に解釈できる.	
DS-036	性質・関係性の把握	性質・関係性の把握	適切な区間データ設定でヒストグラムを作成し, データのバラつき方を説明できる.	必須
DS-037	性質・関係性の把握	性質・関係性の把握	適切な軸設定でクロス集計表を作成し, 属性間のデータの偏りを把握できる.	必須
DS-038	性質・関係性の把握	性質・関係性の把握	量的変数の散布図を描き, 2変数の関係性を把握できる.	必須
DS-039	サンプリング	サンプリング	標本誤差とは何かを説明できる.	必須
DS-040	サンプリング	サンプリング	実験計画法の概要を説明できる.	

DSスキルチェックリスト★レベル

通し番号	大分類	中分類	小分類	種別
DS-041	データ可視化	方向性定義	可視化における目的の広がりについて概略を説明できる（単に現場の作業支援する場合から、ビッグデータ中の要素間の関連性をダイナミックに表示する場合など）	
DS-042	データ可視化	軸だし	散布図などの軸だしにおいて、縦軸・横軸の候補を適切に洗い出せる	必須
DS-043	データ可視化	軸だし	積み上げ縦棒グラフでの属性の選択など、適切な層化（比較軸）の候補を出せる	必須
DS-044	データ可視化	データ加工	サンプリングやアンサンブル平均によって適量にデータ量を減らすことができる	
DS-045	データ可視化	表現・実装技法	適切な情報濃度を判断できる（データインク比など）	
DS-046	データ可視化	表現・実装技法	不必要な誇張をしないための軸表現の基礎を理解できている（コラムチャートのY軸の基準点は「0」からを原則とし軸を切らないなど）	必須
DS-047	データ可視化	表現・実装技法	強調表現がもたらす効果と、明らかに不適切な強調表現を理解している（計量データに対しては位置やサイズ表現が色表現よりも効果的など）	必須
DS-048	データ可視化	表現・実装技法	1～3次元の比較において目的（比較、構成、分布、変化など）に応じ、BIツール、スプレッドシートなどを用いて図表化できる	必須
DS-049	データ可視化	表現・実装技法	端的に図表の変化をアニメーションで可視化できる（人口動態のヒストグラムが経年変化する様子を表現するなど）	
DS-050	データ可視化	表現・実装技法	データ解析部門以外の方に、データの意味を伝えるサインとしての可視化ができる	
DS-051	データ可視化	表現・実装技法	ボロノイ図の概念と活用方法を説明できる	
DS-052	データ可視化	表現・実装技法	1～3次元の図表を拡張した多変量の比較を適切に可視化できる（平行座標、散布図行列、テーブルレンズ、ヒートマップなど）	
DS-053	データ可視化	意味抽出	データの性質を理解するために、データを可視化し眺めて考えることの重要性を理解している	必須
DS-054	データ可視化	意味抽出	外れ値を見出すための適切な表現手法を選択できる	
DS-055	データ可視化	意味抽出	データの可視化における基本的な視点を挙げるができる（特異点、相違性、傾向性、関連性を見出すなど）	
DS-056	分析プロセス	アプローチ設計	スコープ、検討範囲・内容が明快に設定されていれば、必要なデータ、分析手法、可視化などを適切に選択できる。	
DS-057	データの理解・検証	統計情報への正しい理解	ニュース記事などで統計情報に接したときに、数字やグラフの持つメッセージを理解できる	
DS-058	データの理解・検証	データ確認	単独のグラフに対して、集計ミスなどがいないかチェックできる	必須
DS-059	データの理解・検証	データ確認	データ項目やデータの量・質について、指示のもと正しく検証し、結果を説明できる	
DS-060	データの理解・検証	俯瞰・メタ思考	データが生み出された背景を考え、鵜呑みにはしないことの重要性を理解している	
DS-061	データの理解・検証	データ理解	データから事実を正しく浮き彫りにするために、集計の切り口や比較対象の設定が重要であることを理解している	必須
DS-062	データの理解・検証	データ理解	普段業務で扱っているデータの発生トリガー・タイミング・頻度などを説明でき、また基本統計量を把握している	
DS-063	データの理解・検証	データ理解	何のために集計しているか、どのような知見を得たいのか、目的に即して集計できる	必須
DS-064	意味合いの抽出、洞察	洞察	分析・図表から直接的な意味合いを抽出できる。（バラツキ、有意性、分布傾向、特異性、関連性、変曲点、関連度の高低など）	
DS-065	意味合いの抽出、洞察	洞察	想定に影響されず、分布結果の数値を客観的に解釈できる。	
DS-066	時系列分析	時系列分析	時系列データとは何か、その基礎的な扱いについて説明できる。（時系列グラフ、周期性、移動平均など）	
DS-067	課題の定義	KPI	一般的な収益方程式に加え、自らが担当する業務の主要な変数（KPI）を理解している。	
DS-068	課題の定義	スコーピング	担当する事業領域において、市場規模、主要なプレイヤー、支配的なビジネスモデル、課題と機会について説明できる。	
DS-069	課題の定義	スコーピング	主に担当する事業領域であれば、取り扱う課題領域に対して基本的な課題の枠組みが理解できる。（調達活動の5フォースでの整理、CRM課題のRFMでの整理など）	必須
DS-070	データ入手	データ入手	仮説や既知の問題が与えられた中で、必要なデータにあたりをつけ、アクセスを確保できる。	必須
DS-071	言語処理	言語処理	テキストデータに対する代表的なクリーニング処理（小文字化、数値置換、半角変換、記号除去、ステミングなど）をタスクに応じて適切に実施できる。	

DSスキルチェックリスト★レベル

通し番号	大分類	中分類	小分類	種別
DS-072	言語処理	言語処理	形態素解析や係り受け解析の概念を説明できる。	
DS-073	画像・動画処理	画像処理	画像のデジタル表現の仕組みと代表的な画像フォーマットを知っている。	
DS-074	画像・動画処理	画像処理	画像に対して、目的に応じた適切な色変換や簡単なフィルタ処理などを行うことができる。	
DS-075	画像・動画処理	画像処理	画像データに対する代表的なクリーニング処理（リサイズ、バタニング、標準化など）をタスクに応じて適切に実施できる。	
DS-076	画像・動画処理	動画処理	動画のデジタル表現の仕組みと代表的な動画フォーマットを理解しており、動画から画像を抽出する既存方法を使うことができる。	
DS-077	パターン発見	パターン発見	条件Xと事象Yの関係性をリフト値を用いて評価できる。	
DS-078	データ加工	データクレンジング	名義尺度の変数をダミー変数に変換できる。	必須
DS-079	データ加工	データクレンジング	標準化とは何かを知っていて、適切に標準化が行える。	必須
DS-080	データ加工	データクレンジング	外れ値・異常値・欠損値とは何かそれぞれ知っていて、指示のもと適切に検出と除去・変数などの対応ができる。	必須
DS-081	データ加工	データ加工	分析要件や各変数の分布などをふまえて、必要に応じて量的変数のカテゴリライズやビン化を設計・実行できる。	
DS-082	データ加工	フィルタリング処理	数十万レコードのデータに対して、条件を指定してフィルタリングできる。（特定値に合致する・もしくは合致しないデータの抽出、特定範囲のデータの抽出、部分文字列の抽出など）	必須
DS-083	データ加工	フィルタリング処理	正規表現を活用して条件に合致するデータを抽出できる。（メールアドレスの書式を満たしているか判定するなど）	
DS-084	データ加工	ソート処理	数十万レコードのデータに対して、レコード間で特定カラムでのソートができ、数千レコードのデータに対して、カラム間でソートできる。	必須
DS-085	データ加工	結合処理	数十万レコードのデータに対して、単一条件による内部結合、外部結合、自己結合ができ、UNION処理ができる。	必須
DS-086	データ加工	クレンジング処理	数十万レコードのデータに対して、NULL値や想定外・範囲外のデータを持つレコードを取り除く、または規定値に変換できる。	必須
DS-087	データ加工	マッピング処理	数十万レコードのデータに対して、規程されたリストと照合して変換する、都道府県からジオコードに変換するなど、ある値を規定の別の値で表現できる。	必須
DS-088	データ加工	サンプリング処理	数十万レコードのデータに対して、ランダムまたは一定間隔にデータを抽出できる。	必須
DS-089	データ加工	集計処理	数十万レコードのデータに対して、合計や最大値、最小値、レコード数を算出できる。	必須
DS-090	データ加工	変換・演算処理	数十万レコードのデータに対する四則演算ができ、数値データを日時データに変換するなど別のデータ型に変換できる。	必須
DS-091	データ共有	データ出力	加工・分析処理結果をCSV、XML、JSON、Excelなどの指定フォーマット形式に変換してエクスポートできる。	必須
DS-092	データ共有	データ出力	加工・分析処理結果を、接続先DBのテーブル仕様に合わせてレコード挿入できる。	必須
DS-093	データ共有	データ展開	データ取得用のWeb API（REST）やWebサービス(SOAP)などを用いて、必要なデータを取得できる。	
DS-094	データ共有	データ連携	FTP、ファイル共有一などから必要なデータファイルをダウンロードして、Excelなどの表計算ソフトに取り込み活用できる。	必須
DS-095	データ共有	データ連携	BIツールのレポート編集機能を用いて新規レポートを公開できる。	
DS-096	データ共有	データ連携	BIツールの自由検索機能を活用し、必要なデータを抽出して、グラフを作成できる。	
DS-097	環境構築	システム運用	サーバー1～10台規模のシステム構築、システム運用を指示書があれば実行できる。	
DS-098	環境構築	システム運用	数十万レコードを持つデータベースのバックアップ・アーカイブ作成など定常運用ができる。	
DS-099	環境構築	システム企画	データベースから何らかのデータ抽出方法を活用し、小規模なExcelのデータセットを作成できる。	必須
DS-100	環境構築	システム企画	オープンデータを収集して活用する分析システムの要件を整理できる。	
DS-101	データ収集	クライアント技法	対象プラットフォームが提供する機能（SDKやAPIなど）の概要を説明できる。	
DS-102	データ収集	クライアント技法	Webクローラー・スクレイピングツールを用いてWebサイト上の静的コンテンツを分析用データとして収集できる。	

DSスキルチェックリスト★レベル

通し番号	大分類	中分類	小分類	種別
DS-103	データ収集	通信技術	対象プラットフォームに用意された機能（HTTP, FTP, SSHなど）を用い、データを収集先に格納するための機能を実装できる。	必須
DS-104	データ構造	基礎知識	扱うデータが、構造化データ（顧客データ、商品データ、在庫データなど）か非構造化データ（雑多なテキスト、音声、画像、動画など）なのかを判断できる。	必須
DS-105	データ構造	基礎知識	ER図を読んでテーブル間のリレーションシップを理解できる。	必須
DS-106	データ構造	テーブル定義	正規化手法（第一正規化～第三正規化）を用いてテーブルを正規化できる。	
DS-107	データ蓄積	DWH	DWHアプライアンス（Oracle Exadata, IBM Integrated Analytics System, Teradataなど）に接続し、複数テーブルを複合したデータを抽出できる。	
DS-108	データ蓄積	分散技術	Hadoop・Sparkの分散技術の基本的な仕組みと構成を理解している。	
DS-109	データ蓄積	分散技術	NoSQLデータストア（Hbase, Cassandra, Mongo DB, CouchDB, Redis, Amazon, DynamoDB, Cloudant, Azure Cosmos DBなど）にAPIを介してアクセスし、新規データを登録できる。	
DS-110	データ蓄積	クラウド	クラウド上のストレージサービス（Amazon S3, Google Cloud Storage, IBM Cloud Object Storageなど）に接続しデータを格納できる。	
DS-111	プログラミング	基礎プログラミング	小規模な構造化データ（CSV, RDBなど）を扱うデータ処理（抽出・加工・分析など）を、設計書に基づき、プログラム実装できる。	必須
DS-112	プログラミング	データインターフェース	JSON, XMLなど標準的なフォーマットのデータを受け渡すために、APIを使用したプログラムを設計・実装できる。	
DS-113	プログラミング	分析プログラム	Jupyter NotebookやRstudioなどの対話型の開発環境を用いて、データの分析やレポートの作成ができる。	
DS-114	プログラミング	SQL	SQLの構文を一通り知っていて、記述・実行できる（DML, DDLの理解, 各JOINの使い分け, 集計関数とGROUP BY, CASE文を使用した縦横変換, 副問い合わせやEXISTSの活用など）	
DS-115	ITセキュリティ	基礎知識	セキュリティの3要素（機密性, 可用性, 完全性）について具体的な事例を用いて説明できる。	必須
DS-116	ITセキュリティ	攻撃と防御手法	マルウェアなどによる深刻なリスクの種類（消失, 漏洩, サービスの停止など）を常に意識している。	必須
DS-117	ITセキュリティ	攻撃と防御手法	OS, ネットワーク, アプリケーション, データに対するユーザーごとのアクセスレベルを手順に従い設定できる。	
DS-118	ITセキュリティ	暗号化技術	暗号化されていないデータは、不正取得された際に容易に不正利用される恐れがあることを理解し、データの機密度合いに応じてソフトウェアを使用した暗号化と復号ができる。	必須
DS-119	ITセキュリティ	暗号化技術	なりすましや偽造された文書でないことを証明するために電子署名と公開鍵認証基盤（PKI: public key infrastructure）が必要であることを理解している。	
DS-120	ITセキュリティ	暗号化技術	ハッシュ関数を用いて、データの改ざんを検出できる。	
DS-121	行動規範	データ倫理	データを取り扱う人間として相応しい倫理を身に着けている。（データのねつ造, 改ざん, 盗用を行わないなど）	必須
DS-122	行動規範	コンプライアンス	個人情報に関する法令（個人情報保護法, EU一般データ保護規則: GDPRなど）や、匿名加工情報の概要を理解し、守るべきポイントを説明できる。	必須
DS-123	機械学習技法	機械学習	機械学習にあたる解析手法の名称を3つ以上知っており、手法の概要を説明できる。	
DS-124	機械学習技法	機械学習	指示を受けて機械学習のモデルを使用したことがあり、どのような問題を解決することができる理解している。	
DS-125	機械学習技法	機械学習	「教師あり学習」「教師なし学習」の違いを理解している。	必須
DS-126	機械学習技法	機械学習	過学習とは何か、それがもたらす問題について説明できる。	必須
DS-127	機械学習技法	機械学習	次元の呪いとは何か、その問題について説明できる。	
DS-128	機械学習技法	機械学習	教師あり学習におけるアノテーションの必要性を説明できる。	
DS-129	機械学習技法	機械学習	観測されたデータにバイアスが含まれる場合や、学習した予測モデルが少数派のデータをノイズと認識してしまった場合などに、モデルの出力が差別的な振る舞いをしてしまうリスクを理解している。	
DS-130	機械学習技法	機械学習	機械学習における大域的（global）な説明（モデル単位の各変数の寄与度など）と局所的（local）な説明（予測するレコード単位の各変数の寄与度など）の違いを理解している。	
DS-131	行動規範	ビジネスマインド	ビジネスにおける論理とデータの重要性を認識し、分析的でデータドリブンな考え方に基づき行動できる。	必須

DSスキルチェックリスト★レベル

通し番号	大分類	中分類	小分類	種別
DS-132	行動規範	ビジネスマインド	「目的やゴールの設定がないままデータを分析しても、意味合いが出ない」ことを理解している。	必須
DS-133	行動規範	ビジネスマインド	課題や仮説を言語化することの重要性を理解している。	必須
DS-134	行動規範	ビジネスマインド	現場に向いてヒアリングするなど、一次情報に接することの重要性を理解している。	必須
DS-135	論理的思考	MECE	データや事象の重複に気づくことができる。	必須
DS-136	論理的思考	言語化能力	通常見受けられる現象の場合において、分析結果の意味合いを正しく言語化できる。	必須
DS-137	論理的思考	ストーリーライン	一般的な論文構成について理解している。（序論→アプローチ→検討結果→考察や、序論→本論→結論 など）	
DS-138	論理的思考	ドキュメンテーション	1つの図表～枚数程度のドキュメントを理論立ててまとめることができる。（課題背景、アプローチ、検討結果、意味合い、ネクストステップ）	
DS-139	論理的思考	説明能力	報告に対する論拠不足や論理破壊を指摘された際に、相手の主張をすみやかに理解できる。	必須
DS-140	活動マネジメント	プロジェクト発足	ウォーターフォール開発とアジャイル開発の違いを説明できる。	
DS-141	活動マネジメント	リソースマネジメント	指示に従ってスケジュールを守り、チームリーダーに頼まれた自分の仕事を完遂できる。	必須
DS-142	活動マネジメント	リスクマネジメント	担当するタスクの遅延や障害などを発見した場合、迅速かつ適切に報告ができる。	必須
DS-143	ビジネス観点のデータ理解	データ理解	ビジネス観点で仮説を持ってデータをみることの重要性和、仮に仮説と異なる結果となった場合にも、それが重要な知見である可能性を理解している。	必須
DS-144	ビジネス観点のデータ理解	意味合いの抽出、洞察	分析結果を元に、起きている事象の背景や意味合い（真実）を見ぬくことができる。	
DS-145	契約・権利保護	契約	請負契約と準委任契約の違いを説明できる。	
DS-146	事業への実装	評価・改善の仕組み	結果、改善の度合いをモニタリングする重要性を理解している	必須

EDISON DS-BoK

大分類	KA Groups	KA番号	DS Knowledge Areas
A	Data Science Analytics	KA01.01	Statistical methods for data analysis
B		KA01.02	Machine Learning
		KA01.03	Data Mining
C		KA01.04	Text Data Mining
		KA01.05	Predictive Analytics
		KA01.06	Computational modelling, simulation and optimisation
D	Data Science Engineering	KA02.01	Big Data Infrastructure and Technologies
		KA02.02	Infrastructure and platforms for Data Science applications
		KA02.03	Cloud Computing technologies for Big Data and Data Analytics
		KA02.04	Data and Applications security
		KA02.05	Big Data systems organisation and engineering
		KA02.06	Data Science (Big Data) applications design
		KA02.07	Information systems (to support data driven decision making)
G	Data Management	KA03.01	General principles and concepts in Data Management and organisation
		KA03.02	Data management systems
		KA03.03	Data Management and Enterprise data infrastructure
		KA03.04	Data Governance
		KA03.05	Big Data storage (large scale)
		KA03.06	Digital libraries and archives
	Research Methods and Project Management	KA04.01	Research Methods
		KA04.02	Project Management
	Business Analytics	KA05.01	Business Analytics Foundation
		KA05.02	Business Analytics organisation and enterprise management